



Delft University of Technology

From Waves to Shadows

PV systems yield modeling within H2020 Trust-PV project

Alcañiz Moya, A.

DOI

[10.4233/uuid:58085760-3ba5-47b5-94e4-d7dbcc1fc1fc](https://doi.org/10.4233/uuid:58085760-3ba5-47b5-94e4-d7dbcc1fc1fc)

Publication date

2025

Document Version

Final published version

Citation (APA)

Alcañiz Moya, A. (2025). *From Waves to Shadows: PV systems yield modeling within H2020 Trust-PV project*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:58085760-3ba5-47b5-94e4-d7dbcc1fc1fc>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

FROM WAVES TO SHADOWS

PV SYSTEMS YIELD MODELING
WITHIN H2020 TRUST-PV PROJECT

FROM WAVES TO SHADOWS

**PV SYSTEMS YIELD MODELING
WITHIN H2020 TRUST-PV PROJECT**

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
Chair of the Board for Doctorates,
to be defended publicly on
Thursday 15 May 2025 at 10:00

by

Alba ALCANIZ MOYA

Master of Science in Sustainable Energy Technology,
Delft University of Technology, the Netherlands
born in Barcelona, Spain.

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr.ir. O. Isabella	Delft University of Technology, <i>promotor</i>
Em.Prof.dr.ir. M. Zeman	Delft University of Technology, <i>promotor</i>
Dr.ir. H. Ziar	Delft University of Technology, <i>copromotor</i>

Independent members:

Prof.dr. F. Baumgartner	ZHAW School of Engineering, Switzerland
Dr. T. Kjeldstad	Institute for Energy Technology, Norway
Prof.dr. W. van Sark	Utrecht University
Prof.dr. A. Metrikine	Delft University of Technology
Prof.dr.ir. M. Popov	Delft University of Technology, reserve member



ISBN 978-94-6518-038-0

Copyright © 2025 by A. Alcañiz Moya

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

Printed by: Ridderprint | www.ridderprint.nl

Cover design by: CHECKMATE SERVEIS ESTRATEGICS SL.

*As complex as needed
as simple as possible.*

Spirit of science, Dr. rer. Reductionist

CONTENTS

Contents	xi
Summary	xiii
Samenvatting	xv
1 Introduction	1
1.1 Societal acceptance	2
1.2 Political barriers	2
1.3 Infrastructural limitations.	3
1.4 Technical challenges	4
1.5 Aim of this work.	5
1.5.1 Machine learning predictions	5
1.5.2 Physical-based methodologies.	5
1.5.3 Floating PV systems	5
1.5.4 Shading tolerability	6
1.6 Outline of this work.	6
1.7 Novel contributions to the field	6
MACHINE LEARNING PV POWER PREDICTIONS	11
2 Trends and Gaps	11
2.1 Introduction	12
2.2 Machine learning families	14
2.2.1 Key features and trends	16
2.2.2 Algorithms comparison	18
2.2.3 Data preparation.	20
2.3 Special approaches	21
2.4 System characteristics	23
2.4.1 Location	23
2.4.2 Number of systems	25
2.4.3 Timespan	26
2.5 Forecast horizon	26
2.6 Choice of features.	28
2.7 Outlook	29
2.8 Conclusions.	32
Perspective.	32
3 Database of open-source PV systems	35
3.1 Introduction	36
3.2 Data gathering	36
3.3 Website development	38

4	Effect of Climate	41
4.1	Introduction	42
4.2	Data set	42
4.2.1	Data characteristics	42
4.2.2	Data preparation.	43
4.2.3	Data exploration	43
4.3	Methodology	44
4.3.1	Algorithms	45
4.3.2	Hyperparameter tuning	46
4.4	Results	46
4.4.1	Climatic trends.	46
4.4.2	Flexibility to other climates	49
4.4.3	Robustness.	51
4.4.4	Köppen-Geiger-Photovoltaic climate classification	52
4.5	Conclusions.	53
5	Nowcasting for Residential-Scale Systems	55
5.1	Introduction	56
5.2	Dataset	56
5.2.1	Considered features	56
5.2.2	Data preparation.	57
5.2.3	Data exploration	57
5.3	Methodology	59
5.3.1	Feature selection.	59
5.3.2	Models.	59
5.3.3	Loss functions	60
5.3.4	Metrics.	61
5.4	Results	62
5.4.1	Feature Selection.	62
5.4.2	Models performance.	63
5.4.3	Learning curves	64
5.4.4	Error Analysis	66
5.5	Conclusions.	68
6	Fault Detection using Peer Systems	69
6.1	Introduction	70
6.2	Models	71
6.2.1	Performance-to-Peer Model	71
6.2.2	Proposed Model	72
6.2.3	Model optimization	73
6.2.4	Baseline models	75
6.3	Data.	75
6.4	Results	75
6.4.1	Optimization.	76
6.4.2	Models Comparison	77
6.4.3	Use-case Analysis	78

6.5	Fault Detection	80
6.5.1	Categorization	81
6.5.2	Key Examples	83
6.6	Conclusions.	87
	Perspective.	87
ANALYTICAL PV POWER PREDICTIONS		90
7	Introducing the PVMD Toolbox	91
7.1	Introduction	92
7.2	Simulation model.	95
7.2.1	Optical and electrical properties	96
7.2.2	System surroundings.	96
7.2.3	In-plane irradiance	99
7.2.4	Cell temperature	100
7.2.5	Module electrical interconnection model	100
7.2.6	Inverter model	101
7.3	Validation of the horizon reconstruction	101
7.4	Data characteristics	103
7.4.1	System data	103
7.4.2	Weather data.	104
7.5	Results	104
7.5.1	6 systems analysis	105
7.5.2	Single system analysis	106
7.5.3	Sensitivity to input irradiance	107
7.6	Conclusions.	109
	Perspective.	110
8	Irradiance sensor allocation in PV farms	113
8.1	Introduction	114
8.2	Methodology	115
8.2.1	Inputs	117
8.2.2	Sky view factor	117
8.2.3	Irradiance maps	119
8.2.4	Error-based minimization problem	120
8.2.5	Number of sensors.	122
8.2.6	Performance indicators	123
8.3	Databases.	123
8.3.1	Digital surface model	123
8.3.2	Global horizontal irradiance	123
8.3.3	Time-dependent albedo	123
8.4	Results	124
8.4.1	Case studies	124
8.4.2	Simplification	130
8.5	Discussion	131
8.6	Conclusions.	132

FLOATING PHOTOVOLTAICS	135
9 Introduction to floating PV	135
9.1 Why place PV modules on water?	136
9.2 Test system description	137
9.3 Water fluctuations	137
9.4 Dust accumulation	139
9.5 Module temperature	142
9.6 Conclusions.	145
10 Effect of sea waves on the yield of OFPV systems	147
10.1 Introduction	148
10.2 Methodology	150
10.2.1 Waves model.	150
10.2.2 Mechanical model	152
10.2.3 Electrical model	155
10.2.4 AC conversion	155
10.3 Model inputs	157
10.4 Results	159
10.4.1 Validation of the mechanical model	159
10.4.2 Modelled surface elevation	159
10.4.3 Sensitivity analysis on sizing	160
10.4.4 Inclinations due to waves	163
10.4.5 Waves impact on DC yield	163
10.4.6 Module temperature	166
10.4.7 Waves impact on inverter	167
10.5 Conclusions.	169
Perspective.	170
11 Movement and power mismatch losses	173
11.1 Introduction	174
11.2 Methodology	175
11.2.1 Hydroelastic fluid-structure interaction model	176
11.2.2 Link between multidisciplinary models	179
11.2.3 Optoelectrical	181
11.3 Case study description	185
11.4 Numerical results: Base Case	186
11.4.1 Setup of the numerical domain	186
11.4.2 Platform response to monochromatic waves.	187
11.4.3 Response for a wide range of monochromatic waves.	191
11.4.4 Tilt distribution	195
11.4.5 Power losses	196
11.5 Variations to the base case	198
11.5.1 Orientation	198
11.5.2 Young's modulus.	200
11.5.3 Thickness	201
11.5.4 Fill ratio	202
11.6 Conclusions.	203

SHADING TOLERABILITY	206
12 Shading Tolerability	207
12.1 Introduction208
12.2 Shading Tolerability Calculator210
12.2.1 Finding the shading scenarios210
12.2.2 Developing the IV-curve calculator214
12.2.3 Visualizing the Shading Tolerability215
12.2.4 Validating the calculator218
12.3 Sensitivity Analysis218
12.4 Computational burden reduction221
12.4.1 Obtaining the unique scenarios per SAF221
12.4.2 Considering only the most common scenarios222
12.4.3 Optimizing the code224
12.4.4 Reducing the number of scenarios225
12.4.5 Overview227
12.5 Results and discussion227
12.6 Conclusions230
Perspective231
13 Conclusions and Outlook	235
13.1 Conclusions235
13.2 Outlook237
A Reviewed ML forecasting literature	239
B Machine learning families	245
C Metrics	251
D Decomposition models comparison	255
E Sensors allocation software tool overview	259
F Sensors allocation algorithm simplification	263
References	301
Acknowledgements	303
List of Publications	305
Curriculum Vitæ	309

SUMMARY

Solar photovoltaic (PV) energy has the potential to become a major source of electricity production worldwide. However, the deployment of this renewable energy source comes with several challenges. Besides societal, political, and grid-limiting, several technical limitations decrease the trust in this technology. Within this context arises the Trust-PV project, whose objective is to improve the performance and reliability of photovoltaic systems.

This work contributes to this European project by exploring different power prediction models for several types of PV systems. Considering the broadness of the topic, four parts or blocks are identified. The first part deals with machine learning models to forecast the yield of residential PV systems. The second block focuses on analytical models used during the design phase. The third part is dedicated to systems floating on water. Lastly, a metric to assess the tolerance towards shading of different modules is developed in the fourth block.

Starting with the first block of machine learning techniques for PV power forecasting, [Chapter 2](#) introduces the topic by reviewing a large number of manuscripts. The chapter performs a broad classification of the reviewed literature with the objective to identify trends and gaps in the field. Among the identified trends, one can highlight the high percentage of predictions for the day ahead, the generally low number of systems employed to train the models, and the concentration of systems in mild climates.

The latter points may stem from researchers primarily using the systems available within their institutions. To promote collaboration, [Chapter 3](#) presents a developed website that lists PV power open source databases. The website aims to encourage researchers to train and test their models with different data sources.

One consequence of the concentration of systems in mild climates is that the effect of climate on machine learning models remains underexplored in the literature. [Chapter 4](#) addresses this gap by studying how machine learning models behave for systems located in different climatic zones. The results show that weather homogeneity affects the accuracy of the models. Models developed for systems located in uniform climates - like desert areas - achieve in general higher accuracy than the models developed for systems in highly varying climates - like tropical areas.

[Chapter 5](#) addresses another challenge: creating a single machine learning model able to monitor the performance of a large fleet of residential PV systems. The developed model surpasses in accuracy an analytical reference model but is limited by a fundamental characteristic of machine learning methods: the focus on large errors which resulted in the overlooking of smaller systems. Consequently, [Chapter 6](#) develops a different approach based on the peer-to-peer methodology. In this approach, the power output of similar neighboring systems is compared to identify any malfunctions. The method is tested for the residential fleet of PV systems and proves effective for detecting faults.

Moving on to the second block of analytical power predictions, [Chapter 7](#) presents the *PVMD toolbox*, a state-of-the-art analytical simulation framework that can predict the power of systems that do not exist yet. The abilities of the toolbox are tested for residential systems

in the same chapter and the results show the negative influence that inaccurate input irradiance data has on the predictions.

This importance of accurate irradiance data affects all kinds of PV systems, but especially large-scale ones. Therefore, to monitor them, a proper allocation of irradiance sensors is essential. Hence, in [Chapter 8](#), a software tool is developed to identify the optimal number of irradiance sensors and their position in a PV farm. The tool's strengths are more prominent in plants located on terrains with significant elevation changes.

The third block focuses on PV systems that are installed on floating platforms rather than on land. [Chapter 9](#) introduces the topic by examining three factors influenced by proximity to water that can impact the production of a floating PV system in a French quarry lake: movement fluctuations, dust accumulation, and module temperature. The results reveal a limited influence of all factors on the production for the period of study therefore facilitating the deployment of floating systems.

The block continues by studying the effect of sea waves for a system located in the North Sea. The simulation results from [Chapter 10](#) reveal that wave fluctuations can have a negative yet limited effect on the DC and AC yield of floating PV systems. These results are further elaborated in [Chapter 11](#), where the model is improved by considering the fluid-structure interaction. This advanced model enables to study the effect of various platform characteristics on the power mismatch losses. The results reveal a trade-off between mechanical stability and mismatch losses.

Finally, the last part deals with the power lost when a PV module is partially shaded. [Chapter 12](#) develops a simulation tool to efficiently calculate the *shading tolerability* of a PV module given its datasheet. The shading tolerability is a metric that quantifies the resilience towards shading of a PV module, that is how much power is lost when the module is partially shaded. The developed tool is used to create a database of shading tolerability of commercial PV modules, to compare the resilience of different modules towards shading.

SAMENVATTING

Fotovoltaïsche zonne-energie heeft het potentieel om een wereldwijde hoofdbron van elektriciteit productie te worden. Maar de inzet van deze duurzame energiebron komt met meerdere uitdagingen. Naast maatschappelijk, politiek, en stroomnet-beperkend, nemen meerdere technische beperkingen het vertrouwen in deze technologie af. Binnen deze context is het Trust-PV project geboren, wiens doel de verbetering van de prestatie en betrouwbaarheid van zonnepanelen is.

Dit boek draagt bij aan dit Europese project door verschillende stroomproductie modellen voor meerdere soorten zonnepanelen te verkennen. Rekening houdend met de breedte van dit onderwerp, zijn er vier delen of blokken geïdentificeerd. Het eerste deel gaat over machine learning-modellen die de opbrengst van residentiële zonnepanelen voorspellen. Het tweede blok richt zich op analytische modellen die gebruikt worden tijdens het ontwerp. Het derde deel is gewijd aan zonnepanelen die op het water drijven. Tot slot, wordt in het vierde blok een metriek die de tolerantie voor schaduw van verschillende fotovoltaïsche modules beoordeelt ontwikkeld.

Beginnend met het eerste blok van machine learning modellen voor het voorspellen van zonne-energie, introduceert [Hoofdstuk 2](#) het onderwerp door een overzicht van een groot aantal onderzoeksartikelen. Het hoofdstuk geeft een brede classificatie van de beoordeelde literatuur om de tendensen en gaten van het onderwerp te identificeren. Tussen de geïdentificeerde tendensen, kan men de hoogste percentage van voorspellingen voor een dag vooruit, het laagste aantal systemen om de modellen te trainen, en de concentratie van systemen in milde klimaten vinden.

De twee laatste punten kunnen voortkomen uit het feit dat onderzoekers alleen de beschikbare systemen in hun instellingen gebruiken. Om samenwerking te stimuleren, presenteert [Hoofdstuk 3](#) een ontwikkelde website dat een overzicht geeft van open source-databases. De website is bedoeld om onderzoekers aan te moedigen hun modellen met verschillende data bronnen te trainen en testen.

Een consequentie van de concentratie van zonnepanelen in milde klimaten is dat het effect van klimaat in machine learning modellen niet onderzocht is. [Hoofdstuk 4](#) gaat in op dit gat door te onderzoeken hoe machine learning modellen zich gedragen voor zonnepanelen gevestigd in verschillende klimaten. De resultaten tonen dat weersomogeniteit de nauwkeurigheid van de modellen beïnvloedt. Modellen ontwikkeld voor zonnepanelen gevestigd in uniforme klimaten - zoals woestijngebieden - bereiken over het algemeen een hogere nauwkeurigheid dan modellen ontwikkeld voor zonnepanelen gevestigd in sterk wisselende klimaten - zoals tropische gebieden.

[Hoofdstuk 5](#) gaat in op een andere uitdaging: de ontwikkeling van een enkel machine learning model dat de prestatie van een grote vloot van residentiële zonnepanelen kan volgen. Het ontwikkelde model overtrof de nauwkeurigheid van een analytisch referentie model, maar werd door een fundamentele eigenschap van machine learning methodes beperkt: de focus op grote fouten, waardoor kleinere systemen over het hoofd werden gezien. Vervolgens ontwikkelt [Hoofdstuk 6](#) een andere benadering gebaseerd op de peer-to-peer

methode. Met deze benadering wordt de zonne-energie van soortgelijke systemen vergeleken om storingen te identificeren. De methode werd voor de vloot van residentiële zonnepanelen systemen getest en bleek effectief voor het detecteren van fouten.

Doorgaand naar het tweede blok van analytische zonne-energie voorspellingen, presenteert [Hoofdstuk 7](#) de *PVMD toolbox*, een state-of-the-art analytisch simulatiekader dat de zonne-energie van systemen die nog niet bestaan kan voorspellen. De mogelijkheden van het simulatiekader werden in hetzelfde hoofdstuk voor residentiële zonnepanelen systemen getest en de resultaten tonen de negatieve invloed die onnauwkeurige stralingssterkte data op de voorspellingen hebben.

Het belang van nauwkeurige stralingssterkte beïnvloedt alle soort zonnepanelen systemen, maar vooral grootschalige. Daarom is een goede plaatsing van zonnepanelen meters essentieel om ze te controleren. Dus, in [Hoofdstuk 8](#) wordt een simulatiekader ontwikkeld om het optimale aantal en de optimale positie van zonnepanelen meters in een zonnepark te identificeren. De sterke punten van het kader zijn opvallend in zonneparken gevestigd op locaties met aanzienlijke hoogteverschillen.

Het derde blok richt zich op zonnepanelen systemen geïnstalleerd op drijvende platformen in plaats van op grond. [Hoofdstuk 9](#) introduceert het onderwerp door het bestuderen van drie factoren die worden beïnvloed door de nabijheid van water die de productie van een drijvend zonnepanelen systeem in een Frans meer kunnen raken: bewegingsfluctuaties, stofophoping en moduletemperatuur. De resultaten onthullen een beperkte invloed van alle factoren op de productie tijdens de studieperiode, waardoor de implementatie van drijvende zonnepanelen systemen wordt vergemakkelijkt.

Het blok gaat door met het bestuderen van het effect van zee golven op een systeem gevestigd in de Noordzee. De simulatieresultaten in [Hoofdstuk 10](#) onthullen dat golf fluctuaties een negatieve maar beperkte effect op de DC- en AC-opbrengst van drijvende zonnepanelen systemen kunnen hebben. Deze resultaten worden in [Hoofdstuk 11](#) verder uitgewerkt, waar het model wordt verbeterd door rekening te houden met de interactie tussen vloeistof en structuur. Dit geavanceerde model maakt het mogelijk om het effect van verschillende platform eigenschappen op vermogensmismatchverliezen te bestuderen. De resultaten onthullen een trade-off tussen mechanische stabiliteit en mismatchverliezen.

Tot slot gaat het laatste deel over het vermogensverlies wanneer een fotovoltaïsche module zich gedeeltelijk in de schaduw bevindt. [Hoofdstuk 12](#) ontwikkelt een simulatietool om efficiënt de *shading tolerability* met de datasheet van een module te berekenen. De shading tolerability is een metriek die de weerstand tegen schaduw van een module kwantificeert, met andere woorden, hoeveel vermogen de module verliest wanneer het gedeeltelijk in de schaduw is. De ontwikkelde tool wordt gebruikt om een database van shading tolerability waarden van commerciële fotovoltaïsche modules te creëren, om de weerstand tegen schaduw van verschillende soort modules te vergelijken.

Gecorrigeerd door Lune Massop

1

INTRODUCTION

I am glad to write that currently, aside from a few skeptics, climate change is widely recognized as an undeniable reality [1]. And climate change needs to be fought if humanity wants to continue developing as it has been until now. However, this fight will require an effort and a change of habits that most people will be reticent to make.

From an energetic point of view, the electrification of the economy and the massive implementation of renewables have been established as the most cost-effective, feasible, and straightforward path to fight climate change [2, 3]. Solar photovoltaics (PV) will be the cheapest source of electricity almost worldwide in just a few years, see [Figure 1.1](#) [4]. Abundant, affordable, easily scalable, flexible... solar PV has the potential to become a major source of electricity [5, 6].

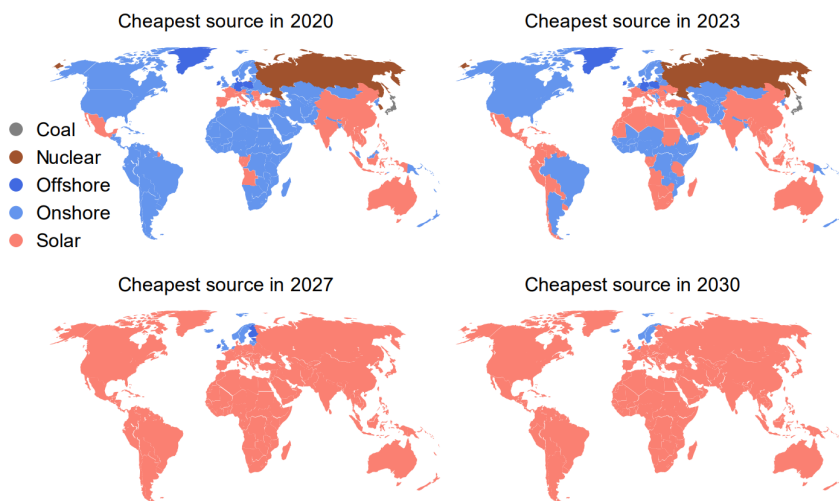


Figure 1.1: World maps showing the energy source with the lowest generalized levelized cost of electricity in 2020, 2023, 2027, and 2030. Figure from [4].

With economic growth being the crucial factor shaping human actions and policies, solar energy can have a profound impact worldwide. However, despite the competitiveness and versatility of PV, it accounted for only 5.4% of the world's electricity generation in 2023 [7]. Several obstacles can explain this low value, categorized into four groups based on their nature: social, political, infrastructural, and technical.

1.1. SOCIETAL ACCEPTANCE

A key aspect of advancing toward a more sustainable energy future lies in comprehending public perceptions [8]. Community acceptance stands out as a main limiting factor in the implementation of renewable energy (RE) projects [9]. Nevertheless, studies on public attitudes towards RE development in the United States and Western Europe reveal robust support, exceeding 90% [10, 11]. Solar power, in particular, is preferred over other sources [8, 9].

However, when projects for large-scale PV installations get specific, discussions arise [10, 12]. The origin of these discussions is not necessarily linked to the known “not in my backyard” (NIMBY) movement, which has faced criticism over the years, but to broader societal concerns [13]. Interestingly, many of the vocal critics often have rooftop PV installations themselves [13] as the local use of residential PV modules is generally accepted and uncontroversial [14]. Public perception differs between large and small-scale projects [9, 14].

Utility-scale PV farms face notable social resistance, especially in rural areas [14], as the installations are seen as something “big and not green” [13]. Key concerns include a perceived loss of local control and equitable benefit distribution, potential threats to the economy, community culture, and tourism, and a general distrust of the motivations of owners and developers [8, 13, 14].

In these cases, citizen engagement is lacking. Due to the nature of the energy transition, citizens are the key actors thus the acceptance of change is an essential element of the success of this revolution [12, 15]. People favor energy projects that support their core values, but they often do not consider the implications for their values once the projects become concrete [16]. By engaging the public in co-constructing the policy in a fair and meaningful way [8], the local ownership and familiarity with the technology can increase, and with them the level of acceptance [8, 9, 15].

Initiatives such as the Social RES project promote and expand the use of energy communities to empower citizens and make them responsible actors in the energy transition [17]. The wind energy installations in Galician common lands are an example of a successful energy community with societal acceptance [18].

1.2. POLITICAL BARRIERS

When discussing the political barriers to PV deployment, the focus lies on the lobbies. Lobbies represent the interests of a group of actors and attempt to influence policy decisions [19, 20]. Interest organizations supply politicians with essential information and legitimacy for decision-making in exchange for access and influence [20]. In the case at hand, the conflict is between the lobby of RE businesses and fossil fuel interest groups. In general, the latter is more resourceful and therefore more influential and impactful, which is a barrier to breaking the status quo of fossil fuel companies [19, 21]. Due to its limited funding, green

coalitions often can only focus on the most pressing and successful issues [21].

Lobbies are a natural part of a democracy [20], but they pose a potential threat as they can exert significant influence on politicians aligned with their interests rather than the country's. A mutual dependency has developed between politicians and interest organizations [20], increasing the risk of manipulation particularly for laws with loose implementation guidelines or legislators close to reelection [21]. A curious example of lobbying manipulation where the solar sector benefited is the Japanese one [22]. As opposed to other large countries, the energy mix of Japan is single-legged to solar which is not optimum economically. The powerful solar lobby twisted the country's RE policy by making the government a stakeholder in its policy.

Similarly, in the 1960s fossil fuel lobbyists were a barrier to RE deployment. The strategic value of the electricity market made it a highly regulated sector, which allowed large electricity-generated companies to hold a monopoly with wide influence and favorable conditions [20].

A big change occurred with the nuclear accident in Fukushima in 2011. In several countries, but especially in Japan, nuclear power had to be quickly substituted. Fossil fuels took over since renewable energy sources (RES) lacked flexibility and rapid implementation. But that substitution left the Japanese government in a bad position after having signed the Kyoto Protocol so they promoted the use of RES [22].

Traditional utilities initially underestimated the threat posed by RES, but that perception changed as RE rapidly expanded [12]. A tug-of-war unfolded between RE lobbyists advocating for cleaner electricity based on climate change concerns and fossil fuel lobbyists pushing to slow down RE deployment by emphasizing potential increases in electricity prices [12, 19].

Nowadays, the increased maturity and economic competitiveness of RES together with the large support discussed in the previous chapter tipped the scale towards climate advocates. Large-size utilities have been unable to stop the RE support in the EU and now even they have RE daughter companies [20]. The policies focus now on increasing investor confidence and reducing the carbon lock-in instead of compensating for market inequalities [19].

However, this increasing political support in the EU is not a common trend worldwide. Governance quality, wealth, energy independence, and security are factors that encourage the deployment of RES [23]. Therefore, in some countries lacking these aspects, PV deployment is still hindered.

1.3. INFRASTRUCTURAL LIMITATIONS

Deploying the whole PV potential, despite being necessary to fight climate change, could be currently impossible due to limitations in the electricity grid. In some countries, this is the biggest barrier that solar energy is facing nowadays [7].

On the one hand, the energy supplied by PV modules is intermittent due to its dependency on weather conditions [24]. This intermittency may lead to voltage instability and increased volatility of the grid [25–27]. Large-scale PV penetration entails a major challenge for electric system operators since it hinders the effective management of the grids and increases the system costs [22]. Additionally, aging grid infrastructures in some areas cannot accommodate variable and decentralized PV systems [7].

On the other hand, the grid has a limited capacity for integrating new generating en-

ergy sources. Most of the Netherlands, for instance, has no transport capacity available, as shown in Figure 1.2, despite continuous efforts to increase the grid capacity and manage the congestion [28]. 20 out of the 26 countries with the highest PV capacity identify grid bottlenecks as barriers to solar deployment [7]. These grid bottlenecks will result in increased PV curtailment and additional challenges as the grid cannot expand fast enough [29].

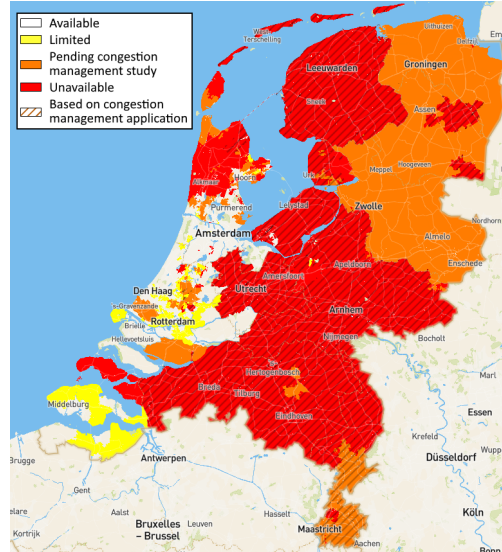


Figure 1.2: Capacity map in the Dutch electricity grid. Figure from [30].

User consumption flexibility, collective self-consumption, adequate interconnections between regions, and energy storage are possible solutions that could help overcome the limited grid capacity [7, 15, 29].

1.4. TECHNICAL CHALLENGES

A way to help system operators manage the grid is to inform them how much electricity will be generated by these renewable sources. Consequently, forecasting the power produced by solar PV is an important factor in facilitating energy transition. Accurate forecasts decrease energy yield uncertainty, therefore reducing generation-load mismatch in the power grid. However, achieving a precise solar power prediction can be extremely difficult considering not only the chaotic nature of weather systems but also uncertainties related to the PV systems' components and location.

A related challenge is the accurate prediction of PV power production before installation. Knowing the expected power production throughout the system's lifetime is essential to support the deployment of solar energy, as accurate and optimized PV system designs decrease the investment risk.

Compared to other sources, solar energy has larger area requirements for equivalent electricity production. The discontinuity in generation and low capacity factor make it a very area-intensive technology. This fact is creating competition with accommodation and agriculture that could slow down the deployment of PV.

PV module datasheets indicate lifetimes of around 25 years. However, some modules experience failures before their end-of-life which can jeopardize the performance of the whole system. Operation and maintenance activities in solar farms excel at detecting these failures, but the same cannot be said for less-monitored residential systems. Moreover, they face an increased risk of failure due to frequent exposure to shading conditions, which can be detrimental to their operation.

1.5. AIM OF THIS WORK

Within this context, arises the Trust-PV project¹ aiming to increase the performance and reliability of PV systems. This four-year-long European collaboration formed by 20 organizations from the industry and research fields focuses on improvements across the entire solar PV value chain.

This work aims to contribute to the objectives of the Trust-PV project. Given the wide variety of the project goals, this dissertation covers as well a wide variety of topics. All fall under the theme of energy yield modeling for PV systems, but four groups can be clearly distinguished.

1.5.1. MACHINE LEARNING PREDICTIONS

As already mentioned, forecasting technologies can help grid operators with the scheduling and dispatching of PV electricity more effectively. However, achieving a precise solar power prediction can be extremely difficult considering not only the chaotic nature of weather systems but also uncertainties related to the PV systems' components and location [31]. Given these shortcomings, a new family of approaches is gaining popularity to obtain these predictions: machine learning (ML) algorithms. This dissertation shows the development and usage of diverse ML methodologies tailored to diverse solar power predictions.

1.5.2. PHYSICAL-BASED METHODOLOGIES

Despite the flexibility and ability to cope with weather- and system-related uncertainties, machine learning has two large drawbacks: the need for large amounts of data and the lack of interpretability. Therefore, when seeking for understanding of PV solar phenomena, researchers employ well-established analytical methods rooted in physical equations. Additionally, analytical methods can yield estimations before the deployment of PV systems for investment purposes. Under this context, the Photovoltaic materials and devices (PVMD) group at Delft University of Technology has been developing for several years an accurate and state-of-the-art toolbox able to model any PV system, even those that do not exist yet. This dissertation explains the functioning of this toolbox and shows its application for different scenarios.

1.5.3. FLOATING PV SYSTEMS

As previously mentioned, one of the current technical challenges of solar energy is its low efficiency-to-area ratio which demands extensive land usage. This issue creates competition with other resources such as agriculture or housing. For this reason, PV is getting expanded to water bodies. This promising expansion entails operational consequences that are not widely documented and this uncertainty increases the investment risk. To decrease

¹Grant Agreement 952957, <https://trust-pv.eu/>

this uncertainty, this dissertation explores the effect that wave movement has on the power performance of floating PV systems.

1.5.4. SHADING TOLERABILITY

PV modules are being installed everywhere, including areas that are not so suitable due to the high shading of obstacles nearby. This is especially a problem for residential systems. Shading can decrease the power production of PV modules and even lead to faults reducing the lifetime of the system. In this dissertation, the development of a metric to quantify the power losses of PV modules when subjected to different shading conditions is performed.

1.6. OUTLINE OF THIS WORK

Following the divisions from the previous section, this dissertation is divided into four blocks.

The first block focuses on machine learning techniques for PV power forecasting. The block starts by reviewing the latest developments in the literature on the topic in [Chapter 2](#). This review highlights some gaps that are covered in consequent chapters: the creation of a database with PV power open sources ([Chapter 3](#)); the effect of climate on the performance of ML models ([Chapter 4](#)), and the generalization of the ML models to a large fleet of PV systems ([Chapter 5](#)). The block finishes with [Chapter 6](#) by exploring an alternative method for PV power forecasting relying on neighboring systems.

The second block deals with PV power predictions but employs physical-based methods instead of ML ones. [Chapter 7](#) presents the PVMD toolbox and applies it to predict the power of residential systems, highlighting the importance of accurate irradiance input data. This importance is extended for commercial systems in [Chapter 8](#), where a software tool for the optimal allocation of irradiance sensors is developed.

The third block presents the developments for floating PV systems. [Chapter 9](#) provides an introduction to the topic and presents the energy yield modeling efforts for an inland floating PV system. [Chapter 10](#) moves offshore to estimate the effect of water movement on the power performance. This work is improved and extended in [Chapter 11](#) where general guidelines are provided to estimate the movement and mismatch losses due to sea waves.

The last block deals with the shading tolerability concept. The concept is explained in [Chapter 12](#), where a tool is developed to rapidly and accurately calculate the shading tolerability of any commercial PV module using only the data in its datasheet.

1.7. NOVEL CONTRIBUTIONS TO THE FIELD

The research developed in this work supports the development of the field of photovoltaics with the following contributions:

- An updated review of the literature on PV power forecasting using machine learning techniques.
- A website listing the open sources of PV power to promote data sharing between institutions.
- A methodology to estimate the power produced by residential PV systems without relying on weather data but on neighboring systems.
- A software solution to allocate irradiance sensors throughout a PV farm on a hilly terrain.

- An extensive evaluation of the effect of wave movement on power losses of a floating PV system and guidelines on how to reduce said losses.
- A method to calculate the shading tolerability of PV modules accurately and effectively.

MACHINE LEARNING PV POWER PREDICTIONS



2

TRENDS AND GAPS

A way to increase the reliability of PV power sources is to know their production at every instant of time. Due to the difficulty of this task, machine learning (ML) techniques have entered the world of PV power forecasting to increase the accuracy of predictions. Researchers have seen great potential in this approach, creating a vast literature on the topic. This chapter intends to identify the most popular approaches and the gaps in this discipline. To do so, a representative part of the literature consisting of 100 publications is classified based on different aspects such as ML family, location, features, etc. Via this classification, the main trends and gaps can be highlighted while offering advice to researchers interested in the topic.

Model	Data	Application	Analysis
<ul style="list-style-type: none">• ML algorithm• Pre-processing• (Post-processing)	<ul style="list-style-type: none">• Location/climate• No of systems• Timespan• Features	<ul style="list-style-type: none">• Forecast horizon	<ul style="list-style-type: none">• Metrics

This chapter has been adapted from the publication A. Alcañiz, D. Grzebyk, H. Ziar, O. Isabella “Trends and gaps in photovoltaic power forecasting with machine learning” *Energy Reports*, 9, 447-471 (2023) [32].

2.1. INTRODUCTION

The PV power forecasting techniques can be classified into three major groups [33]: persistence, physical, and statistical models. The latter group can be divided into time-series-based and machine learning (ML) techniques. Until the surge of ML techniques, none of the approaches rose to the forecasting challenge in terms of reliability, accuracy, and computational economy [33]. Persistence techniques are too simple to provide accurate results except for in the very short term. Analytical equations often fail to predict yield due to incomplete system information or insufficient spatial and temporal resolution of weather predictions [34]. Moreover, in the case of residential PV systems taking continuous measurements of all required parameters in situ is not a common practice due to high associated costs. Finally, the main drawback of statistical methods is their inability to adapt to abrupt changes in the meteorological conditions [35].

For these reasons, in recent years, the number of manuscripts that use ML techniques for PV power prediction has increased exponentially, as depicted in Figure 2.1. Considering this rapid development and the high amount of literature, it is hard to keep track of previous works performed and the recommendations to follow. For this reason, several review papers have already been published on the topic that sum up the progress done so far.

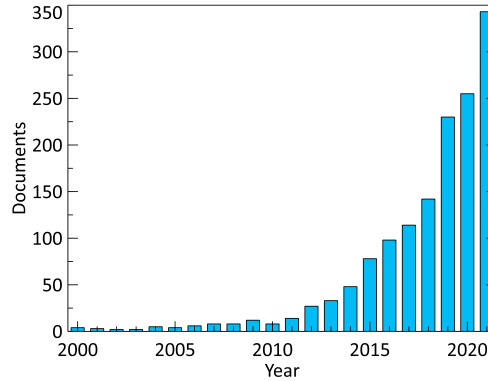


Figure 2.1: Number of documents published on PV power forecasting using ML from 2000 until 2021 as of mid-2022. The data was obtained via an advanced literature search including ML techniques and photovoltaic forecasting in Clarivate Web of Science [36].

Table 2.1 summarizes the reviews done on PV power forecasting where ML techniques have been discussed. Most publications focused on PV power forecasting, although a few also included irradiance estimation methods. A brief explanation of the focus of these reviews is provided so the readers can refer to those references in case the topic is of interest.

Since the first review in 2013, each review has focused on a different aspect of PV power forecasting, showing how ML methods have gained importance and providing a hint at how the main trends have evolved. However, what is lacking from our perspective is an extensive classification of the published literature to highlight what has and has not been done in this discipline. Overall, we want to distinguish from previous reviews in the following points:

- This work has been limited to scientific articles published from 2015 until 2020. By limiting the period, one can better identify the trends while covering a considerable percentage of the literature, instead of selecting the most cited manuscripts in a wider

Table 2.1: Previous reviews on PV power forecasting using ML techniques.

Study	Topic
Inman, 2013 [37]	Theoretical basis for most promising methods for solar forecasting. Large focus on irradiance prediction
Ulbricht, 2013 [38]	Classification of solar forecasting solutions
Ren, 2015 [39]	Wind and solar forecasting with ensemble methods
Antonanzas, 2016 [34]	ML and physical methods have the same importance. Focus on economics in forecasting and the difference between point and regional forecasts
Gandoman, 2016 [40]	PV power forecasting under cloudy weather
Raza, 2016 [41]	Special emphasis on ML algorithms such as ANN (Artificial Neural Network)
Barbieri, 2017 [42]	Very short-term PV power forecasting with cloud modelling
Das, 2018 [43]	Short-term direct PV power forecasting
Sobri, 2018 [44]	Most popular methods for PV power forecasting: advantages and limitations
van der Meer, 2018 [45]	Solar power and load probabilistic forecasting
Mosavi, 2019 [46]	ML applied to energy systems
Yang, 2019 [47]	Provide a set of reporting rules called <i>ROPES</i> to assist scientists in the evaluation of their methods and set guidelines for future research
Ahmed, 2020 [33]	Cloud imaging and online PV power forecasting. Methods for pre- and post-processing of data
Massaoudi, 2020 [48]	PV power forecasting using deep learning techniques
Mellit, 2020 [49]	Focus only on ML techniques
Pazikadin, 2020 [50]	ANNs for solar power generation forecasting
Rajagukguk, 2020 [51]	Deep learning algorithms for solar irradiance and power forecasting
Feng, 2021 [52]	Taxonomical review on the integration of PV using artificial intelligence
Gupta, 2021 [53]	Direct and indirect ML-based PV power forecasting

period. The literature was selected first by choosing recent and most cited manuscripts on the topic and then tracking back the relevant publications referenced in them.

- This work focuses specifically on 1) machine learning employed for 2) PV power forecasting. The considered literature employs only direct PV power forecasting, i.e., the objective is to forecast the PV power produced, not the irradiance [48].
- The 100 reviewed publications on PV power forecasting are classified into ML family, location, climate, number of systems, timespan, forecast horizon, and choice of features. [Appendix A](#) shows this classification for all documents. To the best of our knowledge, no previous review has performed such an extensive categorization, neither has classified by climate nor ML family.
- A section is devoted to discussing special methodologies employed for PV power forecasting, such as online forecasting and global optimization.

- Finally, recommendations are provided based on the extensive analysis. This section aims to be a reference for new researchers who want to explore the world of PV power forecasting and a sea of ideas for future developments for familiarized researchers.

The structure of the review is as follows. [Section 2.2](#) focuses on the ML algorithms by classifying them into families. [Section 2.3](#) explains the highlighted special approaches from the literature. Sections [2.4](#), [2.5](#), and [2.6](#) continue the classification based on system characteristics, forecast horizon, and employed features, respectively. Recommendations are given in [Section 2.7](#), before concluding in [Section 2.8](#). Overall, the aim is to encompass all the choices that a researcher needs to make when developing a PV power forecast model.

2.2. MACHINE LEARNING FAMILIES

This work starts by classifying the publications into ML families. Previous reviews mention only the most relevant algorithms, sometimes disregarding promising but not so extensively used approaches. To avoid that, the algorithms have been grouped by similarity, following the classification performed by Brownlee [54]. The distinctive characteristics of each family are explained in [Appendix B](#), while [Table 2.2](#) offers a graphical summary of it.

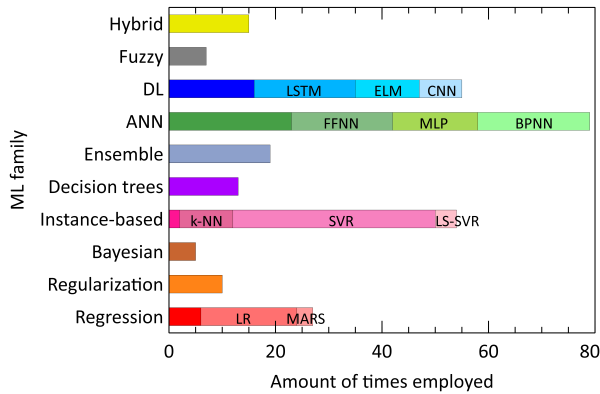
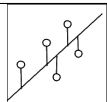
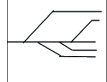
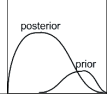

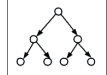
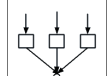
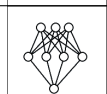
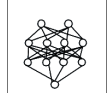
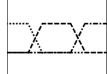
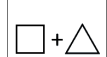


Figure 2.2: Amount of times that an ML algorithm from a certain family has been employed in the literature. Bar areas without a specific algorithm indicate algorithms from the ML family different from the most popular ones within it.

[Figure 2.2](#) illustrates the frequency of usage for ML algorithms within specific families. The most popular algorithms and their occurrence have been indicated within their corresponding family. Consistent with the literature [41], Artificial Neural Networks (ANN) emerge as the most popular algorithmic family in PV power forecasting, with Feed-Forward Neural Networks (FFNN), Multi-Layer Perceptron (MLP), and Back-Propagation Neural Network (BPNN) being the most popular algorithms within it. The popularity of ANNs lies in their structure which makes them very flexible. Nearly equal in second place are the instance-based and Deep Learning (DL) families. From the former, Support Vector Regression (SVR) is the proposed approach from several groups given its good performance [55–59], making it the most popular algorithm. Within DL, Long-Short Term Memory (LSTM) stands out with excellent results in general. Other DL algorithms whose popularity has increased considerably in recent years are Extreme Learning Machines (ELM) and Convolu-

Table 2.2: Machine learning families and their main characteristics. Figures from [54].

	Family	Main characteristics
	Regression	Starting point in machine learning. Composed of linear regression and its improvements.
	Regularization	Improvement of regression methods. Prevent the coefficients from reaching high values, focusing too much on training data (overfitting).
	Bayesian	Based on Bayes' theorem: the probability distribution of the output depends on that of the input.
	Instance-based	Use similarity between the data: build a database and compare new data to it to make the prediction.
	Decision trees	Tree-like model of decisions: data space is partitioned, and a prediction model is fitted within each partition
	Ensemble	Combine weak algorithms (e.g. decision trees) into a strong one.
	Artificial Neural Networks	Sets of nodes (neurons) interconnected via weights, such that signals can travel through them. Very flexible and able to model nonlinear systems.
	Deep Learning (DL)	ANN with complex relations between neurons increasing the problem-solving capacity. Characterized by having more than one hidden layer.
	Fuzzy	Algorithms where binary logic has been substituted by a fuzzy one. Higher flexibility than the original algorithms.
	Hybrid	Combine algorithms from two different groups. Increased capacity for problem-solving.

tional Neural Networks (CNN). The latter is the recommended option when dealing with image data [60].

The next [Subsection 2.2.1](#) dives into each family to highlight the advantages and drawbacks that can explain the reported trends from the previous figure. These arguments will be supported in [Subsection 2.2.2](#), where the findings of selected literature that compare several algorithms are explained. Finally, [Subsection 2.2.3](#) addresses a different application of ML algorithms: data pre-processing.

2.2.1. KEY FEATURES AND TRENDS

This section provides a brief overview of each family by explaining their advantages and drawbacks. Moreover, relying on the authors' impressions, opinions are presented on how popularity trends will evolve. A description of each family's characteristics can be found in [Appendix B](#).

Starting with the regression family, linear regression (LR) is the most common algorithm. Although it does not show the best performance [61–65], it is valued by its simple implementation and ease of interpretation [66, 67]. None of the groups proposed it to forecast the PV power, but it is employed as a benchmark [66, 68–70]. Its main limitation is its approach of constructing a hyperplane to make the predictions, which is a significant oversimplification.

A recently applied regression algorithm that has shown excellent results is Multivariate Adaptive Regression Splines (MARS). This algorithm improves the simple linear regression by including hinge functions (see [Figure 2.3](#)), which allow for more flexibility to fit the data while keeping the interpretability [71]. MARS is the algorithm proposed by some groups [61, 64, 72], and due to its interpretability, we expect that it will play an important role in the future. Currently, interpretability may only seem relevant when discussing high-stakes predictions that deeply impact human lives [73], such as stating that highly polluted air is safe to breathe [74]. However, it can gain importance after an incorrect ML prediction in a grid with high renewable energy penetration.

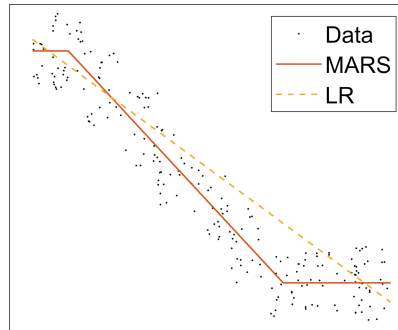


Figure 2.3: Graphical difference between MARS and LR to depict the higher flexibility of the former.

Regularization algorithms aim at solving the overfitting problem of regression ones. Overfitting occurs when the hyperplane is too specific and fits only the training data, being unable to represent unseen data. Regularization solves this issue by ensuring that the hyperplane remains “simple enough”. However, regularization algorithms still fit a hyperplane, which is an oversimplification. Moreover, while overfitting is a concern with insufficient training data, this problem is disappearing in the present era of big data. This family was probably a good actor a decade ago, but it has been displaced.

A similar situation occurs with Bayesian algorithms. Given the ML origins, fitting a probabilistic distribution function to the data was worth exploring. This approach works well on small datasets and it can provide uncertainty measurements [75]. However, some of the probabilistic distributions assume independence between features, which is an oversimplification.

The two main members in the instance-based family are k-nearest neighbour (k-NN)

and SVR. The popularity of k-NN lies in its user-friendliness and adaptability. It is the most popular classification algorithm and identifies clusters of data based on the nearest neighbors. For forecast applications, the algorithm identifies in the training dataset the cluster that resembles the most to each test condition and creates a forecast value based on that. Identified disadvantages are the need to store the whole training set and longer prediction times.

In the case of SVR, and Least Squares-SVR (LS-SVR) which is a simplification in the learning procedure that increases efficiency without losing accuracy, there are facing opinions in the literature. Some defend it as the algorithm of choice [55, 76–78], while others report similar or worse results compared to other algorithms [60, 61, 79–81]. The reason for this disagreement, according to the authors' opinion, is the difficulty of optimization of SVR. SVR has a wide range of parameters, and the performance of the algorithm is highly sensitive to them [82], so results are jeopardized if the optimum parameters are not found. Moreover, like k-NN, SVR also relies on similarity to the training samples, so it is not suitable for large datasets [79], which may be limiting in the future. However, with the help of LS-SVR and the popularity seen so far, we anticipate that SVR will remain relevant for several more years.

The working principle of decision trees is promising due to its flexibility and interpretability, but it presents a huge disadvantage: overfitting. While overfitting solutions did not improve the main limitation of regression algorithms, there is an optimum amendment for decision trees: ensemble algorithms. Therefore, in our opinion decision trees will be every time more irrelevant, and they will only be indirectly used through ensemble algorithms.

Ensemble algorithms sacrifice interpretability compared to decision trees, but, as opposed to SVR, the difference in performance between optimized and unoptimized parameters is relatively small. This family has a high appeal for new researchers wanting to use ML algorithms. We think that ensemble algorithms, together with SVR and MARS, will be the only algorithms able to compete with the hegemony of ANNs and DL that is currently present.

ANNs are the main characters in ML. They are powerful and capable of modeling any non-linearity relation [83]. However, they have the same implementation drawback as SVR: the optimal configuration and learning algorithm are hard to obtain [61]. This may explain the lower performance achieved by some researchers compared to other algorithms [35, 61, 84, 85].

In the explored literature, most researchers focus on and propose an algorithm of the DL family. This fact is not visible in Figure 2.2, since it includes all the algorithms applied in each publication. These multi-layer ANNs have the benefit of including the impact of historical trends, which increases the accuracy of the predictions. The ANNs' drawback of optimal configuration is still present, and interpretability is challenging. However, efforts are being made to address these concerns. We expect that the popularity of this family has only just started and will further grow in the upcoming years.

Two algorithms use fuzzy logic in the explored literature: fuzzy k-means clustering and Adaptive Neuro Fuzzy Inference System. In the first, fuzzy logic is integrated into k-NN, while in the second, it is integrated with ANN. While some researchers may not have differentiated these algorithms, we aimed to highlight them. The flexibility introduced by fuzzy logic can be very powerful, but it may not be advantageous enough. Due to its lack of interpretability, they compete with DL approaches, which are more popular and have demon-

strated accuracy. Hence we think that, despite being promising, fuzzy algorithms will not be relevant in the future.

The final step in complexity is the hybrid family. Most algorithms focus on increasing the accuracy by combining the abilities of two or more ML algorithms. Table 2.3 provides an overview of the combination structures from the reported literature. Ensembles are being combined with strong algorithms, such as ANN, to overcome overfitting and increase accuracy. More examples of hybrid DL methods for PV power forecasting can be found in the review by Massaoudi et al. [48]. The main drawback of this family is the higher complexity and consequent decrease in interpretability. We anticipate ongoing development within this family as researchers combine ML algorithms to overcome drawbacks.

Table 2.3: Hybrid algorithms employed in the literature.

Combination	Objective	Ref.
LSTM + CNN	Combine the temporal ability of LSTM with the feature extraction abilities of CNN	[65, 86, 87]
DL + LSTM/DL	Combine the feature learning of the DL algorithm (an AutoEncoder) with the forecasting ability of LSTM/DL algorithm	[88]
DL + LSTM	Combine the feature learning of a DL (AutoEncoder) with the temporal ability of LSTM	[89]
CNN + SVR	Combine the feature extraction ability of CNN with the forecasting ability of SVR	[90]
DL + LSTM	Increase the accuracy of the forecast	[91, 92]
LS-SVR + ANN	Increase the accuracy of the forecast	[93]
Ensemble + SVR	Combine several SVR to increase the accuracy	[57]
Ensemble + ANN	Create a probabilistic forecast	[94]
Ensemble of LSTM, SVR, ANN, LR, decision tree	Increase the accuracy of the forecast	[95]
Ensemble of ANN, SVR, ELM	Increase the accuracy of the forecast	[96]
Ensemble of instance-based, ANN	Increase the accuracy of the forecast	[97]

2.2.2. ALGORITHMS COMPARISON

To compare the performance of different ML algorithms, they need to be fairly analyzed. The prediction results depend not only on the algorithm employed but also on the amount of data, the solving method, the location, the metric, etc. Therefore, this subsection presents the most relevant studies which have compared the performance of several ML models under the same conditions. Some metrics required for this task are defined in Appendix C.

Three works comparing a broad range of algorithms deserve special attention. Ferlito et al. forecasted the power of a 1 kW grid-connected PV system with 11 algorithms of different complexity and belonging to the regression, instance-based, decision trees, neural networks, and ensemble families [64]. They reported that the best-performing algorithm and optimum dataset length depended on the training and testing years. For example, the

Cubist algorithm (ensemble) showed the best performance in a highly variable year such as 2010 with Root Mean Squared Error (RMSE) of around 93 W, while Random Forest (ensemble) had lower prediction errors than other methods in 2011. They reported that non-linear models showed up to 2% superior performance compared to linear ones. Chen and Koprinska implemented 7 base learners from different families and combined their outputs through several ensemble strategies [95]. One of the ensembles was identified as the best model, while LSTM was the most accurate base learner, showing competitive results. Finally, Carrera and Kim trained 15 ML algorithms to predict the power of a South Korean solar farm [81]. One can observe the reported RMSE for all algorithms in Figure 2.4. In the graph, the red lines represent the medians while the green triangles indicate the means. Ensemble algorithms were the best-performing ones, except in terms of bias when Elastic Net (regularization family) outperformed the rest.

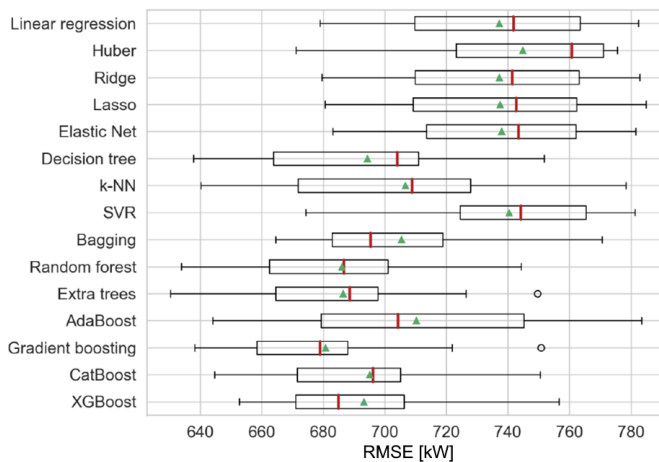


Figure 2.4: Forecasting RMSE of several ML methods for a South Korean solar farm. Figure adapted from [81].

The study performed by Sharadga et al. compares different ANN models [84]. The ranking between models depended not only on the type of ANN but also on the training algorithm, the forecast horizon, and the metric. However, recurrent neural networks generally outperformed feed-forward ones. Li et al. reported that the algorithm with the lowest RMSE was MARS [61]. Although the ranking of these models differed depending on the metric employed, MARS was in the top three position independently on the metric employed and outperformed more complex models such as SVR or ANN in certain metrics.

In [98], several ANN, SVR, and their hybridized forms were compared. In general, the hybrid models outperformed their simpler versions. Without hybridization, RMSE and Mean Absolute Percentage Error (MAPE) showed that BPNN was better than other ANNs. When hybridized, both metrics agreed that the hybrid ANN was the best and hybrid SVR was the worst. Finally, [99] reported that the modified LS-SVR showed the best performance, outperforming ANNs.

All these studies demonstrate the challenge of identifying a superior algorithm or even an entire family. Developing a set of guidelines for researchers would be beneficial in aiding this decision-making process. The authors recommend ensemble algorithms to new researchers due to their favorable accuracy-to-interpretation ratio. They are also a good op-

tion when dealing with a limited amount of data. If the amount of data is high enough and the researcher is experienced, algorithms such as SVR, LSTM, and ELM are recommended. Yet, caution is crucial when optimizing the parameters.

2

2.2.3. DATA PREPARATION

This subsection describes ML-based pre-processing strategies that reduce the computational cost or complexity of the problem without decreasing the accuracy. Post-processing techniques such as reforecasting [100] or corrections [101] can also be employed to improve performance. However, due to its limited use in the explored literature, it was not possible to review them, which can be considered a gap that still needs to be explored. Regardless, it is important to point out that these strategies may demand significant time commitments.

The most common pre-processing step is to reduce the number of features employed, retaining only the relevant ones. These methods are referred to as feature selection and are explained in [Subsection 2.2.3](#). An alternative way to reduce the input data is to group by similarity, explained in [Subsection 2.2.3](#). A third strategy is to decompose the input data, explained in [Figure 2.2.3](#). Pre-processing strategies are quite common, especially when dealing with large amounts of data, a trend expected to grow. Therefore, we anticipate the continuing popularity of these techniques.

FEATURE SELECTION

Feature selection or dimensionality reduction consists of keeping only the most relevant features for an ML model [102]. With this, the performance of the model is increased, the computational cost is reduced, and a better understanding of the underlying process is gained. Feature selection can be done by manual inspection [103–105] or with correlation-based strategies [55, 79, 106, 107]. However, it is more effective to employ more complex strategies.

One of the most widespread methods for dimensionality reduction is Principal Component Analysis, which was applied in [69, 108]. This technique reduces the dimension of the dataset by creating linear combinations of features that have maximal variance and are mutually uncorrelated. An alternative is to employ ML models used for prediction such as Elastic Net (regularization family) [109], Gradient Boost Regression Trees (ensemble) [109, 110], or SVR [111] to perform feature selection. These ML models assign a weight to each feature and features with weights smaller than a certain threshold are removed. The threshold is chosen by the user and depends on the algorithm. For instance, in each split of a gradient boost regression tree, the chosen feature to split on is the one that maximizes the reduction of a certain kind of error [112] (refer to [Appendix B](#) for details on how ensemble algorithms work). By the end of the tree, the divisions are less relevant. Thus, by cutting trees below a certain split, one can identify the most important features.

These techniques can help boost the performance considerably. In [113], feature selection improved the RMSE of the proposed model from an average of 5.7% to 4.6%. Similarly, in [110] the applied feature selection method increased the performance of k-NN by 2%. Alfadda et al. showed how the error evolves as features are incrementally included in the model [58]. They reported that when all features were considered, the RMSE was higher than when a few were left out. Lee et al. observed an increase in performance after removing redundant variables in their model [111]: the average hourly MAPE went from 68.2% to 107.1%.

Overall, the reader is strongly advised to employ a feature selection technique when dealing with a high number of input parameters. Apart from the already mentioned benefits, feature selection can highlight features with high predictive power that are superfluous analytically. However, one also has to be critical of the results, since the selected features could denote errors in the input data. It is recommended to make use of solar engineering knowledge and previous literature to validate the results obtained (refer to [Section 2.6](#) for more information). In case an unexpected feature is selected, one should explore the input data and make an informed decision.

CLUSTERING

When the amount of input data is large, the algorithms become computationally expensive. This data is often repeated and not all of it is needed. For instance, similar weather conditions yield similar PV power output. The data can be grouped with clustering algorithms such as k-NN by meteorological characteristics or location. Since cloudy days are harder to forecast than sunny ones [64, 84, 93, 103, 114], some researchers have developed a different model per type of weather [98, 105, 115, 116] or season [101]. Spatial clustering of PV plants is commonly applied to forecast distributed generation [108, 114], see [Figure 2.5](#).

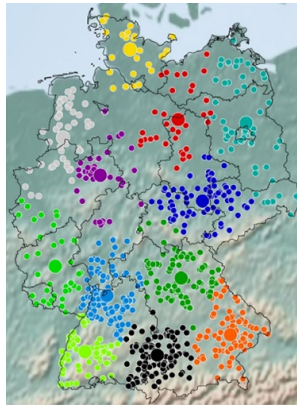


Figure 2.5: k-NN applied to several PV systems spread over Germany. Each color represents a different cluster and the small dots represent the individual PV plants. Figure from [114].

TIME-SERIES DECOMPOSITION

Another strategy to simplify the learning process of ML algorithms is to decompose the input signal into several frequency series so that each series has better outlines and behaviors. The most commonly employed algorithm to achieve this is Wavelet Decomposition, applied in [62, 117–121]. In [113], the effect of decomposing the time series was a reduction of the RMSE from 5.9% to 4.6%. More information on this topic can be found in the review of Ahmed et al. [33].

2.3. SPECIAL APPROACHES

This section introduces variations that several groups have made to the standard forecasting methodology. These novel methodologies are gaining popularity, therefore here the advantages and possible difficulties that each one may present are discussed.

PROBABILISTIC FORECASTS

Probabilistic forecasts can introduce flexibility in the predictions which can be a large advantage for grid operators. There are three main methods to generate a probabilistic forecast: ensemble algorithms, uncertainty estimation, and quantile regression [122]. Ensemble algorithms, which generate probabilistic intervals intrinsically, are used in [94, 122–125]. They are generally computationally intensive, affecting their practical implementation. The second method modifies point forecasting algorithms to provide an estimation of the error distribution, using parametric [108, 122] or non-parametric methods [77, 104]. With this methodology, prediction intervals are usually estimated via a normal distribution, which is not reliable for all values [108]. Finally, examples of probabilistic quantile regression forecasts can be found in [122, 123]. Quantile regression is a regression algorithm where each quantile is separately modeled and a different prediction is fitted within each partition. This method does not report as good results as the other two [122].

Amongst these studies, [77] deserves special attention as the interval forecasts are in 2D, providing not only a range of expected values in power but also in time. A graphical explanation of the difference between point (non-probabilistic), interval (probabilistic in power), and 2D-interval (probabilistic in power and time) forecasts is shown in Figure 2.6. This method can provide even more flexibility to the forecast, which can be especially useful for risk management applications.

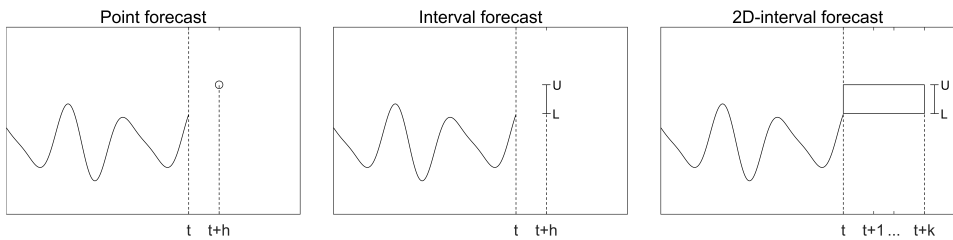


Figure 2.6: Graphical explanation of the difference between point, interval, and 2-D forecasts. Figure redrawn from [77].

More information on probabilistic forecasting can be found in the reviews [45, 122].

META-HEURISTIC ALGORITHMS

As mentioned earlier, some algorithms like SVR or ANN are hard to optimize, which can lead to decreased performance. To overcome this disadvantage, several groups have employed global optimizers to find the algorithm's parameters. Amongst these techniques, meta-heuristic algorithms deserve special attention. These are ML global optimization algorithms, suitable for non-linear parameters. The most popular optimization algorithms are Genetic Algorithm (GA), inspired by the theory of evolution [63, 99, 100, 120, 126]; and Particle Swarm Optimization (PSO), inspired by bird flocks' motion [63, 113, 120, 127]. In the reviewed literature, one can also find the Shuffled Frog Leaping Algorithm [126], Ant Colony Optimization [78], and Multiverse Optimization [127].

The use of these algorithms can considerably improve the performance. Eseye et al. reduced the Normalized Mean Absolute Error (NMAE) of SVR from 0.8% to 0.7% with GA and to 0.5% with PSO [120]. This increase in performance comes again at a cost of simplicity, but we believe that global optimizers can help minimize errors when employing complex algorithms.

More information on this topic can be found in [33].

ONLINE FORECASTING

The last special forecasting approach is online forecasting. In online training mode, the algorithm updates its weights continuously as it receives new data [128]. The training dataset is minimal and the algorithm can rapidly adapt to major changes in the environment. The main drawback is the continuous need for training.

This method was applied in [64] with several ML algorithms. The RMSE of offline training was higher than that of online, and the ranking of the algorithms depended on the training method. Al-Dahidi et al. applied the online training mode to a probabilistic very short-term model for real-time applications [129]. In [80], the model could selectively accumulate or forget knowledge to respond to climate variations. Similarly, the models developed in [130, 131] were proposed for Internet-of-Things systems [132].

More information on online PV power forecasting can be found in [33].

2.4. SYSTEM CHARACTERISTICS

This section focuses on the main characteristics of the PV systems employed, namely location (Subsection 2.4.1), number of systems (Subsection 2.4.2) and timespan (Subsection 2.4.3).

2.4.1. LOCATION

Figure 2.7 represents the number of publications whose studied systems are located in each country. Most of them are in Europe and no PV system has been forecasted in South America. Given the solar energy's intensity across the globe [33], areas such as the Middle East, most of Australia, and the deserts are suitable for large-scale PV installations. However, most of the reviewed literature focuses on areas with relatively small solar intensity (except for Australia).

Characterizing the PV systems' location by country is not representative of the weather conditions that they are subjected to. Therefore, the reviewed papers have been also classified by the type of climate that their systems are subjected to, as reported in Figure 2.8. The Köppen climate classification has been employed for this purpose since it divides the globe with a single metric into five categories: tropical, dry, mild temperate, continental, and polar [133].

Around half of the forecasted systems from the reviewed publications are in mild temperate climates. Roughly 25% of the systems are in dry climates, while only 8 are located in tropical and 11 in snow climates. No system is in polar climate since mostly the Arctic, Antarctica, and tops of mountains are representative of this climate.

Some studies have been left out from one or both figures since they consider more than one system under different climates or countries. Do et al. forecasted two PV systems, one in a tropical climate and the other in a mild temperate one [67]. Due to the seasonal variation of the mild temperate climate, the system there requires a longer training duration. However, once the model is trained, the algorithm shows higher performance in the temperate climate than in the tropical one. This difference in ease of training was also reported in [107, 121]. Zhang et al. forecasted the PV power of systems in the USA, Denmark, and Italy [134]. Although the three systems are in a mild temperate climate, the diverse daily weather distribution influenced the modeling parameters and features that yielded the best forecasting result. Similar conclusions were reported in [130], and in [65, 92, 95, 97] when the

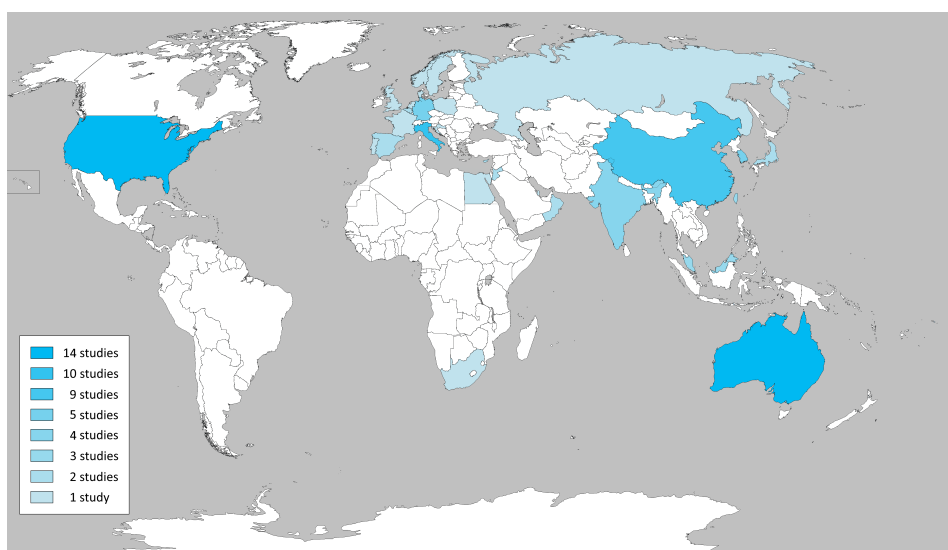


Figure 2.7: World map with the countries highlighted depending on the number of studies published on the topic. Only the mainland has been colored. Since some countries' contribution might not be visible on the map, the raw data is: 14 studies: Australia, USA; 10 studies: Italy; 9 studies: China; 5 studies: Germany, South Korea; 4 studies: India; 3 studies: Malaysia, Taiwan; 2 studies: Belgium, Cyprus, Japan, Jordan, Oman, Spain; 1 study: Croatia, Denmark, Egypt, France, Netherlands, Norway, Poland, Portugal, Qatar, Russia, Singapore, South Africa, Sweden, UK.

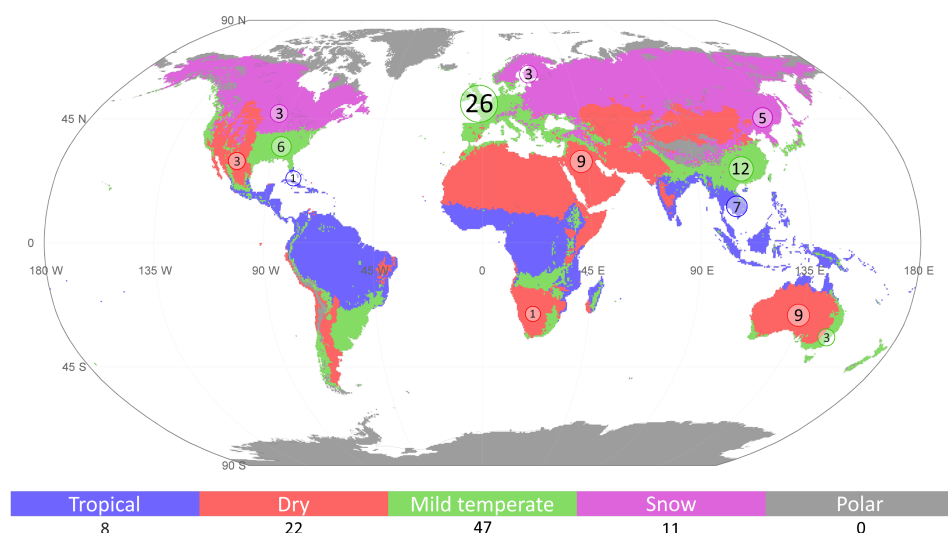


Figure 2.8: World map indicating the major Köppen territories in different colours [133] and the number of systems in each region.

systems were located in the same country. In [80] these climatic differences were considered, so an adaptive learning algorithm was implemented for two different systems, able to

adjust to different climates and periods.

The performance of a PV system is conditioned by the type of climate. Depending on its location, certain features are more important than others and the optimum parameters of the model are affected. Therefore, the high bias for systems in mild temperate climates entails a problem of generalization. If most of the forecasted systems are in mild temperate climates, the model will be limited to those meteorological conditions.

2.4.2. NUMBER OF SYSTEMS

The next classification considers the number of systems used in each publication, [Figure 2.9](#). Most studies used data from only one system, while 10% considered more than 10 systems to forecast the PV power.

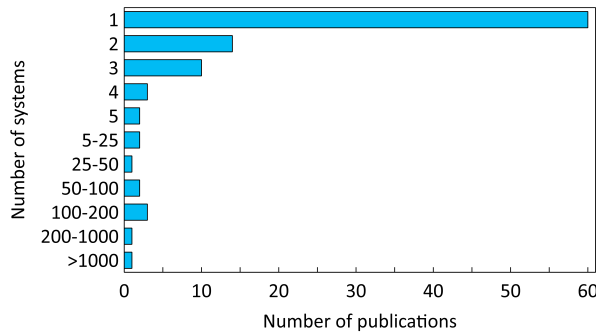


Figure 2.9: Number of publications as a function of the number of systems considered in each one.

When more than 5 PV systems were employed, their PV power production was generally aggregated to perform a regional forecast [76, 99, 131, 135]. In [76], the systems were differentiated between the North and South of Italy before being aggregated. The systems in the South of Italy showed better performance than those in the North, which was attributed to lower weather variability in the southern part of the country.

Wolff et al. grouped 921 PV systems depending on their location and then divided the clusters into sub-clusters depending on the system specifications [114]. In the end, 10% of each sub-cluster was randomly selected, resulting in a test set of 92 PV systems. The predictions obtained from these single PV systems were employed to produce a regional forecast. When considering the difference between the single site and regional forecasts, the errors in the latter were lower due to averaging effects. Spatial clustering was also applied in the 1985 PV systems in [108] to test two different averaging approaches. In the first approach, the power generation was calculated for each cluster and then averaged to obtain the regional prediction. In the second approach, the regional prediction was directly obtained by inputting the data from each cluster centroid. The second model gave slightly better results.

Exceptions to the regional approach were the forecasts performed for 71 systems in [65] and for 21 systems in [89]. In the latter, each facility was considered individually and the forecasting error was higher or lower depending on the target system. Unfortunately, the origins of this unpredictability were not reported. Another exception is the study by Jung et al., whose objective was to evaluate the most suitable locations for PV plants in a wide area considering terrain and weather conditions [92]. They concluded that “the data from

one site might not adequately capture the circumstances at other sites with different geographical and topological features”.

[125] has been left out of Figure 2.9 as this study did not employ any specific PV system, but rather installed PV capacity. Instead, the hourly PV power was estimated for several prefectures in Japan by making use of the monthly installed PV capacities for each type of supplier and weather data.

2.4.3. TIMESPAN

This section classifies the systems by timespan (see Figure 2.10), defined as the period of input data employed to develop a model. Most studies employ a timespan of one year. However, a long timespan is essential for accurate training, since the algorithm needs enough data to identify the seasonal variation throughout the year. This is especially important for places having mild temperature climates [67] where most of the systems are located. It is hence recommended to employ more than one year of data so that the training data covers all seasons. Moreover, a high amount of data is generally required to train a complex algorithm [136].

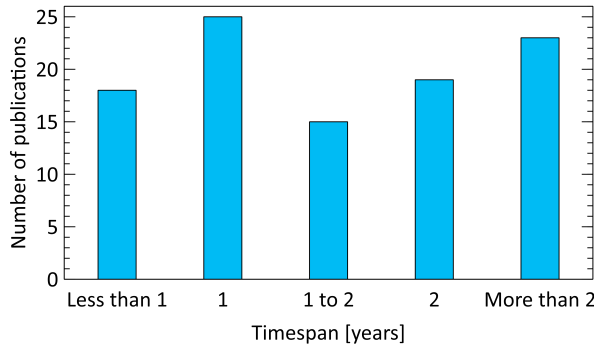


Figure 2.10: Number of publications as a function of the timespan employed for PV power predictions.

The main limitation of this requirement is the high computational power required. This worsens as the time resolution of the input data increases. Training algorithms with such a large amount of data can be unfeasible unless reinforcement or online learning is considered. Alternative workarounds followed in literature are selecting the first period of each month as training data [117, 137], or random days along the period [68, 85]. These alternatives were also employed in [114, 138] to efficiently generalize the developed ML model to unseen systems.

2.5. FORECAST HORIZON

The forecast horizon is the time interval into the future for which PV power is predicted. It depends on the requirements of the decision-making process for which the PV power is forecasted [31]. The forecast horizon affects the performance and choice of the prediction algorithm [139].

There is no standard classification on how to divide the forecast horizons [33, 41, 42, 62]. In this review, four categories are defined: long-term for predictions longer than a week;

day-ahead when forecasting between one week and 5 hours; hour-ahead for predictions between 15 minutes and 5 hours; and nowcasting for forecasts smaller than 15 minutes. These horizons can be related to energy markets: auction (day-ahead), intraday (hour-ahead), and balancing (nowcasting), as presented in Table 2.4.

Table 2.4: Classification of the forecast horizons employed in this review and their most common application.

Category	Time horizon	Application
Long-term	More than one week	Maintenance scheduling Pilot installations Bankability studies
Day-ahead	5 hours to one week	Economic dispatch Day-ahead price determination
Hour-ahead Intra-day	15 minutes to 5 hours	Unplanned outages compensation Unit commitment
Nowcasting	Less than 15 minutes	Power balance Electricity market clearing Early anomaly detection

Following this criterion, the analyzed literature has been classified as shown in Figure 2.11. A fifth group has been included for studies that forecast the PV power in time horizons belonging to more than one category. Most publications focus on day-ahead forecasting, while only a few belong to the nowcasting and long-term categories. Given that the forecast depends on the application, the authors do not recommend any specific category.

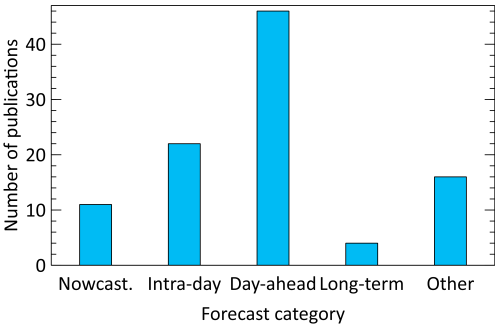


Figure 2.11: Number of publications as a function of the forecast category that they belong to. A new group has been included for publications belonging to more than one category.

Some researchers studied the behavior of ML models under different time horizons. Generally, they found that the higher the time horizon, the higher the error [120, 137, 140–142]. For instance, Giorgi et al. found that LS-SVR achieved an NMAE of around 6% when predicting the PV power 1 hour ahead, and the value increased to around 21% when predicting 12 hours ahead [93]. They stated that when predicting longer times, the PV power peaks were underestimated, while the valleys were overestimated. Similarly, in [108] an RMSE of 5-7% was obtained for the intra-day forecast (1 to 4 hours) while this value increased to 7% and 7.5% for one and two-day forecasts, respectively. One exception was found in [72],

where the day-ahead forecasting had a higher accuracy than when estimating 15 minutes, 1 hour, and 3 hours ahead. They claimed this was obtained for the day-ahead prediction because of compensation along the day.

The forecast horizon also affects the ranking of algorithms. In [84] the performance of several algorithms was tested under one-, two- and three-hour ahead forecasting for a 20 MW grid-connected PV station in China. For one- and two-hour ahead, LSTM showed the best performance, but ANN was superior for three-hour ahead forecasting. Similar results were obtained in [114].

2.6. CHOICE OF FEATURES

This section explores the inputs employed for PV power forecasting, called features in a machine learning model. Figure 2.12 depicts the most common features as a function of the number of publications in which they have been employed. The irradiance bar includes several forms such as plane of array or diffuse irradiance.

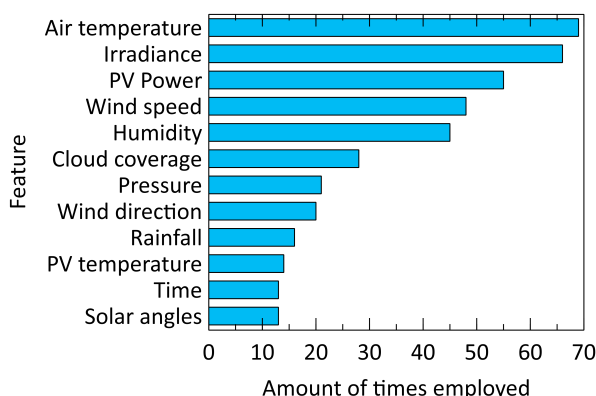


Figure 2.12: Number of times that each feature has been employed for predictions in the publications reviewed.

As expected from a physics-based approach, irradiance and temperature are the most important features. However, none of these two features is essential for ML predictions. Several works rely only on previous PV power outputs [84, 116, 126, 141, 143], which are more flexible to apply in practice. However, the use of previous PV power measurements in combination with weather data can considerably increase the performance of the algorithms [93]. It is also important to consider that some of the ML algorithms consider previous PV power values intrinsically (such as LSTM).

One surprising fact from a solar engineering perspective is the popularity of humidity. According to Figure 2.12, this parameter is more popular than cloud coverage, which is usually essential for accurate analytical predictions. Researchers in [127] found that humidity is negatively correlated with PV power. An increase in moisture implies an increase in the absorption, reflection, and refraction of sunlight, hence it reduces the radiation received by PV panels.

It is commonly presumed that a greater number of features would yield better results. However, as mentioned in Subsection 2.2.3, extra variables can be detrimental to the model's performance and can increase the demand for computational power. The choice of features

depends on the feature selection algorithm employed, on location, and even on specific events [55, 66, 109]. For instance, in [69], wind speed was not as relevant in November as in the rest of the months due to exceptional hurricane-force winds. On the importance of location, Touati et al. predicted the PV power output of a system located in Qatar, a desert-like climate [66]. The feature selection algorithm results highlighted dust as one of the most relevant features, a factor not considered in any of the other studies reviewed.

To conclude, there is no optimum set of features for all cases. When developing a new model, the reader is recommended to gather as many relevant features as possible and check previous literature for guidance. Figure 2.12 can help with the latter. Despite some models being able to identify relevant features, it is still recommended to do at least feature exploration. Less relevant features may offer information indirectly but should only be employed when more pertinent ones are unavailable.

This section does not discuss the weather data origin. The input data can come from different sources such as numerical weather predictions or neighbouring PV systems. The reader is referred to [34, 52] if more information on the topic is yearned for.

USE OF 2D FEATURES

Special attention is required in this section concerning the use of 2D features. A source of uncertainty in PV power predictions is cloud coverage information. Rapidly changing cloud coverage has a significant influence on PV module output and is a big challenge in solar yield forecasting.

To overcome this, some groups have centered their efforts on employing satellite images to make an accurate estimation of cloud movement [103, 108, 114]. This approach, which is extensively used for irradiance forecasting, is now gaining popularity for the prediction of PV power. However, satellite images have either high temporal or high spatial resolution but never both [144], while both characteristics are required for PV power prediction, especially in the short-term [50].

An alternative to satellite images is ground-based sky images, which are gaining popularity. These images can be employed for comprehending cloud movement [85, 145] or detecting the cloud type [146]. Chu et al. uniquely used these images to re-forecast the power prediction [100]. The power predicted by an ANN was integrated with sky images to improve the forecast. The main disadvantage of sky images is that they are local and, hence cannot be employed for wide areas.

More information on the use of satellite and sky images can be found in [33, 52].

2.7. OUTLOOK

This section presents the main recommendations given to the readers based on the discussion performed in previous sections.

MACHINE LEARNING ALGORITHMS

Starting with the ML algorithms, Section 2.2 provides a wide range of families to choose from. Some algorithms are too simple to properly forecast the PV power output, but most of them can yield reasonable results. Indeed, linear regression approaches are unable to correctly model the non-linearities of PV power forecasting, but once this step is overcome, the difference in performance between MARS and a DL hybrid algorithm is not that high

[61, 64]. In these cases, developing a model that would be applied in practice is more relevant. This can imply that the algorithm is interpretable, that the requirements in terms of training data are not high, that the model is not computationally expensive, and that it is flexible enough to be adapted to unseen PV systems with different characteristics and climates.

GENERALIZATION

Regarding the last point, most ML methods are usually tested on a few systems under the same climate so their capacity to adapt to unseen PV plants is unknown. This can be an issue when these models are put into practice.

Starting with the number of PV systems (Subsection 2.4.2) most researchers motivate their study by contributing to improved generator dispatch, power quality effects mitigation, and reducing secondary reserve capacity [147], but if their results are not validated for multiple systems, they are not reliable enough for upscaling. This hinders the practical use of the developed methods since they are not fully tested or verified under different conditions. Moreover, there is a need for more regional forecast studies, which will provide the basis for grid management.

Similarly occurs with the PV systems location (Subsection 2.4.1). Most studies focus on systems installed in the same climate and country. A study is lacking that analyzes the effect of climate on the performance of PV systems and quantifies the added error due to generalization. Large-scale data acquisition poses a significant cost barrier to achieving this goal. Lack of coordination between researchers and the industry might result in an unnecessarily long search for optimal solutions.

DATA ACQUISITION

Another issue related to data acquisition is its availability and quality. There may be enough data, but not with the desired resolution. This affects, for instance, the choice of the forecast horizon. Nowcasting (as short as 10 seconds) is currently the most demanded forecast horizon because of cloud-related instability [34], but the amount of data available with the required time resolution is limited. This explains the small percentage of studies focusing on this forecast horizon (Figure 2.11). Similarly, the lack of sky images and high-resolution cloud coverage information is a major concern.

Data quality is another important factor, even more than the chosen algorithm. Incorrect data will negatively affect the algorithm, without consideration of its complexity [148]. This is also the main source of error in comprehensive analytical prediction models. Unless the target of the forecasting model is a specific PV plant with a meteorological station nearby, we recommend avoiding local weather variables such as wind speed and direction. These lowly correlated parameters [149] can only add noise if they are not measured on-site. Other highly correlated features with lower spatial sensitivity such as temperature will not be so detrimental to the model.

Additionally, up to two hours ahead, the previous PV power is more important than numerical weather predictions [55]. Although we recommend PV power as a feature, with only that the accuracy of the predictions is very limited, especially during cloudy days. From our point of view, the best approach is to complement previous PV power and clear-sky irradiance with cloud movement extracted from images. Unfortunately, the use of imaging is currently limited by the low geographical resolution of satellites and the small coverage and high price of ground-based sky images.

METRICS

Due to the extensive literature covering metrics for PV power forecasting [47, 139, 150, 151] this is not a focus of this work. However, we still want to provide our insights based on Figure 2.13, which represents the popularity of the main metrics. Appendix C defines the main metrics for completeness.

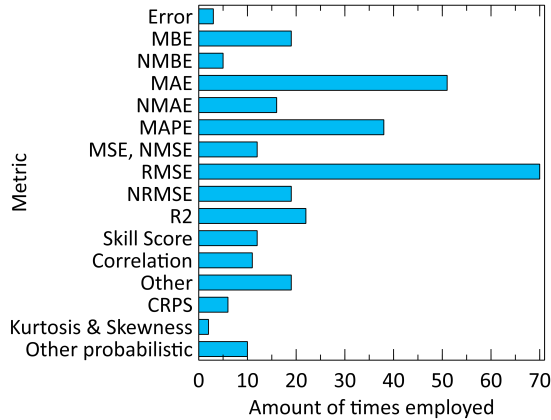


Figure 2.13: Metrics as a function of the number of times employed in the literature explored in this review.

RMSE is by far the most popular measure, commonly preferred in many disciplines as it provides good sensitivity [151]. Other popular metrics are MAE and its normalizations. Although not included, visual inspection is also an important representation commonly employed by researchers. Popularity, however, should not be the determining factor when choosing a metric, although it can be considered for comparison purposes. The most suitable metric depends on the characteristics of the system such as geographic location, the forecast horizon, and the objective of the study [150, 151]. All these factors hinder the existence of a standard metric or group of metrics. As a general recommendation, it is important to employ more than one measure of prediction error and to make an analysis of the type of error made. This will display the flaws of the proposed approach for further improvement. We also recommend the use of normalized metrics for easier comparison and understanding.

A LOOK INTO THE FUTURE

To finalize, we provide our view on future developments. We anticipate that soon, ML-based methods for PV power forecasting will be more extended at the industry level. Some factors such as robustness and interpretability are hindering their application right now. These affect the algorithm choice and could dispel promising models such as neural networks. Interpretability strategies such as explainable AI will most likely flourish since model decisions need to be justified. Probabilistic approaches are likely to play a crucial role as well, especially in energy markets, as they provide flexibility to the predictions. We also hope that the time and spatial resolution of satellite images will increase so that their use is extended shortly.

There are some other aspects to take into consideration regarding the expansion of this field. For instance, the velocity at which PV capacity is deployed. Grid limitations as well

as administrative regulations slow down the installation of new facilities. The lower the importance of PV in electricity generation, the lower the demand for accurate forecasts. Storage development will probably also affect the AI deployment since in the presence of large and cheap storage solutions, supply-demand balancing will become easier, and there will be fewer incentives to have accurate prediction solutions.

Overall, no model is going to yield 100% accurate predictions, mainly due to data collection uncertainties and inaccuracies that are swept along throughout the prediction. The most important is to focus on the problem one is aiming to solve and provide the best possible result accepting some uncertainty in it.

2.8. CONCLUSIONS

This review provides an overview of the state of the art of PV power forecasting employing machine learning algorithms. The main contribution of this work is the extensive categorization of the literature. This has highlighted gaps that have not been reported previously, such as the lack of studies in certain parts of the world or the focus on day-ahead predictions. An overview of all ML families is given, instead of focusing on the most popular algorithms, to provide alternatives to researchers wanting to deviate from the standard. The outlook has summarized the main outcomes of each section while providing recommendations. The review has been made as impartial as possible while expressing the authors' opinions across the document. This can be a valuable contribution to the field.

One of the limitations of this study is the specific range of years that it focuses on. By looking from 2015 until 2020, the most up-to-date literature and long-term trends are not being captured. The process employed to obtain that literature was also biased by the most cited works. However, the large number of publications considered reduces this bias.

PERSPECTIVE: 5 YEARS LATER

At the time of this dissertation's publication, limiting the reviewed articles to the period between 2015 and 2020 makes the work somewhat outdated. This section does not aim to provide a full update of the chapter, but rather offers a reflection on how the field has evolved over the past five years based on recent reviews.

Overall, these reviews reinforce the key conclusions presented in this work. Continuing the trend highlighted in the introduction, the number of studies utilizing machine learning for PV power forecasting has continued to grow, with DL gaining traction [152, 153]. Hybrid and DL models remain highly popular [154] despite their increased complexity [153, 155] and the higher computational burden they impose [156] for only marginal performance improvements [155].

All reviews emphasize the crucial importance of high-quality input data and related pre-processing techniques [152, 153, 155–157], with metaheuristic models helping with data refinement and hyperparameter tuning [152]. Most reviews also highlight the challenge of generalization [153, 156, 157], often attributed to the lack of diversity in input data [154].

It is also worth highlighting the discussions on the importance of interpretability [152], probabilistic forecasting [152], and careful feature selection [157].

However, some points remain open for discussion in relation to this work.

Some researchers emphasize the need for more medium- and long-term forecasts, as these are critical for energy trading and strategic planning [156]. In contrast, this work fo-

cuses on the need of very-short forecasts, a perspective supported by other studies [153]. In our opinion, alternative techniques may be more suitable for medium- and long-term forecasting, where ML capabilities might not be fully exploited.

Dhaked et al. highlight the use of metaheuristic techniques for data augmentation [156]—a set of methods that increase the size and quality of training datasets [158]. This approach can improve ML and, in particular, DL training, making it especially useful for datasets spanning less than a year or for replicating rare events.

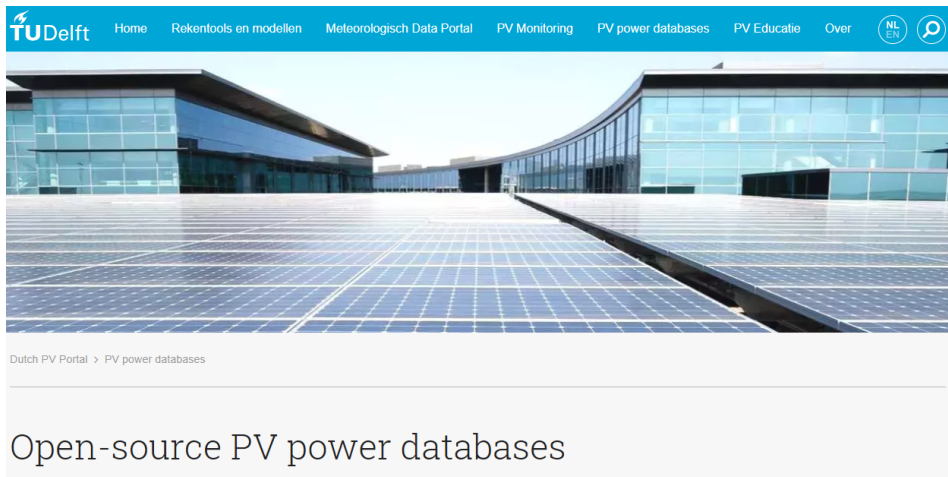
Another topic discussed by Dhaked et al. but not covered in this work is the real-life implementation of ML techniques. The use of DL is associated with challenges related to portability, scalability, and computational cost, which should be addressed to improve robustness and practical applicability [156]. Despite its relevance in real PV systems, the authors stress the limited research on this subject, a gap we acknowledge.

In conclusion, the key findings of this chapter remain valid despite being based on studies from five to ten years ago. However, some recent developments were not covered, underscoring the rapid evolution of the field.

3

DATABASE OF OPEN-SOURCE PV SYSTEMS

Some of the gaps spotted in the literature on PV power forecasting with ML can be solved by increasing the amount and the quality of the employed data. Motivated by this, an extensive data-gathering process was performed in search for open-source PV power databases. A website, www.tudelft.nl/open-source-pv-power-databases, has been created with all found open data sources for ease of use in future research.



This chapter has been adapted from the supplementary material of the publication A. Alcañiz, A.V. Lindfors, M. Zeman, H. Ziar, O. Isabella, “Effect of Climate on Photovoltaic Yield Prediction Using Machine Learning Models” *Global Challenges*, 7(1), 2200166 (2023) [159].

3.1. INTRODUCTION

The previous chapter highlighted that data scarcity, especially the absence of high-quality data, causes some of the current gaps in forecasting PV power output. The lack of generalization of the developed ML models, the low time resolution considered which does not respond to the market needs, and the poor results attributed to low-quality data, are all problems that can be solved by increasing the amount of high-quality data.

Researchers often develop their algorithms relying solely on the data available within their institution. This also complicates the comparison between studies, as pointed out by Yang [47]. Given that ML algorithms are sensitive to many factors such as location and data resolution, it is generally unfair to make a comparison between studies. One solution to this issue would be to employ the same data and be as transparent as possible with the parameters and techniques employed.

Seeking to address this gap, this chapter aims to create a website that contains and links to diverse PV power data sources, making them accessible to anyone interested in utilizing them. This database encourages researchers to utilize the same dataset to improve result reproducibility and motivate collaboration.

3.2. DATA GATHERING

The process of open-source PV power data gathering started by exploring the literature on PV power forecasting [160–162]. The aim was to find researchers who employed open-source data or openly shared their data. After conducting the initial search in the literature, search engines were employed to increase the database.

Unfortunately, most of the open-source data sets found at this stage were located in Europe or the USA. The next step consisted of contacting researchers who were open to sharing the PV systems data located at their institutions and making them open source. Despite having contacted around twenty researchers, only three resulted in positive responses. PV system data from two systems in Finland, one in Costa Rica, and another in Slovenia were achieved using this methodology.

In the following, a brief description of all found sources with individual PV system data is provided:

- PV Data Acquisition (PVDAQ), created by the National Renewable Energy Laboratory (NREL), provides access to PV performance data for systems throughout the US [163].
- The solar center division of Desert Knowledge Australia (DKASC), a non-profit corporation of the Northern Australian territory, offers generated PV power of a wide range of technologies located in two different sites of the desert [164].
- Sunny Portal is a PV monitoring portal that allows PV system operators and installers worldwide to access system data at any time [165]. Despite the privacy of most of the systems, a set of example German PV systems can be downloaded.
- In the 2014 edition of the Global Energy Forecasting Competition (GEFCom), organized by Dr. Tao Hong, one of the addressed problems consisted of forecasting the PV power [160]. The employed data set is publicly available for researchers.
- Marion et al. openly shared their dataset, which consists of data from three American PV systems [161]. In their publication, one can find how to access the data.

- DuraMAT DataHub, also by NREL, is a collaborative framework where the public can provide and access PV data for durability studies [166]. Deep exploration of the website has not been carried out, but a data set of 9 bifacial tracking systems was found for download.
- The London Datastore open data-sharing portal [167] includes the PV Solar Panel Energy Generation data set which contains voltage, current, power, energy, and weather data from domestic sites with PV modules located in the city of London.
- Kaggle is an online machine learning community, a subsidiary of Google, which allows users to find and publish data sets and to build models in a web-based data-science environment [168]. One can find several data sets related to PV power, which have been reported in www.tudelft.nl/open-source-pv-power-databases. For instance, *Horizontal Photovoltaic Power Output Data* set was employed in an Energies publication [162]. It includes PV power and weather data for 12 Northern Hemisphere sites over 14 months.
- The report published by the International Energy Agency regarding the Assessment of Performance Loss Rate of PV Power Systems [169] required of PV system and weather data around the world. They made the dataset public in the OSF open platform which includes time series data of 18 systems located in different climates. For each plant, PV power, plane of array irradiance, ambient and module temperatures, and sometimes wind speed data are available with time diverse resolutions.
- Although generally not recommended since it is meant for software development and version control, GitHub is also employed to store data sets. [170] provides one-minute resolution data for three individual systems in India for over 6 months.
- The *Hathaway Solar Patriot House* is a project which monitors and models the performance of a sustainable house outside of Washington, D.C with a 6 kW_p photovoltaic system [171]. The collected data has been made available and includes a large number of measurements such as solar irradiance, DC electrical measurements from the PV array and battery, energy from and to the grid, and energy consumption from electrical appliances.
- The data provided by researchers Dr. Victor Vega from the University of Costa Rica (UCR) and Prof. Janez Krč from the University of Ljubljana can be directly downloaded from the developed website (more in the next section).
- To access the two Finnish systems, Anders Lindfors from the Finnish Meteorological Institute (FMI) should be contacted.
- Motivated by this initiative, the authors of this work have included the PV power produced by two residential systems, one in the Netherlands (whose data can be directly downloaded from the website) and another one in Spain (which can be accessed via Kaggle).

For researchers interested in cumulative production instead of individual systems, there are also several choices:

- Belgium's electricity system operator, Elia, has had the initiative to provide open access to all of its public grid data [172]. Power generation data including generated PV power can be found for the whole country or by areas.

- The German electricity market information platform SMARD provides actual generation over Germany, Austria, and Luxembourg with a 15-minute resolution [173].
- Outside of Europe, the Electric Power Statistics Information System (in Korean) provides monthly data on power generation sorted by fuel type [174].
- Paul-Frederik Bach had the initiative of collecting data from several European system operators, such as the already mentioned Elia. One can find hourly time series of cumulative production in his blog [175].
- Also for Europe, the Open Power System Data platform offers open data required by energy system models, including time series of solar power generation with up to 15 minutes resolution [176].

Other interesting sources, although not fully open-source, are PV CAMPER and pvoutput.org. PV CAMPER (Photovoltaic Collaborative to Advance Multi-climate Performance and Energy Research) is a collaborative platform of PV institutions that share data within the community [177]. There are entry requirements to access the community, but the platform was founded with the objective of data sharing and collaboration. pvoutput.org is a free service for sharing and comparing PV output data. Through the API, one can retrieve generation data from any system after having shared data from an owned PV system and donated a minimum of 15 AUD per year (as of 2023).

3.3. WEBSITE DEVELOPMENT

All the open data sources explained in the previous section are listed in a website www.tudelft.nl/open-source-pv-power-databases developed by the authors. The website has been included as a section of the [Dutch PV portal](#), a non-profit service provided by Delft University of Technology. This portal's objective is to offer accessible information on solar energy in the Netherlands, which is in line with this chapter's goal.

Open-source PV power databases

This website shows the most extended list of open-source photovoltaic (PV) power databases. It gathers information of databases all around the world which share PV power data.

This was done with the objective to create more open studies on PV power forecasting, and to be able to validate the models with data from different parts of the world. If you **own a PV system** or know of any open-source PV power database that you would like to include in this list, let us know by **filling out the form** at the bottom of the page and we will reach out!

The table below depicts all the PV power databases encountered so far. You can order them and filter depending on your requirements. Links to each source have been provided, although some of them need additional clicking. Description of each header is also shown when hovering over it.

Source	Country	Weather data	Access	Resolution [min]	Data type
PVDAQ NREL	United States	Yes	API	1	Individual
Solar Centre DKA	Australia	Yes	Download	5	Individual
Sunny Portal	Germany	No	Download	15	Individual
GFFCom	Australia	Yes	Science direct, dropbox	60	Individual

TU Delft Home Tools & Models Meteorological data PV Monitoring PV power databases NL EN

Figure 3.1: Screenshot of the top of the developed website.

Figure 3.1 shows a screenshot of the top of the website. After a brief introduction, a large table is presented. Each row in the table is one of the items listed in the previous section. The columns provide the main data characteristics of each source, including the

country, whether weather data is included, how to access it, the time resolution in minutes, and the type of data (individual or aggregated). This table can be filtered based on these characteristics so researchers can select the appropriate set based on their needs.

Below the table, a world map highlights countries with available data in light blue, allowing for data filtering. Clicking a highlighted country reduces the table to only the sources present in that country. For instance, the screenshot in [Figure 3.2](#) shows the result of the table after selecting the Netherlands.

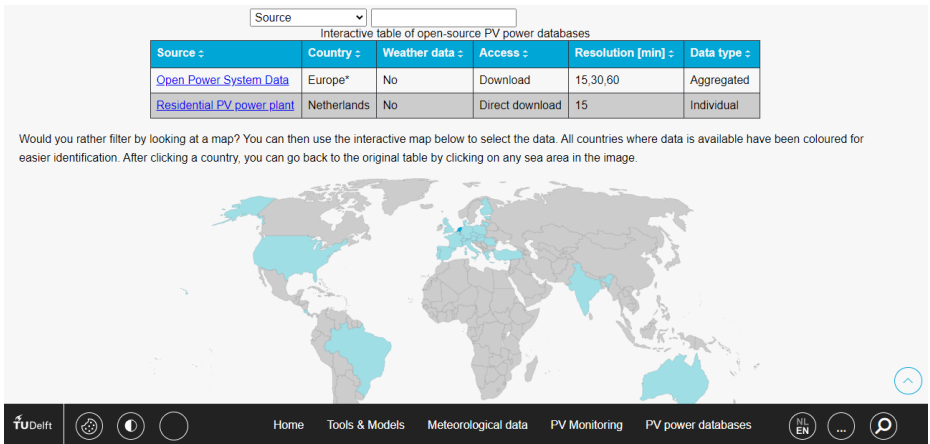


Figure 3.2: Screenshot of the developed website showing the filtered table by the map.

Finally, at the bottom of the website, a form is included for anyone interested in contributing to the database, see [Figure 3.3](#).

Are you willing to share PV power data?

Fill in the form below and we will reach out.

Your name:

Your email address:
email@example.com

Would you like to add a message?

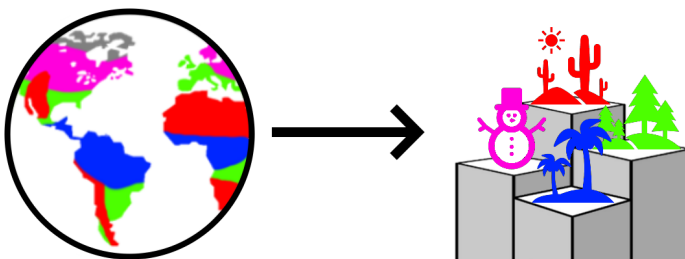
Website created by Alba Alcañiz Moya. Feedback or questions to A.AlcanizMoya@tudelft.nl.

Figure 3.3: Screenshot of the developed website showing the form to contribute to the database.

4

EFFECT OF CLIMATE

Despite the abundant literature on PV power forecasting with ML, the effect of climate on yield predictions is unknown. This work aims to find climatic trends by predicting the power of 48 PV systems around the world, equally divided into four climates. Five machine learning algorithms and a baseline one have been trained for each PV system. Results show that the performance ranking of the algorithms is independent of climate. Systems in dry climates depict on average the lowest NRMSE of 47.6%, while tropical ones present the highest of 60.2%. In mild and continental climates the NRMSE is 51.6% and 54.5%, respectively. When employing a model trained in one climate to forecast the power of a system situated in different climates, it is observed that systems in colder climates tend to exhibit on average a lower generalization error, with an additional NRMSE as low as 5.6% depending on the climate of the test set. Robustness evaluations were also conducted to increase the validity of the results.



This chapter has been adapted from the publication A. Alcañiz, A.V. Lindfors, M. Zeman, H. Ziar, O. Isabella, "Effect of Climate on Photovoltaic Yield Prediction Using Machine Learning Models" *Global Challenges*, 7(1), 2200166 (2023) [159].

4.1. INTRODUCTION

Despite the abundant literature on PV power forecasting with ML, most studies consider less than 5 PV systems which are generally located in the same areas. Few studies have considered PV systems subjected to diverse meteorological conditions to study the effect that climate has on the performance of ML models.

This chapter aims to study the effect that climate has on machine learning predictive models. For this purpose, a data set of 48 PV systems spread around the world is considered. These systems are selected to be equally divided into four climates and as spread as possible. Several ML algorithms are trained on this data and the results are analyzed to identify climatic trends. The best algorithm is selected for further evaluations, which consist of quantifying the increase in error when the climate of the training and test set do not coincide, and of exploring how robust the obtained results are.

Section 4.2 details the data set employed in this work. It explains as well the process of data preparation and exploration. Section 4.3 describes the methods employed, which are mainly the ML algorithms, but also includes the hyperparameter tuning process and the metrics. The main results are analyzed in Section 4.4, where some evaluations are also conducted to show the generalization and robustness of the model, amongst other characteristics. Finally, the main conclusions are presented in Section 4.5.

4.2. DATA SET

This section explains the main characteristics of the data set employed. The first step consists of the data-gathering process, where open-source data sets are prioritized. Chapter 3 elaborates on this open-source data-gathering process. Once the data is selected, its main characteristics are explained. The data is later cleaned and explored before training the ML algorithms.

4.2.1. DATA CHARACTERISTICS

A total of 48 PV systems as spread as possible were selected for this study, situated in 18 different countries. These systems are evenly distributed across four climates, ensuring sufficient statistical robustness. A dataset with a wide range of systems per climate is crucial for accurate analysis. To distinguish between meteorological conditions, the major types from the Köppen-Geiger (KG) climate classification were employed [133]: tropical (A, blue), dry (B, red), mild temperate (C, light green) and continental (D, magenta). The study excludes the polar climate type due to the absence of PV system data in that region.

The arduous task of data gathering severely limited the accuracy of this work. Some of the gathered data included very few meteorological measurements and limited system information. All PV systems needed to have the same meteorological and system information, otherwise it could have affected the robustness of the results. Therefore, it was decided to sacrifice some meteorological parameters in exchange for a higher number of systems that would assure statistical robustness to the climatic trends.

Model inputs consist of hourly resolution clear-sky global irradiance, previous hour measurements of global irradiance, temperature, and wind speed, power produced during the three previous hours, and power produced one day before. No system characteristics were included in the model since very limited data was known from some systems. Since a limited number of descriptive meteorological parameters were available for all sys-

tems, the PV power produced during previous instants of time was included as a feature to provide additional information to the algorithm. The decision to select the power produced during the three previous hours and one day before was taken after looking at the changes in performance (not shown here). Including three previous hours and one day before yielded slightly lower errors than removing previous hours in most algorithms, although some started showing the negative effects of cross-correlated features. The clear-sky global irradiance was implemented using the *pulib* Python library [178]. This variable represents the geographical dependency of the sites and irradiance patterns in time in a more effective way than with features such as latitude, longitude, sun azimuth, or sun altitude.

For most systems, the meteorological data was gathered on-site and obtained from the same source as the PV power. For the German open-source PV system, there was no onsite weather data accessible, so an inverse interpolation method was applied to nearby meteorological stations provided by the German Weather Service DWD [179].

4.2.2. DATA PREPARATION

After gathering the data, it needed preparation for the ML algorithms. Several cleaning steps were therefore performed, consisting of:

- Standardizing the units of all measured data.
- Ensuring that variables were within limits: positive PV power, no irradiance at night, ambient temperatures below 100°C, etc.
- Removing days with no measurements.
- Removing more than two consecutive absent measurements, while interpolating sporadic missing points.
- Removing outliers. For ambient temperature and wind speed, the interquartile range (IQR) was employed. A different IQR constant (usually set to 1.5) was found experimentally for each feature. This measure could not be employed for PV power and irradiance due to their low average. For those variables, values that were 25% higher than the 99% quantile were considered outliers.

Each system's data was individually explored (not shown) to detect any abnormal behavior and ensure that the data-cleaning process was properly implemented. Since the data availability was different for each system, after the cleaning process it was ensured that all systems had at least 75% of a year of data and a maximum of two years of data.

The next step consisted of data normalization, an essential requirement for some ML algorithms like Support Vector Regression that require a similar range for all features. All inputs were scaled between zero and one considering the minimum and maximum of each series. Finally, the data for each system was randomly split into 70% of the training set and 30% of the test set.

4.2.3. DATA EXPLORATION

Before diving into the methodology, exploring potential variations in behavior across different climates was also insightful. The correlation of PV power with each of the model inputs for each PV system was computed. Correlation is a value between -1 and 1 and indicates the similarity between two vectors. The correlation results were grouped based on the climate to find meaningful trends, as shown in Figure 4.1.

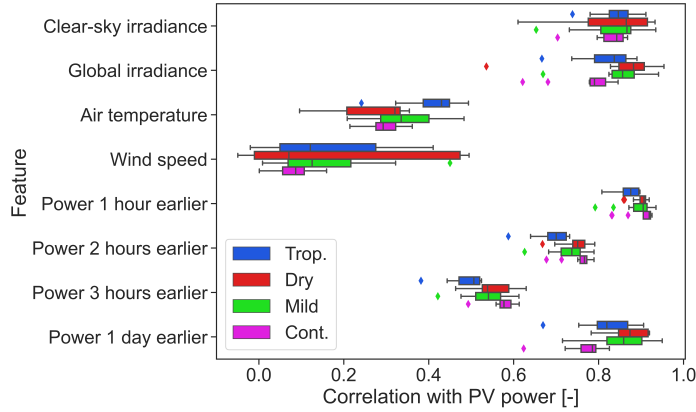


Figure 4.1: Correlation of PV power with each input feature for all 48 systems. The systems have been colored depending on the major KG climate that they are subjected to.

Irradiance is highly correlated with PV power, showing medians greater than 0.84 and 0.79 for clear-sky and global irradiance, respectively. In the case of systems located in dry climates, these two variables are more highly correlated than for the rest: around 0.87 on average for both variables as compared to 0.86 for mild, 0.84 for tropical, and 0.82 for continental systems. As clear-sky irradiance is computed analytically, there is no temporal discrepancy with PV power. This makes it a highly explainable feature of PV power in clear-sky days.

The other two measured variables, air temperature and wind speed, have lower importance for the ML algorithms. Wind speed has on average a correlation of 0.14 with PV power, suggesting no importance for most systems. Seven systems even show a negative correlation, five of which are in a dry climate, but it is so low (minimum of -0.05) that it indicates no relationship between variables. Regarding air temperature, one can observe a high dependency on climate, since the median correlation for tropical systems is 0.43, while in the remaining climates, it drops to around 0.3.

The hourly previous instances of PV power show an expected correlation, dropping as the temporal gap widens. Tropical systems show the lowest correlation of all climates, with a value of around 0.87 while it increases to 0.9 for the other climates. This indicates that the intraday weather variability in tropical areas is very high compared to other climates since the PV power can be very different between hours. Looking now at the correlation with power one day earlier, tropical climates also depict a low correlation (0.82), but continental climates obtain an even lower one (0.78). Following the same reasoning as before, systems in continental climates show a high daily weather variability.

4.3. METHODOLOGY

This section explains the algorithms employed for predicting the PV power ([Subsection 4.3.1](#)), and how their parameters were optimized (hyperparameter tuning in [Subsection 4.3.2](#)).

4.3.1. ALGORITHMS

This subsection briefly explains the working principle of the algorithms employed to predict PV power in this work.

The persistence algorithm assumes that nothing changes between the current time step and the following [180]. It is a simple algorithm usually employed as the benchmark to compare the performance of other models. It is described by Equation 4.1, where $P(t)$ is the PV power produced at time step t . This simple algorithm can achieve low errors when forecasting PV power in the very short term [145].

$$P(t) = P(t - 1) \quad (4.1)$$

The second algorithm is the high-order polynomial regression model. The relation between the independent variable ($P(t)$ in our case) and the features is modeled by a polynomial of degree d . This statistical algorithm can be extended to machine learning by determining the parameters through the training data set.

One issue with regression algorithms is that they tend to overfit the data, as explained in Appendix B. The model focuses too much on the training set and it is unable to generalize to unseen data. To solve this issue, regularization techniques were developed which keep the parameters low. For instance, ElasticNet penalizes the parameters through the 1-norm (absolute value) and 2-norm (root squared). The user can choose the ratio between these two penalization terms l_r as well as the amount of penalization α .

Support Vector Machine (SVM) is an algorithm originally developed to solve classification problems and extended for regression under the name of Support Vector Regression (SVR). The objective of SVM in classification problems is to create a boundary between groups with the highest possible margin. SVR follows the same principle of maximizing the margin: it fits a hyperplane to the data with a margin of tolerance ϵ [181] (see Figure B.1 for a graphical explanation). Usually, a hyperplane does not properly describe the data. In such cases, two simultaneous approaches are taken. First, a penalization term C is introduced to the samples that are outside the hyperplane so that the minimization function has the motivation to fit all points inside the hyperplane. The second approach is to map the data to another feature space, where it may be easier to fit a hyperplane. This is done via a kernel, a function that helps SVR to solve non-linear problems. The most common kernel functions are polynomial, radial basis function, and sigmoid, expressed together with its parameters in Table 4.1. In these equations, x is the real output while x' is the predicted one.

Table 4.1: Possible kernels for SVR.

Kernel	Expression	Parameters
Polynomial	$(\gamma \cdot \langle x, x' \rangle + r)^d$	γ, r, d
Radial basis function	$\exp(-\gamma \cdot \ x - x'\ ^2)$	γ
Sigmoid	$\tanh(\gamma \cdot \langle x, x' \rangle + r)$	γ, r

Random forest is a combination of decision trees. A decision tree is an algorithm that recursively partitions the data space depending on its value and a simple prediction model is then fitted within each partition [182]. To avoid overfitting, a random forest fits several decision trees $n_{estimators}$ on various subsets of the dataset and then averages them.

Several characteristics of the built trees can be controlled, such as the maximum number of divisions (depth) of each tree max_{depth} , the number of features considered for each split $max_{features}$, the minimum number of samples required to be at a leaf (final) node $min_{samples,leaf}$, and the minimum number of samples required to make another division $min_{samples,split}$.

The last applied algorithm is Extreme Gradient Boosting or XGBoost. It also combines several decision trees, but contrary to random forest, XGBoost combines the trees on the go instead of at the end [183]. The first decision tree estimates an output that resembles as much as possible the real one, based only on the input dataset. In the next iterations, new decision trees are built by giving higher attention to large-error predictions. Several parameters optimize this algorithm although we focused only on two: $colsample_{bytree}$ and α . The former adds randomness by determining the fraction of randomly selected inputs employed to train each tree. α is a 1-norm regularization term, as in ElasticNet.

All these algorithms have been implemented in Python using the libraries *sklearn* [184] and *xgboost* [183].

4.3.2. HYPERPARAMETER TUNING

To improve the predictions, a hyperparameter tuning process for each algorithm and PV system has been performed. This process consists of finding the optimum set of parameters for each site. The parameters considered for each algorithm are the ones described in the previous section. The range of each parameter tuned is depicted in Table 4.2. Each algorithm was tuned using 5-fold cross-validation with randomized search.

The results of hyperparameter tuning were analyzed to identify climatic trends. However, no clear relationship between climate and optimum hyperparameters was found.

4.4. RESULTS

Subsection 4.4.1 explains the main results obtained which focus on finding climatic trends. Subsection 4.4.2 explores how well a trained model can predict the PV power of a different system. The robustness of these results is explored in Subsection 4.4.3 to increase credibility. Finally, Subsection 4.4.4 categorizes the obtained results based on a more suitable climate classification: the Köppen-Geiger-Photovoltaic climate classification [185].

4.4.1. CLIMATIC TRENDS

Once all algorithms were trained and optimized for each of the 48 PV systems, their performance was evaluated in the test set. The results for each model in terms of NRMSE, NMAE, and NMBE have been reported in Figure 4.2, color-coded based on climate.

Looking first at the models' performance, their ranking is independent of climate and metric. Persistence shows the highest NRMSE of 70.4% on average, followed by ElasticNet (57.6%) and SVR (51.6%), while polynomial regression, XGBoost, and random forest show similar performance. The latter is selected as the best algorithm with a lower average NRMSE of 46.2% compared to 48.4% and 47.4% for polynomial regression and XGBoost, respectively. Similar conclusions can be reached by looking at the NMAE. The superiority of polynomial regression over ElasticNet suggests that the amount of data is sufficient to prevent overfitting. This is supported by hyperparameter tuning results (not shown), which indicated that the ElasticNet parameter α , representing regularization, was close to zero for

Table 4.2: Hyper-parameters tuned and their explored range for each ML algorithm implemented in this work.

Polynomial regression	
d	1, 2, 3, 4
ElasticNet	
α	0.01, 0.1, 0.3, 0.5, 0.7, 1
l_r	0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99
SVR	
ϵ	0.001, 0.1, 0.3, 0.5
C	0.01, 0.1, 1, 10
kernel	polynomial, radial basis function, sigmoid
γ	0.01, 0.1, 1, 10
r	0, 0.2, 0.5, 0.8, 1
d	1, 2, 3
Random forest	
$n_{estimators}$	100, 500, 1000, 2000
max_{depth}	None, 5, 10, 20
$max_{features}$	auto, sqrt
$min_{samples,leaf}$	1, 2, 4
$min_{samples,split}$	2, 5, 10
XGBoost	
$colsample_{bytree}$	0.3, 0.6, 1
α	0, 5, 10, 100

all sites.

In terms of NMBE, SVR depicts the worst results. This low performance can be due to the cross-correlation of input features. Regarding climatic trends, tropical and mild systems are more prone to being underestimated (median NMBE for all models of -0.21% and -0.42%, respectively) as opposed to those in continental and dry ones (median NMBE of 0.38% and 0.35%, respectively).

Despite depicting the highest NRMSE, persistence can provide interesting information. Low persistence errors indicate steady atmospheric conditions. Climates that show on average low persistence errors indicate low hourly variability in PV power. Continental systems depict the lowest median NRMSE of 62.5% when using persistence, which was already expected considering that the systems in this climate have a high correlation of 0.92 with the previous PV power value (Figure 4.1). Dry and mild systems also show a low median NRMSE of 65.0%, again due to the 0.91 correlation of PV power with its previous value. On the other hand, tropical systems show a high weather variability between hours, hence their high median NRMSE of 74.5% and low correlation of 0.87. For instance, in Costa Rica, the hourly and monthly rainfall patterns can be very fluctuating, without a defined pattern during the diurnal period [186].

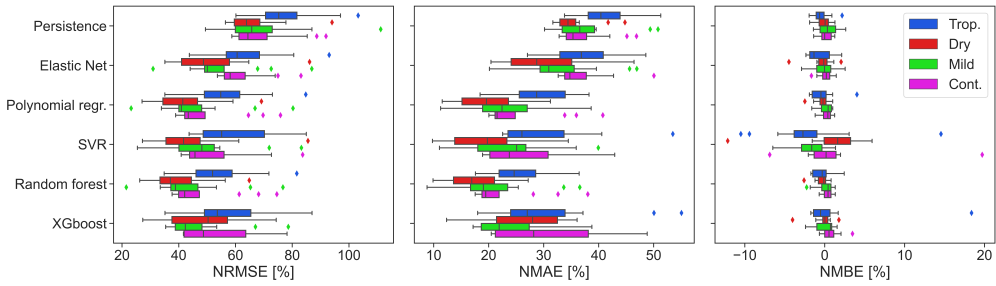


Figure 4.2: Performance of the ML models (and persistence) in terms of NRMSE, NMAE, and NMBE for all 48 PV sites. Each color represents a KG climate.

4

Looking now at climatic trends, on average systems located in dry areas are the ones that report the lowest NRMSE (47.6% for all models), while those located in tropical areas show the highest NRMSE (60.2%). Although both climates are characterized by high ambient temperatures and high irradiation, the higher humidity and precipitation of tropical areas negatively affect not only the performance of the PV systems but also their prediction [185]. Systems in mild and continental climates show similar average NRMSE of 51.6% and 54.5%, respectively.

These results may be due to the absence of climate-specific features in the model, such as humidity, rainfall, snowfall, and dust. This hypothesis could not be confirmed due to insufficient data, but existing literature can help support this notion.

By definition, dry climates are characterized by a lack of available water [187]. Variables such as humidity, rainfall, and snowfall will therefore have a low impact on the prediction of PV power. However, incorporating dust as a feature would likely reduce prediction errors significantly, given its high explanatory power. This is supported by [66], where dust was identified as a relevant feature with a correlation coefficient of -0.56 for a system in the arid State of Qatar. Taking into account global dust accumulation [188], incorporating dust as a variable would reduce errors for systems in tropical and dry climates, though its impact would be minimal for mild and continental systems.

The addition of humidity would also have a positive impact on the prediction accuracy for tropical systems. In [189], the correlation of ambient humidity with PV power for a tropical system was -0.43, which according to Figure 4.1 is higher than the median correlation (in absolute value) with wind speed. Moreover, tropical places usually show seasonal weather variations substantially different from other climates [162]. Finally, by definition tropical climates have significant precipitation [187].

Regarding the two remaining climates, they are only distinguished from one another by temperature [187]. The addition of humidity and precipitation would have a positive effect on the prediction for both climates, although the magnitude of this effect would differ depending on the location of each system. In [162], the variable importance ranking for several systems depended on their location even when they were subjected to the same mild or continental climate. The feature that could explain the lower prediction error of mild systems compared to continental ones is snowfall. By definition, snowfall is more common in continental climates than in mild ones. Awad et al. reported that R^2 increased from 0.93 to 0.96 when including rain and snow in their predictive model for Canadian systems [190].

Snowfall helped the ML algorithm learn the differences in fluctuating conditions between the snow and non-snow seasons. Similarly, Böök et al. reported that, for the two Finnish systems included in this study, the average R^2 of their prediction during snow-free days was 0.995 while it dropped to 0.640 during snow cover periods, which represents roughly 13% of the data points [191].

As a final reflection, the reported errors are quite high. Even when considering the best algorithm, random forest achieves a median of 41.8% NRMSE and a median of 19.7% NMAE. Comparing with the literature, Pujić et al. achieved an NMAE of 8.61% when using random forest [192], while Ferlito et al. achieved an RMSE with random forest ranging between 46.5 W and 82.0 W, depending on the year, for a 1 kW_p PV plant [64]. These results are compared with those of system B3 with an NRMSE of 41.0%, close to the median, and a capacity of 6 kW_p. Therefore, the representative RMSE is 94.8 W/kW_p, while for Ferlito et al., this value was 46.5-82.0 W/kW_p. There may be several reasons for this underperformance. First, the shortage of descriptive information, as already suggested. The algorithms utilized only a few general meteorological features without incorporating any system-specific information. Additionally, the models were not tailored for each PV system, beyond the hyperparameter tuning and a generic cleaning process. A more powerful ML algorithm, such as an ANN, may reduce the error. For instance, in [60] the average RMSE was reduced from 167 to 164 when using LSTM instead of random forest. However, the objective of this study lies in the comparison between climates rather than in obtaining the lowest possible error per system.

4.4.2. FLEXIBILITY TO OTHER CLIMATES

This subsection aims to measure the adaptability of machine learning models to different climates. The goal is to determine the error increase when an ML algorithm, trained with data from a system in one climate, predicts the power of a system in a different climate. The intention is to observe hidden relations and similarities. By no means, it is our objective to substitute this general model with the individual ones.

For this study, five new datasets were generated, and an ML algorithm was trained with each. The first dataset comprises data from all 48 PV systems, forming the *Universal* model. Subsequently, datasets were created for each climate, incorporating data from all 12 systems within that climate. Each dataset underwent randomization and was limited to 5 years of data due to computational constraints.

The selected algorithm for this study is random forest, given its results in the previous subsection. Random forest parameters were first optimized for each set and then the algorithm was trained with each of the five new datasets. The PV power was then predicted for all 48 PV systems using the five newly trained models. Figure 4.3 shows the results, where the colors and legend indicate the model used, while the vertical axis indicates to which climate the test system belongs to. The performance of random forest, when trained with the same PV system (the model used in the previous subsection), has also been added to the graph for comparison purposes with the label of *Self*.

As expected, when the training and the test sets coincide (Self case, yellow color), the performance is optimum (average NRMSE of 46.2%). The second-best performance is obtained when the climate of the test and training sets coincide, which results in an average NRMSE of 47.9%. The third best option is the universal model (grey color, NRMSE of 49.8%), which includes data from all PV systems.

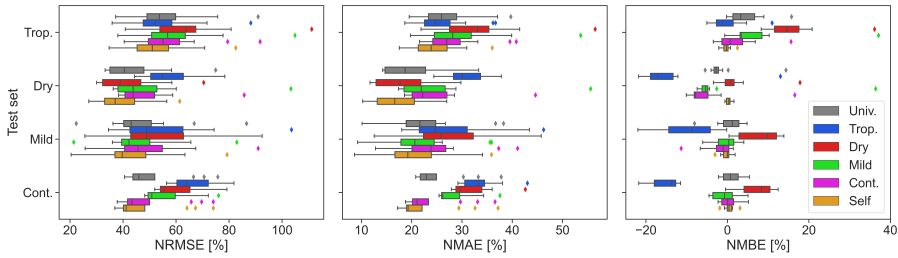


Figure 4.3: Performance of the universal, climatic, and own models for all 48 PV systems.

4

Up to this point, all results are expected. The performance decreases as the similarity degree of the training set with the test set decreases. Similarity degree is the percentage of data belonging to the target system that is included in the training set. In the self case, the similarity degree is the maximum, 100%. When the climate of the test and training sets coincide, the similarity is around 8.3% (one over twelve). For the universal model, a fourth of the PV systems are from the same climate as that of the test set, thus the similarity is decreased to 2.1%. For the remaining conditions, the similarity degree is 0%.

Therefore, the real adaptability measure starts when testing a model trained with data in one climate on a PV system located in a different climate. From the graph, one can note that the continental model is the most flexible of all. In the case of the dry test set, the continental model obtains a mean NRMSE of 48.8%, lower than the mild or tropical models (49.9% and 56.8%, respectively). This higher adaptability of the continental model could be explained because it relies strongly on the PV power produced during the previous hour and because continental systems are subjected to a diverse but stable type of days. With similar reasoning, the mild temperate model also achieves good results. In the case of the dry model, since precipitation rarely occurs, it is not as accurate as continental or mild models during rainy days. Lastly, the tropical climatic model is the one that generalizes worst due to the challenging conditions in which it has been trained.

One could argue that this experiment is of no use since it was already expected that if the training and test sets do not coincide, the accuracy will be lower. However, the main interest is the quantification of this lower accuracy and guidelines for the choice of an alternative training set in case the target PV system does not hold enough training data. Alternative approaches could be employed, such as developing a physical model or using persistence during the first months of operation, however, the aim is to quantify the consequences of using a trained ML algorithm on a different test set.

The method to quantify this error increase is better illustrated with an example. Suppose we aim to forecast the power output of a recently installed system in a tropical climate using an ML model. However, there is not enough data available to train the algorithm and create a new model. Instead, the only available data consists of two PV systems located in dry and mild temperate climates. Based on the discussions in this subsection, lower errors are expected when the model from the mild climate system is employed. In terms of NRMSE, the additional error incurred when training with the mild system would be around 6.5%, whereas with the dry system, the additional error would rise to 10.0%. This increase in NRMSE for all possible combinations is shown in [Table 4.3](#).

These numbers are averages, therefore the actual values in another situation can differ.

Table 4.3: Increase in NRMSE [%] in each of the combinations compared to the performance of the self case.

System's KG	Own climate	Univ.	Trop.	Dry	Mild	Cont.
Tropical	1.4	3.5	-	10.0	6.5	5.6
Dry	2.3	4.3	16.5	-	9.3	9.1
Mild	2.3	3.5	9.8	8.9	-	5.8
Continental	1.5	3.5	17.7	12.1	9.3	-

However, they indicate the expected increase error. Whether this increase is acceptable, depends on the user and especially on the alternative. This path offers a way for PV power prediction using ML techniques with only one day of system operation.

4.4.3. ROBUSTNESS

Although efforts were made to incorporate a diverse range of systems per climate, the results may not be robust enough. Some of the systems are located nearby, and unfortunately, not the same amount of data was available per system. This subsection, therefore, aims to show the robustness of the reported results.

25% of the systems had two years of data available after cleaning, while 37.5% included less than a year. Since the amount of training data can affect the performance of the ML algorithms, the systems' R^2 was compared to the amount of data points per system. Figure 4.4 depicts the results for all 48 PV systems. No meaningful relation can be found between the amount of data points and the accuracy of the prediction, hence the different data availability of systems had no effect in this study.

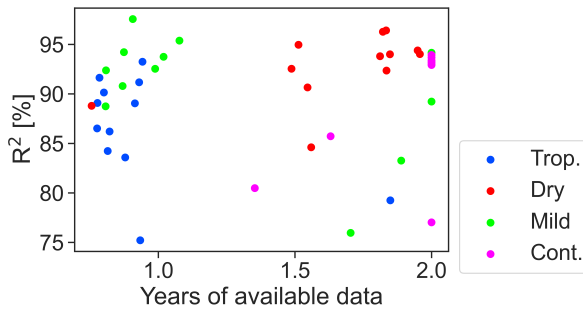


Figure 4.4: Coefficient of determination R^2 as a function of number of years per system. The different colors represent the KG climate that the system is subjected to.

In Subsection 4.4.2, there is some contamination in the created data sets. The system used as the test set is included in the training set corresponding to its own climate, as well as in the universal training set. This could potentially underestimate the prediction error. Additional simulations were conducted to illustrate the impact of including or excluding specific systems within the climate. The process is exemplified with the tropical system A1. To test the robustness of the generalization study, two new data sets were created where A1 is excluded from the tropical and universal datasets. Random forest was trained using these two data sets, and the two models were used to predict the PV power of A1. Metrics were

computed, and later compared with the metrics obtained for system A1 using the procedure explained in [Subsection 4.4.2](#).

This procedure was repeated for two random systems per climate, adding up to a total of 8 PV systems. When comparing the metrics of this procedure with those obtained from the generalization study, the results were best when the system was included, however, the differences were small. In terms of R^2 , NMAE, and NRMSE, the relative difference between the two conditions stays below 1.2%, 2.5%, and 3%, respectively, for the 8 systems. Therefore, despite the existence of some contamination, the error is negligible.

The final aspect is the number of systems per climate. Despite all the data-gathering efforts, 12 systems per climate might not be representative enough. To check this fact, a new set of simulations was conducted, examining whether the trends in performance remained consistent when fewer systems per climate were considered. In the first iteration, instead of having 12 systems per climate, it is assumed that only 11 systems per climate were available. 12 combinations per climate were therefore possible. The average NRMSE per climate of each of these combinations was computed. When comparing the NRMSE of 12 systems with all possible NRMSE of the 11 systems, one can identify if one of the systems is contaminating the results, or if the trends are robust.

This experiment was repeated considering down to 7 systems per climate. [Figure 4.5](#) shows the evolution in NRMSE as a function of the number of systems. Due to the possible combinations when the number of systems was lower than 12, the results are expressed with box plots. In most combinations, the trends reported in [Subsection 4.4.1](#) are consistent: systems in tropical climates show the highest errors while those in dry ones depict the lowest ones. The error range in the box plots shows that increasing the number of systems makes the results more precise but has almost no effect on the ranking.

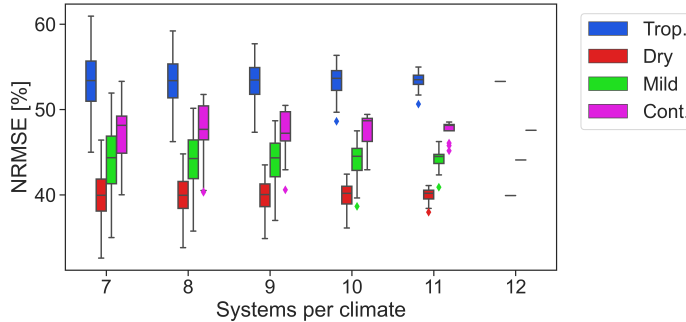


Figure 4.5: Evolution of the NRMSE as a function of the number of PV systems considered per climate.

4.4.4. KÖPPEN-GEIGER-PHOTOVOLTAIC CLIMATE CLASSIFICATION

This study employs the major types of the KG classification to divide the PV systems based on climate. This offers a simple classification into five categories, which is suitable for the problem at hand. However, recently Ascencio-Vásquez et al. updated this classification particularly for PV, creating the Köppen-Geiger-Photovoltaic (KGPV) climate classification [185]. This new classification includes another letter that indicates the strength of the irradiation level received: K-Very High, H-High, M-Medium, and L-Low. It was also based on

an update of the KG classification which includes six major types, not five: A-Tropical, B-Desert, C-Steppe, D-Mild temperate, E-Continental, and F-Polar. Overall, this classification consists of 12 categories after merging combinations.

Considering the nature of this new classification, it would have been more suitable to employ these 12 categories. However, given the data-gathering limitations, it would not have been possible to obtain enough number of systems per category to distinguish between trends.

However, now the results obtained in Figure 4.2 can be classified following the KGPV classification, as in Figure 4.6. One has to be more critical of these new climatic trends since the systems are unevenly distributed between these 12 categories. However, it can be the first step for possible future work.

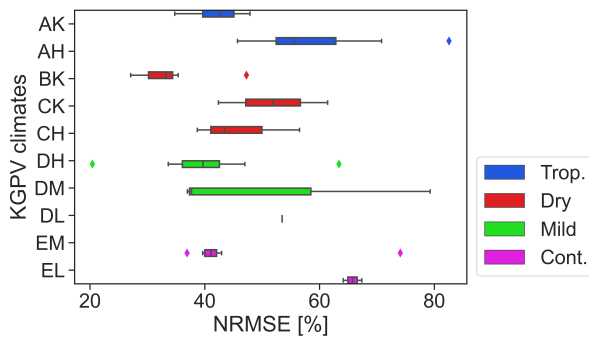


Figure 4.6: NRMSE as a function of the Köppen-Geiger photovoltaic climate classification. The colors represent the KG classification employed so far.

From the graph, one can distinguish almost two perfect continental subgroups: EM and EL. The PV systems under the EL climate are more likely covered during longer periods by snow due to the lower irradiance. These are then harder to forecast since snowfall is not included. Systems in the desert (BK) depict lower errors than those in steppe areas (CK, CH). In general, the higher the irradiance, the lower the error of the PV system. These could even override the previous trends, since from the graph, systems in the DH climate show lower errors than those in the CH climate. This is opposed to the highest accuracy of dry climates found so far in previous sections. Apart from the polar climate (F), which was excluded from this study, there is one KGPV climate missing: BH. This is logical, given that the dry climate has been subdivided into four categories, whereas the others have been divided into two or three. Climates DL, EL, and CK are also highly unrepresented, which hinders any conclusions extracted in this work.

4.5. CONCLUSIONS

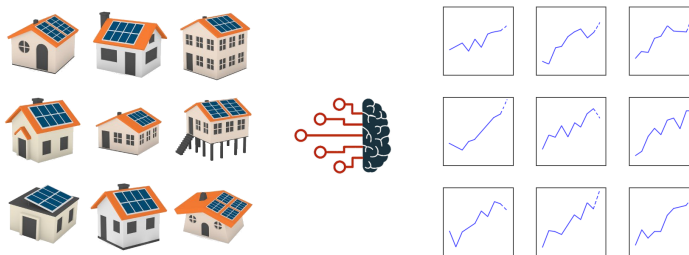
This work studies the effect of climate on the performance of machine learning models for PV power prediction. For this purpose, 48 systems, evenly distributed into 4 climates, were selected and several algorithms were trained for each system. Results showed that random forest is the best algorithm of all tested, with an average NRMSE of 46.2%. Climatic trends depicted how higher errors are achieved for tropical systems (NRMSE average of 60.2% for several models), while lower ones are obtained for systems in dry climates

(NRMSE of 47.6%). The adaptability to the climate of these models is also explored. Systems located in continental climates showed the lowest generalization error, which can be as low as an additional NRMSE of 5.6%, while tropical ones showed the highest. Finally, several robustness experiments were performed to increase the confidence in the obtained results, and the recent Köppen-Geiger-Photovoltaic (KGPV) climate classification was used to further identify more climatic trends.

5

NOWCASTING FOR RESIDENTIAL-SCALE SYSTEMS

As seen in [Chapter 2](#), most of the literature employs less than 5 PV systems during the training process of ML models, which does not ensure generalization to unseen systems. Regional forecasts are the norm when in the presence of a large fleet. Nevertheless, none of these approaches are usable when monitoring residential PV systems. This work proposes a single ML model using XGBoost to predict the individual power of a large fleet of 1102 PV systems. This algorithm obtains an MAE of 0.877 kWh (considering an average system size of 4.44 kW_p) and a MAPE of 23% for hourly data aggregated to daily values. XGBoost predictions are two times better than currently used commercial software on average. During the methodology development, a discussion arose on the lack of a suitable loss function and metric for residential PV yield forecasting.



This chapter has been adapted from the publication D. Grzebyk, A. Alcañiz, J.C. Donker, M. Zeman, H. Ziar, O. Isabella, “Individual yield nowcasting for residential PV systems” *Solar Energy*, 251, 325-336 (2023) [193].

5.1. INTRODUCTION

Most researchers motivate their PV power forecasting studies by contributing to improved generator dispatch, power quality effects mitigation, and reducing secondary reserve capacity. However, if their results are not validated for multiple systems, they might not be reliable enough to upscale. This problem of evaluation on small data sets was also identified by Yang and Theocharides et al. [47, 147]. Few studies employed data from more than 5 PV systems, and when they did, PV power was commonly aggregated to provide a regional approach (check Subsection 2.4.2 for details). Although from a grid manager perspective regional forecasts are preferred over individual ones [194], when it comes to monitoring individual PV systems, a regional forecast is not suitable.

Exceptions found in the literature are Elsinga and Sark, which provided analysis for 202 rooftop PV systems [195], Gensler et al. who forecasted the power of 21 systems [89], and the forecast of 71 systems from Lee et al. [65]. The approach in these cases consisted of creating an ML algorithm for each PV system, which can be computationally intensive during the training procedure and can limit the application of the ML model until enough data is available for a new PV system.

Unlike previous literature, this study aims to utilize a single ML algorithm to rapidly and accurately forecast the performance of a large database of 1102 PV systems. The algorithm improves its training process by learning from multiple systems while keeping the ability to differentiate between systems based on their system information. The developed methodology is applied in a solar software company to show an alternative to the currently employed analytical model. Suggestions for developing a standard metric and a discussion on loss functions are also given.

This chapter is structured as follows. Section 5.2 details the dataset employed, focusing on the considered features, data preparation, and exploration. Section 5.3 describes the methods employed, which include the feature selection strategy and the ML models. Two important characteristics of the method, namely the loss functions and the metrics, are also discussed to highlight their flaws and possible improvements. Results are explained in Section 5.4 in terms of feature selection, model performance, learning curves, and error analysis. Main conclusions are presented in Section 5.5.

5.2. DATASET

5.2.1. CONSIDERED FEATURES

The considered features for the machine learning model can be divided into weather and system features and descriptive parameters. All features are available from July 1st 2018 until June 30th 2019 for 1102 PV systems located in the Netherlands and Belgium. The data set contains more than 4 million points.

The weather parameters considered for the model are: global horizontal irradiance (GHI), cloud coverage, wind speed, precipitation, and ambient temperature. This data was obtained from the Royal Dutch Meteorological Institute with hourly resolution [196]. The sun angles (altitude and azimuth) were included using the Python library *pvl*ib [178].

The system parameters include size, age, type of panels, latitude, longitude, panel inclination, orientation, decay per year, maximal inverter efficiency, and nominal operating cell temperature. Descriptive parameters consist of the day of year and historical yield from 24, 48, 72, 96, and 120 hours before. The decision to use only five past days was based on [197],

but a longer horizon could have been utilized as well. All system data has been provided by the start-up Solar Monkey [198].

5.2.2. DATA PREPARATION

The data set was explored in search of quality issues. For instance, missing cloud coverage values were replaced with mean cloud coverage across all samples equal to 5.83 oktas. Samples corresponding to night values were removed from the data set. Several issues were found in the yield data obtained from inverters. Inverters log hourly yields via wi-fi which might be discontinuous. In case of prolonged disconnection, the amount of data stored in an inverter might exceed its memory capacity, and some information is lost leading to missing yield values. Other issues are constant or lagged yields. Constant yields might be caused by unreported disconnection of the inverter. Examples of these data quality issues can be seen in Figure 5.1 where low outliers are below 0.13 percentile and high outliers above 99.87 percentile of all yield values.

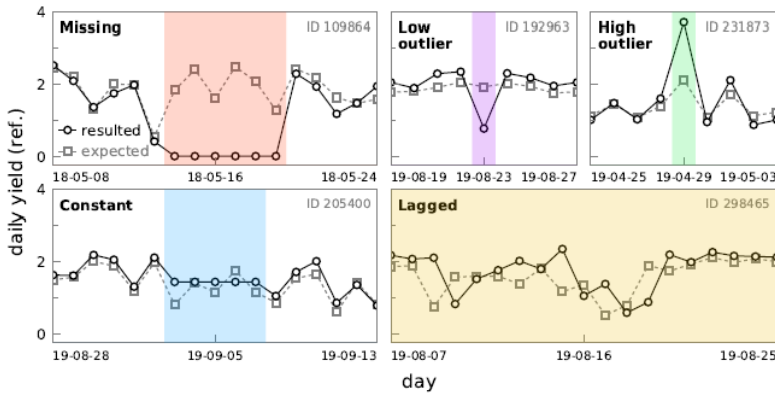


Figure 5.1: Examples of yield data quality issues.

Another step of data preprocessing is normalization used to obtain similar magnitudes of all features. First, presented inputs were processed to contain only numerical features. Then, each of the features was scaled to values between 0 and 1 using minimum and maximum values of all hourly data calculated for every feature.

5.2.3. DATA EXPLORATION

A brief exploration of the data was conducted as well. Nearly 79% of systems consist of mono-crystalline silicon, 18% of poly-crystalline silicon, and around 3% of thin film modules. Systems of size between 2.5 kW_p and 7.5 kW_p dominate, but large systems up to 17.7 kW_p are present. None of the analyzed systems was older than four years.

Regarding weather features, GHI is dominated by small values below 50 W/m² and its maximal values do not exceed 950 W/m². Visualization of GHI against day of the year and hour can be seen in Figure 5.2. According to cloud coverage data, the sky is almost completely overcast for the vast majority of the time (around 70%). Ambient temperatures remain between -9°C and 37°C. Rainfall is present in 13.4% of all samples and its maximal value equals 21.6 mm per hour or 37.14 mm per day.

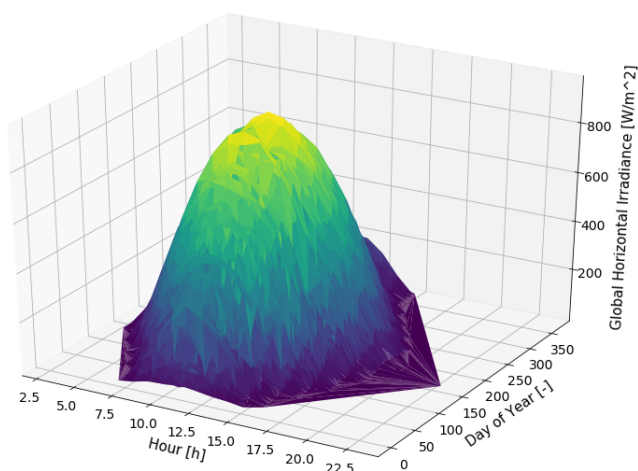


Figure 5.2: Distribution of global horizontal irradiance values along the day and year.

5

Exploration continued by looking for correlations in the data. Features used as inputs to ML models should not be correlated and should have possibly large variance [181]. The correlation was investigated by plotting a heatmap of weather features visible in Figure 5.3. GHI is positively correlated with sun altitude and negatively correlated with cloud coverage. That is intuitive, as clouds cause shading and reduce the incident irradiance on flat surfaces on the ground. Sun altitude has a strong positive correlation with ambient temperature, as the higher the sun is in the sky, the more irradiance reaches Earth's surface and the hotter the air becomes. Ambient temperature is correlated with the day of the year since it incorporates weather seasonality.

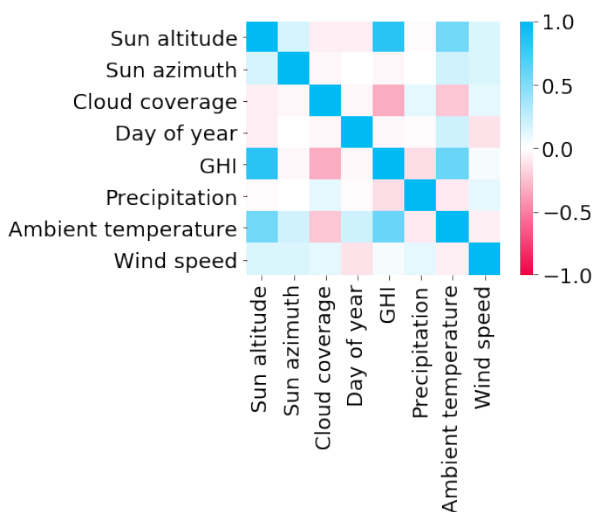


Figure 5.3: Correlation heatmap of the weather variables using hourly resolution.

5.3. METHODOLOGY

This section explains the methods for selecting the best subset of features (Subsection 5.3.1) and the models (Subsection 5.3.2) predicting the PV power. It additionally provides a discussion on loss functions in Subsection 5.3.3 and metrics in Subsection 5.3.4, which are pivotal parameters for machine learning algorithms.

5.3.1. FEATURE SELECTION

Given the high amount of features considered, it is important to determine which of them are relevant for the ML models. Feature selection is a process of eliminating the inputs that do not contribute to increasing the algorithm's performance [128]. Amongst the *sklearn* feature selection functions available, Recursive Feature Elimination (RFE) with 3-fold cross-validation was selected. This technique was used to increase confidence that the selected set of features is the best possible so that maximum performance could be achieved.

RFE eliminates features recursively given a certain metric and algorithm [199]. The function computes feature importance, drops the feature with the lowest importance, and recalculates the metric. The process is iterated until only a user-specific number of features is left. The importance of each feature is calculated by the chosen algorithm [200]. Each training process is repeated three times using different parts of the data to ensure the results are independent of shuffling. Once the best configuration is found, it is implemented in the ML models and used for prediction.

The selected algorithm strongly influences the RFE results. In this case, the algorithm used was XGBoost and the feature was Root Mean Squared Error (RMSE). Since XGBoost is composed of a combination of decision trees, as explained in Subsection 4.3.1, computing the importance of each feature is relatively straightforward. Importance indicates the utility of each feature during the tree construction, determined by the improvement it brings in each split, weighted by split observations, and averaged across all trees [201].

5.3.2. MODELS

This section briefly explains the working principle of Extreme Gradient Boosting (XGBoost), the main model employed in this publication. This proposed model is compared with persistence and Solar Monkey's model. Persistence is a baseline model commonly employed to compute the skill score (see Subsection 5.3.4) and presented in Subsection 4.3.1. Solar Monkey's model is the commercial model employed by the start-up and used here as a reference physics-based model.

The algorithm currently employed by Solar Monkey to compute the power produced by a PV system is based on an analytical approach. It makes use of the skyline profile to account for obstacles surrounding the modules and the sun's movement. Details on the framework can be found in [202].

When selecting the ML algorithm, the aim was to find one with proven performance and capable of efficiently handling large datasets. Additionally, it needed to be low in complexity and easy to optimize. Given these criteria, an algorithm from the ensemble family, known for excelling in ML competitions [183, 203], appeared to be the most suitable option. For instance, Gradient Boosting was employed in GEFCom competition when predicting solar power by three out of five top participants [160]. Inside the ensemble family, the algorithm of choice was extreme gradient boosting (XGBoost), which has already been employed for solar yield forecasting [64, 109, 110]. Ferlito et al. employed XGBoost together with 10 other

prediction models to forecast the yield of a 1 kW grid-connected PV system. Their results indicated that SVR was the best model out of the 11 investigated unless the dataset was limited to less than 60 days, in which case ensemble algorithms were the best alternatives [64]. In [110], a comparison of the performance of several ML models for PV power prediction was performed across five different sites in Sweden. Their main conclusion was that ANNs and Gradient Boosting Regression Trees performed best on average across all sites.

XGBoost starts by building a decision tree that finds a function $f_1(x)$ employing the input dataset x , whose output \hat{y}_1 resembles as much as possible the real output y [183]. In the next iteration, a refined decision tree will fit the pseudo-residuals $f_2(x)$ from the first tree so that the result \hat{y}_2 approaches more to the output, as in Equation 5.1. This process is repeated for a user-defined number of iterations M .

$$\begin{aligned}\hat{y}_1 &= f_1(x) \\ \hat{y}_2 &= \hat{y}_1 + f_2(x) \\ &\dots \\ \hat{y}_M &= \hat{y}_{M-1} + f_M(x)\end{aligned}\tag{5.1}$$

5

The new trees learn from previous trees' mistakes by increasing the weight of data points that were erroneously predicted. This way the algorithm gives higher attention to large-error predictions. The error penalization is determined by a loss function, which should match the problem at hand and is crucial for the model outcomes.

XGBoost has been implemented in Python using the library *sklearn* [184]. The model was tuned for 20 rounds using *RandomizedSearchCV* with 3-fold cross-validation, also available in *sklearn*. Performing cross-validation on all available data would cause *data leakage* [128, 181], which refers to evaluating model performance on samples that the model has already seen, leading to over-optimistic results. To prevent it, all the reported results were calculated on the test set being the remaining 20% of data. Cross-validation was therefore performed on 80% of all the systems, hence two-thirds of this value was the training set (52.8% of all the data) and the remaining third (26.4% of all the data) was used as the validation set. *sklearn GroupShuffleSplit* was employed to randomize the data so that the algorithm would learn the solar annual patterns. Squared error loss function together with RMSE as evaluation metric were used for model training. Outcome predictions were filtered to make all negative values equal to zero.

5.3.3. LOSS FUNCTIONS

Contrary to utility-scale PV forecasting, residential nowcasting requires high-quality predictions for all individual systems. For this task the required loss function should be i) twice differentiable as required by XGBoost [183], ii) sensitive to outliers, and iii) relative to the size of the system. Thorough research and analysis were performed in search of the most suitable loss function for the problem at hand.

Square, log-cosh, and pseudo-Huber losses (see Figure 5.4) are the main options due to their differentiability [204]. Amongst these, square loss is the most sensitive to outliers. However, neither of the three is relative to the size of the system. A relative loss, such as the absolute percentage loss would fulfill this condition, but it is not twice differentiable.

Several attempts were undertaken to develop a loss function that combines absolute and relative errors, yet none proved successful. The main issue is that loss functions focus

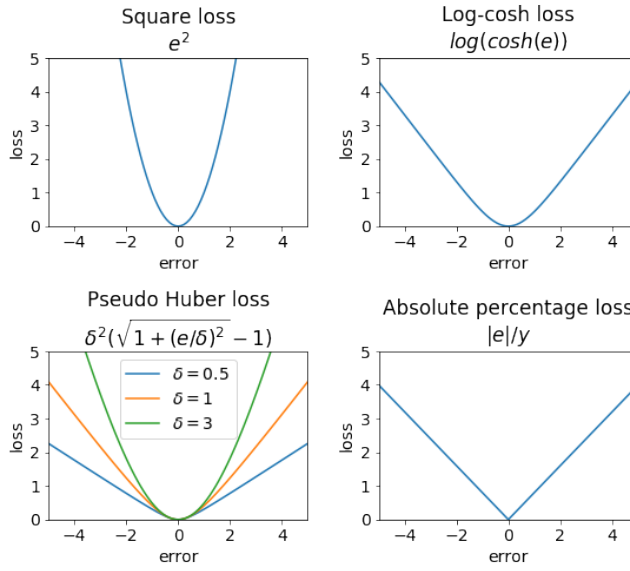


Figure 5.4: Loss functions considered during the search for the most suitable one. e is the error between predicted \hat{y} and real y values.

by definition on large errors. Similar is the case for models utilizing MAPE as the training evaluation metric. Tackling this issue could be a starting point for future studies. Developing individual models for all systems was also considered, but it would not solve this issue, as the models would perform better in summer (higher yields) and worse in winter.

Overall, due to the failure of the aforementioned methods, the sub-optimal square loss function was used.

5.3.4. METRICS

This section describes the metrics employed to evaluate the model performance while exposing metric-related issues and presenting their impact on the quality of predictions.

An extensive overview of available metrics was provided by [34, 44, 150] who pursued both statistical and economic approaches. They described interesting metrics such as skew, kurtosis, Renyi entropy, and Kolmogorov–Smirnov Integral among others. However, the most popular metrics for solar yield nowcasting and forecasting still are RMSE, MAE, and MAPE (see Section 2.7).

For this analysis, a metric that does not increase with system size, such as MAPE, could be the metric of choice. However, it has the drawback of taking high values for small values of yield. Other metrics used for this task are mean bias error (MBE), maximal error (max error), and skill score (SS). MBE allows negative and positive errors to cancel out, maximal error provides no information about error distribution and skill score is RMSE dependent hence influenced by the system size. Moreover, skill score informs about relative improvement compared to the persistence model whose performance highly depends on the employed version of the model and data resolution, as mentioned previously. These characteristics make the usage of skill score insufficient, even though it is promoted as a metric that

allows comparison across projects [47].

An interesting approach was developed in Solar Monkey, where the absolute and relative error were combined in one *E-metric* described by Equations 5.2 and 5.3.

$$E_e = \frac{1}{n} \sum_{i=1}^n f(\hat{y}_i, y_i) \cdot 100\% \quad (5.2)$$

$$f(\hat{y}_i, y_i) = \begin{cases} 1 & \text{if } |\hat{y}_i - y_i| < \epsilon \\ 0 & \text{if } |\hat{y}_i - y_i| \geq \epsilon \end{cases} \quad (5.3)$$

ϵ stands for the threshold of absolute error. E_{10} gives a percentage of predictions with absolute error below 10 Wh. Similarly E_{50} gives a percentage of predictions with absolute error below 50 Wh and so on. The higher the E-metric is, the better the prediction. Its drawback is that it requires different thresholds for different data resolutions.

Another issue is that most models are assessed on the entire test set without investigating into results for individual systems. While this approach may suffice for regional forecasts, it falls short for residential PV owners seeking to optimize their production and consumption [205]. The owners could store solar energy and sell it to the grid during peak hours to maximize their profits. Moreover, early anomaly detection requires precise and accurate yield nowcasting for individual PV systems. The general metric calculation might hide cases for which the model performs badly and does not include the fact that each system usually belongs to a different entity. In the case of Solar Monkey, measuring overall model performance is informative, but insufficient to determine whether all customers receive predictions of high quality. Therefore, each of the previously presented metrics was calculated per system, and then the minimum, maximum, mean, and standard deviation for each metric across the entire fleet of systems were computed. This approach facilitates the identification of the poorest-performing systems, thereby narrowing the scope of error analysis.

5.4. RESULTS

This section presents the main results of the study showing the features selected by RFE in Subsection 5.4.1, the performance of the proposed model compared with other models in Subsection 5.4.2, and the learning curves and error analysis of XGBoost in Subsection 5.4.3 and 5.4.4, respectively.

5.4.1. FEATURE SELECTION

Features selected by RFE with their corresponding weights assigned by the XGBoost algorithm can be seen in Figure 5.5. The most influential feature is GHI followed by historical yields, which coincides with similar results obtained in literature [55, 110]. System size, one of the few system characteristics considered together with orientation, is also a relevant feature. Since the ML model was trained for data from multiple PV systems, system size may be used as a scaling factor. Cloud coverage is considered but does not rank among the top 3 features. This aligns with the literature, which suggests that cloud coverage is particularly important for forecasts less than 15 minutes ahead [33].

All weather and descriptive features were kept by RFE. Out of all PV system parameters, only system size and orientation were selected, showing that most system characteristics are not as influential as expected.

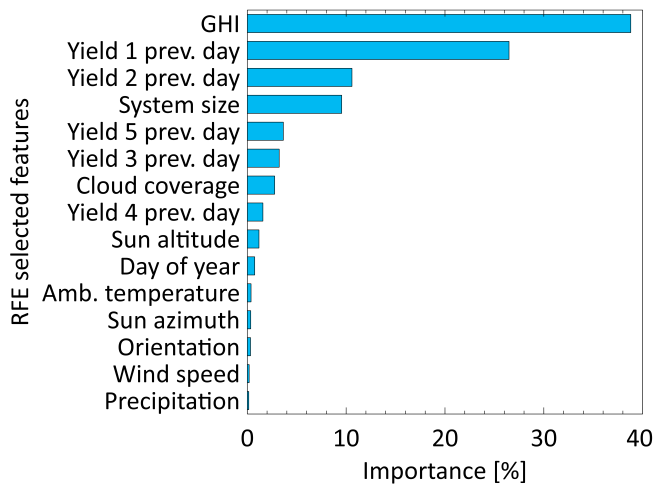


Figure 5.5: Importance assigned by XGBoost to each of the features selected by RFE.

The non-measured features, namely the sun angles and day of the year, have low relevance probably because the periodic information that they introduce is included in other features such as irradiance. One surprising result is the low importance of ambient temperature, which is usually amongst the chosen features [55, 69, 111]. Wind speed and precipitation are the two least relevant features, which coincide with previous results reported in the literature [55, 69, 111].

5.4.2. MODELS PERFORMANCE

After performing feature selection and hyperparameter tuning for XGBoost, the algorithm was trained. Since the objective is to monitor the residential PV systems, the lead time was instantaneous. The produced PV power was compared with the predicted one at the same instant of production to detect any malfunction. This occurred with an hourly resolution.

The performance of several ML algorithms and persistence is presented in Table 5.1 and 5.2. For information on these models, the reader is referred to Appendix B. These models were selected for comparison as they could be appropriate alternatives for commercial applications because they are easy to apply, provide fast results, and can handle large amounts of data. Deep learning methods may obtain lower errors but their lack of interpretability and hard optimization would hinder their use in practice in the start-up.

XGBoost outperforms all other models considered in this analysis in most of the employed metrics. Random Forest has similar performance, with smaller maximum error and MAPE, however, the higher RMSE implies that it makes predictions more often with large error. This is confirmed by the E-metrics in Table 5.2 where Random Forest has higher E_{10} , E_{50} , and E_{100} , but lower E_{500} . To increase trust in machine learning models, one must ensure the absence of significant errors. Therefore, E_{500} is the most important metric in this table and is one of the reasons for choosing XGBoost over Random Forest.

Table 5.3 displays individual PV system metrics calculated for the analytical commercial software and XGBoost. XGBoost exhibits individual system RMSE ranging from 0.4 to 4.6 kWh. While the minimal RMSE of XGBoost aligns with that of the analytical model, its

Table 5.1: Comparison of results calculated for all predictions together.

	R^2 [%]	MAE [kWh]	Max error [kWh]	RMSE [kWh]	MAPE [-]	SS [%]
Persistence	80	0.35	8.3	0.53	71	-
ElasticNet	83	0.30	32.2	0.47	11176	11
Polynomial Regression	90	0.24	30.1	0.37	10504	29
Random Forest	96	0.10	28.8	0.22	418	59
XGBoost	97	0.10	28.9	0.21	478	61

Table 5.2: E-metrics for hourly rough data set.

	E_{10} [%]	E_{50} [%]	E_{100} [%]	E_{500} [%]
Persistence	11.0	22.1	32.9	75.2
ElasticNet	3.2	15.9	30.5	82.1
Polynomial Regression	4.2	20.5	38.2	87.8
Random Forest	27.4	57.4	72.5	96.4
XGBoost	23.6	54.7	71.4	96.7

maximal RMSE is approximately half that of the analytical model. Additionally, the mean per-system RMSE decreased by around one-third. Usage of the ML model caused a drop in relative error (mean MAPE) from around 44% to 23% which is almost a two-fold improvement.

Table 5.3: Comparison of metrics for XGBoost and the analytical commercial software

Individual system metrics	Commercial software	XGBoost
min RMSE [kWh]	0.4	0.4
max RMSE [kWh]	9.4	4.6
mean RMSE [kWh]	1.8	1.4
mean MAPE [%]	43.6	23.0

Distributions of per-system RMSE and per-system MAPE can be seen in [Figure 5.6](#). The distributions corresponding to the XGBoost model exhibit a leftward shift compared to the analytical model which indicates the model's superiority.

5.4.3. LEARNING CURVES

Next to XGBoost metrics, learning curves are presented to depict the error reduction as the model learns patterns in an increasingly large dataset. Learning curves help in assessing the bias-variance trade-off and making an informed decision about the next development step. They also aid in estimating the amount of data needed to decrease the error by a certain value and therefore are important in financial calculations. Additionally, they facilitate the discovery of learning saturation, which is the point above which further training almost

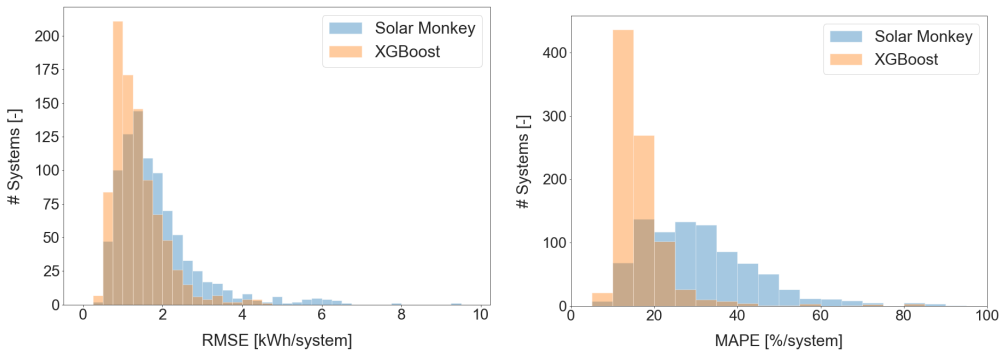


Figure 5.6: Distribution of RMSE (left) and MAPE (right) for individual systems comparing the performance of the commercial software with that of XGBoost.

does not decrease the error.

The learning curve in Figure 5.7 was obtained using 3-fold cross-validation and determines whether full data set potential was utilized. Semi-transparent areas around the lines in the figure correspond to standard deviations of results for all three folds. Standard deviations are significant for small data sets and diminish with increasing training set size. This trend is anticipated as for large data sets the particular shuffling of the training data is less likely to skew the result. Therefore, cross-validation can be neglected for data sets larger than 0.15% of the training set, that is exceeding 60,000 samples, as it significantly increases computational cost and has almost no impact on the results. Until seeing 0.15% of the training set size, XGBoost validation error decreases exponentially and it seems to stabilize later on. Drop in validation RMSE is only around 0.02 kWh for training set size between 40% and 80%. XGBoost learning saturates around 40% of the training set, which is roughly 1.28 million samples.

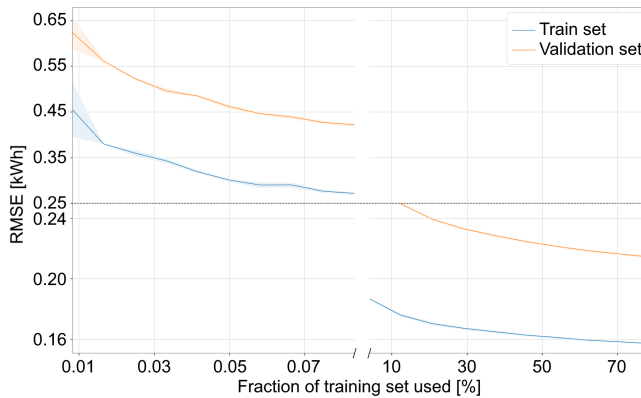


Figure 5.7: XGBoost learning curve during and after the initial learning phase. Be aware of the change of scale in both axes. The curve was plotted for custom training set sizes, with non-constant step values.

Given that each system in the dataset contains 4609 samples corresponding to non-zero irradiance over a year, the XGBoost algorithm saturates with data from approximately 278

PV systems. Further increase in training set size is likely to decrease error, but this small gain is not justified considering the surge in required computational power and the associated financial cost. Other operations, such as data cleaning, are likely to provide larger gains in performance. The usage of fewer systems with larger time horizons is likely to provide similar results, as long as the systems selected for training are representative of the test set.

5.4.4. ERROR ANALYSIS

This subsection focuses on finding the pattern in large XGBoost errors. Based on the literature, it is expected that the model works better under particular weather conditions, such as clear days [103], seasons [101], or for certain systems [89].

Figure 5.8 depicts two graphs for each investigated parameter: distribution of absolute percentage error (left-hand side graphs) and distribution of values (right-hand side graphs). This informs whether the magnitude of the percentage error corresponds to the number of training samples in each bin. XGBoost's performance is independent of wind speed and ambient temperature but depends heavily on GHI and to some extent on cloud coverage. For GHI and cloud coverage, the largest absolute percentage errors overlap with bins containing the largest number of samples. As presented in Figure 5.5, the model did not recognize wind speed and ambient temperature as relevant, indicating their minimal impact on the predictions.

The model's performance as a function of cloud coverage is presented in Table 5.4. The model performs around 50% better during clear sky conditions which supports that modeling clouds' influence is challenging. It also indicates that the utilized cloud coverage data is insufficient to precisely capture real-life changes.

Table 5.4: XGBoost model performance for different cloud coverage ranges.

	Metric	Persistence	XGBoost
Clear sky (<1 okta)	MAE [kWh]	0.55	0.15
	RMSE [kWh]	0.73	0.28
	MAPE [%]	27	6
	R ² [%]	69	95
Partly cloudy (1-7 okta)	MAE [kWh]	0.58	0.24
	RMSE [kWh]	0.77	0.38
	MAPE [%]	28	11
	R ² [%]	60	90
Completely overcast (>7 okta)	MAE [kWh]	0.57	0.23
	RMSE [kWh]	0.76	0.37
	MAPE [%]	32	12
	R ² [%]	46	87

Next to the weather analysis, an error analysis was performed. As previously discussed, the utilization of the squared error loss function favors large yields, that is large PV systems and sunny hours. To further verify this hypothesis, observed yields were plotted against their corresponding absolute percentage error (APE) in Figure 5.9. APE has values far exceeding 100% for yield values below 2 kWh which confirms the initial assumption. APE against the hour of the day and month were plotted in Figure 5.10a and 5.10b, respectively. The relative error is the largest just after sunrise and just before sunset, and in December

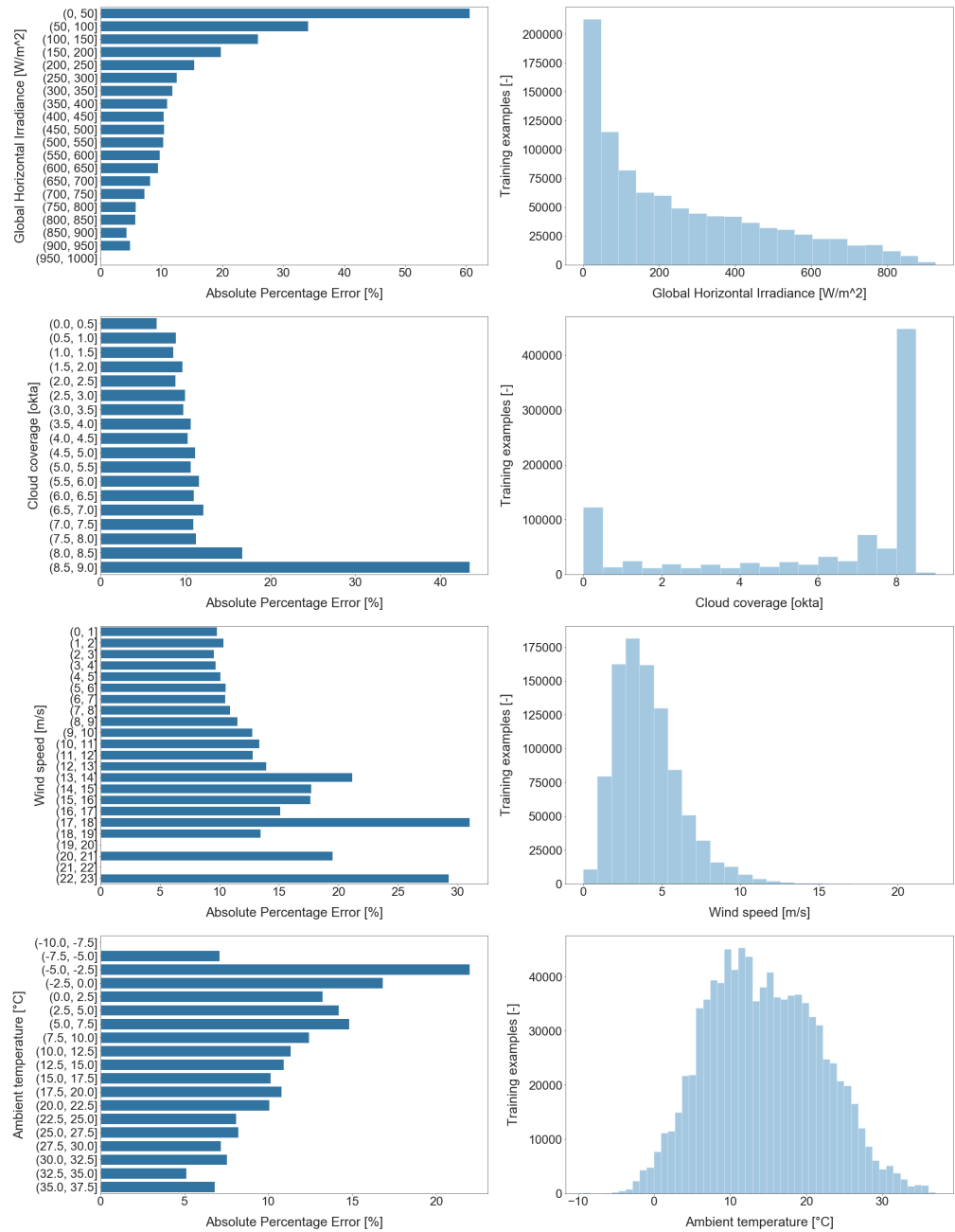


Figure 5.8: Weather features as a function of absolute error.

and January. These time instants receive the least irradiance and therefore produce the smallest yields.

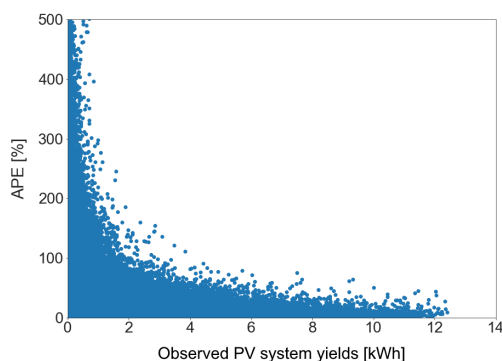


Figure 5.9: Absolute percentage error as a function of observed yield for all systems.

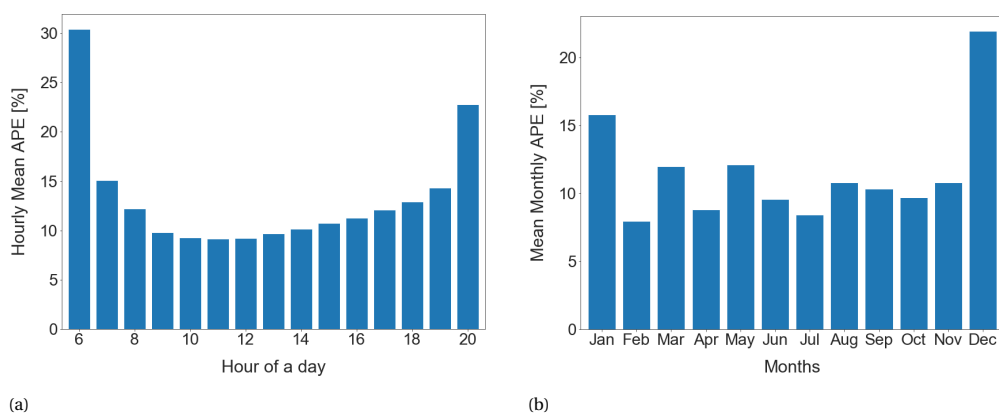


Figure 5.10: Error analysis against (a) time of the day and (b) month of the year.

5.5. CONCLUSIONS

This work presents an effective strategy for monitoring residential PV systems using ML techniques. XGBoost algorithm turned out to be the method of choice for the task of solar yield nowcasting on hourly data surpassing persistence, ElasticNet, Polynomial Regression, Random Forest, and a commercially available analytical model. This work discussed that the currently used metrics and loss functions are insufficient for the evaluation of solar forecasting models, as they fail to combine relative and absolute error. Combining the two is necessary for developing a single model able to learn individual system properties, independently of their size, and to obtain high-quality residential scale solar yield predictions. In this study, the usage of the squared error loss function caused the model to be significantly mistaken for small values of yield occurring in winter, close to sunrise and sunset, and for small PV systems. Attempts to tackle this issue through sample normalization and using MAPE as a training metric have failed. Further research should focus on developing new loss functions and evaluation metrics. Despite the presented issues, XGBoost exhibits a twofold improvement compared to the commercially available analytical model.

6

FAULT DETECTION USING PEER SYSTEMS

As seen in the previous chapter, approaches employing weather data for residential PV power forecasting are unreliable due to the lack of on-site measurements. This chapter proposes a methodology that uses instead data from similar PV systems. This work improves the so-called performance-to-peer approach by adding system characteristics and optimizing with ML techniques. The methodology has been tested in a fleet of more than 12,000 Dutch systems with up to 7 years of data per system. The proposed model achieves an average R^2 of 94.1% and an NRMSE of 0.05, outperforming in terms of R^2 the baseline model by 1.4 points, and the analytical approach by 3.8. The proposed approach has also been assessed for fault detection and categorization and excels in differentiating system mismatch from actual faults and adapting to new situations through retraining.



This chapter has been adapted from the publication A. Alcañiz, M. M. Nikam, Y. Snow, O. Isabella, H. Ziar “Photovoltaic system monitoring and fault detection using peer systems” *Progress in Photovoltaics: research and applications*, 30(9), 1072-1086 (2022) [206].

6.1. INTRODUCTION

The conventional approach employed in monitoring is to use design information along with accurate weather data [34, 44, 207]. This is suitable for large or commercial solar farms where expensive sensors can be installed on-site. However, for residential PV systems, extrapolated weather data has to be used. This rarely represents the local weather as the weather stations are often far away from the actual site of PV system installation. Thus, local phenomena like a cloud passing over the area cannot be accounted for in the calculation of expected energy generation. This problem was observed in the previous chapter.

To overcome the problem of weather data, an alternative approach was developed where neighboring or peer systems are used to monitor each other. A peer system is a system similar to the monitored PV system. It could be physically nearby or have similar design parameters. This way, if a cloud passes over one PV system, it will also pass over a geographically close peer system, and thus, their yields can be compared to detect any fault. Since the approach is based on historical energy yields along with the peer predictions, it will also overcome any monitoring error due to incorrect design information.

Several works can be found where the neighboring PV systems were employed to improve the predictions on the residential scale. The studies [208, 209] employed neighboring systems to determine the irradiance variations due to cloud motion and then forecasted energy production using analytical models. Golnas et al. predicted the output of a PV system from a regional fleet by using data from other systems in the fleet based on historical performance correlation between the systems [210]. Tsafarakis et al. developed a method for fault detection considering that the power produced by a PV system is linearly related to the power produced by neighboring systems [211]. Next, Popovic and Radovanovic presented a methodology for inter-system comparison between correlated PV systems to estimate the operation status of individual modules in urban surroundings [212].

The most relevant approach to using peer systems to monitor the PV power was presented by Leloux et al. They defined a novel performance indicator called Performance to Peer (P2P) which was computed by comparing the energy production of several neighboring PV systems [213, 214]. They proposed it as an alternative to the Performance Ratio (PR), commonly employed for monitoring and automatic fault detection. They showed that P2P was more stable than PR when monitoring 6000 PV installations across Europe with energy output data for approximately 7 years and a temporal resolution of 10 minutes. *Stability* was interpreted by the authors as the ability to easily distinguish between the absence and presence of a fault.

This chapter aims to extend the work performed in [213]. The database will be expanded to more than 12000 PV residential systems located in the Netherlands, with diverse system characteristics and installation ages. System information will be employed beyond just geographical closeness. In contrast to some research papers [195, 208, 209, 211, 212], weather data is not employed for finding similar systems to avoid inaccuracies. Furthermore, in line with the recommendations of [214], ML algorithms are used to optimize the model and increase its accuracy. The application of the developed model for fault detection and categorization will also be demonstrated.

The structure of the work is as follows. Section 6.2 presents the employed models. The data used is briefly described in Section 6.3. Main results are displayed in Section 6.4 where the superior performance of the proposed algorithm is shown, together with its limits. The main application of the proposed model is demonstrated in Section 6.5 by presenting the

fault detection approach and showing the fault categorization performed. Finally, [Section 6.6](#) presents the main conclusions.

6.2. MODELS

This section explains the main models employed in this work. The first [Subsection 6.2.1](#) describes the P2P model developed in [\[214\]](#). With this approach as the base, several improvements are made yielding the proposed model in [Subsection 6.2.2](#). Optimizations for the latter approach are elucidated in [Subsection 6.2.3](#). Finally, [Subsection 6.2.4](#) presents a series of models that will be used for comparison in later sections.

6.2.1. PERFORMANCE-TO-PEER MODEL

In the P2P model, the performance of one system is compared with its peers to monitor the former system [\[214\]](#). The system to be monitored is defined as the *focus* system, whereas all other available systems are referred to as *peer* systems. The model is divided into two steps: identifying the good peers in the fleet and calculating the expected yields by using the data of the good peers.

The energy yield data is the only characteristic employed to find good peers for the focus system. The yields are normalized relative to their total capacity resulting in the Capacity Utilization Factor *CUF* as shown in [Equation 6.1](#). In this equation, E_{PV} is the energy output of the PV system, P is the rated power of the PV system, and T is the time interval of energy output measurement. With the normalized daily yields, the Capacity Utilization Ratio *CUR* is calculated for each focus-to-peer system pair, according to [Equation 6.2](#). Next, the weighting factor is calculated by taking the inverse of the Median Absolute Deviation (*MAD*) of *CUR* and raising it to fourth power ([Equation 6.3](#)). The exponent was determined through a sensitivity analysis performed in [\[214\]](#). This weighting factor is used to determine whether the peer system is a good peer system or not. The higher the weighting factor, the better the peer for the chosen focus system.

$$CUF = \frac{E_{PV}}{P \cdot T} \quad (6.1)$$

$$CUR = \frac{CUF_{focus}}{CUF_{peer}} \quad (6.2)$$

$$weighting\ factor = \frac{1}{[MAD(CUR)]^4} \quad (6.3)$$

The data of good peer systems is used to calculate the expected yields. Once the top 10 peers are chosen as good peers, a weighted median of the peer systems' *CUF* based on their weighting factors is calculated to give the reference CUF_{ref} . CUF_{ref} can be recognized as the normalized expected yields for the focus system. Hence, the true expected yields can be estimated by reverse calculation of *CUF*. Additionally, the model further calculates P2P for each day by taking a ratio of the focus system CUF_{focus} to reference CUF_{ref} . This new metric is proposed instead of the conventional Performance Ratio (PR) for the monitoring of the focus system [\[214\]](#).

6.2.2. PROPOSED MODEL

The proposed model is based on the above Performance-to-Peer model. The main difference between the two is the use of a higher amount of PV system data. Moreover, system design information, daily yields, and system location are used to find good peers for the chosen focus system, not only system yields. Just like P2P, the proposed model is divided into two steps: distance calculation to identify the good peers and expected yield calculation using the data of the good peers.

DISTANCE CALCULATION

The proposed model starts by calculating the similarity between two PV systems. This similarity is computed by using various *distances*, which indicate that the more alike two systems are, the lower their distance. Therefore, the distance should not be necessarily interpreted as the geographical distance between the two. Three distances are computed (one for each characteristic), normalized, and later combined. Normal standardization was applied to all distances to ensure a fair comparison.

To begin with, the *feature distance* d_{feat} considers the system design information. This includes the number of panels n , panel inclination θ , and panel orientation ϕ . These variables were chosen after a feature correlation analysis and experience-based selection. The model calculates the Euclidean distance for each focus-to-peer system pair using Equation 6.4, where sub-index f represents the focus PV system and sub-index p represents the peer PV system. When computing the Euclidean distance, each of these attributes has a different weight w , representing the importance that the features have individually.

$$d_{feat} = \sqrt{w_n \cdot (n_f - n_p)^2 + w_\theta \cdot (\theta_f - \theta_p)^2 + w_\phi \cdot (\phi_f - \phi_p)^2} \quad (6.4)$$

Next, the *yield distance* d_{yield} is calculated as in the P2P model. The daily yields are normalized relative to the actual energy yield in the first year of installation of the system (for systems older than one year) or to the estimated energy yield for the typical meteorological year for the system (for systems less than a year old) to obtain the *CUF*. For each focus-to-peer system pair, the *CUF* is calculated for each day according to Equation 6.2. Since a single value is needed for the distance, the median absolute deviation is calculated from all the daily *CUF* values.

The third and final distance is the *geographical distance* d_{geo} . It is the physical distance between the latitude and longitude coordinates of each focus-to-peer system pair. Due to the curvature of the Earth, haversine distance is used [215].

After the three distances are computed for each focus-to-peer system pair, they are combined into a *total distance* d_{tot} using the weighted sum of the distances according to Equation 6.5. A weighted sum is used to consider a different influence for each distance when finding good peers. Once the total distance is known for each focus-to-peer pair, the pairs with the lowest distances, i.e., the peer systems most similar to the focus system, are chosen as good peers.

$$d_{tot} = w_{feat} \cdot d_{feat} + w_{yield} \cdot d_{yield} + w_{geo} \cdot d_{geo} \quad (6.5)$$

EXPECTED YIELDS CALCULATION

The expected yield of the focus system is calculated using the daily yields of the selected peers. Each of these peer systems has a different level of influence on the expected yields.

The higher the similarity of the peer system to the focus system, the higher its influence. Since the total distance was a measure of the similarity, the weighting factor λ_p per peer system is evaluated as the reciprocal of the corresponding total distance. This is expressed in Equation 6.6, where N_p is the number of selected peers.

$$\lambda_p = \frac{1}{d_{tot,p}} \quad \text{restricted to} \quad \sum_{p=1}^{N_p} \lambda_p = 1 \quad (6.6)$$

These weighting factors along with the CUF of good peer systems are used as a weighted median to determine CUF_{ref} . A weighted median is used over a weighted average as the median is not influenced by an abnormal extreme value that may be present for one of the good peers. This abnormal extreme value could be a fault in a peer system, so using the median essentially helps to prevent that fault from being transferred in the predictions of the focus system. The expected yield is estimated from CUF_{ref} by reverse calculation of CUF .

Figure 6.1 shows the procedure to obtain the expected yield of a focus system with the proposed model. For each peer system, all the distances are computed, as just explained. From the total distance, the weights can be determined, hence the peers can be selected. The variables outside the boxes in this outline need to be determined via optimization, as will be explained in the next subsection.

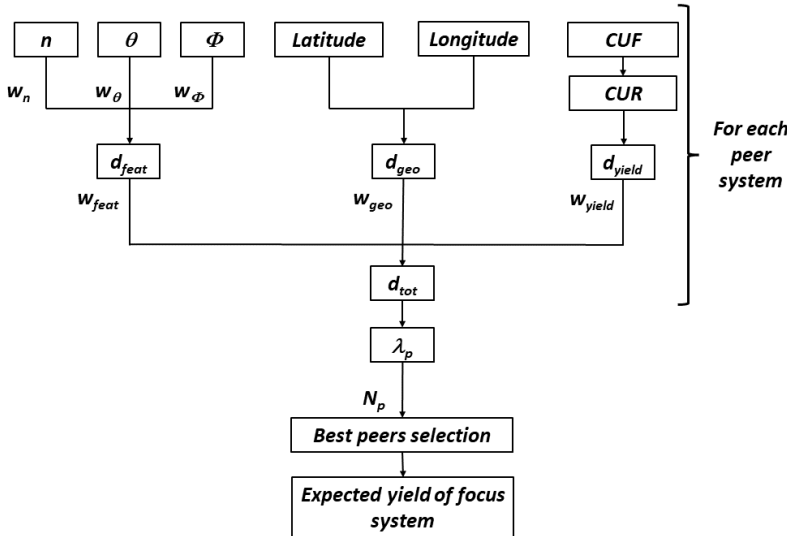


Figure 6.1: General outline of the proposed model for one focus system.

6.2.3. MODEL OPTIMIZATION

While explaining the proposed model in the previous section, several variables emerged that require optimization to ensure the optimal performance of the model. The unknowns are the weights w of each distance and of the individual PV system features, and the number

of peers N_p that need to be chosen for estimating the expected yields. The value of these variables is determined by training the model using an optimization algorithm.

Given the complexity of the problem at hand, local optimization algorithms, such as gradient descent [216] and Nelder-Mead methods [217], were stuck in local minima and unable to find the global optimum. Global optimization algorithms were needed. Amongst the tested global optimization algorithms, evolutionary algorithms gave the best results. Inside this group, particle swarm optimization (PSO) and genetic algorithm (GA) were the most promising ones. Both algorithms are relatively simple and easy to modify, and their performance depends on the problem at hand because they traverse the candidate space rather differently [218].

To select between the two, a literature search was done to find the best algorithm for problems using peer-to-peer strategies. Abraham et al. used PSO and GA for neighbor selection in peer-to-peer networks, and reported that GA obtained better results [219]. Similarly, Rehman et al. employed GA and PSO to optimize the peer-to-peer energy transactions in a decentralized energy trading market and saw that GA results outperformed the PSO ones [220]. GA was also combined with PSO for device-to-device (D2D) communication in advanced communication networks [221]. The authors claimed that PSO alone can be trapped in local optima due to premature convergence, hence a hybrid PSO-GA algorithm was proposed to find the optimum allocation of the D2D communication network's resources and avoid interference with the primary cellular network. Given the similarities of these problems with the one at hand, GA was finally chosen as the optimization algorithm.

Genetic algorithm is a global search engine inspired by natural evolution [222]. GA aims to obtain the optimum value of a fitness or objective function, which represents the performance of the problem. The higher the fitness value, the better the system's performance. In our case, the goal is to minimize the average Mean Absolute Error (MAE) between the actual and expected yields of the focus systems. The optimum solution is found via a generational process that consists of fitness evaluation, selection, crossover or reproduction, and mutation [223] as seen in Figure 6.2.

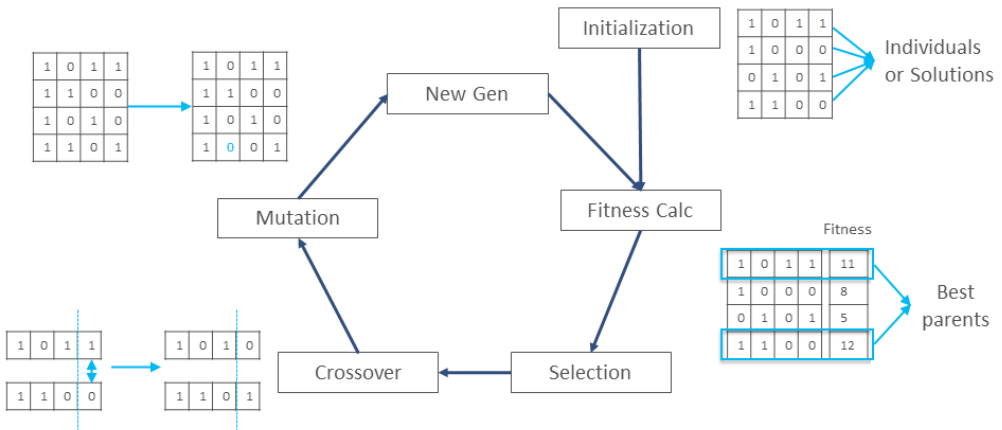


Figure 6.2: Overview of the steps of Genetic Algorithm: Initialization, Fitness Calculation, Selection, Crossover, Mutation, and New Generation creation. The cycle ceases when a termination condition is reached.

The problem is initialized by a set of randomly selected *chromosomes*, which are candidate solutions to the optimization problem. This whole set is called *population*. Each of the variables to be optimized is named *genes* so that a set of genes forms a chromosome. In the first step, the fitness function for each of these chromosomes is computed (fitness evaluation), and the ones with the highest fitness value are chosen (selection). These selected chromosomes will be combined in the reproduction step, generating the crossover population. A mutation is incorporated into the algorithm to prevent the population from stagnating at any local point. It consists of altering one or more genes in a chromosome from its original state. The offspring together with the muted chromosomes will form the next generation. The algorithm repeats the generational process until a termination condition has been reached.

There are several modifications to this GA method with more complex interactions [224, 225], however, here we employed the simplest one with the most popular operators [226]: fitness proportionate selection method for parent selection and one-point crossover for the reproduction step.

6.2.4. BASELINE MODELS

This subsection provides a brief explanation of the algorithms used for comparison with the proposed model in sections 6.4 and 6.5. These are models currently developed at Solar Monkey [198] for PV system monitoring and fault detection.

The current approach of PV system monitoring at Solar Monkey explained in Subsection 5.3.2 is referred to as *analytical* model. This approach is employed for fault detection by using a fraction of the actual yield over the expected yield.

Based also on a physics-based approach, Solar Monkey developed a method for fault detection which consisted of estimating the number of panels that would generate the actual energy yield: the *sizing yields* model. Comparing the estimated amount of panels with the actual value, over and underestimations in the PV systems can be detected.

The third method for fault detection is based on historical yields, referred to as the *year-over-year* model. For systems older than a year, the PV power produced is compared to that one year earlier.

6.3. DATA

The data employed comes from the fleet of PV systems available at Solar Monkey [198]. It consists of the daily energy yield and main characteristics of 12229 roof-top Dutch systems, for a period ranging from 2 months to 7 years. While finer data resolution such as hourly data can provide higher quality results, it was available only for a limited number of systems, hence daily energy yields were employed.

Data cleaning was performed, consisting of removing systems with large amounts of missing data, after which the total number of PV systems was reduced to 9480. A 10:1 system split between the training and testing set was used to reduce the computational burden while ensuring a sufficient number of systems for testing.

6.4. RESULTS

This section provides the results of the proposed algorithm. In Subsection 6.4.1, the main outcomes of the optimization of the proposed model are explained. Then, the performance

of the proposed model is compared to that of the other available PV system monitoring algorithms, namely P2P and the analytical model, in [Subsection 6.4.2](#). Finally, the limits of the proposed model are found and discussed in [Subsection 6.4.3](#).

6.4.1. OPTIMIZATION

Once the model was developed and the optimization algorithm was chosen, GA was trained on a set of random 5000 systems with each having daily yields for up to one year. The whole dataset was not used due to computational and memory limitations, although the peers were found from the whole set of 9480 PV systems. The results of the optimization can be found in [Table 6.1](#).

Table 6.1: Parameters of the proposed model optimized by GA.

w_n	w_θ	w_ϕ	w_{feat}	w_{yield}	w_{geo}	N_p
0.01	0.24	0.75	0.13	0.87	0	12-20

The first three weights correspond to the individual characteristics of the PV systems employed to compute the feature distance. Amongst the three, the panel count has a very low weight, probably due to the normalization of the PV system yields. Panel tilt and panel azimuth have a 1:3 weight distribution. Regarding the weights of the three distances, yield distance has the highest weight among all three, while geographical distance has negligible weight.

The number of peers needed for accurate calculation of expected yields is in the range of 12-20. Upon multiple simulations, it was observed that the fitness value of GA does not change significantly within a given range of number of peers. One possible reason is that the model uses a weighted median when calculating expected yields. Thus, within a certain range, the value of the weighted median does not change the result by a significant amount. This topic will be further discussed in [Subsection 6.4.3](#).

The low weight of geographical distance seemed suspicious considering that location was the only feature employed in previously published peer-to-peer PV monitoring models [227–229]. Thus, further exploration was required. The first hypothesis was that there existed a high correlation between feature and yield distances. The Spearman correlation coefficient [230] was therefore calculated, and resulted in a value of 0.53, which was not significant enough. Despite this, the optimization model was run again without the yield distance. Geographical distance still had a low weight of 0.07, while now most of the weight was skewed towards feature distance. Another hypothesis considered was related to the data itself. The PV systems available are only from the Netherlands, which is a flat and small country. Geographical distance enables the peer systems to account for the changes in very local weather like clouds near the focus system. This is important when dealing with hourly or higher-frequency energy yields. However, for this project, daily energy yields are used and it can be postulated that in the Netherlands, the day-to-day weather is very similar throughout the country. With this logic in mind, it is possible that when dealing with a larger area or with non-uniform weather conditions in a different environment, geographical distance will have a higher impact.

In our case, this low weight of geographical distance indicates that peers are not necessarily located close to each other. As can be seen in [Figure 6.3](#), peers of a focus system can be either closer to it or farther apart from the focus system. This new finding gives more

flexibility to P2P approaches, especially when there are not many systems around a certain focus PV system.

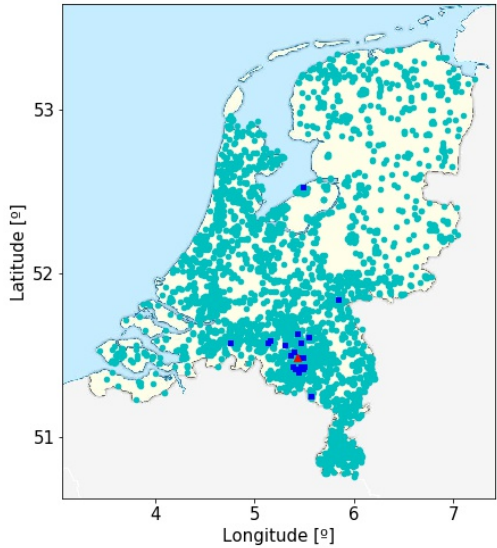


Figure 6.3: Spatial plot of good peer systems for a random focus system. The focus system is indicated by a red triangle, the good peer systems by blue squares, and all available systems by cyan circles.

6.4.2. MODELS COMPARISON

Once the proposed model was optimized and trained, its performance was assessed in the testing dataset. This set was composed of 500 PV systems with data for up to 3 years.

To properly test its accuracy, the performance of the proposed model was compared to that of the P2P and the analytical model using the NRMSE (normalized by the difference between maximum and minimum actual values) and R^2 score. Table 6.2 depicts the mean value of these metrics for all systems in the testing set for the three models. Additionally, the R^2 score was reinterpreted as a new metric: Percentage of Good Systems (PGS). Considering that an R^2 score higher than 85% is a good fit, PGS is the percentage of PV systems with an R^2 score higher than 85%.

Table 6.2: Metrics of the three monitoring models on the testing set.

	NRMSE [-]	R^2 [%]	PGS [%]
Analytical	0.08	90.3	85.2
P2P	0.06	92.7	88.1
Proposed	0.05	94.1	92.5

All metrics show that the cooperative models have higher accuracy than the analytical one. The analytical model underestimates the expected yields, while the P2P-based models are more accurate. This information is also represented in the error plot of Figure 6.4.

Table 6.2 also depicts that the proposed model has better metrics than the previously published P2P model [214]. Hence, the addition of PV system data and the optimization with GA was an improvement.

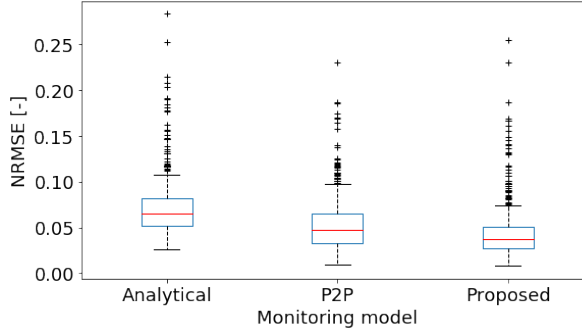


Figure 6.4: Error plot showing the NRMSE for all the test systems for each monitoring model.

6.4.3. USE-CASE ANALYSIS

This subsection elaborates on the limits of the proposed model. Characteristics such as the minimum data requirements for good performance, and the sensitivity to the number of peers employed are explored. These would give a better perspective on the proposed model and show its robustness.

The first experiment explores the minimum data requirements for the model, specifically the minimum number of systems and the minimum number of days of data per system. 56 simulations were performed where both the amount of PV systems and the number of days of each system were tuned. The trained systems still look for peers in the complete dataset of 9480 PV systems. Since cross-validation was not possible due to the high computational requirements, it was ensured that the training and testing sets had a similar distribution as the total dataset to have an accurate representation. The results of the experiment can be seen in a heatmap in Figure 6.5.

The number of systems does not have a strong influence on the model performance. However, if only a few months of data (<6 months) is available per system, the performance degrades strongly. Thus, when it comes to training the model, the number of days per system is important while the number of PV systems used for training is not as long as a minimum of 100-250 are available.

Although the experiment gives an understanding of the minimum data requirement for optimization, it was interesting to observe that even 100 PV systems might be fairly good enough to train. Thus, the results of the experiment were portrayed differently as seen in Figure 6.6.

In this graph, the x-axis shows the total number of years of data for each run. Thus, for grid point (4000, 365) in the heatmap, the corresponding value on the x-axis would be 4000 years of data. A trend line was generated on the experiment results as a help to the human eye. According to the trend line, data higher than 1700 years is more than sufficient for training the model, i.e., achieving R^2 score greater than 90%. This translates to 1700 PV systems with one year of data each or 850 PV systems with two years of data each and so on.



Figure 6.5: R^2 score as a function of number of days and number of training systems for minimum data requirement. The higher the R^2 score, hence the lighter the color, the better.

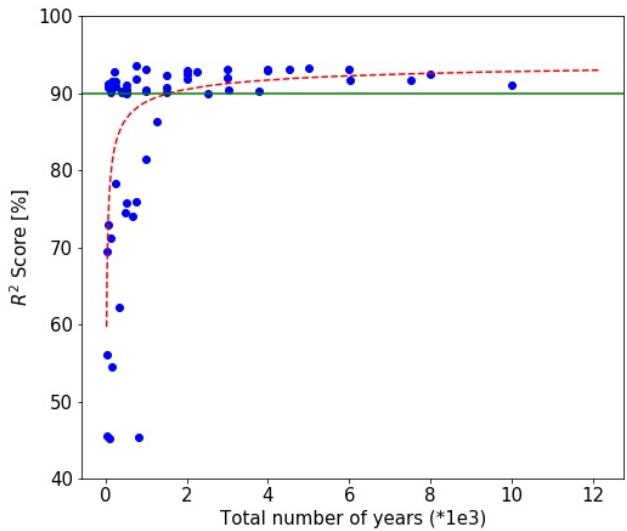


Figure 6.6: Minimum data required: R^2 score as a function of the amount of data employed. The red dashed trend line is a help to the eye, while the green continuous line indicates an R^2 of 90%.

It is important to note that data lower than 1700 years can also lead to good performance, although it is not as reliable. The trend line saturates around 4000 years of data, which is about 25% of the total data available.

The second experiment conducted, also related to the previous one, was to determine

the minimum data requirement when a new element is incorporated into the fleet. When a new PV system is added, energy yield data for the first 30 days of operation is sufficient to locate the good peers. These peers can be used for estimating the expected yields for the next one to two months with fairly good accuracy ($R^2 \approx 85\%$). After this time, good peers need to be located again with the new information available. These new good peers can be used accurately for another few months and so on. This process needs to be repeated until a sufficient amount of yield data is available, at least 6-9 months. After one year of active monitoring, the frequency needed for distance training is low as long as the system parameters stay fairly constant.

The third and final experiment explored the best value for the number of peers. As mentioned in [Subsection 6.4.1](#), while the optimum value of the weights w does not usually change, the variable number of peers varies considerably over different simulations. Thus a learning curve on the number of peers was plotted in [Figure 6.7](#) to find the range of the number of peers that gave high performance.

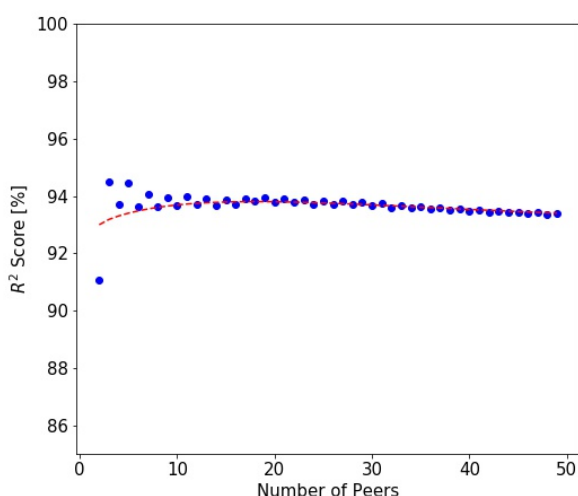


Figure 6.7: Learning curve of number of peers against R^2 score.

The simulation has the highest R^2 score when the number of peers varies between 12 and 20. Less than 12 peers cannot ensure an accurate prediction due to high dependence on only a few systems. More than 20 peers also lead to lower accuracy in predictions most likely due to additional noise resulting from not-so-adequate peers, although this effect is highly mitigated thanks to the use of the median.

6.5. FAULT DETECTION

The main objective of this work was to develop a model for monitoring residential PV systems using a peer-to-peer approach. Monitoring PV systems consists of a two-step process: calculating the expected yield for the PV system and comparing that yield with the actual

one to find any faults in the system. Following the development of the proposed model, this section centers on the second step: fault detection.

The typical approach for fault detection consists of first filtering out poorly performing systems, and then finding out the origin of the poor performance by inspecting the PV system characteristics. The proposed model together with the analytical, system sizing, and year-over-year ones explained in [Subsection 6.2.4](#) are employed in this process. These models are mainly used to detect poorly performing systems and classify the type of fault. To do the former, the performance factor (PF) is calculated for all the expected yields. This metric computes the ratio of actual yield to expected yield, as shown in [Equation 6.7](#). When the PF is outside an experience-determined range, the system is considered faulty, hence it is selected for further inspection.

$$PF = \frac{\text{Actual Yield}}{\text{Expected Yield}} \quad (6.7)$$

An essential step in this process is to distinguish between system design mismatch and fault detection. In the case of a system design mismatch, the design should be simply updated to match the actual installation. Comparatively, in the case of fault detection, there is a drop in performance due to a problem with the system. In this latter case, the owner of the PV system should be notified to solve the issue and act accordingly.

Using the above two-step process along with the expected yield and actual yield temporal plots, some PV systems were scrutinized to diagnose the most common faults. Based on this analysis, a fault categorization framework was developed, whose details are provided in [Subsection 6.5.1](#). Higher focus is given to two of these faults in [Subsection 6.5.2](#) where the added value of using a collaborative approach is highlighted.

6.5.1. CATEGORIZATION

This section categorizes the most commonly encountered faults. 120 randomly chosen PV systems were checked to gain insight into the occurrence and type of faults following the flowchart from [Figure 6.8](#). They were selected to ensure a good distribution of new and old systems. Only the data for June 2021 was used for categorization, although the energy yields for the entire lifetime of the PV system were looked at for inspection. June was selected as in summer months PV systems are expected to produce the most energy yield in the year. A malfunction in summer could lead to significantly larger energy and monetary loss for the owner than a malfunction in winter. Furthermore, fault detection in winter becomes relatively harder due to very low energy yields.

Four criteria were used to separate the systems into groups. The criteria are the expected yields of the analytical model, proposed model, year-over-year, and sizing yields. The fault categories were no fault, missing data, under-performance, over-performance, and false positive as usual ones, and additionally peer-to-peer failure. Each of these categories can have multiple combinations of the four criteria and they may or may not lead to a different fault diagnosis. The occurrence share of each of these faults can be seen in [Figure 6.9](#). An additional category is included in the pie chart for systems whose fault was detected but its origin could not be determined.

To begin with, a baseline of no-fault would be when all four criteria are within an acceptable range, here when PF is between 93% and 120%. Most of the systems (62.5%) fell within this category. There is also a small possibility for false negatives but amongst the

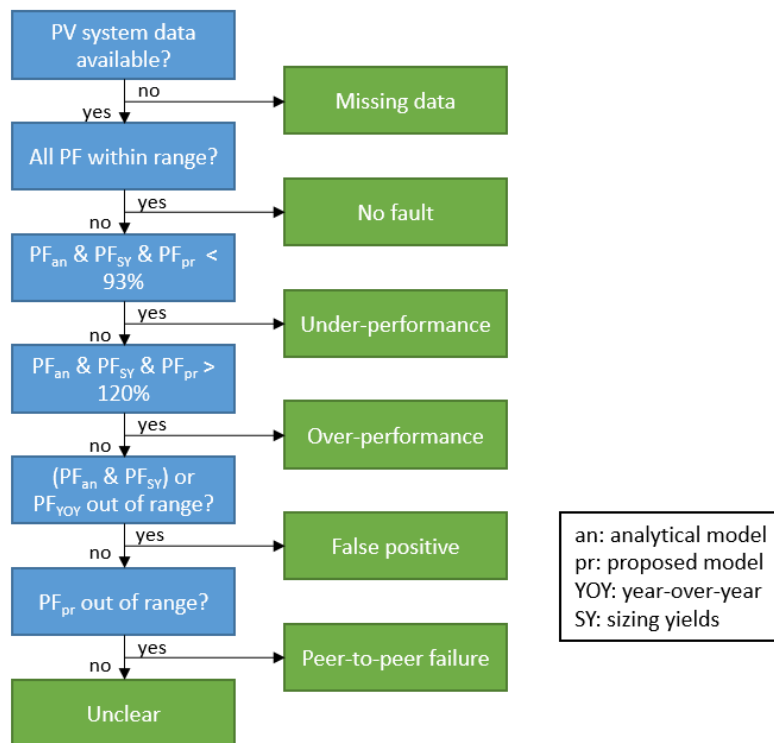


Figure 6.8: Flowchart depicting the main process of fault categorization, where 'an' is analytical model, 'pr' is proposed model, 'YOY' is year-over-year, and 'SY' is sizing yield.

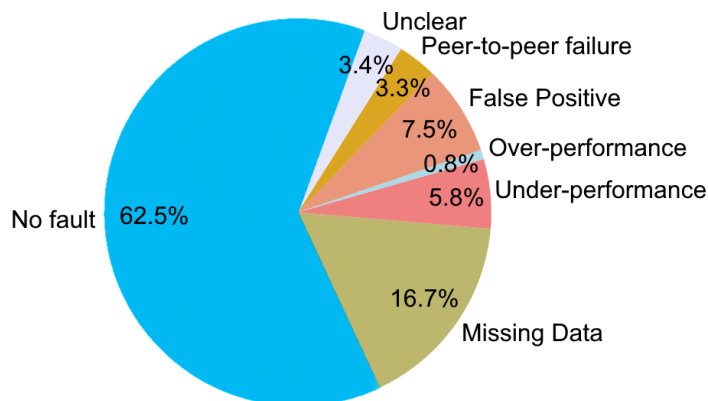


Figure 6.9: Occurrence of faults when categorization tested on 120 PV systems.

120 PV systems considered, none presented a clear case. The case of missing data was the most occurring fault, present in 16.7% of the systems, although it is easy to detect due to automatic notice from the inverter.

In the case of under- (5.8% of systems) and over-performance (0.8% of systems), if the

fault occurred within the last year, it was reflected in all criteria. Yet if the fault persisted for longer than one year, year-over-year yields failed to flag the fault. Depending on the magnitude of under-performance, the diagnosis could either be a small, temporary problem or else it could be due to broken panels or strings. In the case of broken panels, the magnitude of PF should drop by an equivalent amount. Similarly, on the rare occasion of over-performance, depending on the magnitude of the increase in actual yields, it could be a case of system size change. This should always be verified with the latest satellite images.

The fifth category is a false positive when some of the models detect a fault that does not exist. This fault was detected in about 7.5% of the analyzed systems. It usually occurs when there is a mismatch between the actual installation and the system design details in the company database. The analytical and sizing yield models would be more easily fooled in this case due to their high dependence on the system design parameters. On the other hand, since both the year-over-year and proposed models depend more on historical yields, they are more robust against incorrect system data. Another case of false positives is when all except the year-over-year yields are within the acceptable range. Here, the possible diagnosis is that due to, for example, unusually sunny or cloudy days, the actual yields are higher or lower than last year. Since both analytical and sizing yields depend on weather data, they adjust accordingly to the unusual irradiance. Similarly, the unusual weather is experienced by most peer systems thus the proposed model does not flag either.

The final category and the one checked with caution was the case of peer-to-peer failure. In this case, only the proposed model detected a failure. This occurred mainly due to poor distance training related to poor quality or low availability of data. Another cause of failure occurred for systems older than 5 years with degradation in the performance. Since the distance training was done in the early years of the PV system, the model does not adjust to the performance degradation over the years and thus, detects it as a fault. Compared to that, both analytical and sizing yields include system decay, while year-over-year yield experiences degradation in a controlled manner. The occurrence of this fault was not very high, only 3.3%, and it can generally be solved by distance retraining of the systems to ensure proper expected yields.

It was expected to find some examples of inverter-limited systems among the 120 checked. The inverter capacity for a few systems was lower than the total system watt-peak, yet none of them seemed to have an energy generation high enough to be limited by the inverter. Note that while these systems represent the total dataset, the statistics should be taken with a grain of salt. Moreover, hourly yield data would have been more appropriate to properly detect inverter clipping.

6.5.2. KEY EXAMPLES

This subsection presents some examples of PV systems with certain faults. Examining these PV systems will help in understanding how the collaborative approach fits into the categorization framework and can be a useful addition to fault detection. The proposed model plays a key role in the detection of over-performance and false positives, thus examples of PV systems with these faults are presented in this subsection.

COMMON FAULT

This first example serves as a baseline to show how all the models can detect a common fault such as broken strings or panels. The example PV system has 58 panels, formed by

separate sets of 31 panels and 27 panels which differ in both type and orientation.

Table 6.3: Performance factor according to each model for a PV system with broken strings or panels.

	Analytical	Year-over-Year	Sizing	Proposed
PF	37.0%	47.0%	40.1%	46.9%

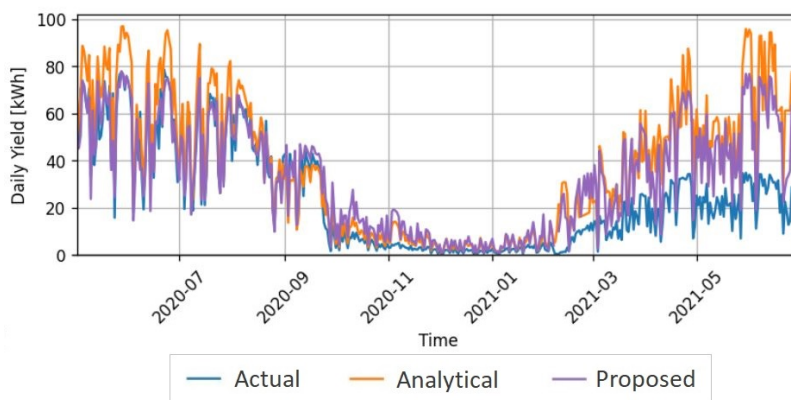


Figure 6.10: Temporal yield plots: example of a PV system with broken strings or panels.

Amongst the models described in [Subsection 6.2.4](#), all the fault detection checks give a red flag, suggesting that the system has been under-performing to around 40% as seen in [Table 6.3](#). There is a sudden drop in the actual daily yield around September 2020 as seen in the yield plots in [Figure 6.10](#). As the fault occurred within the last year, the year-over-year model can flag the issue. In case this fault persists beyond September 2021, year-over-year yields will not be able to flag the fault anymore.

While all these checks simply suggest that the system is under-performing, the fact that under-performance has been consistent, or in other words the performance factor being consistently around 40%, suggests that only a part of the PV system has developed a fault. The sizing yields estimate the system size as 25 panels instead of the 58 that were installed, indicating that part of the system has likely broken down.

OVER-PERFORMANCE

Due to the capital-intensive nature of PV systems, it can be preferred for residential owners to install their systems in steps on their roofs. Thus, a few years after a PV system has been installed, the owner could decide to increase its capacity. This is usually not reported back to the monitoring company which leads to system over-performing. Considering the fault detection checks for this particular example as seen in the first row of [Table 6.4](#), all models except for year-over-year flag the system as over-performing. When the yield plots in [Figure 6.11](#) are checked, it is clear why this is the case. This system has been over-performing since April 2019 by 140% which, looking at the system specifications, is equivalent to a size of 11 panels instead of 8. This was verified when the design image in [Figure 6.12a](#) is compared with the latest satellite image in [Figure 6.12b](#).

Table 6.4: Performance factor according to each model for an over-performing PV system before and after retraining.

	Analytical	Year-over-Year	Sizing	Proposed
Before retraining	143.0%	106.9%	136.2%	139.3%
After retraining	143.0%	106.9%	136.2%	99.5%

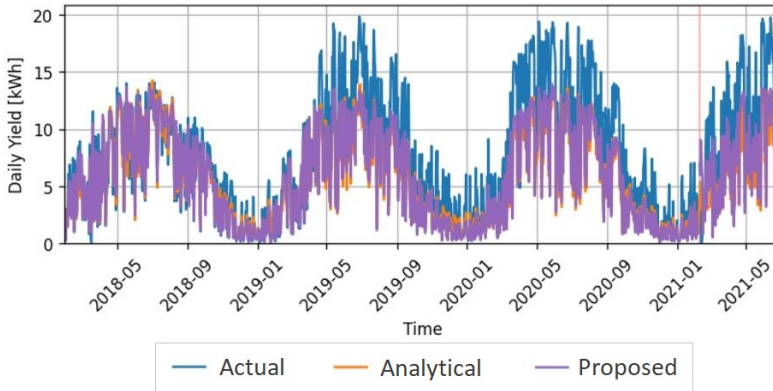


Figure 6.11: Temporal yield plots: example of over-performing PV system due to an increase in system size.

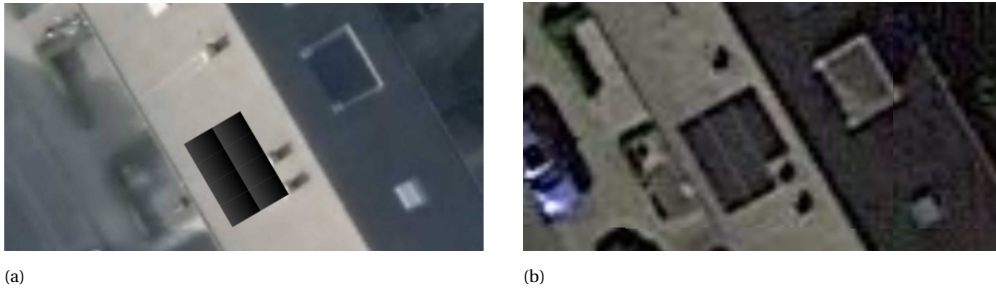


Figure 6.12: Images of a PV system (a) during the design phase with 8 PV panels, and (b) from the satellite showing 11 panels.

A particular advantage that the proposed model has in such a scenario is that it can be tweaked by changing the period over which the distance training is done. In the above case, distance training was done before the system change. Since system size change is not a malfunction in the system, this system can be retrained with the new PV system size to absorb this change in the system. Hence, distance training can be done after the system change, to find different peer systems that indicate this focus system is performing as expected. The results of retraining can be seen in Figure 6.13, and in the second row of Table 6.4. While the analytical and sizing yields still suggest that the system is over-performing, the proposed model does not flag any faults. This distance retraining is especially important if a true fault occurs after repowering since it will be flagged by the proposed model while the analytical one might suggest only a lesser over-performance.

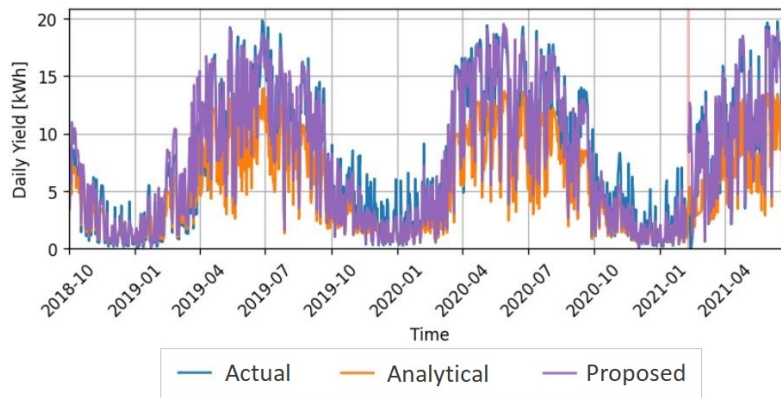


Figure 6.13: Temporal yield plots: over-performing PV system after retraining to absorb the increase in system size.

FALSE POSITIVE

One of the strengths of using peer-to-peer yields for fault detection is the ability to detect false positives. A false positive occurs when the analytical model suggests an issue with the system despite the system not having any particular malfunctions. As an illustrative example, the chosen PV system has 14 panels according to the database. From the fault detection checks in Table 6.5, most models suggest that the system is underperforming. When these three checks give a red flag, the first conclusion would be that there might be a few broken panels. However, the proposed model suggests that the system is performing perfectly fine with a 97% performance factor. To make sense of this discrepancy, the yield plots are inspected.

Table 6.5: Performance factor according to each model for a PV system when a false positive is detected.

	Analytical	Year-over-Year	Sizing	Proposed
PF	53.7%	82.0%	55.8%	97.0%

Figure 6.14 indicates that the system has been under-performing since the installation day. Since the proposed model was trained during this alleged underperformance, it does not consider this to be any fault. From the yield plot, one can deduce a mismatch between the system design details and the actual installation. This suggests that the under-performance flag by the analytical models is a false positive as the system has been performing as historically expected according to the proposed model and there is no malfunction in the system that might degrade its performance.

This kind of mismatch fault is the cause of many persistently under-performing systems. These systems are time-sink monitoring companies and thus it is key to simplify the diagnosis. The proposed approach can be highly advantageous in such scenarios.

The categorization framework and the examples given are based on theory, discussions, and experience. Unfortunately, there was no possibility to validate the results as this would entail physically visiting 120 installation sites all over the country or contacting the corresponding homeowners. Nonetheless, the objective for undertaking fault detection was to

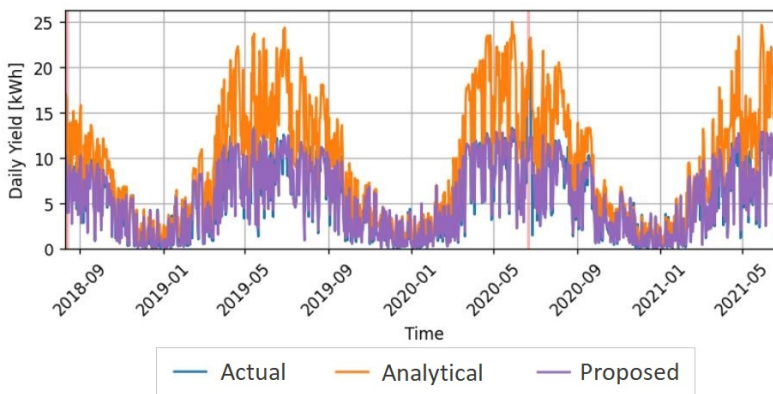


Figure 6.14: Temporal yield plots: example of false positive PV system.

show how the proposed approach can be used for fault detection.

6.6. CONCLUSIONS

This chapter has developed a model based on comparisons between neighboring and similar installations for PV system monitoring and fault detection. The model was based on a previously developed performance-to-peer approach, which has been improved by the addition of PV system parameters and the use of machine learning techniques.

The viability of the proposed model has been demonstrated in the fleet of Solar Monkey consisting of more than 12000 residential PV systems with up to 7 years of data per system. The developed model showed an average R^2 score of 94.1% and normalized root mean squared error of 0.05 on all tested PV systems. This implied an improvement in terms of R^2 score of 1.4 percentage points compared to the baseline performance-to-peer model and of 3.8 points compared to the analytical one. This superiority over analytical models stemmed from its independence from inaccurate weather data and reduced dependence on PV system parameters. A use-case analysis was also performed to find the limits of the proposed model. It was discovered that 1700 years of data were required for proper model training, with a minimum of 6 months of data per system and 100 PV systems.

The usage of the developed model for fault detection and categorization has also been demonstrated. This model has the strength of distinguishing between incorrect PV system information and actual faults. Moreover, distance and peer retraining provide the flexibility to adapt the model to changes occurring in the PV systems. Although validation of these faults was not possible, the proposed model has proven to be a valuable tool when combined with other developed models for fault detection and diagnosis.

PERSPECTIVE: FURTHER DEVELOPMENTS WITHIN TRUST-PV

The developments of this chapter were continued by Solar Monkey within the Trust-PV project [231]. They explored the limits of the methodology applied in a real-world scenario and extended its application to systems outside the Netherlands.

The first limit they found was the unreliability of PV system characteristics data, es-

pecially regarding the number of modules and the inclination angles. Therefore, they discarded the feature distance when selecting the peers. They also implemented an alternative normalization method to estimate the relationship between peer and focus yields more accurately. Another limitation was the high computational requirements when finding good peers, with the calculation of the yield distance being the limiting process. They solved it by considering only the 500 geographically closest systems as potential peers.

With data from outside of the Netherlands, they reported that accurate results can be obtained only if more than 100 potential peers within 100 km are available. Peers at distances larger than 500 km are unreliable.

Regarding the fault detection procedure, they improved the underperformance detection. To avoid setting a threshold, they generated probabilistic forecasts using confidence intervals derived from the yields of the good peers.

Finally, they explored deeper into the false negatives flagged by P2P and reported three main issues limiting the methodology. The first one is that P2P cannot detect mistakes during installation. If a system is underperforming from day one, the training period will include that fault. The second one is if there is an outlier during training, which can jeopardize the proper selection of peers. The third one is if there are no peers of quality within the fleet, for instance, due to significant geographical or system characteristics differences.

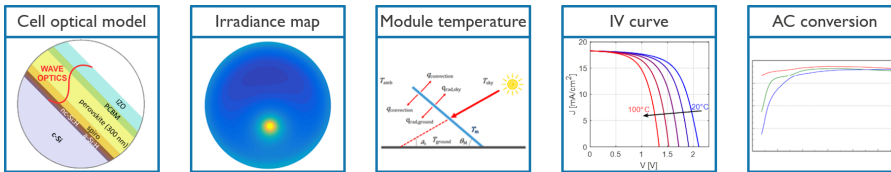
ANALYTICAL PV POWER PREDICTIONS



7

INTRODUCING THE PVMD TOOLBOX

This chapter introduces the PVMD toolbox, a novel simulation tool developed at Delft University of Technology by the Photovoltaic Materials and Devices (PVMD) Group. The tool can calculate the energy yield of a PV system based on its fundamental material properties and using self-consistent models. Thus, this simulation model can operate without measurements of a PV device. As a case study, the toolbox is tested using data from 15 residential systems. The results were disappointing, with NRMSE ranging from 42.2% to 109.5% and NMAE from 16.3% to 41.3%, varying by system and study period. Two other analytical frameworks produced even worse results, suggesting that the issue lies with the quality of the input data. Further analysis indicated that interpolated irradiance data was a significant contributor to the error, potentially accounting for up to 53.3% of the NRMSE.



As a collaborative framework, many people have contributed to the PVMD toolbox. My main contributions were: adding the horizon reconstruction ability; code refactoring, commenting, and testing of the in-plane irradiance computation, cell temperature calculation, and DC to AC conversion steps; and providing tools for better coding. These contributed to the paper written by M.R. Vogt in [232] which has been slightly adapted as Section 7.2. The remaining sections were partly published in the Trust-PV reports I.T. Horvath, A. Alcañiz, F. Segata, J. Lemmens, T. Hall, W. Vanheusden, “Automated PV digital twin-based yield simulation framework” (2022) [233] and A. Alcañiz, J.C. Ortiz Lizcano, H. Ziar, M. Dallapiccola, Y. Snow, I. Kaaya, “Residential-scale integrated energy yield simulation and reliability modelling framework” (2023) [234] .

7.1. INTRODUCTION

The previous block of chapters has explored different methodologies to predict PV power using machine learning approaches. However, these approaches have limitations, the main one being their high data requirements. An ML model cannot be properly trained without enough data. This is limiting for instance when predictions are required from the first day of installation, or during the project development for financial assessments. Additionally, ML is not explainable, and a deep understanding of the processes occurring during solar energy conversion and how the weather parameters can affect the power output are needed to find the optimal conditions for installation.

For this reason, several institutions have developed energy yield modeling software programs based on analytical models. Some of these programs are for internal use only but others are commercialized. These software tools aim to create a digital twin used for monitoring PV plants.

Each software tool has different objectives regarding the target system, intended audience, usage requirements, and computational limitations, among others. Depending on the objective of the simulation, the priorities may include accuracy, ease of use, or computational time; the focus is placed on existing or theoretical modules; and different aspects of a PV system are optimized. Each simulation framework is designed with specific objectives in mind, and their excellence lies in addressing the particular tasks they aim to achieve. This implies that the software tools have different levels of detail for each step involved in calculating the energy yield modeling. These steps can be divided into:

1. **Optical properties:** Determine the optical response of the layers of materials composing the PV cell.
2. **System surroundings:** Determine the surroundings of the PV system for the shading calculations.
3. **In-plane irradiance:** Calculate the plane-of-array (POA) irradiance incident on each cell or module of the PV system.
4. **Module temperature:** Determine the temperature of the modules at the cell or module level.
5. **DC power:** Compute the DC power output produced by the system at module, string, and/or system level.
6. **AC conversion:** Convert the DC power into AC power considering the inverter characteristics and the system architecture.

A level of detail (LoD) can be assigned to each of these steps measuring their complexity (and hence accuracy and requirements). The following explains what each level of detail signifies, using module temperature as an example.

- **LoD 0:** The step is not modeled. A fixed value is assigned independently of the conditions. A module temperature fixed to the Nominal Operating Cell Temperature (NOCT) is a LoD 0.

$$T_{mod} = T_{NOCT}$$

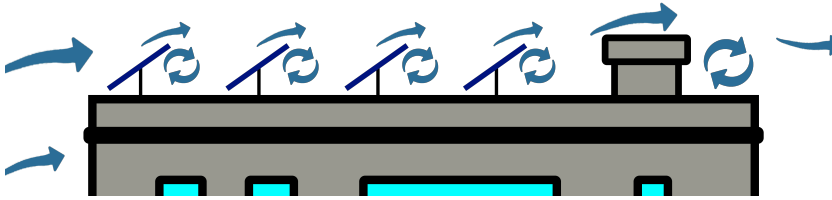
- **LoD 1:** A simple model is considered for the step, where simple is defined as being representable by a single equation. Faiman's temperature model [235] would be an example of LoD 1.

$$T_{mod} = T_{amb} + \frac{G_{POA}}{U_0 + U_1 \cdot ws}$$

- **LoD 2:** Achieving LoD 2 involves employing a complex model, defined here as represented by a system of equations. Fuentes' fluid dynamic model [236] is an example of LoD 2.

$$\begin{cases} f_1(T_{mod}) = 0 \\ f_2(T_{mod}) = 0 \\ \vdots \\ f_n(T_{mod}) = 0 \end{cases}$$

- **LoD 3:** An accurate (usually 3D) system representation is employed to consider all the phenomena impacting the step. It can be considered to be the ultimate digital twin. Achieving a LoD 3 for the module temperature would entail modeling the whole PV system and its surroundings in 3D to study the air movement and consequent thermal flows and conductive properties of each material.



Considering these 6 steps and 4 LoD levels, several commercial software solutions have been compared in Table 7.1. The optoelectrical modeling of the PV modules is generally simple since most software tools focus on existent modules, hence the datasheet provides enough optoelectrical information. A LoD 3 is only achieved by some software tools when modeling the system surroundings. For this step, only the deterministic sun position is required after the 3D modeling of the surroundings is obtained. For the remaining steps, achieving such a high level of detail would be too costly in terms of both time and resources compared to the gain in accuracy. Despite the significant impact of module temperature on power output and the challenges in accurately predicting it, many software tools opt for simple models. This is because the accuracy gained by using more complex models is often marginal.

In terms of accuracy, the PVMD toolbox stands out by surpassing or matching the LoD of other software tools. Its determination of optical and semiconductor behavior makes it unique. The toolbox can simulate the AC energy yield of a non-fabricated module. This allows for the energy yield prediction of, for example, tandem devices, a topic that is growing in interest amongst the research community.

The toolbox has been extensively validated across several studies during its development [237–243]. This chapter aims to evaluate the toolbox's capabilities in a real-world scenario by predicting the power output of residential-scale PV systems. The goal is not to validate the framework but to identify potential areas for improvement.

Table 7.1: Level of detail of several simulation frameworks for each step in the energy yield modeling process.

Framework	Optical and electrical properties	System surroundings	In-plane irradiance	Module temperature	DC power	AC conversion
Sisifo ¹ [244]	0	0	1	1	1	1
PVWatts ¹ [245]	0	0	1	1	1	1
Solar PathFinder ¹ [246]	0	0	1	1	1	1
pvDesign ¹ [247]	0	1	1	1	1	1
Solar Monkey [198]	0	3	2	1	2	1
Dutch PV portal [248]	0	0	2	2	1	1
Plant Predict ¹ [249]	0	1	2	1	2	2
System Advisor Model ¹ [250]	0	1	2	1	2	2
PV*Sol ¹ [251]	0	2	2	1	2	1
pvlb python ¹ [178]	0	1	2	2	2	2
BIM Solar ¹ [252]	0	2	2	1	2	2
Solar Farmer ¹ [253]	0	2	2	1	2	2
Eurac research [254]	0	3	2	1	2	0
Archelios Pro ¹ [255]	0	3	2	1	2	2
Helioscope ¹ [256]	0	3	2	1	2	2
PVSyst ¹ [257]	0	3	2	1	2	2
Imec [258], pvCase [259]	1	3	2	2	2	2
PVMD Toolbox (this work)	2	3	2	2	2	2

¹ These LoDs have been assumed from the information provided by the organizations about their software tools.

Several distinctive challenges arise when monitoring residential PV systems compared to utility-scale ones. One major issue is the lack of on-site weather data, as installing meteorological sensors near residential PV systems is typically not cost-effective. Additionally, there are often greater discrepancies between the planned and actual installations due to less thorough site assessments, which can lead to deviations in the final setup, caused, for instance, by unexpected obstacles on rooftops. Consequently, information about these systems is often limited and inaccurate. Furthermore, the shading environment is more complex in urban settings, where numerous obstacles can impact the system's performance.

This chapter presents the toolbox in detail in [Section 7.2](#). The toolbox is composed of several blocks, each independently and collectively validated. [Section 7.3](#) exemplifies such a process by focusing on the validation of the horizon reconstruction. Next, the case study begins by presenting the system and weather data in [Section 7.4](#). [Section 7.5](#) dives into the results, which highlight the sensitivity of the framework to the input irradiance. Finally, [Section 7.6](#) concludes this chapter.

7.2. SIMULATION MODEL

The flowchart of the simulation model for energy yield calculation in the PVMD Toolbox is shown in Figure 7.1, with the main models in blue, the information flow and intermediate results in green, and the inputs in black.

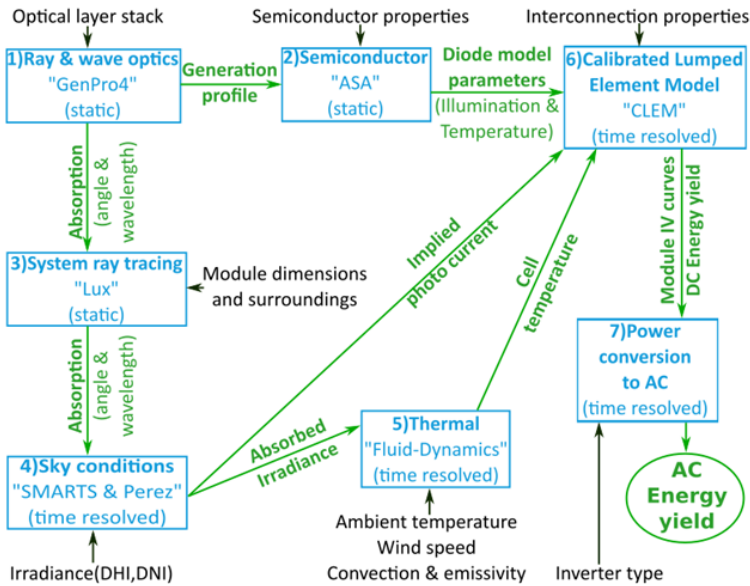


Figure 7.1: Flowchart of PVMD Toolbox describing the simulation methodology. The input data is indicated in black. The main models and (intermediated) results are in blue and green, respectively.

The PVMD toolbox starts by considering ray and wave optics in the module layer stack described by the materials' complex refractive indices. It calculates spectral, depth, and angle-dependent absorption for each layer in the cell. This generation profile serves as an input for the advanced semiconductor analysis software, which is used to calculate the cell IV curves at different irradiances and temperatures. An equivalent circuit model of the solar cell is then fit to match all the IV curves, resulting in a calibrated lumped-element model with temperature and illumination-dependent parameters. This initial block is what sets the PVMD toolbox apart from other software tools. The system surroundings are then considered using a system-level ray tracing, which simulates the PV module mounting conditions and setting allowing for shading and albedo effects to be considered, resulting in a sensitivity map of the PV system.

The first time-resolved step is the creation of a distribution of the radiance across the sky based on the Perez model and SMARTS to add spectral information. Afterward, integrating the sensitivity and sky maps results in the irradiance absorbed by the module and the cell photocurrents. The former is used to calculate the temperature of each cell in the module together with ambient temperature, wind speed, convective heat loss, and emissivity. Next, a lumped-element model is used to simulate the module interconnection in combination with the calibrated lumped-element model parameters and the resulting module IV-curve for all hours of the year considering the temperature and implied photocurrent for each cell at each simulated timestep. Finally, the DC power is converted to AC considering the

voltage and current-dependent efficiency of the inverter for each timestep.

Each of these steps is explained in detail in the following subsections.

Two additional blocks, not present in the flowchart, were recently added to the toolbox. The first one considers the yearly degradation of the PV modules. This block is based on a degradation model that considers three precursors: hydrolysis, photodegradation, and thermomechanical degradation [260]. The second block is that of the economic analysis which is based on the levelized cost of electricity (LCoE). This concept refers to the cost of generating 1 kWh of electricity from a power generation facility [261].

7.2.1. OPTICAL AND ELECTRICAL PROPERTIES

The solar modules' optical properties are simulated using the GENPRO4 software [262]. GENPRO4 implements the net radiation method [263] and further extended versions of that [264, 265]. The original net radiation method approximates the cell structure as a 1-D multilayer stack along the depth axis of the cell. The thickness and complex refractive index of every layer are input parameters. The extended models allow for interference and scattering effects to be considered. Random surface textures are simulated using the scalar scattering model developed by Jäger et al. [266], which utilizes a 3D scan by an atomic force microscope to model a realistic texture.

GENPRO4 calculates spectral and angle-resolved reflection as well as absorption which is also depth-resolved for each layer in the module including the cell with all its layers. In the case of bifacial structures, this calculation is performed for front as well as rear irradiation. This approach has demonstrated very good agreement with measurements [262, 264–266].

The resulting outcomes for each material are input variables for the optoelectronic simulator program ASA [267, 268]. The Advanced Semiconductor Analysis (ASA) solves the semiconductor equations in one dimension, specifically the Poisson equation and two continuity equations for electrons and holes. The physical background is based on the drift-diffusion model consistently coupled with recombination and tunneling models. ASA accounts for semiconductor electronic properties such as bandgap, electron affinity, density of states, mobility, doping profiles, and trapped charges in the localized states within the band gap to simulate one-dimensional devices under various illumination or dark conditions, all as a function of temperature. Figure 7.2a shows an example of resulting band diagrams for a perovskite top and Figure 7.2b a silicon heterojunction bottom cell.

ASA is used to calculate the cell IV curves at different irradiances and temperatures. To save computational effort, an equivalent circuit model of the solar cell is then fit to match all the IV curves. This way a calibrated lumped-element model (CLEM) is obtained with temperature and illumination-dependent parameters of the one-diode model for the solar cell. Figure 7.2 shows the comparisons between ASA and CLEM. The one-diode model parameters for the top and bottom cells are extracted separately and then later combined in an equivalent circuit model for each tandem solar cell. These are later combined in series connection or independent operation depending on the number of terminals of the cell.

The toolbox offers the flexibility to avoid this block if a commercial PV module is to be modeled. In that case, the datasheet parameters are employed as input to the model.

7.2.2. SYSTEM SURROUNDINGS

The in-house developed forward Monte-Carlo ray tracing software LUX is used to simulate the PV module mounting conditions and surroundings, allowing for shading and albedo ef-

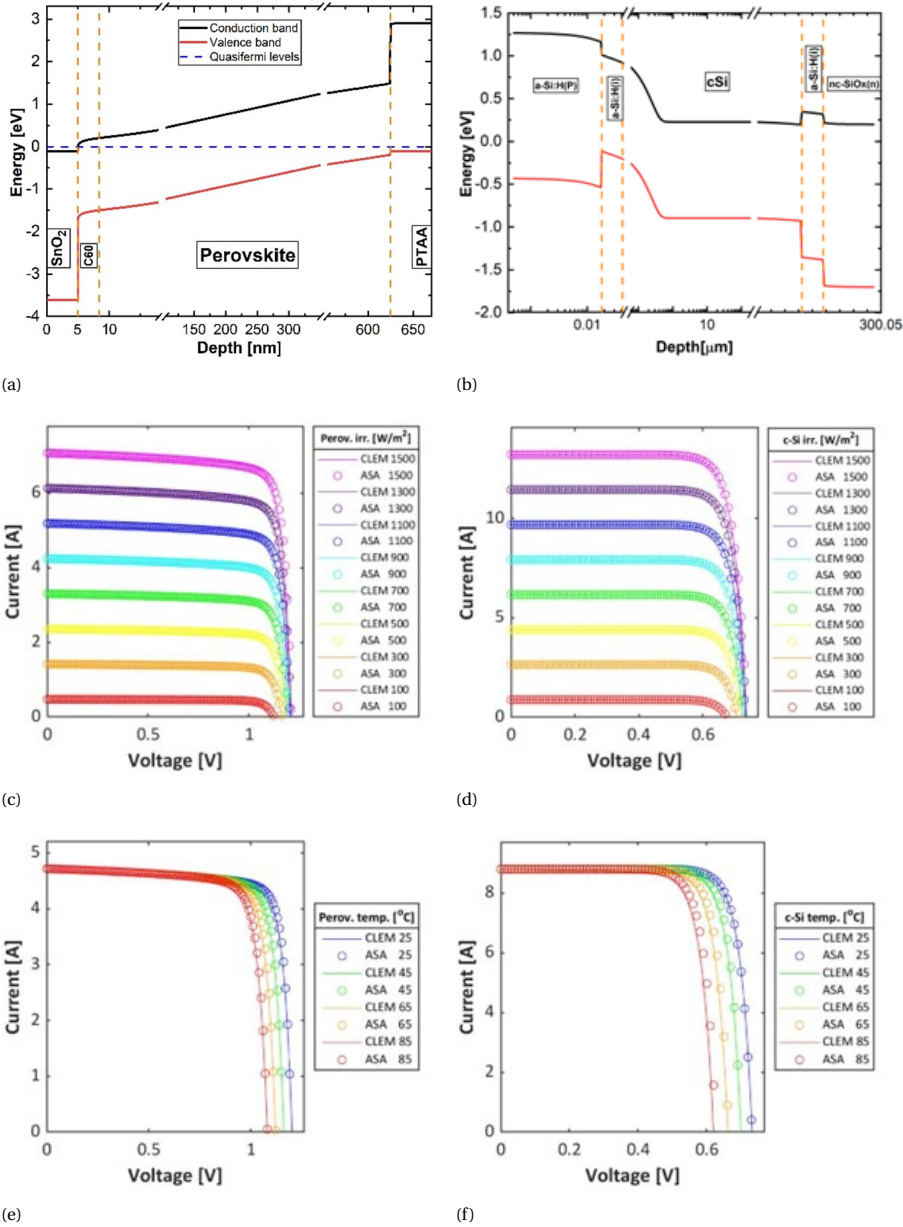


Figure 7.2: The perovskite top cell is always shown on the left, while the silicon heterojunction bottom cell is always shown on the right. Band diagrams as simulated by ASA for both cells are shown in the top row. Irradiance dependence of the IV curves for both ASA and CLEM are shown in the middle row. Temperature dependence of the IV curves for both ASA and CLEM are shown in the bottom row.

fects to be considered. The ray-tracing method was chosen for its flexibility as it can treat complex illumination situations, including shading and light scattering from nearby ob-

jects. The optical properties of these nearby objects, such as spectral reflectance and diffuseness of reflection, can be chosen freely, to closely mimic the real-world situation. Note that commonly used methods based on view factors require less computation time but are based on the assumption that surfaces reflect light in a perfectly diffuse way.

The simulation domain is shown in Figure 7.3a. It simulates a single PV module with periodic boundary conditions on four sides representing a module in a large plant. The module is modeled considering the size of each cell, the tilt, the mounting height, the frame, the distance between cells and the frame, and the distance to other modules. The reflection and absorption properties of the PV module are based on the simulation as described in the previous subsection.

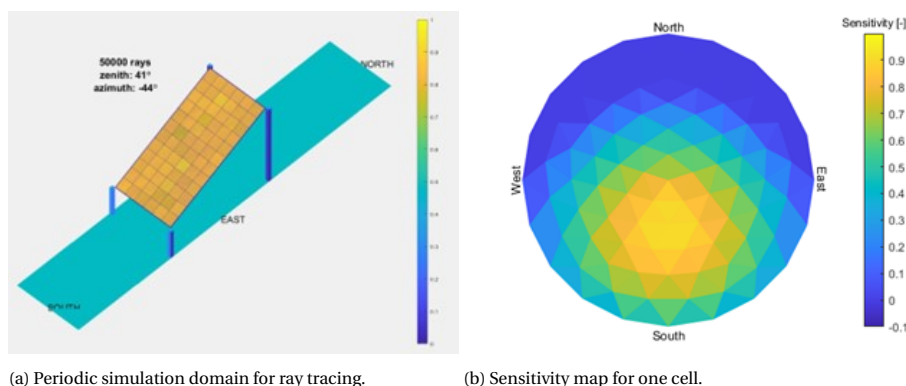


Figure 7.3: Periodic domain and resulting sensitivity map for one of the cells.

The haze and diffuse exponent of the ground reflection are adjusted in the model, which results in sensitivity maps [269] as shown in Figure 7.3b. Typically, one sensitivity map is calculated for each junction material as a basis to calculate their current and another one for the whole module to derive the input for the module temperature model.

To consider elements casting shadows on the modules such as mountains or large buildings, the shape of the terrain and characteristics of the surface can be obtained via LIDAR data. LIDAR (Light Detection and Ranging) uses a pulsed laser that scans the Earth's surface to provide a digital height map that includes both the built environment and vegetation [270].

With the height data, the skyline profile is extracted, which is a 2D projection of the surrounding landscape. To extract it from the height map, an *observer* is placed in the location of the PV module which finds, for each azimuth step, the highest altitude angle [271, 272]. By sweeping all possible azimuth values, a skyline can be obtained such as the one in Figure 7.4.

Once the skyline profile is obtained, the sky view factor (SVF) and shading factor can be computed. The SVF is a constant value ranging from 0 to 1 and indicates the portion of visible sky considering both the tilt of the PV module and the horizon [273]. The shading factor depends on time and indicates where there is an obstacle blocking the sun. This is included in the next step of the methodology as a factor multiplying the direct normal irradiance (DNI) with zero when the sun is behind the horizon. This is however an approximation because when the sun is blocked, not only the direct component is affected but also

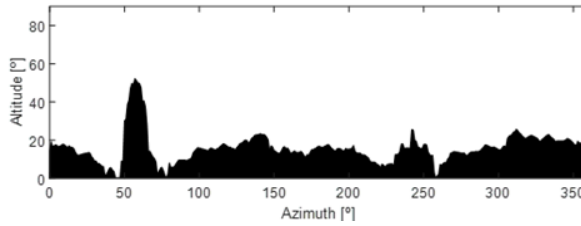


Figure 7.4: Raised horizon of a PV module placed in the center of the Markt of Delft, the Netherlands.

the diffuse one, due to the anisotropic nature of the latter.

This methodology was initially integrated into the toolbox with all PV modules sharing the same skyline. However, nearby objects like chimneys were not adequately considered. To address this, the framework was adapted to offer two options: a different skyline per PV module or a 3D reconstruction of the system's roof or environment. Users can choose between the two options based on accuracy and time constraints. The impact of these options on the accuracy of the predictions is discussed in [Section 7.5](#).

7.2.3. IN-PLANE IRRADIANCE

The first time-resolved calculation step is the creation of a distribution of the irradiance across the sky in the time domain based on the Perez model [274]. [Figure 7.5](#) shows an example of such an irradiance map. Then, SMARTS [236, 275] adds the spectral composition of the incoming data.

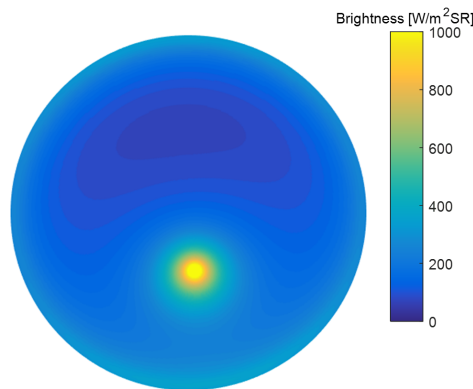


Figure 7.5: Irradiance map showing the irradiance distribution across the sky for a one-time instance.

Afterward, the sensitivity and irradiance maps are integrated over all wavelengths, zenith, and azimuth angles. This outputs the irradiance absorbed by the module block enveloping each cell including half of the inter-cell gap, and the implied cell photo-generated currents for each cell, time instance, and junction. The former is an input variable for the module temperature calculation, while the latter is an input variable in the electrical model.

7.2.4. CELL TEMPERATURE

The PVMD Toolbox uses a fluid dynamic model [236] to calculate the temperature of each cell in the module. It considers natural as well as forced convection and radiative heat transfer for all time instances. The model outputs the temperature of each cell in the module at each time instance. This fluid dynamic model relies only on non-empirical parameters and is thus able to directly include the impact of a different efficiency or other property on the operating temperature whereas previous works are based on empirical temperature models [250, 256, 257]. This allows for simulations in which the empirical temperature parameters are not known.

7.2.5. MODULE ELECTRICAL INTERCONNECTION MODEL

Next, the PVMD Toolbox uses a lumped-element model to simulate the module interconnection with bypass diodes to calculate the resulting module IV curves for all hours of the year. The first step is calculating the time-resolved IV curves for each cell. Using the Lambert W-function [276], the simulation process is speeded up to 100,000 IV curves per minute. The inputs are the implied cell photo-generated currents, the cell temperature, and the temperature- and illumination-dependent parameters of the one-diode model.

To simulate a 60-cell module for one year in hourly resolution, several hundred thousand IV curves need to be calculated, even after excluding night times with zero irradiance. However, as shown in Figure 7.6, many solar cell operating conditions in terms of cell temperature and implied photo-generated current are repeated several hundred times. Thus, the operating conditions are organized in discrete cell temperature steps of 0.3 K and implied photo-generated current steps of 0.4 A/m^2 , and those discrete condition steps are only simulated once if they appear in the climate.

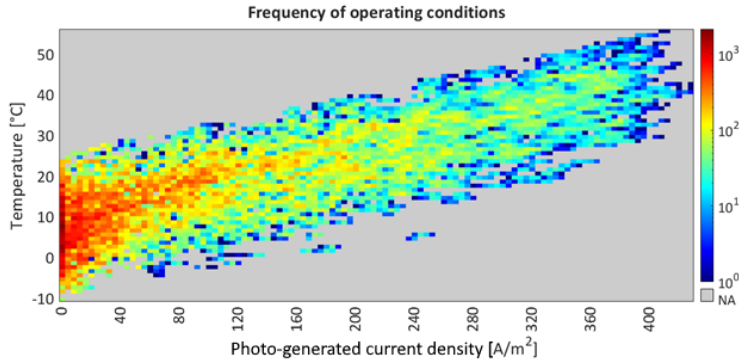


Figure 7.6: Frequency of solar cell operating conditions as a function of cell temperature and photogenerated current for one year in hourly resolution in Delft. The colored rectangles show the organization of these operating conditions. For illustration purposes, these rectangles are 33.3 times larger than the ones typically used.

For simulating one year in Delft in hourly resolution, this strategy reduces the number of simulated IV curves for a 60-cell module by 86% and the simulation time to about 2 min on a desktop PC (3 GHz, 8 GB RAM). The specific reduction depends on the climate conditions at the location and the number of cells in the module. The computational effort reduction scales extremely well with higher time resolution or longer periods of the climate data, as the number of unique operating condition steps hardly increases, only the frequency

in each step. In this example, this discretization creates an overestimation of the yearly module yield of 0.23%. This is due to a slight reduction in current mismatch losses. The overestimation scales with step size as larger step sizes decrease current mismatch losses, thus there is always a trade-off between speed and accuracy when using this approach of organizing operating conditions in discrete steps.

7.2.6. INVERTER MODEL

The DC power is converted to AC considering the voltage- and current-dependent efficiency of the inverter for each time step. The input variables are the DC current and voltage of the module for each time step from the previous calculation. The input parameters are the number of modules in series and parallel connection, and the DC and AC cable and inverter properties. The PVMD Toolbox can simulate PV systems with central inverters, string inverters, microinverters, and power optimizers. For the three inverter types, conversion efficiencies are calculated based on the SNL model [277]. For power optimizers, our own model is applied based on efficiencies measured as a function of input power at different input voltage levels [242].

7.3. VALIDATION OF THE HORIZON RECONSTRUCTION

Subsection 7.2.2 has detailed the process of obtaining a skyline profile from LIDAR data. This methodology offers the clear advantage of process automation, but its accuracy needs to be assessed. The skylines obtained with LIDAR data can be compared with those obtained by a HoriCatcher for this purpose.

A HoriCatcher is a tool that digitalizes the local horizon of a certain location [278]. It accurately measures the skyline profile at the placed location. The instrument consists of a camera attached to a tripod pointing down to a horizon mirror. The result of HoriCatcher measurements is a *fisheye* image which can be converted into a rectangular 360° panorama image, with the y-axis representing the elevation. This can be directly compared with the skyline obtained with LIDAR data [272].

Figure 7.7 compares the skyline measured by the HoriCatcher to that obtained via LIDAR data at 7 considered locations. Locations 1 to 5 are several points on the campus of Delft University of Technology and were considered in a previous publication [272]. They are surrounded by tall buildings which considerably decrease the horizon. The last two measurements are taken on residential roofs, which are more representative of the horizon that PV modules are subjected to.

Several discrepancies are detected between the measured and simulated skylines. One of the largest appears at Location 4, where the effect of an inaccurate skyline has been translated to DC power output difference. For a hypothetical PV system at this location, the annual DC power was 60% higher with the simulated skyline than with the measured one.

These discrepancies are evident even on residential roofs with no tall buildings nearby. For instance, in the skylines of Roof 1 from Figure 7.7, the simulations show a large flat object at around 330° that does not appear in the HoriCatcher measurement. The characteristics of that element were explored in the height data, and it was identified as a chimney. That chimney appeared in the HoriCatcher image but it was overrepresented in the simulated horizon, similar to the effect of the tall building at Location 4.

The origin of these discrepancies was deviations in the location. If the coordinates at which the HoriCatcher measurements are not accurately determined, a considerable dis-

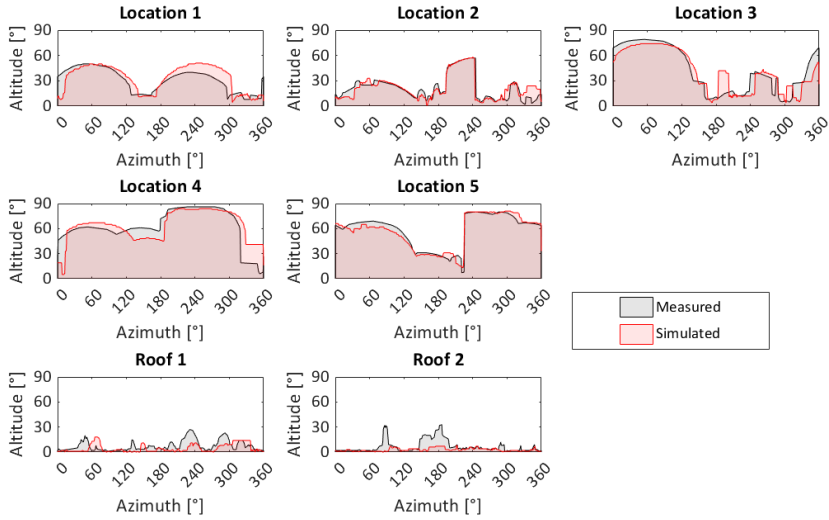


Figure 7.7: Measured and simulated skylines for seven selected locations.

crepancy can arise between simulated and measured skylines. This effect is shown in [Figure 7.8](#), which plots 4 skylines shifted by 1 meter to each cardinal direction relative to the location of Roof 1. The skylines are quite similar in the 5 graphs, except for the chimney. It disappears when moving to the East (also due to a higher roof height) and it considerably widens when taking one step to the West. Depending on the location and the proximity of the objects, they appear at different positions in the skyline. The object positioned around 110° depicts this phenomenon.

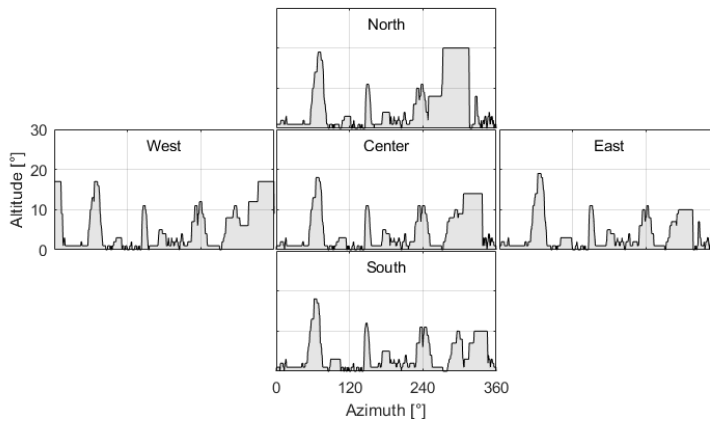


Figure 7.8: Skylines resulting from the sensitivity analysis on the location obtained by displacing the simulated point by one meter in each cardinal direction. The title of each subplot indicates the shift in direction.

The resolution of the employed height data also introduces errors in representing small obstacles on the roof. The smoothing conducted to generate the digital surface model removes noise, but this could potentially hide actual obstacles, especially in the case of chimneys with longer tubes. Higher resolution data would be needed to properly represent nearby obstacles.

To conclude, using LIDAR data is the recommended method due to its higher feasibility and availability. Remote reconstruction, low permission requisites, and the possibility of updating are the main advantages of the methodology. Amongst the pitfalls, one can highlight the lower accuracy compared to HoriCatcher measurements, the dependence on external sources, and the limited resolution.

7.4. DATA CHARACTERISTICS

7.4.1. SYSTEM DATA

In this case study, we aim to predict the performance of 15 residential-scale Dutch PV systems provided by Solar Monkey to identify potential improvements to the simulation software. Solar Monkey has a large database of more than 11000 residential systems. Out of those, they selected 15 high-quality systems, meaning that they did not show large periods of unavailability or abnormal behavior. These systems employed SolarEdge power optimizers thus obtaining yield data with a 15-minute resolution. They are also systems of different sizes and technologies so that different flaws can be tracked in the simulation framework.

Each system is identified by a system number which will be used throughout the report. The systems' location with their numbers can be observed in [Figure 7.9](#). The system size is around 4 kW, except for system 182210 which has an installed capacity of 44.25 kW. The SVF of these systems lies in the 66.8% to 94.9%, mostly caused by obstacles in the surrounding city environment.

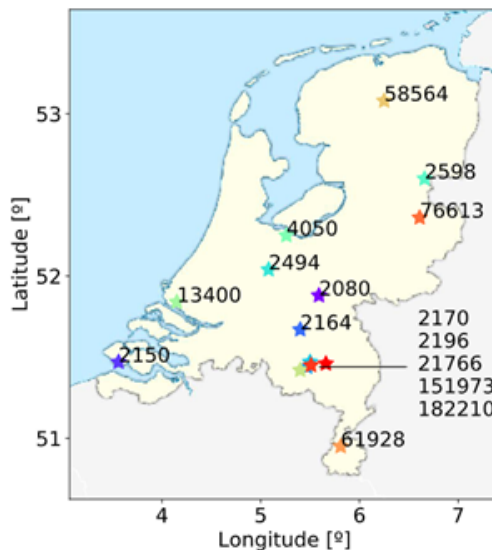


Figure 7.9: Location of the residential systems provided by Solar Monkey.

The initial challenge for the validation was the limited availability of data. The PVMD toolbox requires detailed system information starting from the cell level. However, most of that information is not accessible by Solar Monkey. The main limitations are listed below to motivate future developments.

- Limited cell types. The PVMD toolbox only has a few default solar cell technologies constructed for the cell module. PV systems using poly-cSi cells were too time-consuming to create, therefore they were discarded.
- Unknown information. Several parameters were not available from the datasheet or the data provided by Solar Monkey, therefore, approximated values were taken.
- Time-consuming process. Gathering all the necessary information and formatting it appropriately was a time-consuming process.
- Memory issues. Some sections of the simulation framework were not properly optimized, leading to memory issues. Simulating the data for the whole year with a 15-minute resolution was not possible so the results were run per season and merged afterward.
- Lack of automation. Some steps of the PVMD toolbox required user interaction which impeded the automation and increased the possibility of error.

These limitations hindered the simulation of all 15 systems for the whole period, so the study was limited to only 6 systems.

7.4.2. WEATHER DATA

System data consisted of yields and information, so weather data was sourced from the open database of KNMI, the Royal Dutch Meteorological Institute [196]. This weather forecasting service offers measurements from several meteorological stations spread all over the Netherlands over a wide range of years. The KNMI data, available at a 10-minute resolution, was interpolated to match the time format of the yield data.

Since the measurements were not taken on-site, the inverse distance methodology was employed to create more accurate inputs. The weather at the location of the system of interest was calculated with a weighted average of the measurements recorded by each station within a 50 km radius. The weights of each station were proportional to the inverse of the distance to the PV system, so that the closer a meteorological station was, the higher the importance it had [279].

The PVMD toolbox requires the three irradiance components, wind speed, and ambient temperature. The two latter weather variables were recorded in most meteorological stations, however, only the global horizontal irradiance (GHI) was measured in some stations. The BRL decomposition model [280], the best model for the Dutch weather [279], was employed to obtain the diffuse and direct components from the global irradiance.

7.5. RESULTS

To describe the results of the use case analysis, three metrics will be employed: the normalized root mean squared error (NRMSE), the normalized mean absolute error (NMAE), and the normalized mean bias error (NMBE). Table C.1 provides the definition of these metrics as well as the importance of using them appropriately.

7.5.1. 6 SYSTEMS ANALYSIS

The PV power of six systems was predicted for the first three months of 2021 with a resolution of 15 minutes. The winter period was chosen specifically to capture the occurrence of higher errors. Table 7.2 displays the real and predicted yields as well as the error metrics for all six systems, identified with their system number. The PV power is underestimated for all systems and the error metrics are considerable. Figure 7.10 shows the instantaneous errors normalized by system size in a boxplot. The normalized error ranges from -0.2 to 0.2 kWh/kW_p, which indicates that for a system of e.g. 5 kW_p of capacity, the prediction deviates up to -1 kWh and 1 kWh from the real yield.

Table 7.2: Real and predicted yield, and error metrics for six Solar Monkey systems using 15-minute resolution data for the first three months of 2021.

System number	2150	151973	13400	2080	61928	4050
Real [kWh]	731.2	348.6	726.2	550.5	479.7	350.2
Predicted [kWh]	722.0	317.1	616.3	492.0	355.9	297.2
NRMSE [%]	42.2	80.8	86.0	108.4	109.5	86.6
NMAE [%]	16.3	30.7	33.1	39.3	41.3	33.4
NMBE [%]	-1.3	-9.1	-15.1	-10.6	-25.8	-15.1

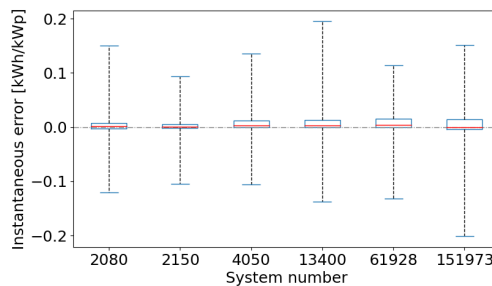


Figure 7.10: Instantaneous errors boxplots normalized by system size for the six Solar Monkey systems explored.

The skyline was identified as one of the contributing factors to the high errors in some systems. Initially, a single skyline per system, representing the most characteristic module, was used to account for long-distance obstacles. However, for some of these residential systems, the skylines varied significantly between PV modules.

To overcome this, the code was adapted to consider a skyline matrix instead of a single skyline. This matrix considers all skylines, assigning each point on the horizon a value between 0 and 1, representing the percentage of modules with a skyline at that specific point. The metrics were generally improved after considering the skyline matrix (see Figure 7.11), although the results depend on the considered system.

In pursuit of identifying the source of error in some systems, we examined various aspects:

- The correlation between system characteristics and error was explored. For instance, older systems may depict higher errors than young ones, but nothing was found.

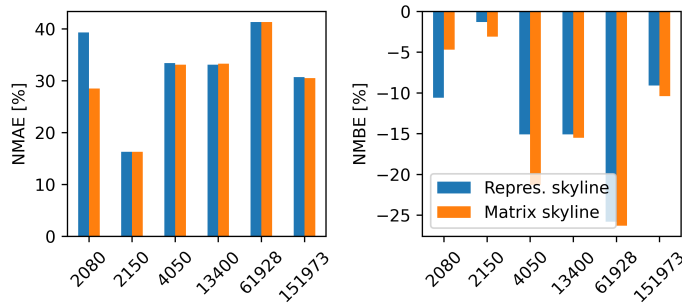


Figure 7.11: Difference in NMAE and NMBE when using the most representative skyline of the system or the skyline matrix.

- The correlation between weather characteristics and error was also performed. Larger errors were present during higher irradiance periods, but that is expected. No correlation was found for temperature or wind speed.
- The distance to the weather stations and the number of stations employed were also analyzed. Since the weather parameters were computed using inverse distance interpolation, we hypothesized that systems that relied on meteorological stations located closer would depict lower errors. However, the results did not support this hypothesis.

7.5.2. SINGLE SYSTEM ANALYSIS

A detailed investigation of system 13400 for the year 2020 was conducted to identify the error source. The most representative skyline was chosen as it yielded slightly improved results compared to using the skyline matrix.

The annual metrics were NRMSE = 66.0%, NMAE = 29.9%, and NMBE = -20.7%. As a comparison, the metrics obtained during the first three months of 2021 were NRMSE = 94.3%, NMAE = 36.3%, and NMBE = -21.0%. In winter, the average error is slightly higher, and larger outliers occur than during the rest of the year. However, these outliers are not responsible for the underestimation of the results since this underestimation occurs throughout the year.

In some instances of time, the prediction error was almost as high as the actual yield. A deeper exploration into these values shows that the top five instances of time when the error is the highest occur on days when the irradiance is high and intermittent. These days, the irradiance as well as the PV power production, have sudden drops and rises that are not well modeled. That is probably because the fast movement of clouds is not well represented by the interpolated weather irradiance. However, because of the compensation, the daily error on those days is very low. The days that show the highest daily error are days with high irradiance but not very intermittent. The predicted curve follows the shape of the real one but is considerably lowered. However, those are not the days with the highest relative daily error. Those days occur during winter when the irradiance and PV production are low, but the predicted PV power is even lower. Figure 7.12 exemplifies the power production and the irradiance components during these three situations.

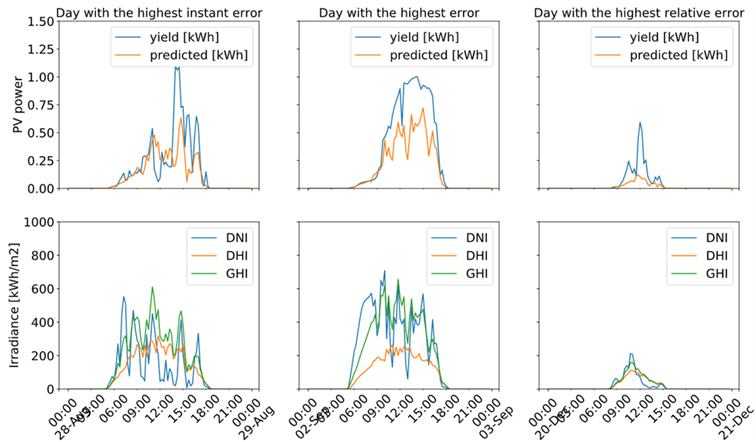


Figure 7.12: PV production and prediction (top row) and irradiance components (bottom row) on days where the highest errors were reported. Each column displays a different type of error.

COMPARISON WITH OTHER FRAMEWORKS

To identify whether the errors stemmed from the simulation framework, the results were compared with Solar Monkey and PVSyst estimations. PVSyst is a simulation software used to model PV systems, from small residential size up to large utility-scale [257]. It allows to inclusion of horizon shading as well as user input data, so it has the flexibility required for the comparison.

Both of these simulation frameworks employ hourly resolution, so the real yield and the predicted one by TUD were resampled to match the resolutions. Table 7.3 shows the results for the three simulation frameworks. For completeness, the yield produced by the system is 4883.8 kWh.

Table 7.3: Comparison of yields predicted by TU Delft's, Solar Monkey's, and PVSyst's simulation frameworks.

Framework	Predicted [kWh]	NRMSE [%]	NMAE [-]	NMBE [-]
PVMD Toolbox	3875.2	66.0	29.9	-20.7
Solar Monkey	3574.8	110.1	51.3	-26.8
PVSyst	3355.6	103.7	52.0	-31.3

The errors made by PVSyst are more spread: while the PVMD toolbox under-predicts quite steadily, the error performed by PVSyst has a higher deviation and considerably over-predicts during some hours. Solar Monkey and PVSyst depict similar errors, being Solar Monkey slightly better. The monthly energy graph of Figure 7.13 shows that PVSyst generally performs worse than TU Delft's and Solar Monkey's frameworks throughout the year.

7.5.3. SENSITIVITY TO INPUT IRRADIANCE

Given the reported results, we suspected that the issue with these simulations might stem from the input data, particularly the interpolation of weather data. The decomposition of global irradiance, a typically inaccurate step, could further contribute to the error.

To test the first hypothesis, we estimated the error caused due to the interpolation of

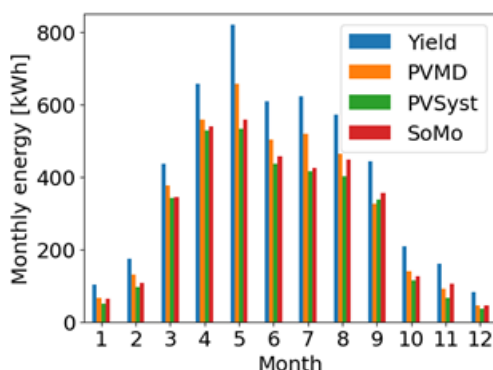


Figure 7.13: Real and estimated monthly energy yield by the PVMD toolbox, PVSyst, and Solar Monkey (labeled as SoMo).

the weather parameters. We located a hypothetical PV system right next to a meteorological station. First, the yield of that hypothetical system was estimated using the data from the weather station. Then, the yield of that same system was estimated by assuming that there is no meteorological station in that location. By comparing the difference in yield between the two situations, one can estimate the error due to the interpolated weather data. The characteristics of system 2150, which shows the lowest error, were the ones considered for the hypothetical system. Three scenarios were considered depending on the location and number of meteorological stations surrounding the target meteorological station:

- Best-case scenario: the meteorological station with 6 stations within 50 km, the most surrounded KNMI station, being the closest station at less than 1 km.
- Close-by scenario: the meteorological station is the closest to PV system 2150, the target PV system. It is surrounded by two other stations within 50 km at distances around 24 and 34 km.
- Mean dist. scenario: the meteorological station is surrounded by two equally-distanced stations within 50 km, each at an average distance of 25 km.

The results are reported in [Table 7.4](#). The errors exclusively due to interpolated weather conditions are high. Even in the best-case scenario, the NRMSE is almost 23%. Looking at these values, and considering that the NRMSE for system 2150 is 41.7%, it is highly probable that the origin of most of the error is the input data.

Table 7.4: Metrics comparison for the three different made-up scenarios.

	Best-case	Close-by	Mean dist.
NRMSE [%]	22.9	53.3	52.6
NMAE [%]	5.6	20.2	19.6
NMBE [%]	-2.7	-3.9	-4.2
Measured [kWh]	4319.6	4803.7	4684.0
Interpolated [kWh]	4202.4	4618.0	4488.9

To test the second hypothesis, that the error is caused by the global irradiance decomposition, a similar approach is followed. A hypothetical PV system is located next to a meteorological station. Using the measured weather data, the yield is computed, which is considered the real one. Next, we artificially alter each irradiance component (GHI, DHI, DNI) in the range of -5% to 5%, achieving 30 scenarios. This new weather data is used to estimate the PV power and compare it with the baseline scenario, the assumed real yield. The error made in each of these 30 scenarios can be seen in Figure 7.14:

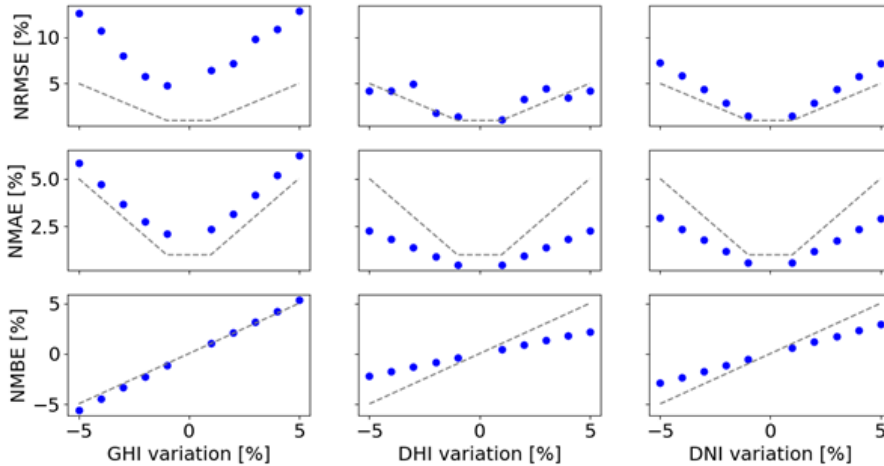


Figure 7.14: Error metrics as a function of the variation of each of the three irradiance components.

The variation in GHI is the one that propagates the most due to the decomposition step. A 5% variation in GHI can increase the NRMSE to 12%. Even a small GHI variation of 1% translates to around 5% in PV power. The NRMSE in power when modifying the other irradiances is only slightly higher than the artificially added irradiance error, but lower in terms of the other two metrics. Therefore, errors in DNI or DHI are not highly propagated in terms of PV power.

7.6. CONCLUSIONS

This chapter introduced the PVMD Toolbox, the first simulation framework capable of modeling the AC energy yield of PV systems using fundamental material parameters without requiring input from fabricated devices. The toolbox achieves this by integrating state-of-the-art models into a modular framework, ensuring the flexibility to incorporate the latest advancements and continuously improve the tool. As a case study, we validated one aspect by reconstructing the horizon using LiDAR images, which is crucial for residential-scale PV systems. During this process, we discovered that accurate geo-location of the PV modules is essential for correctly representing the surrounding horizon.

The toolbox abilities for power prediction were tested with residential system data. The analysis revealed significant errors, with NRMSE ranging from 42.2% to 109.5% and NMAE from 16.3% to 41.3%. A primary cause of the high error rates was insufficient and inaccurate input data. This is a limitation common to all analytical simulation frameworks, as both Solar Monkey and PVSyst recorded even higher errors than the PVMD toolbox. Notably,

interpolated irradiance data emerged as a major source of error, contributing up to 53.3% of the NRMSE, which could account for a significant portion of the reported discrepancies. These findings underscore the critical importance of accurate irradiance data for reliable PV power predictions and the need to identify alternative irradiance sources.

PERSPECTIVE: HOW MACHINE LEARNING COULD HELP

This chapter has shown that interpolated irradiance data from meteorological stations is an unreliable input for PV power predictions. Alternative sources of irradiance should be considered. To achieve this, one could make use of the machine learning knowledge acquired in the previous block of research.

Using machine learning techniques, one can extract the value of irradiance from satellite images. Previous research has been successful in this task [103, 281–286] with convolutional neural network (CNN) being the key algorithm. Combining CNN for image analysis with an algorithm for temporal analysis such as Long Short Term Memory (LSTM), one can even predict the irradiance. Integrating deep learning-forecasted irradiance as input into the toolbox would merge the strengths of machine learning with analytical methods, combining the best of both worlds.

Inspired by these previous works, we contributed to creating a forecast of irradiance maps for the Netherlands in the coming half hour with higher spatial and temporal resolution than currently available with meteorological stations [287]. Specifically, we aimed for a spatial resolution of 1 km² and a temporal resolution of 5 minutes. First, preprocessing techniques were applied to the satellite images to separate clouds from the ground. Then, the CNN-LSTM algorithm aimed at predicting the satellite images for the coming half an hour. These predicted images were converted to irradiance maps by using an Artificial Neural Network (ANN) model. Irradiance measurements of meteorological stations were used as ground truth for the latter model.

Visual inspection is an essential tool to assess the accuracy of this kind of methodology. Figure 7.15 exemplifies the method for a specific instant of time. The top left shows a satellite capture, while the bottom left displays the predicted outcome. Comparing the prediction with the actual images reveals that the predicted images are blurred out. Moreover, small clouds are often clustered together or ignored. The images on the right show the irradiance maps obtained from the real and the forecasted satellite images. The irradiance map on the bottom right is the overall outcome of the model.

In terms of metrics, the first methodology, the CNN-LSTM algorithm, achieved MSE, MAE, and MBE of 0.0257, 0.0905, and -0.0176, respectively. These values correspond to normalized pixel intensities between 0 and 1 and indicate how much, on average, the forecasted pixels differ from the actual values. The second model, the ANN, achieved a MAE of 119.4 W/m² and MBE of -2.6×10^{-2} W/m². The accuracy of the overall model could not be assessed due to the lack of an irradiance map for the Netherlands.

An additional issue was that the background of the Netherlands was considered a cloud in many of the captures, highly reducing the accuracy of the method. This suggests that the simple preprocessing methods applied did not completely fulfill their objectives. Using another CNN for the preprocessing would likely yield a better outcome. Additionally, the methodology could not predict the creation and disappearance of clouds. This could be included by integrating more meteorological variables into the forecasting process. Further work is needed before this irradiance map can substitute the interpolated irradiance,

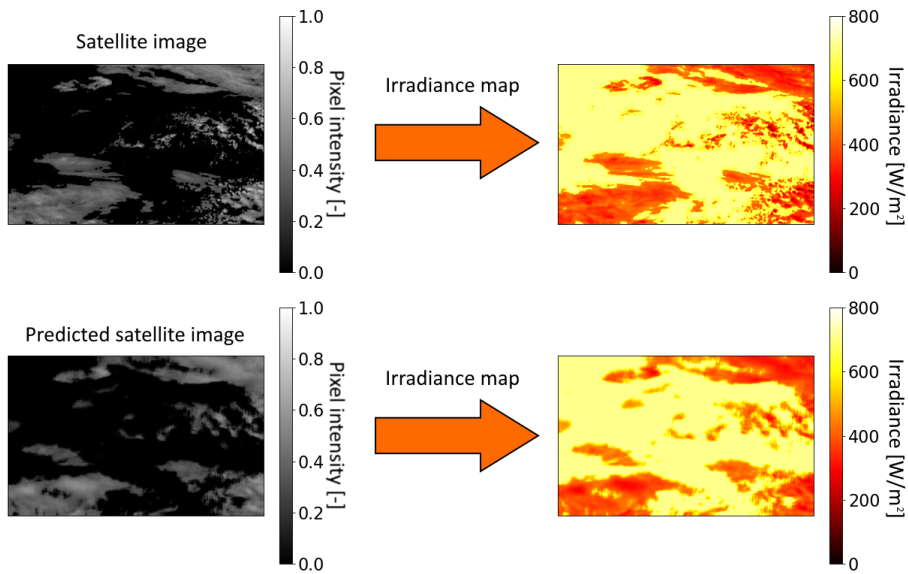


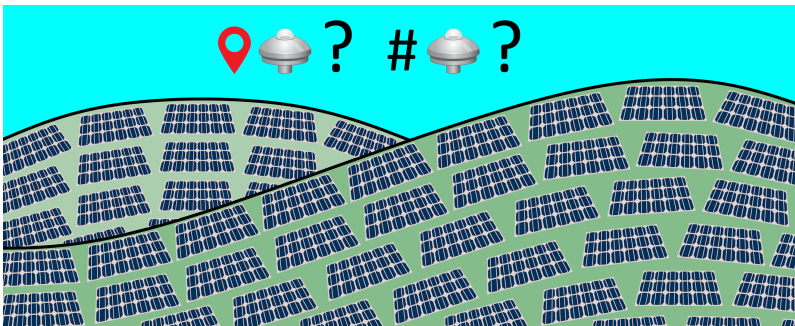
Figure 7.15: Example of the methodology to obtain irradiance maps from satellite images. The top row consists of a true preprocessed satellite image (left) and an irradiance map (right) created by converting each pixel intensity to an irradiance value. The bottom row contains a forecasted satellite image (left) and the converted irradiance map (right). All images are from 14:00 on May 28th 2021.

although it is a promising first step.

8

IRRADIANCE SENSOR ALLOCATION IN PV FARMS

Solar farm installers generally struggle with the allocation of irradiance sensors throughout the plant area, which are essential for monitoring purposes. Despite the International Electrotechnical Commission guidelines for PV plant monitoring, no specific guidance is provided when it comes to allocating sensors. This can be especially problematic for solar farms in hilly terrain. In this work, a software tool is built to allocate horizontal and in-plane irradiance sensors. Additionally, advice on the optimum number of sensors and the prevented error is provided based on the layout of the farm. The software has been applied to two case studies of existing solar farms in hilly areas in Greece and Germany, showing its applicability for real case scenarios in different climates and geological landscapes.



This chapter has been adapted from the publication A. Alcañiz, M. I. van Kouwen, O. Isabella, and H. Ziar “Wide-area sky view factor analysis and Fourier-based decomposition model for optimizing irradiance sensors allocation in European solar photovoltaic farms: a software tool” *Solar Energy*, 286, 113139 (2025) [288].

8.1. INTRODUCTION

The previous chapter has stressed the importance of employing accurate irradiance data when predicting the PV power output. This problem is presented for residential systems, which are generally not equipped with irradiance sensors. Since the power production of these systems is relatively small, one could claim that their monitoring is not essential. That is not the case, however, for large power producers like PV farms, which typically measure the irradiance onsite.

The purpose of these measurements depends on the type of irradiance measured and the time resolution [289–291]. One can distinguish between global horizontal irradiance (GHI) and plane-of-array (POA) irradiance G_{POA} measurements. While the former is placed horizontally, the latter is at the same inclination as the PV modules. Horizontal sensors are employed to connect with satellite data or for forecasting purposes when measuring at high temporal resolutions. Forecasting can also be performed with high-resolution POA sensors, which are also employed for power loss analysis (PLA) on less than a daily basis and performance ratio (PR) calculations on an annual basis. This latter variable can be significantly affected by irradiance deviations [292], highlighting the need for highly accurate irradiance measurements.

To reduce the uncertainty, the standard 61724-1:2021 from the International Electrotechnical Commission (IEC) “serves as guidance for monitoring system choices” [293]. This standard requires class A pyranometer measurements of in-plane irradiance and GHI throughout the plant. The standard also provides the number of sensors that should be placed horizontally and in-plane depending on the PV plant layout. Guidelines are also provided regarding the location of these pyranometers:

- The location shall be chosen as representative.
- Shading on the sensor shall be avoided. If it occurs, it may only be within a half hour from sunrise and sunset and it shall be documented.
- GHI sensors should be leveled to within 0.5°. G_{POA} sensors should be aligned with the intended plane within 0.5° tilt and 1° azimuth angles.

These IEC guidelines can be however confusing due to the absence of clarification on what constitutes a *representative* location. A meticulous allocation approach is crucial when planning the monitoring infrastructure, especially in PV farms with spatial differences in irradiance due to altitude variations or horizon obstructions [294]. The usage of few sensors or incorrectly allocated sensors can lead to, among others, incorrect evaluations and fault assessments [289, 290, 295].

In the literature, some methods for sensor allocation have already been proposed. In [294], an optimization algorithm was presented based on minimizing the error induced by different cloud conditions. However, the proposed method was developed only for a flat rectangular PV plant with an unobstructed horizon. The results reported in [296], show that the optimal sensor placement is away from the edges of the PV farm and at roughly 25% and/or 75% across the module. These results, however, were obtained for an unobstructed tracked bifacial PV farm. Another allocation algorithm used an interpolated irradiance map [297] which was obtained from the sensor data using the Kriging interpolation technique [298]. However, the methodology requires sensor data information hence it cannot be used before the sensors are installed.

A workaround would be using satellite data, making ground-based irradiance measurements redundant. However, satellite irradiance data currently offers low spatial resolution and is less accurate than ground-based measurements [299, 300]. Therefore, most PV power plant commissioners prefer using on-site irradiance sensors [301]. In conclusion, no universal and scalable algorithms exist for irradiance sensor allocation.

This work aims to fill this gap by developing a software tool that provides the optimal sensor allocation for PV farms considering their elevation profile. The irradiance is calculated at all the possible sensor locations of the PV farm, and the most *representative* location is proposed based on an error minimization process. The error definition depends on the purpose of irradiance measurement. Additionally, advice on the optimal number of sensors is provided based on the layout of the farm.

This work is structured as follows. The methodology of the software is explained in detail in Section 8.2. The databases required for the software are presented in Section 8.3. The main results consisting of the software overview and its application to two case studies are presented in Section 8.4. These outcomes are discussed in Section 8.5 before concluding in Section 8.6.

8.2. METHODOLOGY

This section explains the methodology employed to determine the optimal sensor allocation. It assumes that the user knows the location and layout of the future PV plant. The user can or cannot know how the sensors will be used and the number of sensors to be placed.

So far, this methodology has been developed for monofacial fixed-tilted PV farms. Nevertheless, a similar framework applies to the increasingly widespread bifacial and/or tracking PV installations.

The strategy of the methodology consists of calculating the irradiance at each point of the PV plant area and selecting the most representative points as those with values closest to the average. In particular, it consists of the following steps, which are graphically presented in Figure 8.1:

1. Determine the sensor distribution amongst groups and planes.
2. Import Digital Surface Model (DSM) data.
3. Perform Sky View Factor (SVF) calculations for relevant plane-of-arrays and interpolate grid.
4. Import GHI and albedo α .
5. Generate a G_{POA} map for the relevant plane-of-arrays using an improved Boland-Ridley-Lauret (BRL) decomposition model and the Perez transposition model [302].
6. Determine the reference irradiance G_{ref} and calculate the error map for every sensor.
7. Suggest the best sensor locations.
8. Create output report including figures and tables.

In the following, each of these steps is thoroughly explained. However, before diving in, some assumptions need to be taken:

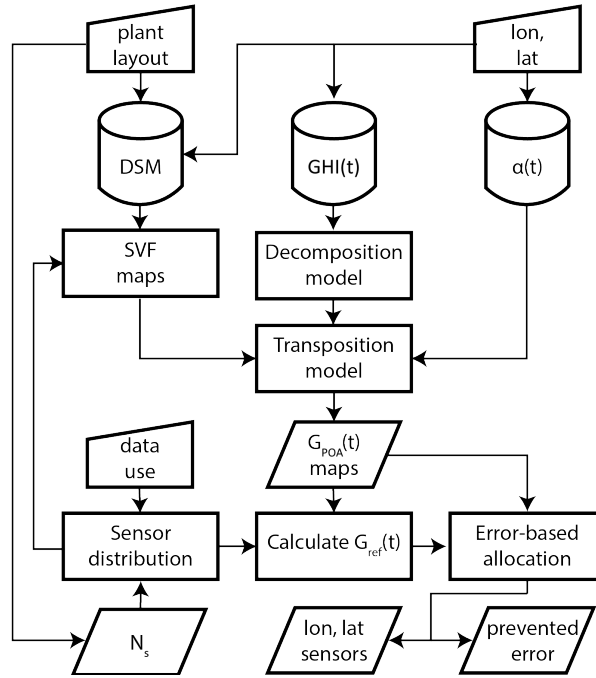


Figure 8.1: Flowchart of the methodology followed in this work. The trapezoids are user inputs, the cylinders denote imported data, the rectangles are predefined algorithms, and the parallelograms denote (intermediate) outputs.

- Pyranometers are placed within the PV plant area, defined here as the terrain to which the PV farm operator has access. Within that terrain, all obstacles such as trees or water streams are disregarded.
- POA sensors are always placed maximally 10 m away from the PV group for which they measure. This step expands the boundaries of the PV group to increase accessibility with the assumption that no significant alterations occur in this additional terrain.

PV group is defined here as a cluster of modules with the same characteristics. To ensure that the sensors capture all the variability across the farm, the user-defined PV groups should be selected by grouping modules that are expected to perform similarly. For instance, by separating modules that could be subjected to different shading levels or that are connected to different inverters.

The difference between PV group and PV plant area is visible in [Figure 8.14](#) for one of the case studies (marked with blue and red dots and labeled as plant area and panel area, respectively).

- PV modules are evenly distributed within user-input PV groups.
- GHI and its decomposed components are spatially constant in the whole PV plant (justified by the use of hourly averaged data).
- For computational efficiency, the SVF calculations for the PV panels and the pyranometers are done at 1.5 m height, approximately the middle point of modules from

the ground [295].

- Albedo is spatially constant but temporally variant.

Subsection 8.2.1 starts by specifying the required inputs. The creation of sky view factor and irradiance maps is explained in Subsection 8.2.2 and 8.2.3, respectively. The core of the methodology, the error-based minimization problem, is presented in Subsection 8.2.4. Subsection 8.2.5 elaborates on the tactic to find the optimal number of sensors for each PV farm. Finally, outputs and performance indicators are described in Subsection 8.2.6.

8.2.1. INPUTS

The user inputs are:

1. Purpose of the irradiance sensors, which influences the sensor type and its optimal spatial distribution. The defined purposes are [289–291]:
 - (a) PR: Yearly calculation of performance ratio using G_{POA} .
 - (b) PLA, daily: Daily power loss analysis with G_{POA} .
 - (c) PLA, real: Power loss analysis in real-time (hourly) with G_{POA} .
 - (d) POA, for: Forecasting models using hourly G_{POA} .
 - (e) Hor, for: Forecasting models using hourly G_{hor} .
 - (f) Sat: Connection with satellite data via GHI.

where G_{hor} is the horizontal irradiance considering horizon obstructions, that is the irradiance measured by horizontally placed pyranometers. It is different from satellite readings of GHI that are based on a free horizon.

When selecting multiple data usages, the user is prompted to rank them so the software tool prioritizes purposes in case of conflicts.

2. PV farm location.
3. Location, tilt, and azimuth of modules, clustered into groups of equal inclination.
4. PV plant capacity.
5. Number of sensors, if known.

8.2.2. SKY VIEW FACTOR

After defining the inputs, the sky view factor (SVF) is computed, a parameter necessary to estimate the incident irradiance. The SVF is defined as the portion of the sky visible from the PV module. It ranges from 0 to 1 and is affected by the inclination of the module and altitude differences in the terrain or horizon obstructions. Digital Surface Model (DSM) data is employed to consider these terrain differences. This data provides the height for equidistant points across an area. The spatial resolution of the SVF map is therefore determined by that of the DSM data. Linear interpolation can be employed to increase the spatial resolution and with it the number of possible sensor locations. The SVF is computed for the entire PV plant area obtaining an SVF map.

The calculation of the SVF is based on the work of Keijzer [271] and further implemented in the work of de Jong [303]. The sky hemisphere is divided into altitude bands and azimuth

slices. The maximum obstructed altitude angle is determined for every azimuth slice. Only DSM grid points at a distance lower than a certain radius of evaluation are considered.

The number of altitude bands is already optimized in [271] to be 90, but the azimuth slices and radius of evaluation are optimized here by using data from the 3797 European solar parks that are registered in the Global Power Plant Database (GPPD) [304] (see Figure 8.2a). Figure 8.2 shows part of the results of the optimization process as a trade-off between accuracy and computational resources. 30m-resolution DSM data from Sentinel Hub is employed in this analysis [305].

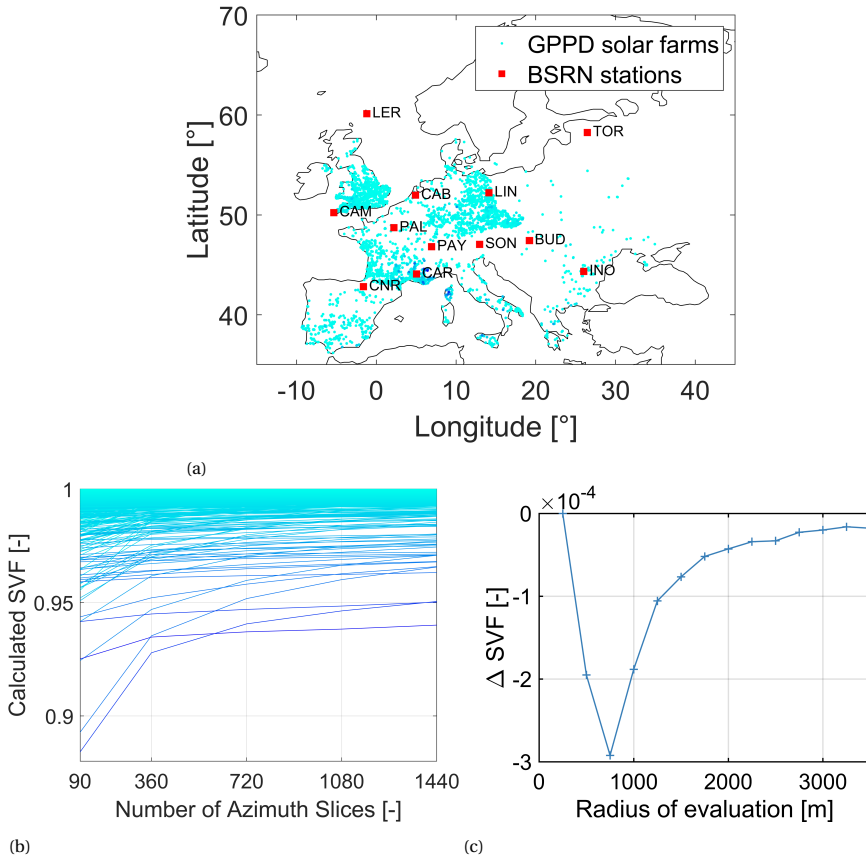


Figure 8.2: Optimization of SVF parameters for a horizontal plane for the GPPD European PV farms. a) Location of the GPPD solar farms in cyan and of the Baseline Surface Radiation Network (BSRN) meteorological stations in red squares with the corresponding labels. b) Effect of increasing the azimuth slices. Each colored line corresponds to a PV farm of the matching color as depicted on the map. c) Mean change in SVF for increasing radii of evaluation compared to the SVF obtained with a radius of 250 m.

The incremental increase in SVF is plotted against the number of azimuth slices for all locations in Figure 8.2b. The results in the figure are in line with what is expected: a higher amount of azimuth slices leads to a more accurate and higher SVF. In this figure, one can also observe how most existing solar plants are in areas without large local height differences since most PV plants have an SVF of 0.97 or higher, and the mean SVF is close to

0.999. 1080 azimuth slices are selected as optimal, given that the SVF barely changes after this point. This corresponds to an azimuth resolution of 0.3° .

Using this optimized value, the SVF was computed for varying radii of evaluation on four planes with different tilts. Figure 8.2c illustrates the mean change in SVF compared to the value at 250 m for the horizontal plane. As expected, the SVF reduces with increased radius. A too-small radius overestimates the SVF due to the underestimation of horizon obstruction. The SVF decreases the most in the 500 m to 750 m range, as most PV plants have few obstructions in their direct surroundings. At large radii, only large hills, mountains, or high-rise buildings will significantly impact the SVF. Similar behaviors were observed for all tilted planes and orientations. Based on this analysis, an optimum value of 2000 m was selected. It should be noted that these results are obtained for the imported DSM data thus they may not apply to data with different resolution.

8.2.3. IRRADIANCE MAPS

Once the SVF is available, the POA irradiance (G_{POA}) can be calculated. This represents the irradiance received by the modules. Several inputs are required, including GHI, Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), the ground albedo α , the PV module angles, and the solar angles. For definitions of these parameters, the reader is referred to [261].

At every time step, the solar angles are calculated from the plant location [261]. The albedo and the GHI are imported as explained in Subsection 8.3.3 and 8.3.2, but obtaining DHI and DNI measurements is challenging due to the need for costly equipment. However, decomposition models can be employed to calculate DHI from GHI. Subsequently, DNI can be derived using the closure equation [306].

The decomposition step is typically the main source of uncertainty in G_{POA} evaluation [307]. Therefore, insight into the best-performing decomposition models is valuable. Despite the several comparisons between decomposition models in Europe available in the literature [279, 308–315], no model stands out.

Therefore, a comparison of the most common decomposition models is performed in Appendix D. Among all the models tested, BRL proves to be the overall best performer, exhibiting superior results in terms of normalized root mean squared error (nRMSE) and an average bias. The bias exhibits a seasonal pattern consistent across all models, with an overprediction of the diffuse component in autumn and an underprediction in spring. This seasonal normalized mean bias error (nMBE) trend is mitigated in this work for the BRL model through a Fourier-mediated adjustment, resulting in the BRL-MvK model, a BRL tailored for European climates.

Once GHI, DHI, and DNI are obtained, G_{POA} can be computed. This irradiance comprises three components: diffuse, direct, and ground-reflected. Whereas the latter two are calculated deterministically using equations 18.18, 18.20, and 18.22 from [261], the diffuse component is determined by a diffuse transposition factor R_d [306]. Different sky diffuse models exist for calculating this diffuse fraction. Comparison studies based in Europe [311, 316–325] show that the Perez model [302, 326] is superior most often.

The time-varying $G_{POA}(t)$ is calculated for each location of the plant over two years. Since only two years are selected, the influence of solar cycles, which have an 11-year periodicity, is not considered. Nevertheless, their impact on yearly incident irradiance is minor and can be disregarded. Since the $G_{POA}(t)$ depends on the PV module angles, this process

is performed for every PV group or plane. When horizontal sensors are to be placed, the process is repeated considering a 0° tilt.

8.2.4. ERROR-BASED MINIMIZATION PROBLEM

Once the irradiance maps are ready, the pyranometers can be allocated. As mentioned in the introduction, the objective is to find the most *representative* location. At that location, the irradiance time series reflects the irradiance incident on the group of PV modules to which the sensor is assigned. A representative time series can be defined as a series in which the values are close to the reference time series at each instant. Since each module in the group is subjected to a different irradiance profile and all modules have the same weight, the reference time series is defined as the mean irradiance.

Using the time-dependent irradiance map, the reference irradiance $G_{ref}(t)$ is constructed by averaging the irradiance for each PV group at every instant. That value is then compared to the calculated $G_{POA}(t)$ for every potential pyranometer location l in the PV plant area $l \in L$. L is the number of possible locations in which a sensor can be placed. $G_{POA,l}$ is the POA irradiance at a location l . The location with the minimal error is chosen as the most representative one. The error $\epsilon_{p,l}$ is calculated as the sum of the absolute differences for every time instant, see Equation 8.1. $p \in [h, d, y]$ as the error can be calculated on an hourly, daily, or yearly basis. The time resolution of the error is determined by the monitoring purpose, as the minimization problem is defined by the period $t \in T$ over which no bias should occur.

$$\arg \min_{l \in L} (\epsilon_{p,l}) = \arg \min_{l \in L} \left(\sum_{t \in T} |G_{POA,l}(t) - G_{ref}(t)| \right) \quad (8.1)$$

G_{ref} varies depending on the sensor orientation and PV plant distribution. The following subsections explain the procedure for each possibility.

SINGLE GROUP

First, the objective is to place an in-plane pyranometer representative of a group of PV modules that reside in a single area with identical tilt and orientation (same plane). Assuming equal distribution across the PV group, G_{ref} is the average across all irradiance time series within the PV group, Equation 8.2.

$$G_{ref}(t) = \frac{1}{L} \sum_{l=1}^L G_{POA,l}(t) \quad (8.2)$$

MULTIPLE GROUPS, SAME PLANE

In this second case, all PV modules have identical tilt and orientation (same plane), but they are not clustered, therefore belonging to multiple groups of modules. G_{ref} is defined as the average of the irradiance of each group weighted by the number of modules in each group. Equation 8.3 expresses this weighted average, where N_{gp} are the number of groups in the PV farm, N_n are the number of modules in each group n , and L_n is the number of locations in which a sensor can be placed in group n .

$$G_{ref}(t) = \frac{1}{\sum_{n=1}^{N_{gp}} N_n \cdot \sum_{n=1}^{N_{gp}} L_n} \sum_{n=1}^{N_{gp}} \sum_{l=1}^{L_n} G_{POA,l}(t) \cdot N_n \quad (8.3)$$

Since the spatial resolution of the imported data may not coincide with the distance between PV modules, the possible sensor locations L_n need to be considered. The schematic in Figure 8.3 can help understand this. A PV plant is represented by three groups of modules ($N_{gp} = 3$), all with the same orientation. Sensor locations are marked by colored dots: red for group 1, orange for group 2, purple for group 3, and black for boundaries. Considering this, $N_1 = 9$, $N_2 = 5$, and $N_3 = 12$, but $L_1 = 4$, $L_2 = 3$, and $L_3 = 8$. In the software tool, some of these boundaries would also be considered as pyranometer locations (as long as they are within 10 m of a PV module) but they are discarded for ease of understanding in this example.

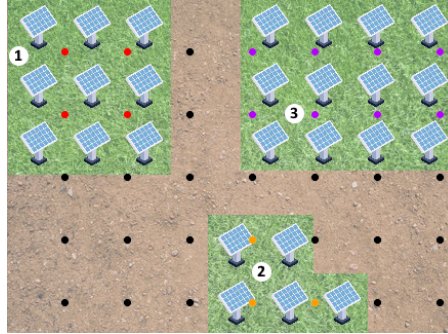


Figure 8.3: Example of a PV farm sketch with 3 groups of PV modules and matrix of sensors' locations to help understand the calculation of G_{ref} .

HORIZONTAL SENSORS

For the horizontal sensors, the following adaptations occur in the minimization procedure:

- The SVF grid covers the whole plant area.
- G_{POA} is replaced by G_{hor} for a forecasting sensor and GHI for a satellite connection sensor. This distinction is because local differences should be included only for forecasting purposes.
- The hourly bias is always minimized.

MULTIPLE SENSORS

When multiple sensors within the same plane and group are to be placed, the minimization procedure changes:

1. The grid points that are within the lowest 10% of the error map are identified.
2. Using k-means clustering [327], k clusters are formed, where k is equal to the number of to-be-placed pyranometers.
3. The minimization is performed for each cluster.

If simply the k best locations are chosen, all locations will be next to or near the global minimum and thus spatially very close. k-means clustering ensures that the pyranometer locations are spatially spread.

8.2.5. NUMBER OF SENSORS

If the user indicates no number of sensors, advice is generated by the software considering no (financial) limitations. The rules are that when at least one in-plane data usage is chosen, a sensor is advised for every plane and an additional sensor per PV group. One sensor is advised for each use of horizontal irradiance measurements. Figure 8.4 graphically summarizes this decision process.

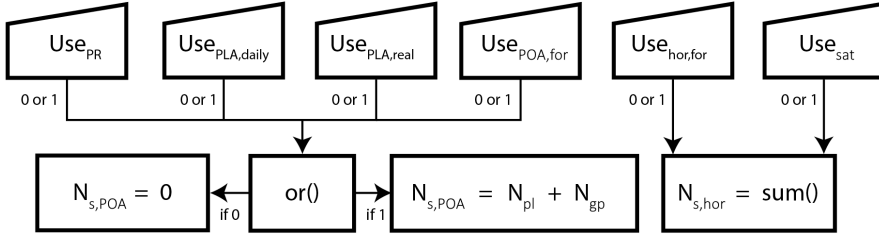


Figure 8.4: Flowchart illustrating how the advice for the number of horizontal and in-plane sensors follows from the foreseen data collection purposes as input by the user. The trapezoids denote binary variables that indicate whether a certain data purpose is entered by the user. The subscripts in the figure denote the six data-gathering purposes.

The number of sensors advised by the software tool is from now on referred to as the *ideal number of sensors*. The eventual distribution of the ideal number of sensors N_s when all future purposes are chosen is:

- $N_{s,POA,pl} = N_{pl}$
- $N_{s,POA,gp} = N_{gp}$
- $N_{s,hfor} = 1$
- $N_{s,sat} = 1$

where N_{pl} stands for the number of planes in the PV plant (defined as the unique combinations of tilt and orientation), $N_{s,hfor}$ denotes the number of horizontal sensors for forecasting, and $N_{s,sat}$ denotes the sensor for satellite connection.

When the user inputs a number of sensors that deviate from the ideal, the following rules, implemented via nested if-else loops, ensure the optimal distribution:

1. When $N_s = 1$, and both in-plane and horizontal sensors are required based on the data purposes, the in-plane sensor gets priority if $N_{pl} = 1$. Otherwise, $N_{s,hfor} = 1$.
2. When $N_s = 2$, and more in-plane and horizontal sensors are required to accomplish all intended data purposes, $N_{s,POA} = 1$ and $N_{s,hfor} = 1$, as long as it does not violate another rule.
3. When $N_s > 2$ and one or two horizontal sensors are to be placed: $N_{s,hfor} > 0$.
4. Distribution of in-plane sensors follows:
 - $N_{s,POA,pl} = N_{pl}$ or 0 (all planes or no planes have a sensor).
 - $N_{s,POA,gp} = N_{s,POA} - N_{s,POA,pl}$.
 - In-plane group sensors are distributed based on the group size.
5. When conflicts occur, the user-input priorities determine how the pyranometers will be placed.

8.2.6. PERFORMANCE INDICATORS

Next to providing allocation advice, the software tool intends to show the benefits of its use via performance indicators. The main used performance indicator is the prevented measurement deviation (PMD) $\Delta\epsilon_p$ (see Equation 8.4). It indicates the accuracy improvement of the allocated sensor compared to the worst-case location, thus it can be interpreted as the prevented error made. It can be defined for every time resolution, with $p \in [y, d, h]$.

$$\Delta\epsilon_p = \epsilon_{p,l_w} - \epsilon_{p,l_b} \quad (8.4)$$

Where l_w and l_b are the worst-case and the advised sensor location, respectively. These values are directly retrievable from the error map.

Another performance indicator is the relative PMD:

$$rPMD_p = \frac{H \cdot \Delta\epsilon_p}{H' \cdot \sum_{h=1}^H G_{ref}(h)} \cdot 100\% \quad (8.5)$$

With $H' = 8760$ and H the number of hourly time steps employed, as the methodology is commonly performed for two years of data.

8.3. DATABASES

This section presents the databases needed for the software tool. Subsection 8.3.1 justifies the selection of the digital surface model data. Subsection 8.3.2 briefly presents the retrieval of the GHI satellite data, and Subsection 8.3.3 explains the choice and preparation for the time-dependent albedo data.

8.3.1. DIGITAL SURFACE MODEL

The digital surface model (DSM) data is used in this work to calculate the SVF. To access DSM data in the entirety of Europe and to prevent bulk downloads of the height data, an online tool is chosen that combines globally available DSM data: Sentinel Hub [305]. Sentinel Hub is a big data satellite imagery service. Users can quickly and easily access satellite data from complete archives using an application programming interface (API). The Sentinel Hub service offers at least 30m resolution DSM data for Europe.

8.3.2. GLOBAL HORIZONTAL IRRADIANCE

Since the possibility of using an API and having a high resolution was a priority in this work, the PV-GIS database was selected for importing historical measurements of GHI [328]. Using the BSRN ground-based measurements [329], the PV-GIS imported satellite-based measurements are validated for the year 2016. The RMSE is in the range of 90 to 140 W/m² for all stations. The source of this inaccuracy likely arises from the cumulative error propagation throughout all the modeling steps. For this work, however, ensuring unbiased data is more crucial than historical accuracy. The average absolute nMBE is only 3% and that is considered good enough.

8.3.3. TIME-DEPENDENT ALBEDO

The effective ground albedo α fluctuates during the day due to the reflectivity of the ground being dependent on the incident angle of radiation [330, 331], and throughout the year due

to vegetation changes, snow coverage, and other factors [331, 332]. When albedometers are unavailable, satellite data can be used to determine the ground albedo [333, 334]. Nevertheless, the spatial resolution of this data is too low to be used on a local scale. The available albedo databases with a spatial resolution of up to 1 km sacrifice temporal resolution, but their spatial resolution is still insufficient for local-scale applications [335].

Considering this, satellite data is used as a source of time-dependent ground albedo. A resolution of one hour is needed, as the data will be combined with the hourly GHI. The only database with hourly albedo in Europe is the NASA POWER database [336]. The spatial resolution of these albedo measurements is only $1^\circ \times 1^\circ$, which is around $110 \text{ km} \times 110 \text{ km}$. Although this is not a high spatial resolution, using a spatially inaccurate time-dependent albedo is considered closer to reality than assuming a constant albedo.

Many stochastic climatic processes induce a high variability in the yearly albedo pattern. Therefore, as recommended by the literature [337, 338], hourly albedo data from 2017 until 2021 are averaged to create one sample year of data.

8.4. RESULTS

The developed methodology was implemented in MATLAB in the form of a software tool. The software is presented in [Appendix E](#). Its application in two case studies is explored in [Subsection 8.4.1](#). A simplification of the current version of the software is explained in [Subsection 8.4.2](#).

8.4.1. CASE STUDIES

This section presents the use of the tool on two existing PV plants located in hilly terrain as case studies. For each of them, the resulting pyranometer allocations are presented, together with all preliminary results that have led to an informed decision on the most representative locations.

The 6 MW_p solar park in Kolindros, Greece, is situated close to the mountain range where the mountain Olympus is. As seen in [Figure 8.5a](#), it is a solar park with only two groups of PV modules, oriented south. The tilt is estimated to be 30° which is the optimal tilt for south-oriented panels at that location [339].

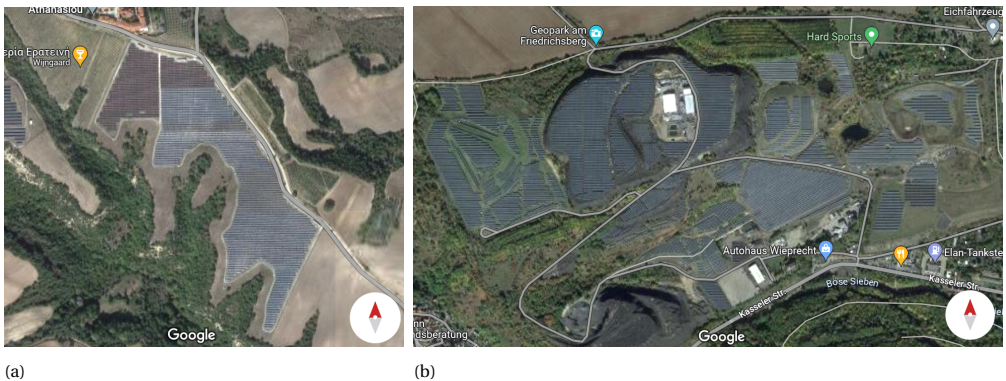


Figure 8.5: Satellite images of the solar park a) in Kolindros (Greece) and b) in Eisleben (Germany) used as case studies. Images from [340].

The Solar Park Krughütte is a 29.1 MW_p plant in Eisleben, Germany. It consists of small groups of PV modules south- and southeast-oriented, as seen in Figure 8.5b. The tilt is estimated to be 33°, which is optimal for that location [339].

SOLAR PARK IN KOLINDROS

The tool starts by calculating the SVF for all grid points within the possible sensor allocation area. Since the SVF depends on the tilt and orientation, SVF maps are created for the horizontal plane and the 30°-inclined and south-oriented plane. The results are shown in Figure 8.6. The SVF for horizontal planes and tilted planes decreases towards the southern part of the PV plant. This is a direct result of the local height differences in the PV plant, as the altitude is relatively low in the south compared to the northwest corner. The tilted SVF is lower than the horizontal one, as seen by the range of values of the color bars. For horizontal planes, the SVF ranges between 0.98 and 0.995, while for tilted planes, it varies from 0.895 to 0.93. The maximum difference in SVF across the PV plant for the tilted plane is 0.035. Although a small number, it can lead to noticeable G_{POA} differences over the year.

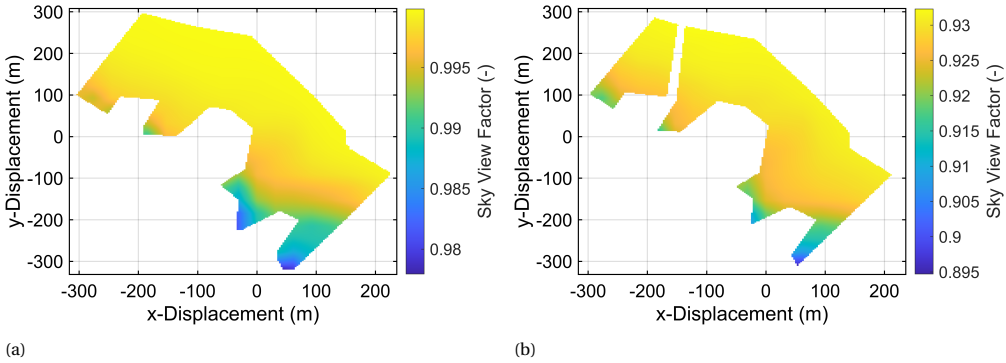


Figure 8.6: Surface plots of the SVF map for the horizontal (a) and in-plane (b) sensors of the Kolindros case study. Please note that the gradient scales are tuned for each subfigure.

Using the SVF maps, G_{POA} maps are determined with the imported GHI time series for the years 2018-2019. For every grid point, the error is calculated. The error definition and time resolution depend on the intended data usage, as explained in Subsection 8.2.4. Figure 8.7 displays the hourly error maps for horizontal sensors for forecasting and connection with satellite purposes.

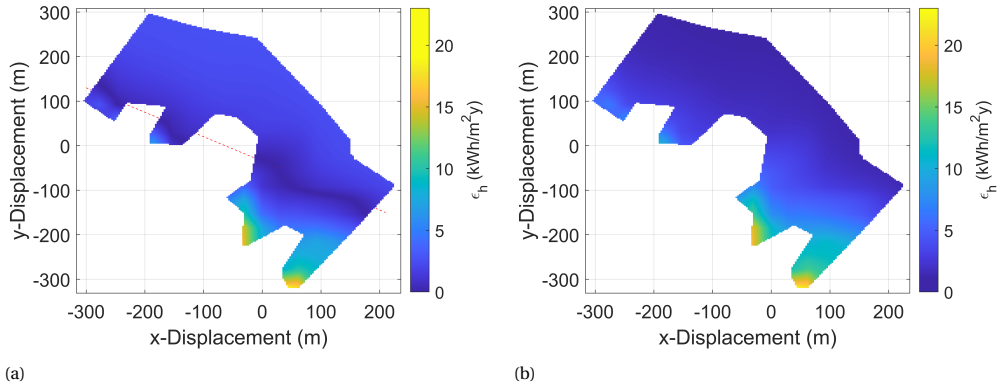


Figure 8.7: Surface plots of the hourly error map for the horizontal sensors to be placed for a) forecasting purposes and b) satellite connection purposes in the Kolindros case study. The red dashed line in (a) is included to help the eye identifying the band of points with minimal error.

As seen in Figure 8.7a, the error for the sensor used for horizontal forecasting is minimal at a band of grid points that crosses diagonally the PV plant (indicated by the red dashed line in the figure). Thus, the most representative locations are clustered around this line. Figure 8.7b shows the error map for sensors that will measure close to GHI. Therefore, the optimal point is found at locations where the horizontal SVF is closest to unity.

Moving on now to the error map for in-plane sensors in Figure 8.8, the lowest error values are once again concentrated within a band that diagonally traverses the PV plant from northwest to southeast. The group sensors are placed using error maps for all separate groups, created using the reference irradiance of that group.

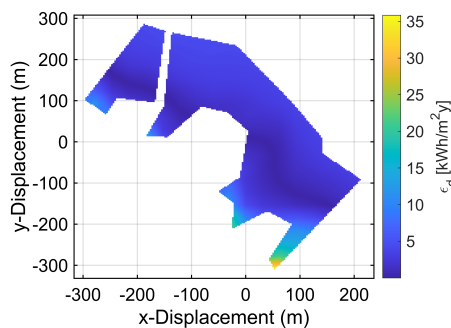


Figure 8.8: Surface plots of the daily error map for in-plane sensors in the Kolindros PV plant.

The software tool was tested for the Kolindros case without inputting a predetermined

number of sensors. The data usage input by the user was prioritized as follows:

1. Daily power loss analysis
2. Forecasting using G_{POA}
3. Forecasting using G_{hor}
4. Connection to satellite data

The resulting monitoring infrastructure, therefore, needed in-plane sensors, a horizontal sensor for measuring G_{hor} , and a horizontal sensor for measuring GHI. Three in-plane sensors were advised because there are two groups and one plane in the PV plant. In conclusion: $N_s = 5$. The in-plane sensors were optimized using the yearly error since PR calculations have the highest priority amongst the G_{POA} uses. Figure 8.9 shows the sensors locations. These results align with the information on the error and SVF maps. For instance, the horizontal satellite sensor is positioned where neighboring obstacles are minimal (highest SVF) while the horizontal forecasting sensor is allocated in the middle of the farm, where SVF is a good representative of the whole farm.

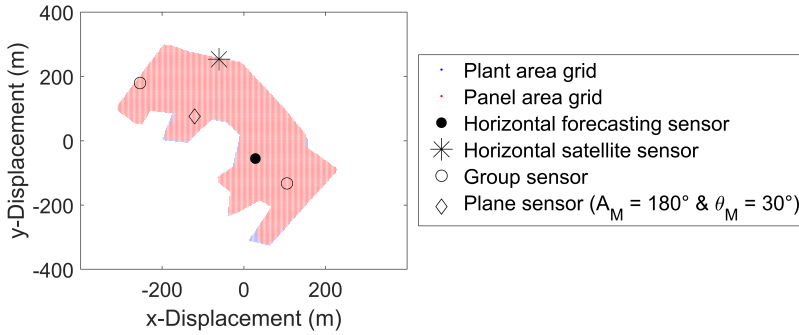


Figure 8.9: Allocation of five sensors as advised by the software tool for the Kolindros case study.

The hourly measurement deviations of the sensors (Equation 8.4) are $0.2 \text{ Wh/m}^2\text{y}$ for the G_{hor} sensor, $3.2 \text{ Wh/m}^2\text{y}$ and $4.5\text{e-}4 \text{ Wh/m}^2\text{y}$ for the group sensors, and $0.1 \text{ Wh/m}^2\text{y}$ for the plane sensor. This is a negligible measurement deviation which demonstrates the efficacy of the allocation algorithm for these sensors. For the GHI sensor, the measurement deviation is $1.1 \text{ kWh/m}^2\text{y}$. This is higher than for the other sensors because the SVF at that location is not exactly unity, leading to a deviation of GHI. This measurement deviation, however, is still small compared to the annual incident irradiance of $3.2 \text{ MWh/m}^2\text{y}$, which is the irradiance measured by a hypothetical horizontal sensor placed on top of the PV farm with no horizon obstacles. The average relative measurement deviation over all allocated sensors is 0.8%. Although the measurement errors throughout the year are small, the instantaneous relative measurement error can be up to three times larger. The largest errors occur at hours with large DHI and low albedo.

When a predefined number of sensors lower than five was input, the same locations for a part of the sensors in Figure 8.9 were found:

- $N_s = 4$: same results without the left group sensor (due to this group being the smallest).

- $N_s = 3$: same results without the group sensors.
- $N_s = 2$: in-plane sensor and forecasting sensor (based on priority).
- $N_s = 1$: only the in-plane sensor (based on priority).

For $N_s > 5$, the additional sensors are used as extra group sensors, distributed amongst the two groups. The largest group always has an equal or larger number of sensors than the smaller group. In Figure 8.10, $N_s = 8$ was used to observe the effect of the k-means clustering. As expected, the group sensors are well distributed over the module area. The fact that all minima in the error map are clustered around a line leads to all cluster minima, and thus group sensors, are positioned around that line as well.

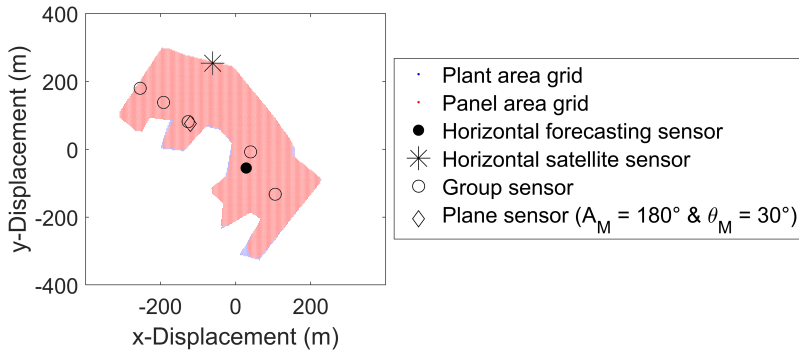


Figure 8.10: Map of the allocation of eight sensors for the Kolindros case study.

SOLAR PARK KRUGHÜTTE IN EISLEBEN

For the second case study, since there are two distinct PV module orientations, the SVF maps for the southeast- and south-oriented planes are shown separately in Figure 8.11. The SVF pattern results directly from the DSM. The western (left) part of the PV plant is on a hill, up to 50m higher than the eastern part. This is reflected by a high SVF in the west and a low SVF in the east. The maximum SVF differences seen within planes are 0.02, significantly lower than at the Kolindros plant.

The resulting hourly error maps for horizontal sensors in the Eisleben plant are seen in Figure 8.12. The maximum measurement deviation is 8 to 10 kWh/m²y, over two times lower than in the Kolindros case. This results from the lower SVF difference.

The error maps for allocating the southeast- and south-oriented in-plane sensors at the Solar Park Krughütte are shown in Figure 8.13. One can observe a high correlation with the SVF maps shown in Figure 8.11.

The assumed user input for prioritized data purposes is:

1. Real-time fault detection
2. Forecasting using G_{POA}
3. Forecasting using G_{hor}
4. Connection to satellite data

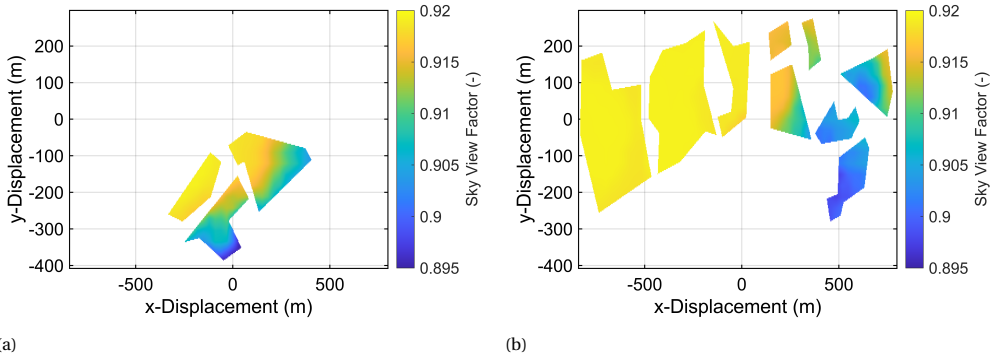


Figure 8.11: Surface plots of the SVF map for the a) southeast-oriented and b) south-oriented planes of the solar park Krughütte.

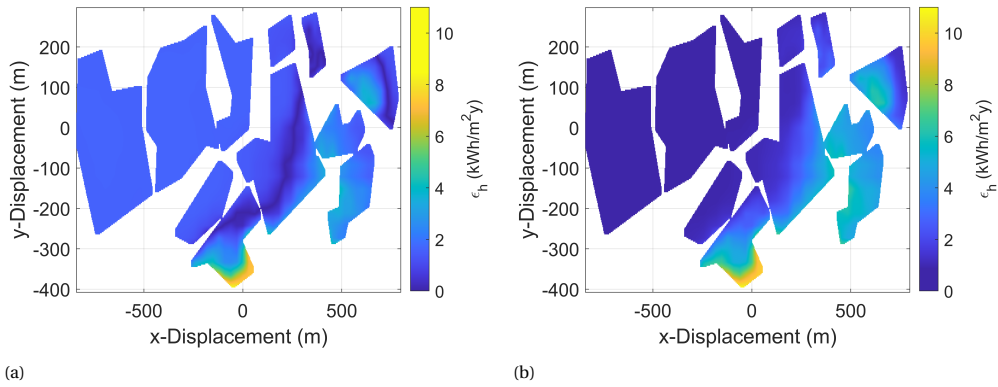


Figure 8.12: Surface plots of the hourly error map for horizontal sensors to be placed for a) forecasting purposes and b) satellite connection purposes in the Eisleben PV plant.

Due to these data purposes, the resulting monitoring infrastructure needed in-plane sensors, a horizontal sensor for G_{hor} , and a horizontal sensor measuring GHI. Fourteen in-plane sensors were advised since there are twelve groups and two planes in the PV plant. Overall, $N_s = 16$. To have sixteen pyranometers in the PV plant might be unnecessary and too expensive. Some groups can easily be combined into one larger group with only one in-plane sensor. However, the software tool is currently only capable of using panel groups as defined by the user. The in-plane sensors are optimized using the hourly error since the real-time PLA has the highest priority amongst the G_{POA} uses.

In Figure 8.14, the advised sensor allocation in Eisleben is visualized. The horizontal satellite sensor is at the maximum SVF location, the far west of the solar park. The horizontal forecasting sensor is in the middle of the plant as expected from Figure 8.12b, and the plane sensors are also in line with the error maps in Figure 8.13. Two group sensors appear as one due to the close proximity of their group perimeters.

The average relative prevented measurement bias of in-plane sensors for this case study

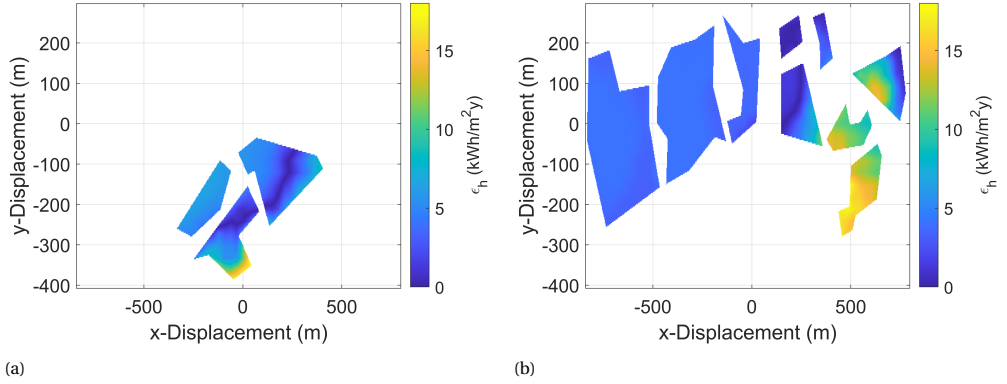


Figure 8.13: Surface plots of the hourly error map for in-plane sensors of the a) southeast-oriented and b) south-oriented planes in the Eisleben PV plant.

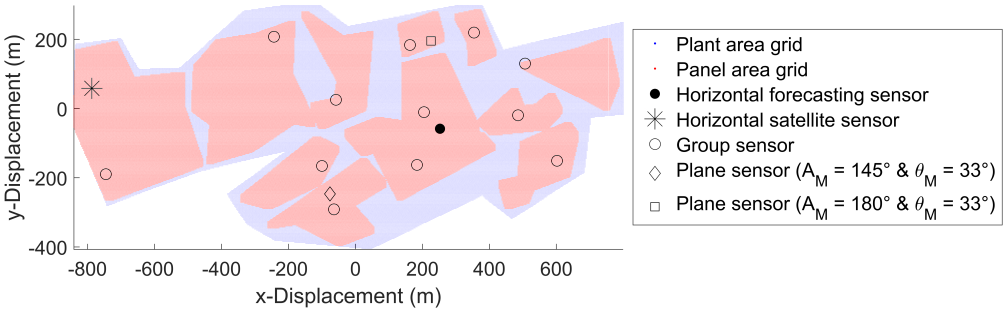


Figure 8.14: Allocation of sixteen sensors as advised by the software tool for the Eisleben case study.

8

is 0.3%. This is lower than the Kolindros case, as expected from the smaller local SVF differences in the Eisleben case.

After analyzing the two case studies, one can compare the results with the IEC 61724 recommendations. Considering the size of the PV farms, their guidelines suggest at least two horizontal irradiance sensors and two in-plane ones. This recommendation is followed for the farm in Kolindros, where the software suggests 5 sensors, but not for the Krughütte farm where the software suggests 16 sensors. As already mentioned, 16 sensors may be unnecessary and financially unfeasible. However, having only 4 may be inaccurate considering that two distinct orientations are present. That can lead to inaccuracies affecting the operations of the PV farm.

8.4.2. SIMPLIFICATION

The sensor allocation was found to have a high dependency on the SVF map. The horizontal satellite sensor is allocated at the point with the highest SVF, and the other sensors are all allocated at the contour lines of the SVF.

It has been proven mathematically in [Appendix F](#), by exploring the definition of the terms involved, that the minimization problem of all the error functions reduces to [Equa-](#)

tion 8.6. In the equation, \widetilde{SVF} is the average SVF weighted by the number of modules per group. Thus when the measurement deviations are minimized over all the locations, the difference between the SVF of locations and the weighed SVF is minimized. This equality means that the software tool does not have to distinguish between the time resolution of the errors when minimizing. This explains that all minimum errors are located along a contour line with the value of the weighted average SVF on the map.

$$\begin{aligned} \argmin_{l \in L} (\epsilon_{h,l}) &= \argmin_{l \in L} (\epsilon_{d,l}) = \argmin_{l \in L} (\epsilon_{y,l}) = \min_{l \in L} |\Delta SVF_l| \\ \Delta SVF_l &= SVF_l - \widetilde{SVF} \end{aligned} \quad (8.6)$$

Therefore, the whole allocation algorithm can be done based on solely the SVF maps. For all planes and groups, \widetilde{SVF} can be calculated, and the location at which the SVF is closest to this value is the most representative. Since the horizontal sensor used for connection to satellite data is placed at the highest SVF, the SVF map is also sufficient here.

These simplifications are based on the assumptions that the GHI, DHI, and DNI time series are location-independent, at least within the geographical scope of the installation. Additionally, the found simplification is based on the assumption that the albedo is spatially constant. This is valid for the current input possibilities of the software tool. Moreover, the simplification is valid only when the angle of incidence amongst all modules is constant. That is true because, in the current version of the software tool, all modules in a group are bound to be equally tilted and oriented. When discontinuously tilted groups are included as input options, the simplification no longer holds, and the G_{POA} and error maps should be included again. Finally, the behind-the-horizon effect is neglected in the current algorithm. This implies that the solar radiation when the sun is below the horizon is ignored, such as during sunrise or sunset. If this is included in the software tool, an extra location-dependent variable is added, making the simplification invalid.

8.5. DISCUSSION

The reported average relative prevented measurement bias for the two presented case studies is below 1%. This value is lower than the uncertainty of pyranometers, which can be 1.3 to 1.7% in economically relevant hours of the day [296, 341]. Therefore, one could question the significance of this work. However, one also needs to consider that the accuracy of the employed data, especially the DSM data, can underestimate the bias. If more accurate terrain data was obtained, for instance from a scanning drone [233], the height differences could be more precisely captured and the detailed shading profiles could be included in the code. The former may also impact the variation in tilt due to the terrain of the modules, assumed to be constant in the case studies. The latter could be used to distinguish between the irradiance profiles of the modules located at the edges of the plant compared to those at the center. All of these aspects would impact the prevented error.

The geographical scope of this work has been limited to Europe. However, most of the methodology can be extended to other geographical regions by employing databases covering those regions. The only point that requires special attention is the choice of the irradiance decomposition model. Considering the importance of this step in the methodology, an analysis similar to the one performed in [Appendix D](#) should be undertaken before extending the software tool to new regions.

Related to this latter point, the accuracy of the newly presented BRL-MvK model should be tested for other locations. It would be interesting to observe whether the seasonal variation of the bias reported in this work is present in regions outside of Europe.

Finally, the software tool should be upgraded to cover the increasingly common bifacial and tracked PV systems. To achieve this, special attention should be given to the estimation of the irradiance.

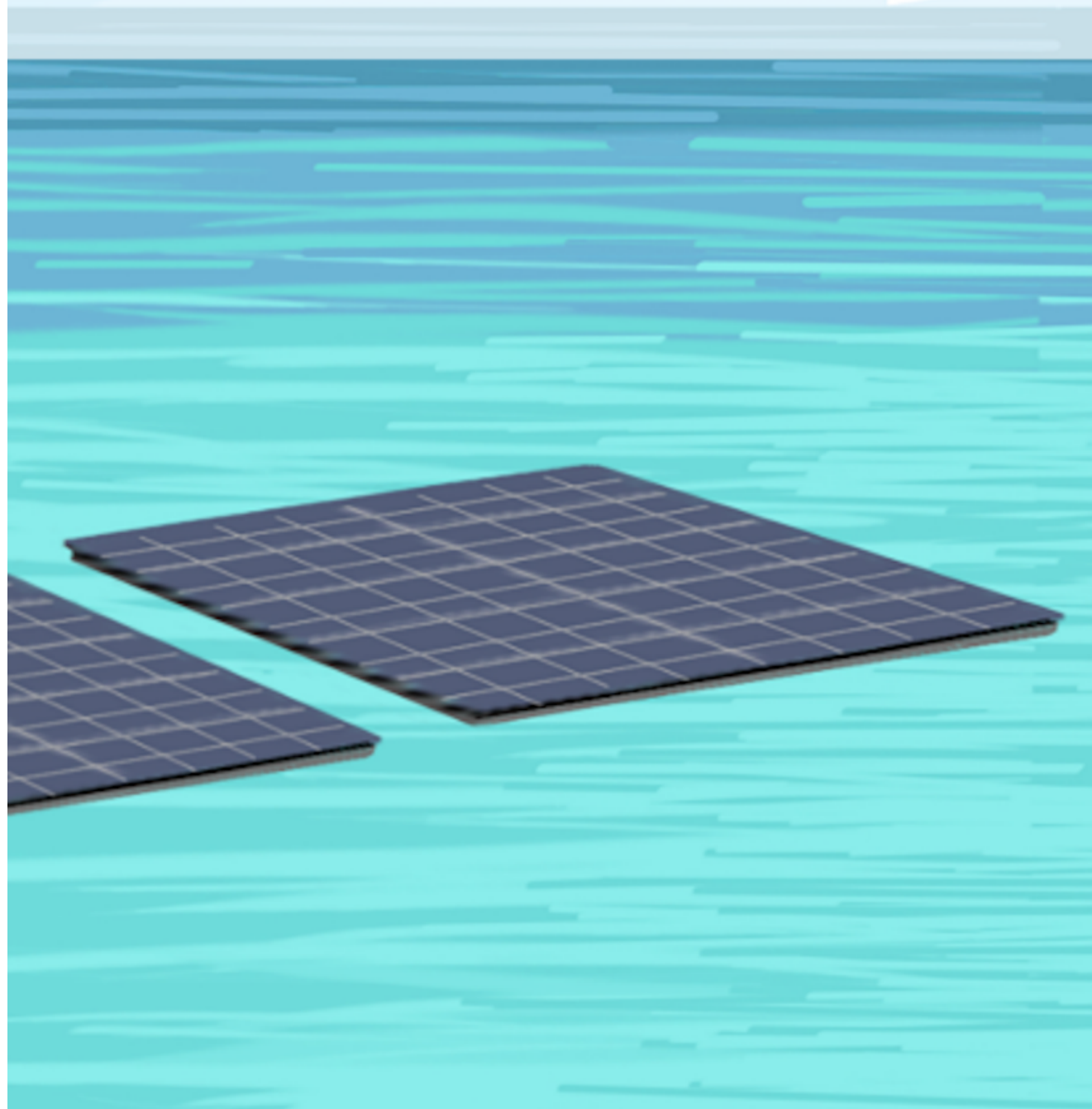
Out of the two, tracked PV systems may be the simplest to implement. The yearly energy yield should be calculated considering the angles of the tracker instead of the irradiance on a fixed plane. If the tracking mechanism is analytically implemented, this extension can be relatively straightforward. More consideration is required for trackers that adjust the angles in real-time to maximize the overall irradiance harvesting.

Bifacial systems may require an additional sensor to measure the rear irradiance, therefore a sensor facing the ground. The guidelines provided in the IEC 61724 [293] could be used as a starting point when adding this feature to the tool. Additionally, an accurate modeling of the albedo should be included, that is not only time-dependent but also location-dependent. However, the lack of available albedo data may currently hinder this implementation. Efforts in the literature to estimate the albedo from satellite data may help overcome this limitation [342]. If the albedo is finally made location-dependent throughout the farm, the algorithm simplification presented in Subsection 8.4.2 would no longer hold.

8.6. CONCLUSIONS

In this work, an irradiance sensor allocation algorithm for Europe has been proposed and implemented in the form of a software tool. The strategy consists of finding representative sensor locations following IEC guidelines. Additionally, the software proposes the optimum number of sensors in a PV farm based on the IEC guidelines and the purpose of data usage and provides metrics to show the benefits of using the developed software. In the process of the software tool creation, several steps were optimized. By employing data from existing farms, the optimum parameters for the sky view factor calculations using digital surface elevation data are found to be a radius of 2 km and an azimuth resolution of 0.3° . Considering the high bias of the irradiance decomposition models of the literature, a custom model has been proposed which is based on the Perez model. The software was applied to two case studies of existing PV farms in hilly terrain. Although the method has been currently developed for PV farms composed of monofacial fixed PV modules with similar orientation (i.e. no East-West configuration), the methodology can be extended to bifacial and/or tracking PV installations.

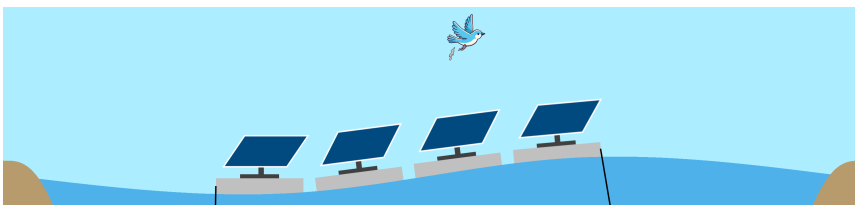
FLOATING PHOTOVOLTAICS



9

INTRODUCTION TO FLOATING PV

The expansion of photovoltaic energy often requires land to be occupied, creating competition with other activities such as agriculture and residency. To address this challenge, solar PV is increasingly expanding to alternative environments, such as water bodies. The electrical performance of photovoltaic systems on water can differ from that on land due to the distinct environmental conditions. This chapter investigates the impact of various factors on the performance of a floating photovoltaic (FPV) system situated on a French quarry lake. The study examines the effect of water level fluctuations, dust accumulation, and proximity to water on module temperature during the winter months. The analysis revealed that water fluctuations and dust accumulation had a minimal impact on the system's power output over the study period. Furthermore, contrary to some claims in the literature, no significant thermal benefits were observed from the system's proximity to water when comparing the module temperatures of the FPV system to those of a land-based system. These findings suggest that, for certain FPV systems, the aquatic environment may have little to no impact on power performance, which could facilitate their broader adoption.



Parts of this chapter have been published in the Trust-PV reports S. Tonnel, M. Ikhennicheu, R. Le Failler, A. Alcañiz and H. Ziar “Guidelines for design, procurement and O&M friendly concepts for floating PV” (2023) [343] and S. Tonnel, F. Huntingford, M. Ikhennicheu, R. Le Failler, F. Gorintin, A. Alcañiz and H. Ziar “Demonstration of solution and challenges on floating systems” (2024) [344].

9.1. WHY PLACE PV MODULES ON WATER?

The previous chapters have mostly focused on predicting the power of residential-scale PV systems. Placing PV systems on the ground is the most intuitive approach as the electricity will be generated close to where it is consumed. However, these solar installations occupy a large area creating competition for land between the photovoltaic market and other essential needs such as agriculture or accommodation [345].

In this scenario, PV technology may expand to a relatively new environment: water. The application on water bodies has enormous potential [346]. If only 10% of all land-based water reservoirs were covered by PV, they would generate electricity equivalent to almost 23% of all global electricity consumption from 2019 [347].

Beyond the strong reduction of land occupancy, systems located on the water have a generally lower operating temperature due to the increased wind speed and the sea acting as a heat sink [348]. Researchers from the Indian Institute of Technology reported that an FPV module was on average 6°C cooler than a land-based one during a reported measurement period of 17 months [349]. However, the cooling effect depends on the reference system and the climate, and the effect is relatively minor in mild temperate countries like the Netherlands [350].

Other benefits of FPV are the lack of obstacles and the low influence of the systems on the radiation balance [351, 352]. The latter refers to the modification of the albedo by land-based photovoltaic (LPV) systems, which influences the local temperature. The albedo is the part of solar radiation reflected from the ground and depends on the environment's reflectivity [261]. The water albedo of about 6% [353] is very similar to that of the PV modules,

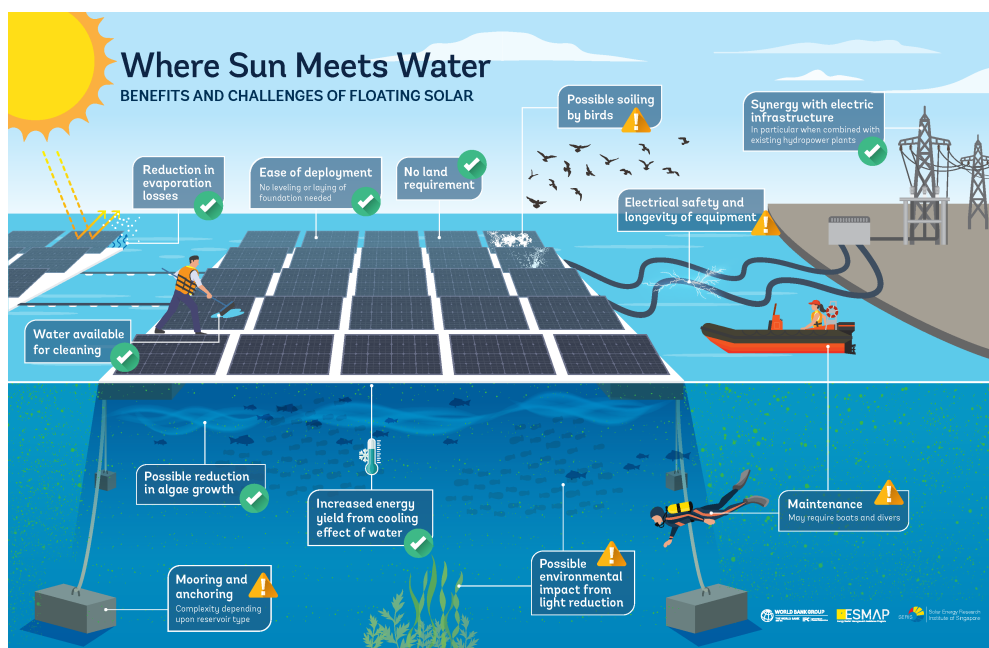


Figure 9.1: Benefits and challenges of floating PV. Figure from [347].

so a floating installation does not impact the albedo notably. However, this albedo value fluctuates depending on the weather conditions, the surrounding geometry, the position of the sun, the wavelength of light, wind speed, and turbidity [348, 354] increasing the modeling complexity.

Despite the benefits, several challenges restrain the growth of FPV. FPV systems are more difficult to maintain given their higher inaccessibility [351]. Although water is available for cleaning, it has to be actively employed to be effective. Moreover, the humid conditions that the modules and balance of system components need to work under increase the degradation rate of the system [349, 351]. Additionally, the need for floats, anchoring, and mooring increases the cost of inland FPV systems by around 4-8% compared to that of LPV ones [355]. Fluctuations of the water surface may also negatively impact the performance and stability of the systems [352].

Since this technology is still in its developmental stage, these challenges have yet to be thoroughly addressed. This chapter aims to enhance the existing literature by examining three different aspects of an inland floating PV system situated on a lake in France, as outlined in Section 9.2. The aspects explored are the impact of water level fluctuations on the performance of the FPV system (Section 9.3), dust accumulation on the system (Section 9.4), and module temperature, compared to a similar land-based PV system (Section 9.5). Section 9.6 wraps up with the main chapter's conclusions.

9.2. TEST SYSTEM DESCRIPTION

The test system examined in this chapter is an inland floating PV farm located on a quarry lake in Peyssies, in the South of France. This lake has a small fetch, which limits wave generation, and is primarily surrounded by forests, with no large cities or industrial complexes nearby.

The plant has an installed capacity of approximately 6 MW and is operated by Urbasolar. It consists of polycrystalline silicon half-cell modules mounted on floaters. Commissioned in June 2022, the facility was mechanically and electrically instrumented from early February 2024 to early April 2024, although this report focuses solely on the power production aspect. Supporting the instrumentation data of the system, water temperature is obtained via on-site measurements performed by CNRS for the Solake project [356].

9.3. WATER FLUCTUATIONS

In a floating PV system, waves can modify the inclination and orientation of the modules deviating them from the installation angles [357]. The magnitude of the oscillations depends not only on the height of the waves but also on the type of floating structure that the modules are mounted on. Considering the typically limited movement in water basins, the impact of these fluctuations is anticipated to be minimal, unlike what would be observed in an offshore system. However, since this effect depends on the specific water basin, it is worthwhile to examine the case under study.

To study the effect of water fluctuations on the performance of the floating PV system, three inclinometers were placed aside from three different PV modules along the same string. These sensors measure the pitch ψ (rotation along the x-axis) and the roll ϕ (rotation along the y-axis). The combination of the two determines the tilt θ and azimuth A_M , which are commonly used angles in the PV field. Equation 9.1 shows the relation between these

angles.

$$\tan \psi = \frac{\tan \theta}{\cos A_M} \quad \tan \phi = -\frac{\tan \theta}{\sin A_M} \quad (9.1)$$

The plots in [Figure 9.2a](#) show the time series of the roll and the pitch for the three inclinometers for the whole period of study. The variation of the angles is not large. The daily standard deviation of the angles is below 0.1° for all three sensors. The only visible variation in the trends occurs on March 25th when the pitch of the first inclinometer drops by around 0.5° caused by the cleaning of the sensors (more about it in [Section 9.4](#)). Another insight from this figure is the variance in installation angles among the three modules. This deviation most likely occurred during the PV farm deployment and can cause a loss in power if the modules are connected in series.

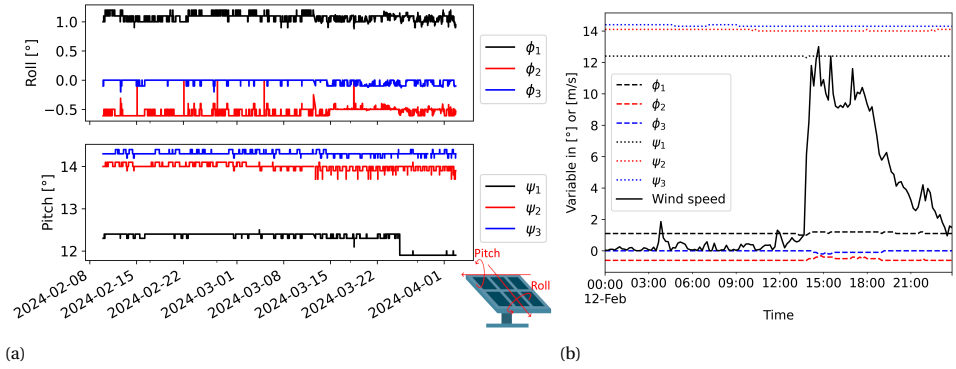


Figure 9.2: Roll ϕ and pitch ψ for the three inclinometers (a) throughout the whole measurement period and (b) for the windiest day together with the wind speed.

Even on the windiest day, the variation in angles is negligible, as seen in [Figure 9.2b](#). The deviation is maximum 2° for the pitch and 1.5° for the roll. Only the roll of the second sensor seems to be slightly affected, but that is unlikely to be translated into any significant power loss. Therefore, the conclusion is that the oscillations experienced by this FPV system do not affect the performance.

The reason for such low oscillations is probably twofold. First, water oscillations in lakes are generally created by either wind or an atmospheric pressure gradient [358]. That wind needs to be persistent to result in gravity waves. Since in the lake under study the wind is not persistent, most oscillations will be capillary waves. These are regular, small, uniformly developed waves that appear as a ruffling of the surface and create minor oscillations [358]. Second, the FPV system impacts these oscillations. On the one hand, the waves need to cause oscillations not only in the PV modules but also in the platform supporting them, thereby increasing the force required to shake the modules. On the other hand, the wind's interaction with the water surface is significantly reduced, as a large portion of the lake is covered by the modules.

9.4. DUST ACCUMULATION

After analyzing the water fluctuations, this section aims to evaluate the dust accumulation on the FPV modules and its impact on irradiance. Soiling is the second most significant factor affecting PV yield [359]. However, the impact of soiling depends strongly on the location of the PV system, being it more impactful in arid regions. In the case of a floating PV system in a Mediterranean climate, the suspended particle density will be low. However, in high humidity conditions, cementation of soiling particles is intensified [359].

The literature suggests the higher the dust accumulated, the lower the irradiance received by the module [360]. For instance, in Málaga, a Southern Spanish city, the mean of the daily loss of irradiance due to dust was 4.4% [361]. For these types of studies, the general approach is to compare the irradiance output of an uncleaned sensor to one that is regularly cleaned [359, 361, 362]. However, that was not possible for the plant under study, so the benchmark of a clean module was lacking. Two methodologies were considered to overcome this issue. During them, extra care is needed to ensure that the difference in irradiance is due to the accumulation of dust and not other phenomena since the impact of dust on irradiance is expected to be small.

The first alternative was based on the fact that dust tends to collect at the base of the PV installations [363]. Therefore, three reference cells were placed aside the same PV module at different heights: on the top, the middle, and the bottom of the module. There were however two problems with this methodology. The first one is that reference cells are individual units. More dust accumulates on the bottom of the module than on the top because dust glides down the module due to gravity. However, since the reference cells are separated units, the dust of the top cell cannot glide to the bottom one.

The second problem was deviations in the installation angles of the reference cells, which hindered their comparison. Although small and with no significant power loss, these deviations add noise to the dust accumulation study as seen in the previous section. An attempt was made to find the actual installation angles and overcome this issue but it was unsuccessful.

The second alternative to estimate the effect of dust accumulation consisted of using rain as a cleaning agent. Literature suggests that dust accumulates until it rains [363, 364]. Therefore, the irradiance received by the sensors after the rain would be higher than that before it. This approach entails its difficulties as the cleaning power of rain depends not only on its intensity but also on the accumulated dust [363]. Hence the first step consisted of obtaining dust and precipitation data. The dust deposited hourly on a horizontal surface at the location of study is obtained from MERRA2 [365]. Precipitation data is next gathered from [366]. Rainfall is used to categorize the data into dry periods (no rain in the last 24 hours), light rain periods (0 to 10 mm of rain in the last 24 hours), and moderate rain periods (10 to 25 mm of rain in the last 24 hours) following [360]. Rain higher than 25 mm in 24 hours has not been registered in the period under study. The literature rarely provides clear results on how much rain is needed to clean the sensors, hence both light and moderate periods are considered for comparison. With these data, one can estimate the dust accumulated over time since the last light or moderate rain period, as seen in Figure 9.3a. This is a simplification since rain is not 100% efficient in cleaning the modules [363] and the dust accumulated on a tilted surface is not the same as on a horizontal one [360]. However, it can provide a fair estimate.

Once having estimated the dust accumulated on the sensors, one can compare the ir-

radiance on days with different accumulated dust. Special care needs to be taken for this comparison since the irradiance depends on the amount of clouds, the hour of the day, and the day of the year. The reduction of irradiance by dust also depends on the irradiance intensity and the angle of incidence [361]. To overcome this, only clear-sky days are considered. The time dependency of the measured irradiance can be removed by dividing it by the clear-sky irradiance. Overall, the scatter plot in Figure 9.3b shows the ratio of incident G_{POA} to clear-sky G_{CS} irradiances as a function of the dust accumulated considering the two different rain periods for the seven clear-sky days of the measurement period.

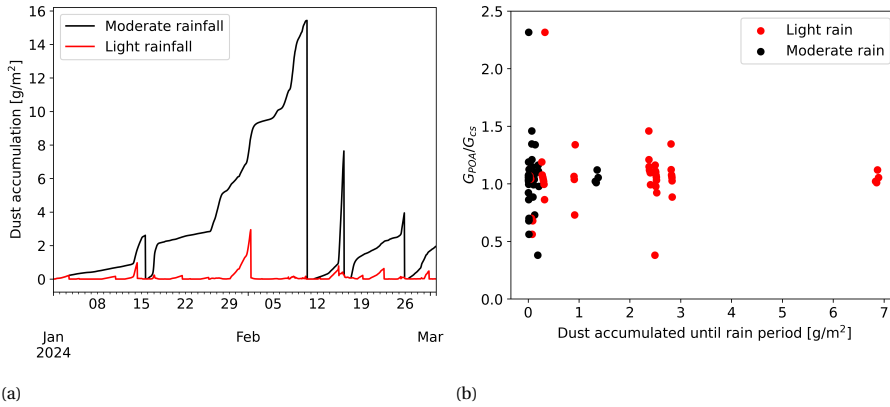


Figure 9.3: (a) Accumulated dust on a horizontal surface in the PV farm until light or moderate rain cleans it completely. (b) Ratio of incident irradiance to clear-sky irradiance as a function of dust accumulated until light or moderate rain period.

The expected trend of Figure 9.3b should be the higher the dust, the lower the G_{POA} to G_{CS} ratio. However, that is not visible. All values are within the same range independent of the amount of dust. It is not possible either with this figure to assess whether moderate rain cleans more than light. The problem with these inconclusive results may be that the data is not enough or clear-sky irradiance is not a good control variable. Alternative irradiance sources were considered but they all entailed issues. The irradiance from a nearby pyranometer would also be subjected to dust. The error of irradiance obtained from satellite images would add noise to the estimation. The physical distance to the controlled pyranometer with periodic cleaning would also add noise.

This study was supported by a technical visit to the PV farm in which the three irradiance sensors were cleaned. Figure 9.4a shows the sensors before being cleaned, while Figure 9.4b shows them right after. No significant dust accumulation is observed in Figure 9.4a, which is in line with the obtained results. However, higher dirtiness is visible at the bottom of the PV module and of each reference cell. This shows the cementing phenomena discussed in [363].

The clean sensors can be now a benchmark for the dust accumulation study. Therefore, to assess whether the dust visible in the photographs affects the incident irradiance, the irradiance incident on the clean sensor will be compared to that of an uncleaned one located beside a nearby PV module. For distinction purposes, the uncleaned sensor is labeled sensor B while the clean one is labeled sensor A. We will analyze the ratio of irradiance incident on sensor A G_A to that on sensor B G_B during the week preceding and the week following

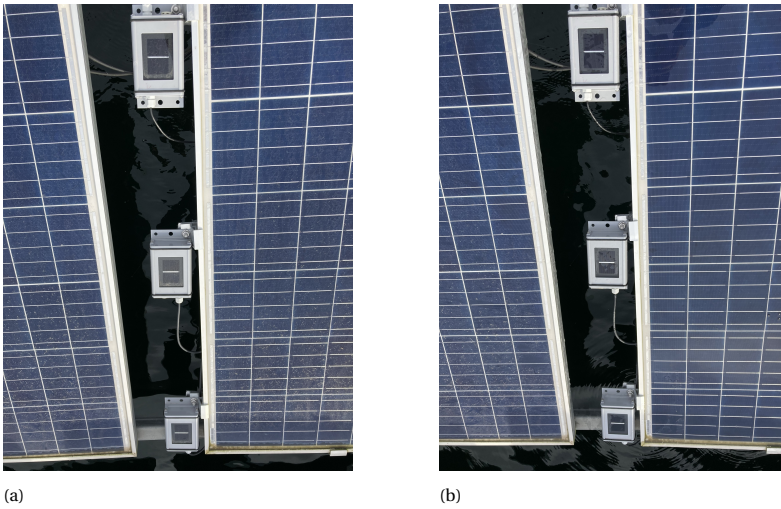


Figure 9.4: Reference cells before (a) and after (b) being cleaned by the plant operator.

the cleaning event. Figure 9.5 plots the results.

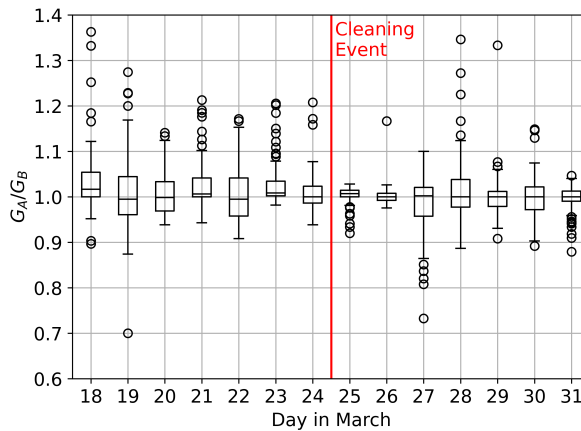


Figure 9.5: Irradiance difference between the middle and bottom sensors as a function of the irradiance incident on the middle sensor for the days before and after the sensors were cleaned.

The irradiance incident on sensor A is expected to be higher than that on sensor B after the cleaning event. Therefore, the ratio of irradiances should be higher after the cleaning event than before it. However, the daily boxplots do not show this trend. The median irradiance ratio is constant throughout the days. This indicates that dust for this plant has no apparent effect in the duration of observation, which is in line with the inconclusive results obtained previously.

9.5. MODULE TEMPERATURE

The operating temperature of FPV systems is one of the aspects more studied by researchers due to their potential benefits. Higher wind speeds and the effect of water as a heat sink in the summer months reduce the module temperature of floating systems. In mild temperate climates like the Netherlands, the irradiance-weighted temperature difference of floating PV is around 3.2°C [350]. However, this value depends on the climate, and the same researchers reported that in Singapore this difference is generally 14.5°C lower than similar modules mounted on rooftops. This can lead to an increase in the energy yield of 11% [367] or 20% [368] for inland FPV systems, and of 18% for offshore ones [357] compared to LPV systems. On the same line, Liu et al. reported that FPV systems in Singapore achieved a 5 to 10% higher PR (Performance Ratio) in 2018 than rooftop ones [348].

However, achieving the previously claimed increases in electricity production has not been consistently observed. Oliveira-Pinto and Stokkermans highlighted that the actual increase in production falls within the range of 0.31% to 2.59% [369]. The authors attribute this discrepancy to the lack of a comprehensive simulation tool, particularly in terms of addressing thermal modeling, which makes it challenging to estimate the exact electricity output accurately.

Thus, while the cooling effect of water on solar cells in FPV systems does enhance performance, the precise extent of the increase in electricity generation remains a topic of ongoing research and debate, with various factors such as location, environmental conditions, and modeling tools influencing the observed outcomes [357, 368, 369]. This section intends to contribute to this discussion by comparing the module temperature of the FPV system under study with that of a nearby LPV system. However, before performing that comparison, the data is analyzed in search for trends.

The initial analysis examines the correlation between module temperature and relevant weather variables. As opposed to what is generally expected [261], the wind speed showed a low yet positive correlation with module temperature (0.02 during daylight and 0.36 at night). The wind speed was very low in the studied period, which could explain its limited effect. Additionally, when an influence was observed during the night, it was not of cooling the modules but of warming them, as the air temperature was generally higher than the module one.

Water temperature also showed a low correlation with module temperature (0.23 during the day and 0.56 at night). Since the PV modules were mounted at an angle on top of the floaters instead of in direct contact with the fluid, the cooling effect of water was limited. Therefore, although the effect of water temperature was probably positive, it was also highly limited. When compared with the incident irradiance, the module temperature increases at almost the same rate, but then it decreases more slowly. This results from the thermal inertia of the modules. This inertia is also observed when a cloud causes a sudden drop in irradiance which is smoothed out in the module temperature.

Several models have been proposed to estimate the module temperature of FPV systems. The review by Micheli provides a good overview of them [370]. In the following, the accuracy of some of these temperature models for LPV systems will be computed for the case at hand. The considered models are: Fluid-dynamics (FD) [236], Duffie-Beckman (DB) [371], Faiman (Fai) [235], Tina et al.-derived Faiman for FPV (Fai-Tin) [372], Kjeldstad et al.-derived Faiman for FPV (Fai-Kje) [373], Liu et al.-derived Faiman for FPV (Fai-Liu) [348], Dörenkämper et al.-derived Faiman for FPV (Fai-Dor) [350], Sandia (San) [374], Charles

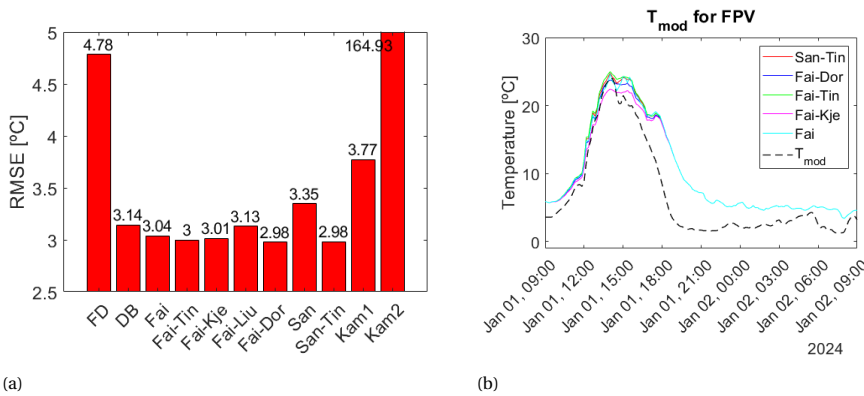


Figure 9.6: (a) Root Mean Squared Error (RMSE) of all the tested temperature models. (b) Performance comparison of the models with the lowest RMSE.

Lawrence Kamuyu et al. first empirical equation (Kam1) [375], and Charles Lawrence Kamuyu et al. second empirical equation (Kam2) [375]. Figure 9.6a shows the Root Mean Squared Error (RMSE) of each temperature model for two months. The performance of the five models with the lowest error is shown in Figure 9.6b.

Most models perform equally except for the two models proposed by Charles Lawrence Kamuyu et al., the fluid-dynamics model, and Sandia's model. The models proposed by Charles Lawrence Kamuyu et al. were empirically found to match the module temperature of a Korean system. Such tailored models are usually hard to generalize. The difference between the first and the second proposed model is the inclusion of water temperature in the second one. As the authors experienced, including that variable negatively impacted the temperature estimation. In our case, it resulted in a significant RMSE increase from 3.8 to 165.0°C. The fluid-dynamics model can be highly accurate, however it is also complex, which complicates its usage. Its low accuracy (RMSE of 4.8°C) most likely stems from errors in the estimation of its parameters. Finally, the comparatively high RMSE of 3.4°C in the Sandia model could be because that model is derived for LPV systems. However, the Duffie-Beckman and Faiman models are also intended for LPV systems but they achieve a similar error compared to the adapted models. In fact, the Liu et al.-derived Faiman model for FPV systems performs around 0.1°C worse than the original model aimed at LPV systems.

The reason why some of the adapted models perform better than others could lie in the similarity between the FPV system under study and the system for which the model was adapted. However, that does not seem to be always the case. For instance, one would expect that the Kjeldstad et al.-derived Faiman model would perform badly since its parameters are obtained for a system on a floating membrane. However, it performs very similarly to the best model. On the other hand, the FPV system studied by Liu et al. matches more closely the case at hand, but its performance is worse. It is also interesting how all the adapted Faiman models perform similarly (in the range of 2.98 to 3.13°C), while the difference in error between the original and adapted Sandia models is 0.37°C.

In Figure 9.6b, one can observe that all the models perform very similarly hence they are not able to grab the thermal dynamics of the system. Due to the thermal inertia observed in this system, none of the literature models are satisfactory when predicting the module

temperature. They do not properly represent the dynamics of the problem at hand, which is something already defended in the literature, still under discussion, and with many researchers devoting time to the topic.

Finally, the temperature of the FPV system was compared to that of an LPV system. Firstly, the mean and standard deviation were explored. Looking at all the values, the average temperature of the FPV system is 8.4°C, while that of the LPV one is slightly higher, 8.5°C. The standard deviation is also higher for the LPV system (7.7 and 7.0°C). When looking separately at day and night, the difference in average temperatures is higher. While the FPV system is at 13.1°C, the LPV one is 0.3°C higher. However, at night, the FPV system is slightly warmer (4.8°C) than the LPV one (4.7°C). Water is likely acting as a heat sink, which cools the modules during the day and warms them at night. The lower standard deviation can also be explained by this phenomenon. However, that effect is on average highly limited and not likely to have a significant impact on the performance.

Moreover, when comparing the module temperatures over time, the FPV system is not consistently cooler than the LPV one. A random day is presented in Figure 9.7 to show this trend, but it was present in more days of the dataset. This is in line with previous findings [351].

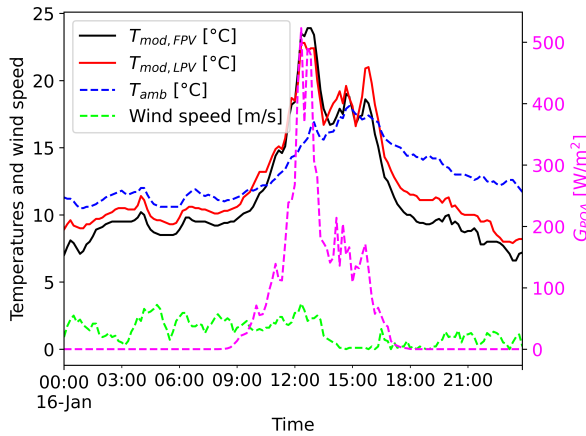


Figure 9.7: Module temperature of the FPV and LPV systems for one random day with relevant weather parameters: ambient temperature, wind speed, and incident irradiance.

The main reason for the similar temperature in both systems stems from their comparable heat transfer balance with the environment. When modules lie horizontally on a floater, which is in thermal contact with water, a conductive heat transfer mechanism occurs between the water and the back of the module. When the modules are at an angle on top of a floater, the only thermal contact with water occurs through the contact points of the mounting structure. That area is relatively small and its contribution is neglected even in detailed thermal models [261]. Therefore, the heat transfer mechanisms in the tilted FPV system under study do not differ considerably from those of a tilted LPV system as long as they experience a similar wind profile that drives the convection cooling.

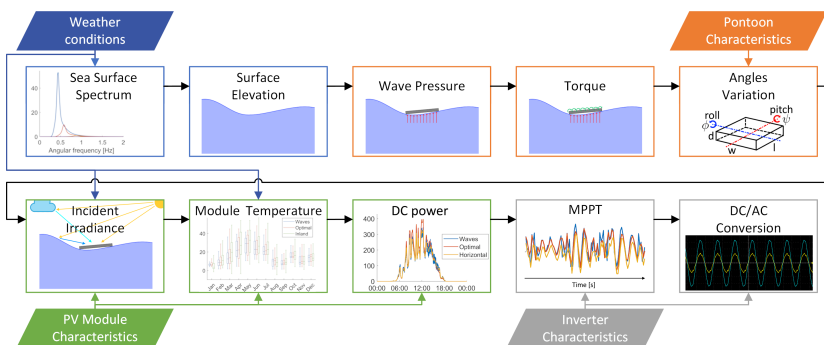
9.6. CONCLUSIONS

This chapter introduced the topic of floating photovoltaics by examining three key factors influencing the performance of a floating PV system in a quarry lake. Our analysis revealed a limited impact of water level fluctuations and dust accumulation on power production for the system during the study period. Moreover, the thermal benefits often attributed to floating PV systems due to their proximity to water were not observed when comparing the module temperatures of the floating system with those of an inland system. While more definitive results could likely be achieved with a longer study period during warmer months, this was beyond the scope of our current research. In the following chapters, the focus shifts to offshore floating systems, where we will explore similar phenomena in a more challenging maritime environment.

10

EFFECT OF SEA WAVES ON THE YIELD OF OFPV SYSTEMS

The previous chapter examined the challenges of placing PV modules on water and analyzed some of them for a floating system in a quarry lake. This chapter shifts to a harsher environment by evaluating the power output of an offshore FPV system in the North Sea. A validated 3D mechanical movement model is developed to analyze the impact of waves of performance and how floater dimensions affect fluctuation size. The study finds that a heavy, wide floater aligned with prevailing winds reduces angle variations. DC power simulations show that sea fluctuations slightly reduce PV power production by 0.1% annually. However, an optimally-tilted water-based system could still generate 14.6% more DC power than the floating one. On the AC side, laboratory tests reveal that rough sea conditions decrease inverter efficiency by over 2 percentage points compared to a still system.



This chapter has been adapted from the publication A. Alcañiz, N. Monaco, O. Isabella, H. Ziar, “Offshore floating PV–DC and AC yield analysis considering wave effects” *Energy Conversion and Management*, **300** 117897 (2024) [376].

10.1. INTRODUCTION

The previous chapter examined various factors influencing the performance of an FPV system on a quarry lake. However, the sea offers a significantly larger area for renewable energy installations. Offshore floating PV (OFPV) has an immense potential, with the capacity to cover the energy needs of over 11 billion people if only the most favorable locations are employed [346]. In contrast to the well-established offshore wind technology [377], OFPV remains an emerging technology, with the first pilot OFPV farm deployed in 2019 [378]. This slow development is because freshwater installations cannot simply be transferred to marine environments [379].

Offshore FPV (OFPV) has additional drawbacks compared to FPV, such as the effect of sea salt on the corrosion rate of the elements in the PV system, the difficulty of maintenance, the limited mechanical stability, and the fluctuating PV power [362, 380–383].

An accurate yield prediction of OFPV systems is needed to assess the economical feasibility of the floating projects and investigate the technical aspects of these systems. One can find in the literature several works focusing on thermal predictions [370, 373, 384, 385], but the literature on yield modeling does not seem that extensive. Table 10.1 shows the most recent studies in which the yield performance of a floating system has been modeled. The approach developed in this work has also been included for comparison purposes.

Table 10.1: Previous works that estimate the yield produced by floating PV systems.

Study	Water body	Waves model	Optoelectrical model
Goswami, 2019 [386]	Dam	None	PVSyst [257]
Suh, 2019 [387]	Dam	None	SAM [250]
Golroodbari, 2020 [357]	Sea	2D linear wave theory	Physics-based equations
Sukarso, 2020 [388]	Lake	None	SAM
Tina, 2021 [372]	Artificial basin	None	PVSyst, SAM
Ziar, 2021 [238]	Pond	None	In-house software [232]
Ghigo, 2022 [389]	Sea	WEC-Sim [390]	Fixed efficiency
Kumar, 2022 [391]	Reservoir	None	PVSyst, SAM, HelioScope [256]
Ravichandran, 2022 [392]	Sea and lake	None	HelioScope
Rahaman, 2023 [384]	Lake	None	Physics-based equations
This work	Sea	3D linear wave theory	In-house software [232]

The studies in the literature that compared the performance of land-based and FPV systems report a higher energy yield of FPV systems. Choi compared the normalized performance of a 100 kW FPV system with a nearby 1 MW land-based PV system in South Korea [393]. Assuming that the irradiation and ambient temperature were the same in both locations, their results showed an 11% higher average efficiency of the FPV system. Similarly, the study performed in Singapore reported that the FPV system had a performance ratio of FPV around 10% above that of the rooftop system [348].

Most literature focuses on land-based water reservoirs in which the effect of waves is

negligible due to their low height. Other researchers have explored the dynamic response and robustness of OFPV platforms due to wind and waves but have not studied how that movement affects the energy yield of the systems [394–397]. The only exception is the work performed by Golroodbari and Sark [357]. They modeled the performance of a PV system on land and at sea considering the effect of sea waves, wind speed, and relative humidity. They reported an almost 13% increase in the relative annual average output DC yield of offshore systems compared to land-based ones. Although a good starting point, several aspects can be put forward in their modeling approach and have been applied in this work. For instance, Golroodbari and Sark considered only one axis of rotation for the offshore floating PV, which may underestimate the effect of wave movement. In this work, the rotation of the floating structure is modeled along two axes. Moreover, the developed mechanical model has been validated with data from a real OFPV system. These two aspects allowed the creation of guidelines for optimal floater dimensions that minimize angle fluctuations.

In Table 10.1, one can also observe that most of the electrical modeling has been conducted using commercial software tools. In this work, we employ a highly accurate in-house developed toolbox presented in Chapter 7 that has already been applied to floating systems [238]. This does not allow, as often done in the literature, to easily adapt the simple thermal models for the floating case [370]. However, the fluid dynamics-based thermal model employed in this work tends to outperform simpler thermal models [232].

Finally, it was not possible to find in the literature any study regarding the losses in DC/AC conversion due to the fluctuating power produced by PV modules.

Overall, this work pursues the quantification of the power output of an offshore floating PV system considering the effect of waves on the tilt and azimuth of the modules. The wave movement will be modeled to estimate the interaction between water and the system. A mechanical model is developed and validated with real data to obtain the inclination of the system along the two axes. These inclinations will be input into a simulation framework to estimate the DC yield, comparing several cases. Moreover, this work aims to experimentally investigate the conversion losses due to this fluctuating DC output. Therefore, laboratory experiments will be performed to study the effect of waves on inverter efficiency. Overall the contributions of this work are as follows:

- 3D mechanical model of sea waves which results in two axes of rotation for the floating body
- Validation of the developed mechanical model using data from an OFPV system
- Development of guidelines for the optimal dimensions of a floater that decrease the angle fluctuations
- Accurate optoelectrical model using an in-house developed toolbox
- Estimation of inverter conversion losses via hardware-in-the-loop laboratory experiments

This work is organized as follows. Section 10.2 explains the methodology followed through the document. Section 10.3 determines the inputs chosen in this study. The results will be presented and discussed in Section 10.4, before concluding in Section 10.5.

10.2. METHODOLOGY

Figure 10.1 provides an overview of the methodology followed in this work. The blocks have been colored according to the subsection to which they belong. Therefore, the first three blocks correspond to the waves model (Subsection 10.2.1), orange blocks to the mechanical one (Subsection 10.2.2), the green blocks deal with the DC yield estimation (Subsection 10.2.3), and the last two blocks deal with the AC conversion (Subsection 10.2.4).

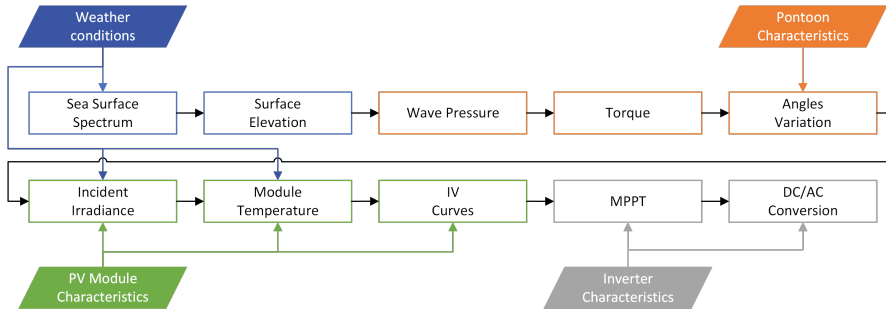


Figure 10.1: Flowchart of the simulation process conducted in this work. Each color represents a different subsection.

10.2.1. WAVES MODEL

The first step of this model consisted of deciding the type of waves to simulate. Depending on the originating mechanisms and characteristics, one can identify several wave types [398]¹. The criterion used for selection was the impact of fluctuations on the floating body. The oscillation level of a floating body is linked to the wavelength of the oscillations. Therefore, only waves with wavelengths of the same order of magnitude as the dimension of the floating body will cause a significant disturbance [399]. Amongst all options, gravity waves were selected for the modeling. Gravity waves are generated by winds that have been consistently blowing over a substantial distance [398]. They have high specific energy content [400], which facilitates wave energy modeling. Moreover, they have wavelengths between 1.5 and 900 meters, which are of the same order of magnitude as the floaters in OFPV systems.

Once the type of wave is selected, its elevation and energy content need to be computed to model the interaction with the OFPV system. The interaction of sea undulations and floating bodies is highly challenging and non-linear. However, this problem can be simplified through the linear water wave theory, whose assumptions are as follows [399]:

- The floating body is smaller than the wavelength. Large floating structures reflect and scatter waves. Nonetheless, if the wavelength is longer than the object dimensions, the wavefield is only slightly modified and wave diffraction is negligible. The object is then passively driven by the oscillations, and its presence does not notably affect the waves.
- Water is incompressible.

¹ Capillary waves, infragravity waves, gravity waves, long-period waves, and tidal waves.

- Viscosity is negligible, so there is no energy loss due to water's resistance to movement.
- Coriolis force is ignored.
- The floater is at a location where deep water conditions apply.

Using the linear water wave theory, the ocean surface can be relatively simply described, while considering that the sea surface is composed of random waves of various lengths and periods. To factor in this randomness, a sea spectrum is employed to describe waves with different heights, durations, and shapes with limited predictability [401].

The idea of a spectrum is based on Fourier's theory that any function can be represented as the sum of an infinite series of sine and cosine functions with harmonic wave frequencies [402]. Several spectra have been suggested to represent the ocean surface, which can depend on the location. Therefore, the location of the OFPV system needs to be set. No recommended practice was found in the literature on the suitable location for an offshore floating farm. Hence, the first offshore pilot PV farm in the world was taken as a reference, which is placed on the North Sea [378]. Considering this location, the Joint North Sea Wave Observation Project (JONSWAP) spectrum was employed to model the ocean surface [403].

The system of equations explained in [404] was implemented to obtain the spectra as a function of wind speed. Figure 10.2 shows the curves obtained at this step as a function of wind speed with a unique direction. The intensity of the peaks decreases with the wind speed, following a hyperbolic behaviour, typical of these spectra.

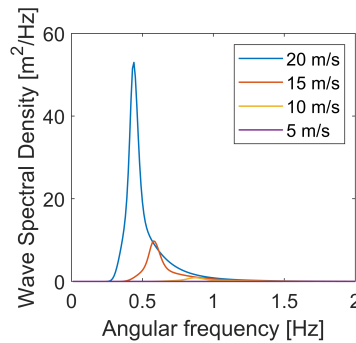


Figure 10.2: JONSWAP spectra for different wind speeds.

From the spectra, one can obtain the surface elevation. Recalling Fourier, the sea surface is the superposition of waves of different wavelengths and amplitudes at certain times. However, this model is developed for infinite values, which need to be discretized in order to implement it numerically. Therefore, the approach by Tucker et al. is employed, in which the amplitude components are random variables [405], as seen in Equation 10.1 [406].

$$\eta(x, t) = \sum_{n=1}^{N/2} (a_n \cdot \cos(k_n \cdot x - \omega_n \cdot t) + b_n \cdot \sin(k_n \cdot x - \omega_n \cdot t)) \quad (10.1)$$

$$a_n = r n_{a_n} \cdot \sigma_S$$

$$b_n = r n_{b_n} \cdot \sigma_S$$

Here, $\eta(x, t)$ is the surface elevation [m] at point x and time t , ω_n is the angular frequency [Hz] of wave n , related to the wave number k_n [m^{-1}] through the dispersion relation for deep water conditions, Equation 10.2.

$$\omega_n^2 = g \cdot k_n \quad (10.2)$$

Where g is the acceleration of gravity equal to 9.81 m/s^2 . a_n and b_n represent the Fourier amplitudes [m]. They are equal to the product between normally distributed random variables with zero mean rn_{a_n} and rn_{b_n} , and the standard deviation of the spectrum σ_S . N represents the discretization of the integral and is the number of angular frequencies considered, 800 in this work, from 0.01 to 8 Hz with a 0.01 Hz step.

For a more structured analysis, a sea state classification based on the surface elevation is employed. This allows to explore the effect that different sea states have on the system yield and to quantify this influence over the year. The sea will be classified based on the Douglas scale for the sea state [407]. The original scale employs 10 categories. However, this classification was too itemized for this research, so only three categories are included: calm, moderate, and rough. Ranges are displayed in Table 10.2. The considered height is the significant wave height, an often used statistical wave measure which is the average of the highest one-third of all the waves present in an area of the sea surface [408].

Table 10.2: Sea state scale adopted in this study.

Sea State Scale	Height [m]	Description
1	0 - 1.25	Calm
2	1.25 - 2.5	Moderate
3	> 2.5	Rough

10.2.2. MECHANICAL MODEL

This subsection aims to describe the model that explains the interaction between the floating solar plant and the sea waves. For this purpose, the simplified Froude-Krylov theory will be applied. The theory estimates the force that waves exert on a structure by determining pressure variations around it [409]. This approach was originally developed for submerged cylinders but can be adapted for the case of a floating cuboid.

The application of the simplified form of this theory requires a few assumptions:

- The wave field is not affected by the floating structure.
- The floating body is considered rigid. This allows the implementation of an analytical approach, without the use of specific software packages.
- Movements on the plane parallel to the sea surface are neglected. Generally, water only moves vertically and very slightly in the direction of the wave motion [410]. It is assumed that the anchoring system reduces completely motions on the plane parallel to the sea surface as well as any rotations around the z -axis.
- The floating body is located at a height $z = 0$ relative to the seawater level.

As previously mentioned, the Froude-Krylov approach employs the wave pressure. The pressure p [Pa] under a progressive wave can be calculated as shown in Equation 10.3. In

this equation, the first term refers to the dynamic pressure while the second one represents the static one.

$$p(x, z, t) = -\rho \frac{\partial \phi}{\partial t} - \rho \cdot g \cdot z \quad (10.3)$$

Where ρ is the seawater density of 1029 kg/m³ and ϕ is the velocity potential function [m²/s], which is related to the surface elevation via [Equation 10.4](#).

$$\eta(x, t) = -\frac{1}{g} \frac{\partial \phi}{\partial t} \Big|_{z=0} \quad (10.4)$$

Knowing the surface elevation and making use of the assumption that the floating body is located at a height $z=0$ m, the pressure acting on a floating body can be simplified as in [Equation 10.5](#).

$$p(x, y, z=0, t) = \rho \cdot g \cdot \eta \quad (10.5)$$

The force F [N] acting on the floater can be obtained by integrating this dynamic pressure over the area that this pressure acts on, the cuboid area ([Equation 10.6](#)). The integration range of the x -axis in the equation determines the location of the angle of rotation of the floater.

$$F = \int p dA = \int_{-w/2}^{w/2} \int_0^l p dy dx \quad (10.6)$$

Where w and l are the width of the cuboid in the x - and y -axis, respectively (see [Figure 10.3](#)). The upward force will be different at each location because of its dependency on the sea surface elevation. Therefore, when calculating the torque τ of this force [N·m], this has to be considered as in [Equation 10.7](#).

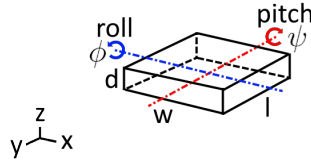


Figure 10.3: Schematic of the floating body indicating the dimensions and the axes of rotation. Figure inspired by [357].

$$\tau = F \cdot x = \int_{-w/2}^{w/2} \int_0^l p \cdot x dy dx \quad (10.7)$$

Lastly, the rotational angle of the floater θ [rad] can be obtained by using the relation between the torque, the moment of inertia [kg·m²], and the angular acceleration α [rad/s²], written in [Equation 10.8](#).

$$\tau = I \cdot \alpha = I \cdot \frac{\partial \omega_\theta}{\partial t} = I \cdot \frac{\partial^2 \theta}{\partial t^2} \quad (10.8)$$

When solving [Equation 10.8](#), a constant term would appear because of the indefinite integral. This term represents the value of the inclination angle at the time of the simulation.

This initial inclination angle has been assumed to be 0° in the simulations since there is no information regarding its initial value. Although this can result in a slight deviation in the inclination angle, it has no statistical effect and therefore, it does not hinder the analysis.

This methodology has been developed for a single axis. However, since gravity waves follow the wind direction, which changes over time, waves hit the floater from all directions. To consider this fact, the wind is decomposed according to the North-South (NS) and East-West (EW) directions, so they are aligned with the axes of the floating body. The operations explained in the previous subsection as well as in this one are repeated for the two components so that two independent spectra are generated, from which two independent surface elevations, η_{NS} and η_{EW} , are derived and eventually two rotational angles are obtained, the roll ϕ whose rotation axis is the x-one, and the pitch ψ rotating around the y-axis. These follow common naming conventions from the literature on floating structures.

The moments of inertia around both axes are to be calculated. The moments of inertia around the x- and y-axes passing through the center of the cuboid shown in [Figure 10.3](#) are written in [Equation 10.9](#) [411].

$$\begin{aligned} I_x &= \frac{1}{12} m(l^2 + d^2) \\ I_y &= \frac{1}{12} m(w^2 + d^2) \end{aligned} \quad (10.9)$$

Where m is the total mass [kg] and w , l , and d are the floating body dimensions [m] as illustrated in [Figure 10.3](#). In this schematic, the names of the rotational angles along each axis are also included.

FAILED MODELS

The mechanical model just explained reached its final form after several discussions. This section intends to explain the methods that were considered and even implemented but were in the end rejected since they were not suited approximations.

Instead of using the pressure variations, the initial methodology relied on the wave energy to obtain the inertia and angle of rotation of the system. First, the energy of the wave was computed at every instant of time [412]. Then it was assumed that this energy was transferred to the floating object through the moment of inertia. In this transfer, the floater would undergo a rotation from which the angle of rotation of the floater θ could have been obtained. The main problem with this approach was the assumption that all the energy was absorbed by the floater. That would be equivalent to having a vertical floater not allowing waves behind the floater, similar to a cliff whose height would be equal to the sea water level. This resulted in strong overestimations as the model did not properly represent a floating element. The inclusion of an attenuation factor to decrease the overestimation proved to be unsuccessful.

Another tested solution to decrease the overestimation consisted of considering the differential of energy across the floater. The intention was to evaluate the energy that the floater opposes due to the movement of the waves. However, the results were still unsuccessful. Including the “added mass effect”, a method known to increase the stability of fluid-structure interaction numerical simulations [413], was also considered but never implemented.

An alternative to using the wave energy was to solve the geometrical problem. The floater would be modeled as a rectangle located on top of the wave crest. By looking at

the difference in elevation between the extremes of the floater, the rotation angle was obtained. The main issue with this model is the lack of interaction between the floater and the water. In fact, no floater characteristics were considered in this approach hence it was discarded.

10.2.3. ELECTRICAL MODEL

This section models the waves' impact on the DC yield produced by an OFPV system. For this purpose, several scenarios are simulated. Apart from the OFPV oscillating scenario, two offshore systems are simulated with a fixed tilt at 0° and 34°. The former corresponds to the installation tilt of most offshore solar farms while the latter is the optimal tilt angle that maximizes the energy incident over a year at the North Sea [414]. A fourth scenario is considered inland instead of offshore with an optimal tilt as well.

The simulation framework used to reproduce the different scenarios is the PVMD Toolbox, explained in Chapter 7. Given its high accuracy, the simulation framework is computationally intensive. The computational time needed to replicate the DC output of the modules with a 1-second resolution is 4-12 hours/simulated day, depending on the number of sunshine hours. One day per sea state per month is considered to reduce computational time. The performance for the overall month is obtained by multiplying the yield of each simulated day by the number of days with the same sea agitation state in that specific month.

Until now, the implemented methodology provides the roll and pitch inclination angles and assumes a constant azimuth towards the South. However, the simulation framework requires as input the tilt and azimuth at every time step. Therefore, a change of coordinates is needed to go from roll ϕ and pitch ψ to tilt α and azimuth A_M . This was done based on the generic case of panels mounted on an inclined roof [261]. The rotation that needs to be performed is expressed in Equation 10.10.

$$\begin{aligned}\tan A_M &= \frac{\sin \psi}{\tan \phi} \\ \cos \alpha &= \cos \psi \cdot \cos \phi\end{aligned}\tag{10.10}$$

Finally, the angles were discretized to optimize the computational time by creating bins of 2° for the tilt and 5° for the azimuth.

10.2.4. AC CONVERSION

This subsection explains the tests performed on a commercial string inverter connected to an in-lab-emulated OFPV system. The objective is to explore wave movement's effect on the inverter efficiency.

The input for these experiments is the current-voltage (IV) curves generated by the PVMD toolbox with a 1-second resolution, represented by the Maximum Power Point (MPP) coordinates P_{DC} , V_{DC} and I_{DC} . Voltage drop by cables' losses is neglected.

These curves are input to a Chroma 62100H solar array IV simulator. The device can monitor output parameters ($P_{operating}$, $V_{operating}$, and $I_{operating}$), thus allowing a dynamic study of the Maximum Power Point Tracking (MPPT) algorithm. The emulator is capable of loading up to 100 different I-V curves. Therefore, at each round of the experiment, a maximum of 100 consecutive I-V curves were fed to an inverter.

The inverter considered in this work is an SMA Inverter Sunny Boy 2.5 [415], a single-phase string inverter. Output rated power is 2.5 kVA.

A Tektronix TBS2000B Digital Oscilloscope was operated to observe and study the output signals from the inverter. Figure 10.4 shows a stamp of this oscilloscope during operation. A differential and a current probe were used to monitor voltage and current, respectively. A Digilent Analog Discovery 2 data logger was used for the data extraction of P_{AC} , V_{AC} and I_{AC} . Figure 10.5 shows a picture of the setup during operation.

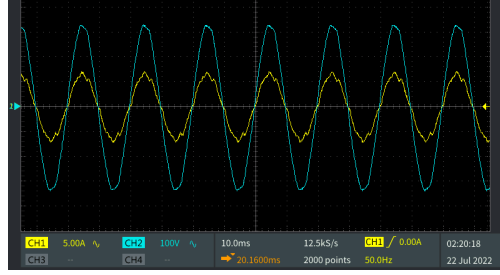


Figure 10.4: Stamp of the oscilloscope during operation.



Figure 10.5: Photo of the setup in operation in the laboratory.

With this setup, the inverter efficiencies can be explored. The MPPT efficiency η_{MPPT} is the ratio of the energy after the MPP tracker to the energy generated by the PV modules. The conversion efficiency η_{conv} is the ratio of the AC energy output to the DC energy input. The product of these two numbers is the inverter's total efficiency η_{tot} . Equation 10.11 displays the expression for these efficiencies.

$$\begin{aligned}
 \eta_{mppt} &= \frac{\int_0^T P_{operating}(t) dt}{\int_0^T P_{DC}(t) dt} \\
 \eta_{conv} &= \frac{\int_0^T P_{AC}(t) dt}{\int_0^T P_{operating}(t) dt} \\
 \eta_{tot} &= \eta_{mppt} \cdot \eta_{conv} = \frac{\int_0^T P_{AC}(t) dt}{\int_0^T P_{DC}(t) dt}
 \end{aligned} \tag{10.11}$$

10.3. MODEL INPUTS

Once the methodology is explained, this section details the inputs to the model. The first input is the location of the study. Within the North Sea, the first assumption in the waves model needs to be fulfilled: the location is at deep water conditions and subjected to high winds so that wavelengths are larger than the floating body. Moreover, the position for the development of the model is affected by the availability of data. With all these considerations, the study is located at a latitude of 53.0085° and longitude of 3.8498° , approximately 20 km from the coast (deep water condition), as shown in [Figure 10.6](#).



Figure 10.6: Selected location for the OFPV system in the North Sea, and for the inland meteorological station. Image from [\[340\]](#).

This research uses meteorological data from the Royal Netherlands Meteorological Institute, which is the Dutch national weather centre [196]. It provides hourly undisturbed wind and temperature data for the offshore location based on more than 40 years of collection. This research selected 2017 as a representative year after a brief exploration of the wind profiles in the latest years. The data were interpolated using the piecewise cubic method [416] to get the desired second resolution.

Figure 10.7 provides the wind rose of the studied year at the target offshore location. The percentages indicate how frequently that type of wind is repeated during the year, while the colours suggest wind speed levels. The wind blows mostly from the southwest, with the most frequent direction being 236°. Most of the year, the wind speed is between 5 and 10 m/s, reaching a maximum value of 23 m/s.

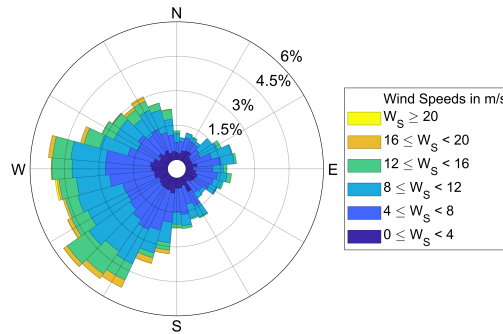


Figure 10.7: Wind rose of 2017 for the location of this research, with hourly resolution.

Unfortunately, no irradiance data were measured in the offshore meteorological station, so global horizontal irradiance was taken from an inland meteorological station located at 53.4117° 6.1992°, around 160 km from the offshore location (see Figure 10.6). This data was available with a 10-minute resolution, which was interpolated to get per-second data. Global irradiance was decomposed into direct and diffuse components using the BRL model [417]. These components are needed as inputs for the simulation framework. The wind speed and ambient temperature values for the inland scenario were also taken from the inland meteorological station.

An additional input to the simulation framework is the albedo. Although the state of agitation influences the sea reflection [418], the value of ocean albedo was assumed to be a constant of 0.06 [419]. In contrast, in the inland scenario, a 0.25 grassland albedo was assumed [420].

Moving on to the floating structure, there are several options available [351]. A key criterion in this research was the ease of modelling. In particular, the rigid body assumption had to be held. Therefore, a pontoon was simulated consisting of floating cubes of high-density polyethylene (HDPE). In particular, they were Sunnydock cubes manufactured by Dock Marine Europe [421]. Their dimensions are 50 cm × 50 cm × 40 cm with a weight of 6.5 kg. Each cube can withstand 87.5 kg (floatation). Although in reality these cubes allow some movement at the connection points, this is neglected in order to satisfy the rigid body assumption. As mentioned in Subsection 10.2.2, this pontoon is anchored to the sea ground and the mooring lines allow rotation of the pontoon but not displacement. A schematic of the pontoon, with the final dimensions already optimized, can be seen in Figure 10.12.

Regarding the PV modules, LG400N2W-A5 monocrystalline silicon module [422] with a rated power of 400 W was selected. Considering the maximum input voltage and the rated power of the selected string inverter Sunny boy 2.5 [415], 6 modules in series can be connected per string. Since the irradiance conditions are homogeneous throughout the string, the PV modules' bypass diodes are not activated.

10.4. RESULTS

After explaining the methodology and its inputs, this section reports the results obtained at each step. The first [Subsection 10.4.1](#) explains the outcome of the mechanical model validation using data from an OFPV system. [Subsection 10.4.2](#) focuses on the surface elevation obtained as a result of the waves model. In [Subsection 10.4.3](#), a sensitivity analysis on the model is performed to determine the suitable size of the system. [Subsection 10.4.4](#) shows the interaction between the floating body and the waves through the angles' variation. Finally, results regarding the effect of waves on DC yield, module temperature, and inverter performance are explained in [Sections 10.4.5, 10.4.6 and 10.4.7](#), respectively.

10.4.1. VALIDATION OF THE MECHANICAL MODEL

The applied approach, in particular up to the derivation of the rotational angles, was validated with data from an OFPV system. The sea surface elevation was also statistically validated by comparing the modelled with the measured significant wave height.

The methodology explained in [Subsection 10.2.1](#) and [Subsection 10.2.2](#) was applied in order to emulate the OFPV system located in the North Sea. Given the stochastic nature of waves and the methodology employed, direct point-by-point comparisons of values are not reasonable. Only statistical analyses can be performed. Therefore, [Figure 10.8](#) shows two overlapped scatter plots of the modelled (in translucent red) and measured (in blue) roll and pitch for the whole measurement period. One can observe how the angles are similarly distributed in both plots. The majority of pitch and roll values are within 5° of the flat state. The distribution of modelled pitch values appears more dispersed compared to the roll due to the floater's rectangular shape. The most spread measured values could be attributed to random phenomena, not accounted for in this model, such as disturbances caused by local wind, marine life, or vessels, or to assumptions made on the dimensions and weight of the real floater.

10.4.2. MODELLED SURFACE ELEVATION

This section explores the resulting waves that could have been statistically formed in 2017 with the assumed wind data and 1-second resolution. [Figure 10.9](#) displays examples of these results before the two axes decomposition. Sea elevation is shown for three days, one per sea state. One can observe a non-linear relation between wind speed and wave elevation: an increase in wind speed in the higher range results in a greater rise in surface elevation compared to an increase in the lower range. The lack of symmetry around 0 for the surface elevation has its origin in the random factors of [Equation 10.1](#) that can shift up and down the final summation value. Although the average monthly elevation values are relatively low (always less than 1 m), on some stormy days, the elevation exceeded 10 m.

This elevation was employed to classify the sea state into calm, moderate and rough categories for the year 2017, as explained in [Subsection 10.2.1](#). Most of the time (66%), the

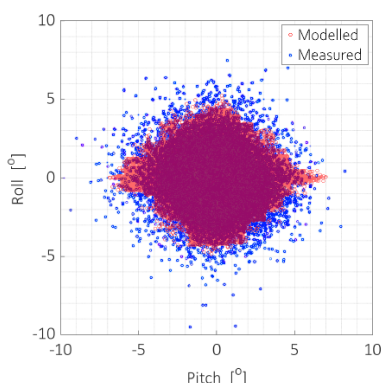


Figure 10.8: Comparison between modelled and measured values of pitch and roll for an OFPV system. Figure adapted from [423].

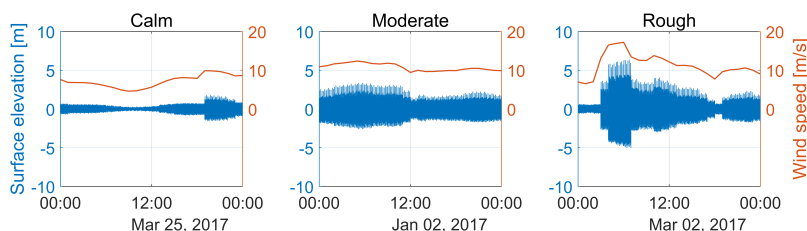


Figure 10.9: Relation between surface elevation and wind speed shown for one day per sea state.

sea was classified as calm, while only 13% of the days were rough. Figure 10.10 shows the distribution of states throughout the year, where days were classified based on daily median significant wave height and then counted for each month. The elevation is generally lower in summer and higher in winter, following the wind speed trend. No day is ranked as rough in April, May and July, while December has more rough days than calm ones.

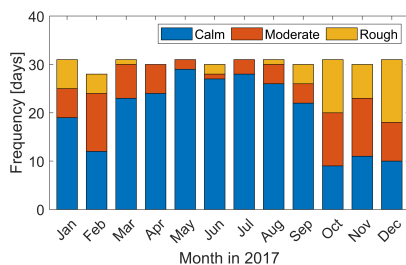


Figure 10.10: Distribution of sea states over the year 2017 on daily resolution.

10.4.3. SENSITIVITY ANALYSIS ON SIZING

As seen in Equation 10.9, the inertia depends on the structure dimensions, which have a significant effect on the system dynamics. Therefore, a sensitivity analysis is performed to

find the optimal size of the OFPV system. This analysis consists of exploring how sensitive the final angle variations are to the moment of inertia of the rigid floater. The objective is to map this relation in order to understand which is the optimized design for the pontoon of this offshore PV system.

Consequently, the model was run for various pontoon sizes, exploring the impact on the mean and standard deviation σ of all angles in the model. Starting with the roll and pitch, their average value was always close to 0° , hence independent of the pontoon size. Their standard deviation showed, however, more interesting trends. As observed in the top left of Figure 10.11, the pitch oscillates the least when the moment of inertia of the x-axis is small, while that of the y-axis is large. That is, when the pontoon is wide but not long (the schematic of Figure 10.3 is shown in Figure 10.11 as well for ease of understanding). Square-shaped floaters achieve similar values although the larger and heavier the pontoon, the smaller the standard deviation. The trend of the standard deviation of the roll (top right of Figure 10.11) is not so intuitive. One would have initially expected a trend similar to that of the σ pitch (smaller deviations when the pontoon is long but not wide), but that is not the case. The origin of this difference is the wind profile of the selected location. Looking at the wind rose of Figure 10.7, one can see that the waves in the EW direction (y-axis of the pontoon) are larger than those in the NS one (x-axis of the pontoon). This makes the standard deviation of the roll in general higher than that of the pitch because it is affected by a combined effect of shape and mass, not only shape as happens with the pitch. The heavier the pontoon, the smaller its oscillations.

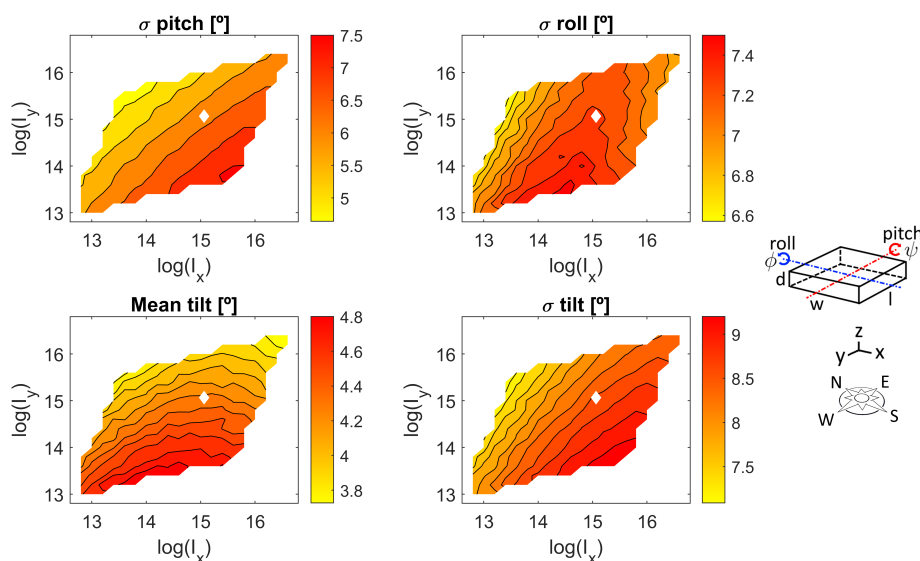


Figure 10.11: Contour plots of the standard deviation of roll (top left) and of pitch (top right), mean of tilt (bottom left), and standard deviation of tilt (bottom right) as a function of the logarithm of the moment of inertia of each axis. Figure 10.3 with a compass rose is shown again for ease of understanding. The pontoon modelled in this work has been represented as a white diamond in the plots.

The relation of the moments of inertia with the angles is also analyzed after the coordinates transformation from roll and pitch to tilt and azimuth. Few conclusions can be

derived from the analysis of the azimuth since the mean is always around the South and the periodicity of the variable hinders the proper interpretation of its standard deviation. However, some information can be obtained from the mean and standard deviation of the tilt. Looking at the mean of the tilt (bottom left of [Figure 10.11](#)), it is more affected by the width, although the minimum can be found for a squared and very heavy pontoon. Finally, the standard deviation of the tilt (bottom right of [Figure 10.11](#)) shows a combined effect between the trends of the standard deviations of the pitch and the roll.

In summary, in this particular location, the effect of sea waves can be reduced by placing a heavy, short, and wide pontoon. This is the optimal design for two reasons. First, the heavier the floater, the smaller the fluctuations because it is harder to incline it. Second, the largest side of the pontoon is aligned with the most common wind direction. Actually, for this location, a pontoon that was aligned not with the cardinal directions but with the most common wind direction of 236° would show even lower fluctuations. This can be applied to any offshore location where the most common wind direction is known. Despite a rectangular pontoon being optimal, it was finally decided to model a squared pontoon so as to consider the intermediate scenario.

In order to determine the dimensions of this pontoon, one of the assumptions of the wave modelling needs to be considered. As stated in [Subsection 10.2.1](#), the wavelength is assumed much longer than the dimension of the object so that the wavefield is only slightly modified by the body. But this condition is not respected for any pontoon area. Considering that 6 modules need to be connected per string, as explained in [Section 10.3](#), two strings were installed per row. That is 12 modules and two string inverters for each row. [Figure 10.12](#) shows a schematic and the dimensions of the final layout. The pontoon has been added to [Figure 10.11](#) as a white diamond. Finally, floatation calculations were performed to ensure the stability of the model.

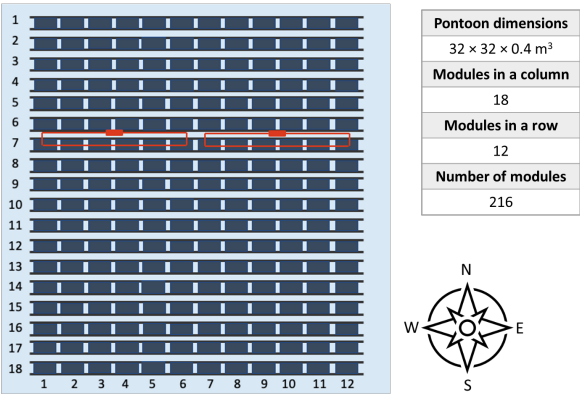


Figure 10.12: Layout of the floating body facing South. The red boxes represent the two string inverters per row.

With this configuration, 3.8% of the time waves have a wavelength smaller than the width of the float. This percentage can be considered reasonable, especially considering that the elevations corresponding to such short wavelengths are minimal, and therefore have a marginal effect on the floating structure.

10.4.4. INCLINATIONS DUE TO WAVES

As explained in Subsection 10.2.2, the tilt and azimuth of the floating body were obtained from the sea elevation, the Froude-Krylov approach, and a coordinate transformation. The histograms in Figure 10.13 show the variation in tilt probability for different wind speeds. As expected, high wind speeds increase the likelihood of significant tilting, with the tilt distributions being sharper around zero in the presence of low winds. Simulations showed that the average tilt for calm days is 0.2° , for moderate ones 1.8° , and for rough ones 4.5° . The average value over the whole year is 1.1° and the maximum simulated tilt is 52.2° .

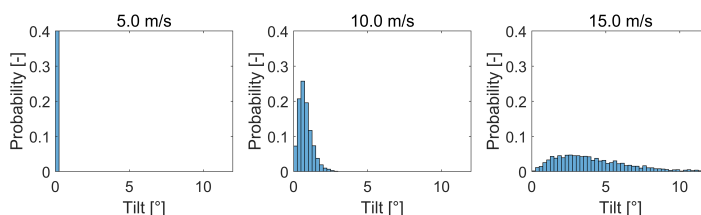


Figure 10.13: Histograms of hourly tilt for different wind speeds.

Figure 10.14 shows the connection between azimuth and wind direction. Starting with the calm sample, until around 19:00 the wind comes from the West so the pontoon oscillates between East and West without almost any disturbances. At 19:00, when the wind slowly shifts towards the North, the pontoon shows a rather irregular behavior and does not calm until 22:00 when the pontoon oscillates between North and South. On rough days, it is harder to reach an equilibrium oscillation even when the wind is aligned with one of the axes. Higher disturbances can be observed on December 28th after 7:00 compared to the oscillations during the calm sample. This behavior also indicates a faster change in the oscillations of the pontoon.

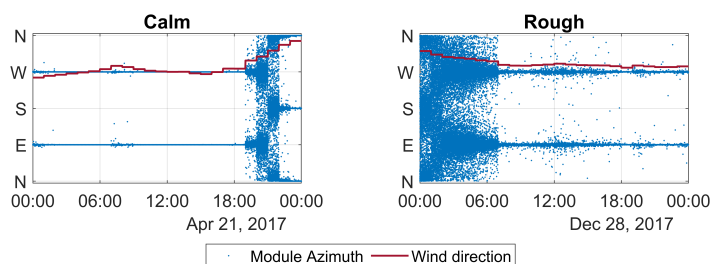


Figure 10.14: Relation between wind direction and azimuth for a calm and rough day.

10.4.5. WAVES IMPACT ON DC YIELD

This section focuses on the impact that waves have on the DC yield. Offshore waves influence the OFPV system yield by inclining the modules and affecting the irradiance incident on them.

As a first analysis, the difference in yield is compared for three offshore scenarios: impacted by waves, fixed horizontal tilt, and fixed optimal tilt. Installing offshore solar panels with the optimal tilt is very challenging due to the mechanical stresses, so this comparison

is only a theoretical analysis. Figure 10.15 compares the generation of a single PV module (installed in a pontoon similar to that sketched in Figure 10.12) for the three scenarios under the three sea states.

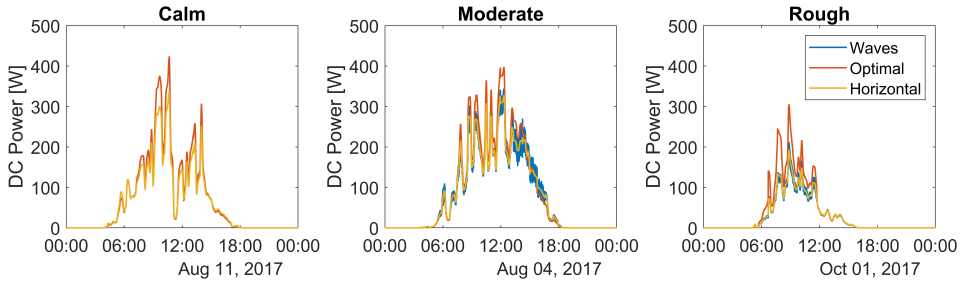


Figure 10.15: DC power trend of a single PV module (rated power of 400 W) for three different sea states. Three offshore scenarios are investigated: with waves, fixed optimal, and fixed horizontal.

On the calm day, because of the low tilts occurring, the wave and horizontal cases follow an identical trend. As expected, the optimal tilt scenario produces more than the other scenarios. At the end of the day, the undulating scenario produces 3.62 kWh/kW_p, the same as the horizontal case and 12.7% less than the optimal tilt.

On the moderate sea condition, the optimal situation also produces more. However, in some instances of time during the early morning and late afternoon, the wave case presents more significant power peaks than the optimal scenario. During these instances, the waves cause the PV panels to pose in a better orientation for that specific day and hour. Throughout the day, the fluctuations of the module with respect to the horizontal case result in an increased generation in some moments and decreased in others. Overall, the PV module with variable orientation generates 4.67 kWh/kW_p, about 0.2% less than the one horizontally mounted. The module installed optimally produces +10.5% compared to the waves scenario.

Finally, a day in October is shown to investigate a rough sea. Due to the generally lower incident irradiance during rough sea conditions, the fluctuations in power generated by the case with variable orientation are of the same order of magnitude as those experienced during moderate sea conditions. On that day, the wave scenario produced 1.58 kWh/kW_p, 0.4% less than the horizontal case and 42.7% less than the optimal one.

The difference in performance between the diverse sea states is investigated on all the simulated days, hence representing a whole year. The results are visible in Figure 10.16. The case under the effect of the waves is always disadvantageous compared to the case with the optimal tilt, reaching an average loss of 34.1% on days with rough seas. Compared to the horizontal scenario, even in the worst-case scenario of rough sea conditions, the energy losses are impressively low, averaging just -0.4%. Considering this, the energy losses due to wave movement can be neglected in practical terms in many scenarios, as already suggested in [372].

Looking now at the yearly energy yield, the OFPV plant modelled in this study would produce 975 kWh/kW_p. The fixed horizontal system would generate 976 kWh/kW_p, an increase of only 0.1% compared to the fluctuating system. This value represents the overall losses due to the wave movement. As expected, the fixed optimum case is the most productive one by generating 1141 kWh/kW_p, 14.6% more compared to the system affected by

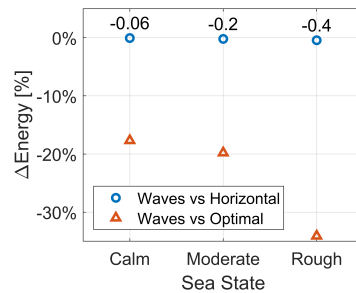


Figure 10.16: Percentage difference in energy yield between the three scenarios according to the sea state.

waves.

These trends can be graphically seen throughout the year in [Figure 10.17](#), which provides the monthly DC production. No difference can be observed between the system subjected to waves and the horizontal one, while the optimum one consistently produces more.

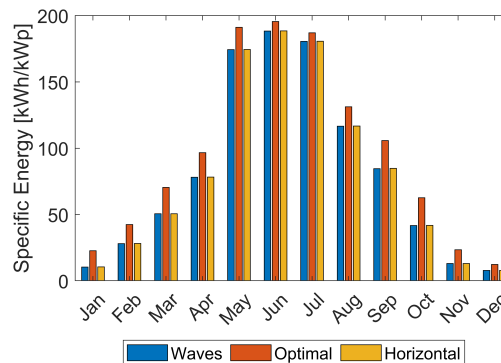


Figure 10.17: Energy yield in 2017 for three offshore scenarios.

In summary, although the energy yield from the floating scenario is lower than that of the horizontal scenario, given the simulated conditions in this work, waves generally have a negligible impact on power production.

This study has not considered mismatch losses due to series-connected PV modules receiving different irradiance. The rigid body assumption taken in [Subsection 10.2.2](#) subjects all PV modules to the same weather conditions. Considering however that the pontoon is composed of floating cubes, the sea agitations will most likely create movement between them leading to mismatch losses.

Using results from the literature, one can roughly estimate the mismatch losses of the system as if the PV modules in the same string received different irradiance. Kumar et al. linked the mismatch losses to the wave height based on experimental work [424]. They reported that for wave heights of 0.25, 0.5, and 1 m, the mismatch losses are around 0.1, 0.3, and 0.7%. Assuming a linear relation between these wave heights and mismatch losses, one can fit a linear interpolation between these two variables (R^2 of 95.9%) and estimate the mismatch losses for every hour using the significant wave height. Results show that the system's yearly average mismatch losses would be 1.1%, consistent with an average significant

wave height of 1.5 m for the selected location. Using the results from another study obtained by Dörenkämper et al., the loss in yield due to wave-induced mismatch losses would be around 9% [350]. The variation in mismatch losses is likely attributed to the distinct conditions simulated, including differences in floaters and locations. In both cases, the mismatch loss is higher than the loss in DC power resulting from the changing irradiance due to wave motion. These losses could be reduced during the DC to AC conversion step with a converter that employs a dynamic mismatch loss mitigation algorithm [425]. This dual input dual output converter for solar PV systems enhances power production when series-connected modules are subjected to different shading patterns, such as those created by different tilts.

10.4.6. MODULE TEMPERATURE

In order to study the difference in module temperature, a comparison is made with an inland scenario. In this subsection, three scenarios are compared: stationary optimal inland, stationary optimal offshore, and offshore subjected to waves. Ambient temperature, wind speed, and albedo differ between the inland and offshore scenarios.

Figure 10.18 shows the daylight module temperature for the three investigated scenarios. Due to computational restrictions, the same approach as for the DC yield is employed: one day per month and sea state is selected as representative and it is extrapolated for each month considering the sea state distribution. The wave case shows slightly lower and less varying module temperatures than the offshore optimal one, due to the reduced irradiance incident on the PV modules. The inland system is subjected to a significantly higher temperature during the central months of the year. This difference is, in median terms, up to 5°C in some months. This trend is reversed for the coldest months of the year. This is due to water acting as a heat sink, which reduces the modules' temperature in summer but keeps them warm during winter. Additionally, the inland temperature temporal variation is more significant than the offshore one.

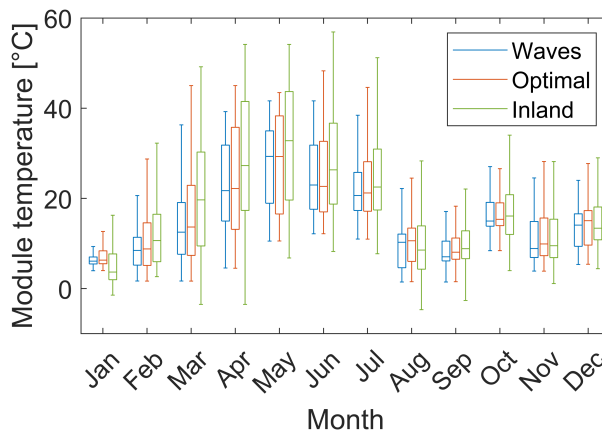


Figure 10.18: Daylight module temperature in 2017 for three different scenarios: offshore considering waves, offshore steady optimal tilt, and inland steady optimal tilt.

This difference in temperature has a significant effect on the yearly DC yield. The inland system would produce 1123 kWh/kW_p, i.e. -1.7% compared to the optimal offshore appli-

cation and +13.2% compared to the wave case. When contrasted with the results observed by Golroodbari and van Sark, who noted a nearly 13% increase in the DC yield for a land-based scenario compared to an offshore one [357], this study presents similar findings. This may indicate that considering two axes of rotation is not necessary in certain conditions.

Figure 10.19 provides the monthly trends in 2017. From September to April, the production difference between the optimal inland and offshore scenarios is minimal. In summer, however, the advantage of installing at the sea is more evident due to the temperature effect.

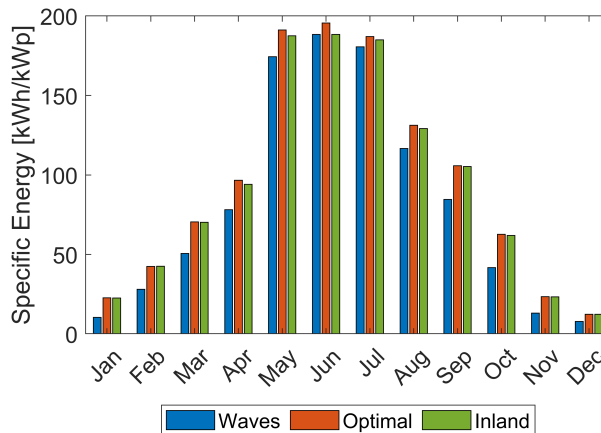


Figure 10.19: Energy yield in 2017 for two offshore and one inland scenarios.

Ultimately, from a technical point of view, the benefit of the offshore temperature in the ideal case leads to a gain of less than 2% in the yearly yield. Discovering if this margin can cover the higher investment cost for an offshore plant and the degradation of the components due to humidity, algae, and salt is an economic analysis that this work will not consider. However, the literature indicates that FPV can have similar or lower costs than land-based PV if there is a decrease in the cost of floaters and hybridization with hydropower [426]. The feasibility of each project will depend on the land market, installation tilt, and the drop in the price of materials for offshore installations.

10.4.7. WAVES IMPACT ON INVERTER

This final section analyses the inverter results obtained in the laboratory. The objective is to quantify the inverter losses due to the fluctuating output of an OFPV system. Since the oscillations depend on the sea state condition, each condition will be studied separately and compared with a fixed horizontal OFPV system. Since the inverter efficiency depends on the input power produced by the modules, 9 time windows for each condition were considered in order to cover a wide range of output power values.

Figure 10.20 shows one curve per condition (all with the same average power of around 1000 W) for exemplification. One can observe the undisturbed profile of DC, operating, and output power when the system is fixed or the sea is calm. The fluctuations in power are only in response to variations in irradiance. The operating and DC power curves overlap, which indicates a high MPPT efficiency. Fluctuations in power can be observed in the moderate and rough samples, with greater variability observed in the latter. These variations decrease

the MPPT efficiency compared to the stationary and calm cases, although the tracker can follow quite closely the changes in DC power. The conversion efficiency does not seem much affected either by the fluctuations in power.

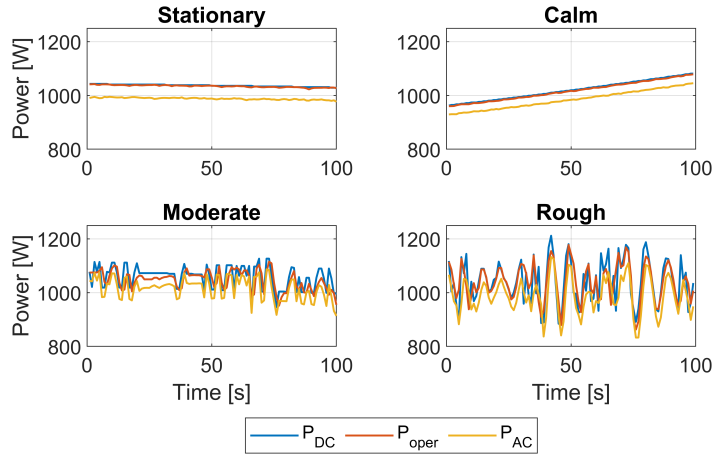


Figure 10.20: P_{DC} , $P_{operating}$ and P_{AC} samples for each of the four cases analysed.

The average MPPT, conversion, and total efficiency have been computed for all the samples. After grouping the samples by condition, the average and standard deviation of each group have been reported in Table 10.3. Here, one can observe similar trends as in Figure 10.20. The MPP tracker has no difficulty in finding the maximum power when there are no or low fluctuations. Even when the sea state is moderate, the MPPT efficiency is barely affected. In the presence of higher fluctuations, the average MPPT efficiency drops only half a percentage point, and the standard deviation doubles. Therefore, the inverter is able to track the constantly varying maximum power point although with difficulties.

Table 10.3: Average values and their standard deviation for total, MPPT, and conversion efficiencies resulting from the experimental investigation.

	Total efficiency [%]	MPPT efficiency [%]	Conversion efficiency [%]
Stationary	96.0 ± 0.5	99.5 ± 0.5	96.4 ± 0.5
Calm	96.4 ± 0.3	99.3 ± 0.6	97.1 ± 0.5
Moderate	96.4 ± 0.4	99.3 ± 0.4	97.0 ± 0.2
Rough	94.1 ± 2.1	98.8 ± 1.1	95.3 ± 1.3

The conversion efficiency is affected just like the MPPT efficiency. Very similar mean and standard deviation values have been measured for the stationary, calm, and moderate samples, while the rough samples report an average conversion efficiency of about 1.5 percentage points lower and a standard deviation more than twice as large. Overall, the observed behavior leads to an average total efficiency of 94.1% with a standard deviation of 2.1% for the rough samples. This is approximately 2 percentage points lower than the efficiency observed in the remaining cases, which have a mean efficiency ranging from 96.0% to 96.4% with a standard deviation of 0.3% to 0.5%.

To observe this phenomenon better, the instantaneous inverter efficiency of all the experimental data as a function of the input DC power is shown in Figure 10.21. One can observe how the moderate but especially rough samples have a high variation and oscillate between high and low power values. In contrast, the stationary and calm samples are cluttered instead of dispersed. The less turbulent the sea state, the smaller this dispersion.

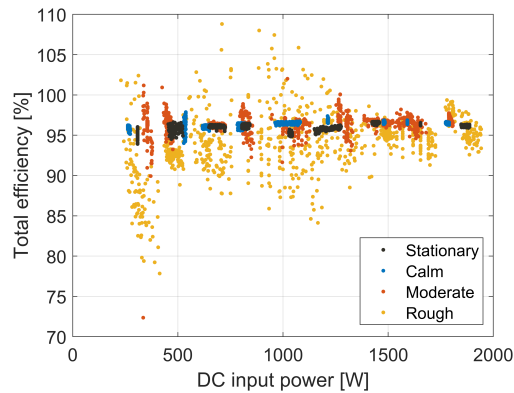


Figure 10.21: Total efficiency of the inverter as a function of input DC power for all emulated samples and classified according to the four cases analysed.

In summary, the inverter's total efficiency on an offshore calm or moderate day is the same as for a stationary system. An average loss of around 2 percentage points is expected during rough days compared to stationary applications. This loss is caused mostly during the DC to AC conversion process. Nevertheless, it should be noted that rough days occur especially in the winter months, with low irradiance and lower power losses compared to a clear summer day. In view of these results, one can conclude that the overall inverter losses of an oscillating system will only be slightly higher than those of a stationary one.

10.5. CONCLUSIONS

This chapter has simulated the performance of an offshore floating PV system with the objective to explore the effect of waves movement on DC yield and inverter efficiency. The waves have been modeled using the JONSWAP spectrum and the interaction with the floating structure has been simulated considering two axes of rotation, an improvement in accuracy compared to previous literature. This methodology has been validated with real data, resulting in a statistical match between modeled and measured data. A sensitivity analysis conducted on the optimal shape of the floater indicates that a heavy and rectangular floater with the widest side aligned with the most common wind direction reduces angle variations. The modeling results indicate that waves have a negative yet negligible effect on production, leading to a decrease in the yearly energy yield of 0.1% compared to a stationary horizontal scenario. This production is however 14.6% lower than if the system was optimally tilted offshore. Compared to an optimally-tilted inland scenario, the decrease in production is 13.2%. This difference is due to the heat sink effect of water which can decrease the median daylight temperature by about 5°C during the summer months. Additionally, the effect of the fluctuating PV power due to waves on the inverter efficiency has been studied for the first time in the literature, as far as the authors are aware. The experi-

mental results show that the effect of waves on the inverter efficiency is detrimental during a rough sea, with a loss of 2 percentage points in total efficiency compared to a stationary system. Throughout the year this loss is however small considering that only 13% of the time the sea is rough.

PERSPECTIVE: MORE ACCURATE METHODOLOGY

After publishing the article which is the basis of this chapter [376], a better approximation of the mechanical model was identified by Dr. Antonio Jarquín Laguna from the Delft University of Technology.

The equations of the mechanical model were derived independently for each axis. This resulted in two Jonswap spectra, one per axis. However, these spectra are not independent as the superposition rule does not apply to the spectra. This can be seen as yet another assumption of which we were not aware since this approximation did not affect the validation performed in Subsection 10.4.1. This could have been because the sea elevations were small. However, implementing the new approximation is relatively straightforward and its process is detailed in this section.

Instead of obtaining two spectra, one per axis, a single spectrum is computed using the same equations but with the wind speed measured by the meteorological station. This spectrum will be directional as the employed wind speed changes direction over time. This directionality is expressed in the surface elevation η by making it dependent on both x and y , as in Equation 10.12:

$$\begin{aligned}\eta(x, y, t) = & \sum_{n=1}^{N/2} (a_n \cdot \cos(k_n \cdot (x \cdot \cos(\theta(t)) + y \cdot \sin(\theta(t))) - \omega_n \cdot t) + \\ & b_n \cdot \sin(k_n \cdot (x \cdot \cos(\theta(t)) + y \cdot \sin(\theta(t))) - \omega_n \cdot t)) \\ & a_n = r n_{a_n} \cdot \sigma_S \\ & b_n = r n_{b_n} \cdot \sigma_S\end{aligned}\quad (10.12)$$

where $\theta(t)$ is the wind direction [°].

This additional dependency on y is translated to the pressure $p(x, y, z, t)$. Aside of that, Equation 10.4 and Equation 10.3 still hold.

Equation 10.7 would result in equations for the x - and y -axes, as in the following Equation.

$$\begin{aligned}\tau_x = F \cdot x &= \int_{-w/2}^{w/2} \int_0^l p \cdot x \, dy \, dx \\ \tau_y = F \cdot y &= \int_{-l/2}^{l/2} \int_0^w p \cdot y \, dx \, dy\end{aligned}\quad (10.13)$$

Similarly, one obtains the two expressions shown in Equation 10.14 for the rotational angles.

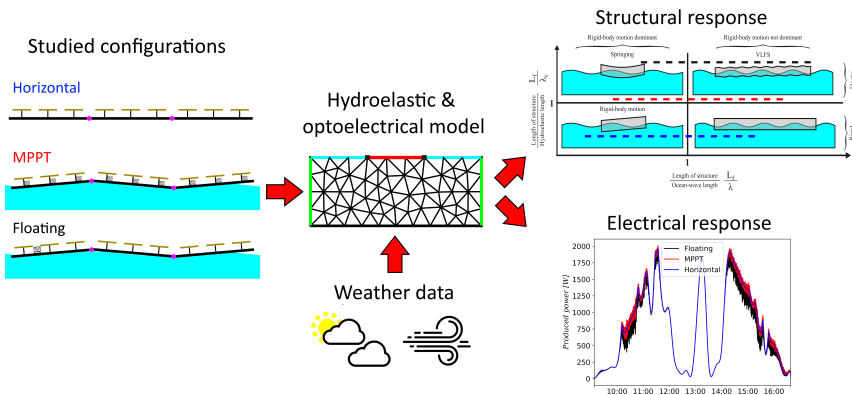
$$\begin{aligned}\tau_x = I_y \cdot \alpha_x &= I_y \cdot \frac{\partial \omega_\psi}{\partial t} = I_y \cdot \frac{\partial^2 \psi}{\partial t^2} \\ \tau_y = I_x \cdot \alpha_y &= I_x \cdot \frac{\partial \omega_\phi}{\partial t} = I_x \cdot \frac{\partial^2 \phi}{\partial t^2}\end{aligned}\quad (10.14)$$

Both Equation 10.13 and Equation 10.14 are also applied in the initial methodology. They were expressed for a single axis to increase compactness. However, the resulting equations after solving the differentials using the new expression for the elevation will be different.

11

MOVEMENT AND POWER MISMATCH LOSSES

The previous chapter's OFPV model overlooked fluid-structure interactions. This chapter addresses this assumption by developing a multi-physics framework integrating the mechanical model of a floating structure with an optoelectrical model of PV modules. It analyzes a hypothetical OFPV design with various floater configurations, from one large floater to multiple small floaters connected by free hinges. Results show a trade-off: fewer, longer floaters reduce power mismatch loss but experience higher structural stress. Young's modulus influences longer floaters with dominant elastic responses, while cross-section fill ratio affects shorter floaters, where rigid-body response prevails.



This chapter has been adapted from the publication A. Alcañiz, S. Agarwal, P. Tiwald, O. Isabella, H. Ziar, and O. Colomès "Assessment of power losses and structural response of offshore floating solar platform" under review in *Energy*.

11.1. INTRODUCTION

While informative, the model developed in the previous chapter made several assumptions that did not represent the complex interaction between the floater and the PV system. For instance, the model simplified the inherent fluid-structure interaction (FSI), i.e., the interaction of the floating platform with ocean waves. This interaction is particularly relevant for the truly offshore deployment of floating PV. The consequences of this FSI are twofold: 1) the FSI results in structural loads within the floaters and joints of the floating island, depending on both the configuration and properties of the floater and sea waves; 2) the FSI results in a dynamic change in the tilt of the individual PV panels, resulting in potential power losses due to power mismatch and movement. Therefore, the hydroelastic response of OFPV platform has significant multi-physical structural and opto-electrical implications.

So far in the literature, the structural analysis of OFPV platforms has largely been done using simplified approaches, owing to the complexity of the problem. The structural deformation of OFPV platforms has been investigated using the Froude-Krylov method for estimating the wave force [427, 428]. This basic method has several limitations, such as assuming undisturbed waves, disregarding diffraction, lack of hydrodynamic coefficients for thin-floating platforms, and inability to resolve the difference in loads between boundaries and inner regions of the floating PV platform. Alternative simplified methods involve estimating the rotational motion of a single rigid panel based on undisturbed wave energy density [357], however, their rudimentary assumptions restrict their applicability. Recent studies on hydro-elastic interaction between small amplitude linear waves and perforated elastic plates using the arbitrary Lagrangian-Eulerian methodology show limited accuracy for regular-wave tests compared to the corresponding experiments [429], thus highlighting the challenges in simulating OFPV structures. Wave and wind loads can exert significant pressure on the floating and mooring structure, resulting in loads in the order of mega-Newton as suggested by a quasi-static analytical model [430]. The stress experienced by the connection pins and the structural components is addressed by Sree et al. [429]. These studies contribute to a better understanding of the mechanical aspects and structural integrity of OFPV systems.

The literature currently lacks a thorough investigation into the effects of hydroelastic responses on power generation in OFPV systems. In [357], the performance of an OFPV system is compared with that of a land-based one by using a simplified representation of the sea. In the previous chapter, that study was extended but key assumptions remained, including that the wave field was unaffected by the floating structure, the wave force on the floating body was estimated using the Froude-Krylov approach, and the floating body was considered rigid. Additionally, power mismatch losses were not addressed in either of the manuscripts. The measurements conducted by Dörenkämper et al. address the mismatch issue in [431]. They reported power mismatch losses of 3%, 7%, and 9% for median wave heights of 0.1, 0.4, and 1.5 m, respectively. These losses differ from the measured ones in [424], where the mismatch losses were 1.6%, 2.4%, and 3.7% for average wave heights of 0.25, 0.5, and 1 m, respectively. These differences may stem from variations in floater characteristics. However, none of these studies tackle the origin of this disparity, making it challenging to identify the source of the discrepancies.

Since our analysis is quite focused on the hydroelastic response, we simplify the problem with several assumptions. We limit external loads to wave forces, omitting current-induced loads. This is because currents primarily exert in-line friction and drag forces on

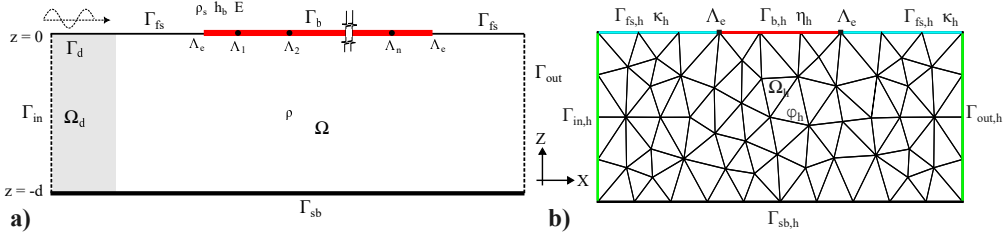


Figure 11.1: a) Schematic of the boundary value problem, consisting of the fluid domain Ω with constant water-depth d , bounded by the bottom Γ_{sb} , inlet Γ_{in} , outlet Γ_{out} , free-surface Γ_{fs} and the floating PV platform Γ_b . Here Λ_j represent the free-hinges in the floating PV platform. b) Sketch of the discretised geometric entities, with Ω_h representing the discretised fluid domain, and $\Gamma_{sb,h}$, $\Gamma_{in,h}$, $\Gamma_{out,h}$, $\Gamma_{fs,h}$ and $\Gamma_{b,h}$ representing the discretised boundaries. Consequently, ϕ_h , η_h and κ_h are the discrete solutions to the velocity potential, free-surface elevation and beam deflection, respectively.

thin elastic structures, thereby affecting station-keeping [427], rather than power mismatch losses, which are mainly influenced by transverse loads. Furthermore, while the mooring and anchoring systems are vital to station-keeping [430], and structural fatigue is essential for durability assessment [432], these aspects lie beyond this study's scope since our emphasis is on the fluid-structure interaction and associated power mismatch losses.

Additionally, the designs presented in this work are academic and are not affiliated with any commercial entities. Therefore, we do not investigate a specific OFPV platform design. Instead we formulate a systematic engineering analysis and then investigate the impact of the structural properties in isolation on the power-mismatch losses. For example, through our analysis, we will answer fundamental questions such what happens when you double or half the thickness of the floater. By narrowing the focus to power mismatch related performance measure, this study aims to bridge the gap in the literature on the the fundamental correlation between the hydroelastic response of OFPV platform to irregular sea-states and their power mismatch losses.

This work is structured as follows. The following Section 11.2 presents the methodology employed to solve the problem at hand. Section 11.3 outlines the inputs to the model. The structural and optoelectrical results for a base case are presented in Section 11.4. The variations to this base case and their effect on the results are discussed in Section 11.5. Finally, Section 11.6 concludes this work.

11.2. METHODOLOGY

This section explains the problem at hand and the methodology developed to solve it. The flowchart presented in Figure 11.2 presents the steps undertaken. The following sections will dive into its details.

The first step consists on defining and solving the structural problem in Subsection 11.2.1. The output of this step will be the elevation at every position of the floater. It will be used to compute the tilt in Subsection 11.2.2, which serves as a link between the structural and optoelectrical models. The methodology of the latter will be presented in Subsection 11.2.3.

11.2.1. HYDROELASTIC FLUID-STRUCTURE INTERACTION MODEL

The structural analysis of the floating PV platform involves a complex hydroelastic problem characterised by a two-way interaction between the fluid and the elastic structure. We conduct this analysis using a monolithic finite-element model for fluid-structure interaction, referred to as the FE-FSI model throughout this work.

The monolithic formulation and the FE model used in this work are based on the detailed methodology presented in [433]. In this work, we focus on a frequency domain analysis, because we are primarily interested in the steady-state solution. This section briefly outlines the final frequency domain governing equations and the FE model. The convergence, conservation, and validation analysis for this FE model, when applied to a similar problem involving a floating beam with joints, was previously presented in [433].

DEFINITION OF THE BOUNDARY VALUE PROBLEM

Figure 11.1a presents the schematic of the boundary value problem. Here, the fluid domain is denoted Ω . It is bounded by the bottom boundary Γ_{sb} , the fluid free-surface Γ_{fs} , inlet Γ_{in} and outlet Γ_{out} . The floating PV platform is represented by the boundary Γ_b . The end points of this platform are denoted by Λ_e . This platform may be composed of multiple floaters, connected by joints, denoted by Λ_j . This boundary value problem involves three unknowns, 1) velocity potential ϕ defined on Ω , 2) free-surface elevation κ defined over Γ_{fs} and 3) transverse beam deflection η defined over Γ_b .

In the frequency domain formulation, the unknown quantities ϕ , η , and κ are assumed to be harmonic. Here, a space and time-varying quantity, say the beam deflection $\eta(t, \mathbf{x})$, is defined in terms of the excitation frequency ω and a time-independent quantity $\bar{\eta}(\mathbf{x})$ as

$$\eta(t, \mathbf{x}) = \bar{\eta}(\mathbf{x}) \exp(-i\omega t). \quad (11.1)$$

Following this harmonic assumption, the governing equations for the time-dependent ϕ , η and κ get simplified to governing equations for time-independent $\bar{\phi}$, $\bar{\eta}$ and $\bar{\kappa}$. To simplify the notations in this section, we drop the over-bar hereafter, i.e., $\phi \stackrel{\text{def}}{=} \bar{\phi}$, $\eta \stackrel{\text{def}}{=} \bar{\eta}$ and $\kappa \stackrel{\text{def}}{=} \bar{\kappa}$.

We now present the coupled governing equations for the system in frequency domain formulation. Assuming inviscid, irrotational, and incompressible flow, the governing equation for the fluid flow within Ω is given by the potential flow theory Equation 11.2.

$$\Delta\phi = 0 \quad \text{in } \Omega. \quad (11.2)$$

The velocity potential ϕ is a scalar measure of the fluid velocity vector $\mathbf{u} = \nabla\phi$.

We formulate a boundary value problem by specifying conditions at the boundaries Γ of this fluid-flow domain Ω . We impose a no-flow boundary condition Equation 11.3a on the bottom boundary Γ_{sb} . The ocean wave input is given through the inlet boundary Γ_{in} using Equation 11.3b, where \mathbf{n} is the outward unit normal to the boundary and u_{in} is the inlet velocity prescribed using the Airy's wave theory [434, ch. 3]. At the outlet boundary Γ_{out} , we impose a radiation boundary condition Equation 11.3c for releasing the waves downstream of the floating PV platform.

$$\mathbf{n} \cdot \nabla \phi = 0 \quad \text{on } \Gamma_{sb} \quad (11.3a)$$

$$\mathbf{n} \cdot \nabla \phi = u_{in} \quad \text{on } \Gamma_{in} \quad (11.3b)$$

$$\mathbf{n} \cdot \nabla \phi = i k \phi \quad \text{on } \Gamma_{out} \quad (11.3c)$$

Here k is the angular wave-number for the ocean wave with frequency ω , obtained using the dispersion relationship [Equation 11.4](#), where d is the water depth at the inlet boundary.

$$\omega^2 = g k \tanh(kd) \quad (11.4)$$

The free-surface Γ_{fs} and beam Γ_b boundaries involve a kinematic and a dynamic boundary condition. The kinematic free-surface boundary condition [Equation 11.5a](#) on Γ_{fs} implies that the free-surface elevation moves at the vertical fluid velocity. The dynamic free-surface boundary condition [Equation 11.5b](#) on Γ_{fs} is used to impose atmospheric pressure $p_a = 0$ at the fluid free-surface.

$$\phi_z + i\omega\kappa = 0 \quad \text{on } \Gamma_{fs} \quad (11.5a)$$

$$i\omega\rho\phi - \rho g\kappa = 0 \quad \text{on } \Gamma_{fs} \quad (11.5b)$$

This is obtained using the linearised dynamic Bernoulli's equation, assuming a small free-surface elevation.

$$p = i\omega\rho\phi - \rho g\kappa \quad (11.6)$$

Here g is the acceleration due to gravity and ρ is the density of the fluid. Therefore, the dynamic boundary condition [Equation 11.5b](#) is obtained by setting the pressure $p = p_a = 0$ along Γ_{fs} .

The kinematic boundary condition [Equation 11.7](#) on Γ_b is obtained assuming no gap between the deforming beam and the fluid-free surface underneath the beam.

$$\phi_z + i\omega\eta = 0 \quad \text{on } \Gamma_b \quad (11.7)$$

The dynamic boundary condition on Γ_b requires the introduction of the beam governing equations. The beam is modelled using the Euler-Bernoulli beam theory. Consider a beam with mass per unit area m_b and bending stiffness D_b . The transverse deflection η of this beam, under the action of transverse load p is given by [Equation 11.8](#).

$$-m_b\omega^2\eta + D_b\nabla^2\eta = p \quad (11.8)$$

For a beam having a solid cross-section of thickness h_b , and made of material with density ρ_s and Young's modulus E , the mass per unit area is $m_b = \rho h_b$ and the bending stiffness is $D_b = EI_b$, where $I_b = \frac{1}{12}h_b^3$ is the second moment of area assuming a unit-width beam. For the floating beam, the transverse load is due to the pressure from the fluid underneath the beam. This is imposed by equating the transverse load in Euler-Bernoulli equation

Equation 11.8 with the pressure in the dynamic Bernoulli-equation Equation 11.6, thus resulting in the dynamic boundary condition Equation 11.9 on Γ_b . Here $m_\rho = m_b/\rho$ is the normalised mass and $D_\rho = D_b/\rho$ is the normalised rigidity of the beam.

$$-m_\rho \omega^2 \eta + D_\rho \Delta^2 \eta - i\omega \phi + g\eta = 0 \quad \text{on } \Gamma_b. \quad (11.9)$$

The beam governing equation requires the specification of its own boundary conditions on its endpoints Γ_e . In this work, we restrict the analysis to free-edge boundary conditions. This is imposed by setting zero-moment and zero-shear conditions at these boundaries.

$$D_b \Delta \eta = 0 \quad \text{on } \Lambda_e \quad (11.10a)$$

$$\nabla(D_b \Delta \eta) \cdot \mathbf{n}_\Lambda = 0 \quad \text{on } \Lambda_e \quad (11.10b)$$

Here \mathbf{n}_Λ is the outward unit normal for the beam edges Λ_e . In case the platform is composed of multiple floaters, we join the floaters using joints Λ_j . At these joints, we impose a free-joint condition by setting the bending moment at the joints to zero.

$$D_b \Delta \eta = 0 \quad \text{on } \Lambda_j. \quad (11.11)$$

This completes the definition of the multi-physics boundary value problem. We additionally place a damping zone Ω_d adjacent to Γ_{in} for absorbing waves reflected by the floating PV platform. Refer to [433] for a detailed description of the wave generation and absorption.

NUMERICAL FORMULATION

The above-mentioned boundary value problem is numerically solved using a monolithic finite-element scheme, as defined in [433]. Here, the monolithic term implies that the coupled fluid-structure governing equations are formulated in a single coupled weak-form.

For the sake of completeness of the work, we describe the formulation of the FE problem. We refer to [433] for a more detailed description of the steps to reach this final form. Consider a functional space \mathcal{V} defined in the domain Ω , with trace space $\mathcal{V}_{\Gamma_{fs}}$ on the free-surface Γ_{fs} , and trace space \mathcal{V}_{Γ_b} on the beam boundary Γ_b . The weak form reads as follows: Find $[\phi, \eta, \kappa] \in \mathcal{V} \times \mathcal{V}_{\Gamma_b} \times \mathcal{V}_{\Gamma_{fs}}$ such that:

$$B_\omega([\phi, \eta, \kappa], [w, v, u]) = L_\omega([w, v, u]) \quad \forall [w, v, u] \in \mathcal{V} \times \mathcal{V}_{\Gamma_b} \times \mathcal{V}_{\Gamma_{fs}}, \quad (11.12)$$

where, w, v, u are the test functions corresponding to ϕ, η, κ , respectively, and

$$B_\omega([\phi, \eta, \kappa], [w, v, u]) = (\nabla \phi, \nabla w)_\Omega + (-i\omega \phi + g\kappa, \beta(u + \alpha w))_{\Gamma_{fs}} - (-i\omega \kappa, w)_{\Gamma_{fs}} + (m_\rho \eta_{tt} + \phi_t + g\eta, v)_{\Gamma_m} + (D_\rho \nabla^2 \eta, \nabla v)_{\Gamma_b}, \quad (11.13)$$

$$L_\omega([w, v, u]) = (u_{in}, w)_{\Gamma_{in}} + (u_{out}, w)_{\Gamma_{out}}. \quad (11.14)$$

In Equation 11.13 and Equation 11.14, we have used the notation $(a, b)_\Gamma$ to denote the integral operation, i.e., $(a, b)_\Gamma = \int_\Gamma a b d\Gamma$. The weak-form is solved in a discretised domain, as shown in Figure 11.1b. We consider a triangulation Ω_h of the domain Ω , with conformal triangulation of the domain boundary, $\Gamma_{sb,h}$, $\Gamma_{in,h}$, $\Gamma_{out,h}$, $\Gamma_{fs,h}$ and $\Gamma_{b,h}$. We define the

finite-dimensional finite element spaces for the velocity potential, free surface elevation, and beam deflection, \mathcal{V}_h , $\mathcal{V}_{\Gamma_{fs},h}$ and $\mathcal{V}_{\Gamma_b,h}$, respectively, as follows:

$$\mathcal{V}_h = \{w_h \in \mathcal{C}^0(\Omega) : w_h|_K \in \mathbb{P}_r(K), \forall K \in \Omega_h\}, \quad (11.15)$$

$$\mathcal{V}_{\Gamma_{fs},h} = \{w_h|_E : w_h \in \hat{\mathcal{V}}_h, \forall E \in \Gamma_{fs}h\}, \quad (11.16)$$

$$\mathcal{V}_{\Gamma_b,h} = \{w_h|_E : w_h \in \hat{\mathcal{V}}_h, \forall E \in \Gamma_bh\}, \quad (11.17)$$

In this work, $\mathbb{P}_r(K)$ is the space of Lagrange polynomials of degree 4 in an element K in the 2-dimensional space of fluid domain Ω . On the other hand E are 1-dimensional facets, corresponding to boundaries Γ_b and Γ_{fs} .

Using this notation, the final discrete form in the frequency domain reads: *Find* $[\phi_h, \eta_h, \kappa_h] \in \mathcal{V}_h \times \mathcal{V}_{\Gamma_b,h} \times \mathcal{V}_{\Gamma_{fs},h}$ *such that*

$$B_\omega([\phi_h, \eta_h, \kappa_h], [w_h, v_h, u_h]) = L_\omega([w_h, v_h, u_h]) \quad \forall [w_h, v_h, u_h] \in \mathcal{V}_h \times \mathcal{V}_{\Gamma_b,h} \times \mathcal{V}_{\Gamma_{fs},h}, \quad (11.18)$$

with B_ω and L_ω as defined earlier.

The finite element formulation was numerically implemented in the *Julia* programming language [435], using the *Gridap* finite element library [436]. This library allows for a high-level implementation of the formulation, closely resembling the mathematical notations in weak form. The *Gridap* library efficiently assembles internal loops based on this notation-based expression, optimised for the high-performance *Julia* Just-In-Time (JIT) compiler. A comprehensive software implementation of *Gridap* can be found in [437]. These features facilitate rapid prototyping of the finite-element program and offer excellent computational performance. Notably, *Gridap* can assemble finite-element formulations over mixed dimensions and mixed orders. This capability is crucial because the monolithic floating beam problem requires a formulation over both a fluid domain and the attached beam surface. The program used to generate the results for this work is similar to the registered *Julia* package [438].

The presented FE model has been thoroughly validated for problems involving floating beams with joints in our previous work [433]. This includes validation against a standard test case of a floating platform composed of beams with different properties, connected by either free or elastic joints [439]. The model was also validated against results for floating beams in irregular bathymetry [440] and against experiments involving a 2D rectangular plate floating in a 3D domain [441]. These validations demonstrate that the FE model is particularly suitable for studying the floating PV platforms considered in this work.

11.2.2. LINK BETWEEN MULTIDISCIPLINARY MODELS

The FE-FSI model provides the hydroelastic response of the floating solar platform to incoming ocean waves. This model yields complex-valued solutions for the unknowns, assuming unit amplitude waves, thereby encompassing information about both the amplitude and phase. However, to conduct the optoelectrical analysis, it is necessary to evaluate the time series of the tilt of each panel on the platform. This section describes the process of obtaining this time series of tilt for each photovoltaic (PV) panel. It is summarized in Figure 11.2.

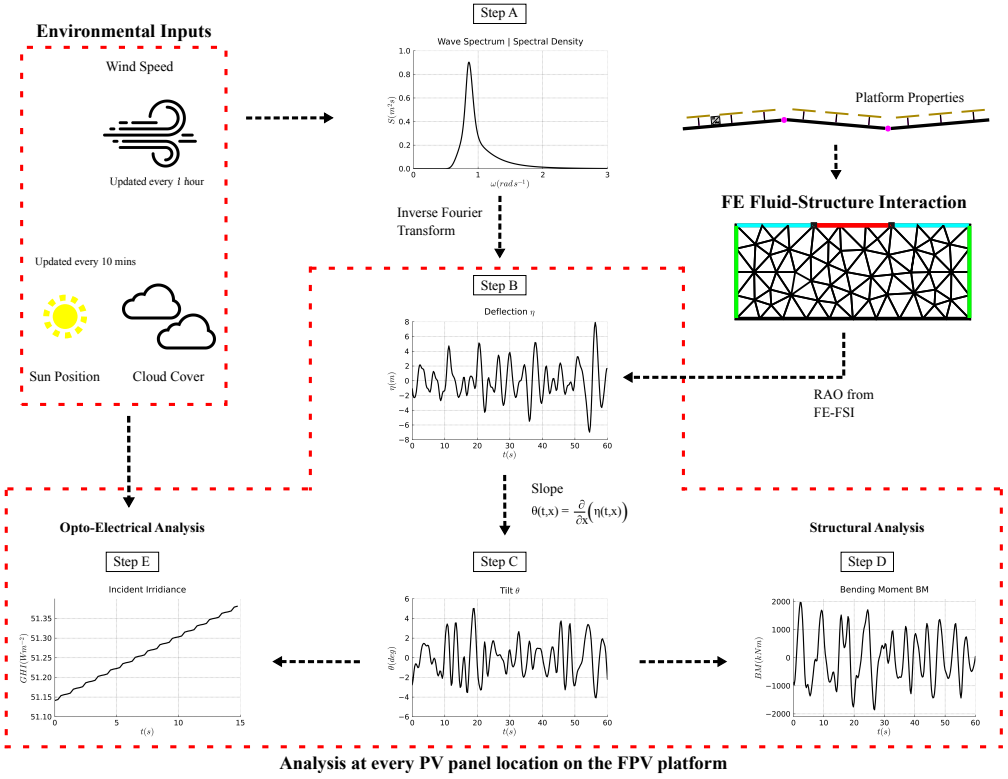


Figure 11.2: Schematic of the solution procedure for the multi-physics problem.

The inputs required for the FE-FSI model include the platform properties and the description of the fluid domain. The model returns complex-valued solutions of the velocity potential ϕ throughout the fluid domain, free-surface elevation κ along the ocean free-surface, and transverse beam deflection η along the entire floating PV platform. The solution is obtained for the entire range of the ocean wave spectrum, assuming unit-wave amplitude. This solution is referred to as the complex-valued response-amplitude operator (RAO) $RAO(\mathbf{x}, \omega)$ for each point \mathbf{x} and frequency ω . This RAO is computed for the given floating PV platform properties and sea-bed topography.

We next use these RAOs with the environmental conditions to compute the time series of the tilt at each panel. We source the environmental conditions from KNMI, the Dutch Meteorological Institute [196]. The hourly mean wind speed is obtained from reanalysis dataset of meteorological measurements provided by the KNW KNMI North Sea Wind atlas [442]. This time resolution corresponds to the highest one we could find for the Dutch North Sea. Using the hourly mean wind speed, we calculate the hourly sea state, a standard offshore analysis procedure [443]. From the hourly sea state, a time series input wave elevation is obtained at a resolution of 0.25 s. This input wave elevation along with the RAOs from the structural model are used to evaluate the tilt of each panel at a time resolution of 0.25 s.

Since the structural results are obtained in the frequency domain and the optoelectrical

analyses are performed in the time domain, a conversion from frequency to time is performed for the tilt with a 0.25s resolution to accurately capture the wave movement. To represent the annual variations of the sea conditions, the sea spectrum varies hourly depending on the wind speed value, using the same approach as in the previous chapter. The algorithm followed for the calculation is described in Algorithm 1.

11.2.3. OPTOELECTRICAL

Once the time series data for tilt measurements at each position along the beam is collected, the optoelectrical modeling phase begins. As varying tilts will affect the incident irradiance on the PV modules, the initial step consists of calculating this irradiance in [Subsection 11.2.3](#). [Equation 11.2.3](#) will continue by implementing the most suitable module temperature model from the literature. Finally, the power output and system losses can be obtained using the approach in [Equation 11.2.3](#).

INCIDENT IRRADIANCE

The change in tilt affects the incident plane-of-array irradiance G_{POA} of a PV module [$W m^{-2}$], which was obtained with the isotropic Perez model [\[326\]](#) implementation of *pvlb* [\[178\]](#). It consists of three components: the direct G_{dir} , diffuse G_{diff} , and ground-reflected G_{ref} irradiances. Their relation is expressed in [Equation 11.19](#).

$$G_{POA} = G_{dir} + G_{diff} + G_{ref} \quad (11.19)$$

These variables depend on the various components of solar irradiance: Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI), and Direct Normal Irradiance (DNI). GHI represents the total irradiance received by a horizontal surface. This irradiance can be split into direct and scattered components. The direct component DNI is the irradiance that arrives straight from the sun to a surface normal to it. The sun altitude α_s tackles this difference in angle between the plane normal to the sun and the horizontal one. The diffuse component DHI represents the irradiance received by the horizontal surface scattered due to clouds and particles in the atmosphere. These three components are related as in [Equation 11.20](#).

$$GHI = DHI + DNI \cdot \sin \alpha_s \quad (11.20)$$

MODULE TEMPERATURE

One key advantage of OFPV, leading to increased energy yield and system lifetime, is the lower module temperature caused by higher wind speeds and the heat sink behavior of the sea [\[370\]](#). The operating efficiency of PV modules is higher when they function at a lower temperature.

The increase in yield depends on the local conditions and the system design [\[348\]](#). This is reflected in the wide variety of results obtained by several researchers, in which the maximum difference in module temperature between FPV or land-based PV can range from 1.4°C [\[444\]](#) to 12°C [\[386\]](#). Considering the impact of module temperature on PV power output and the varying results documented in the literature, special attention needs to be paid to modeling this quantity.

In literature, the most commonly used model to estimate the temperature of FPV systems is pvsyst's module temperature model [\[257\]](#) (see [Equation 11.21](#)). This model is derived from the heat balance proposed by Faiman [\[235\]](#) and its success lies in its simplicity.

Algorithm 1 Algorithm for conducting the structural and optoelectric analysis of the floating PV system.

Input Data: Wind speed U_w updated every hour. Consequently, the algorithm is repeated every hour. Irradiance and temperature updated every 10 minutes.

function CALCULATE THE OCEAN-WAVE INPUT FOR THE HOUR(U_w)

$$U_w \rightarrow (H_s, T_p) \rightarrow S(\omega)$$

Calculate the values of significant wave height H_s and peak time-period T_p from the wind speed U_w by interpolating the expected wave characteristics for fully developed seas with various wind velocities [410].

Calculate a JONSWAP spectrum [403] for these values of H_s and T_p to describe the wind-driven sea state. This spectrum is the power-spectral density function $S(\omega)$ [$\text{m}^2 \text{s}$], which measures the intensity of each wave frequency in the sea state.

end function

function COMPUTE THE TRANSVERSE DEFLECTION IN THE PV PLATFORM($S(\omega)$, RAO)

$$\eta(\mathbf{x}, t) = \sum_{j=1}^{j=n} \text{RAO}(\mathbf{x}, \omega_j) \sqrt{2S(\omega_j)\Delta\omega} \exp(i(k_j\mathbf{x} - \omega_j t + \alpha_j))$$

Compute the transverse deflection time series from the input spectrum and RAO using inverse Fourier transform. Evaluate the time series for 1 h at the resolution of $\Delta t = 0.25 \text{ s}$. Here, k_j is the wave-number for frequency ω_j , calculated using the dispersion relationship, while α_j is the uniformly randomized initial phase.

end function

function GET TILT TIME-SERIES AT EACH PANEL LOCATION($\eta(\mathbf{x}, t)$)

$$\theta(\mathbf{x}, t) = \frac{\partial}{\partial x} (\eta(\mathbf{x}, t))$$

Compute slope time series at each location on the PV platform by taking the derivative of the deflection time series.

end function

function STRUCTURAL ANALYSIS($\theta(\mathbf{x}, t)$, $\eta(\mathbf{x}, t)$)

Use the tilt and deflection time series to compute the bending moment, shear force, and longitudinal stress within the platform.

end function

function OPTOELECTRICAL ANALYSIS($\theta(\mathbf{x}, t)$, GHI, T_{amb})

Use the tilt time series, along with the global horizontal irradiance (GHI) and ambient temperature T_{amb} , to compute the module temperature and the power produced by the modules.

end function

$$T_C = T_{amb} + \frac{\alpha \cdot G_{POA} \cdot (1 - \eta_{mod})}{U_c + U_v \cdot ws} \quad (11.21)$$

Where T_C is the cell temperature [°C], which will be used as module temperature T_M , T_{amb} is the ambient temperature [°C], α is the absorption coefficient [-] (the default value of 0.9 will be used in this work), η_{mod} is the module power efficiency [-], U_c is the combined heat loss factor coefficient [$\text{Wm}^{-2}\text{K}^{-1}$], U_v is the combined heat loss factor coefficient affected by wind [$\text{Wsm}^{-3}\text{K}^{-1}$], and ws is the wind speed measured at a 10 meter height [ms^{-1}].

An overview of the U_c and U_v values of FPV systems in the literature has been provided in the review by Micheli [370, Table 1]. Considering the similarity in conditions and the OFPV system design, the study by Dörenkämper et al. [350] has been selected as a reference.

Dörenkämper et al. extracted the temperature coefficient of 5 FPV systems: 2 in the Netherlands and 3 in Singapore. Amongst these systems, the one denoted as NL2 is the most similar to this work. This system is closed (meaning that the back surface of the PV modules is not directly exposed to the water surface), has a high footprint, can be categorized as a pontoon made of HDPE or similar material, and is located in the Netherlands. For this system, U_c is $25.2 \text{ Wm}^{-2}\text{K}^{-1}$ and U_v is $3.7 \text{ Wsm}^{-3}\text{K}^{-1}$.

POWER OUTPUT

The power output of a PV module may deviate from the value indicated on the datasheet, which is specified under standard test conditions (STC). In operating conditions, the incident irradiance, temperature, and wind speed experienced by the PV module may vary. As a result, the power output will differ from the STC value. This environmental effect can be modeled by considering how the main electrical characteristics of a PV module are impacted. Therefore, Eqs. 20.4-20.5 and 20.21-20.22 from [261, ch. 20] have been implemented to model the effect of irradiance and temperature on the open circuit voltage V_{OC} [V], and short circuit current I_{SC} [A].

The power produced by each module is the maximum power point (MPP) of the current-voltage (IV) curve. This curve can be modeled using the ideal one-diode model expressed in Equation 11.22 [261, ch. 20]. This equation ignores the effect of the series and shunt resistances.

$$I = I_L - I_0 \cdot (\exp v - 1) \quad (11.22a)$$

$$v = \frac{V}{V_{th} \cdot n \cdot N_S} \quad (11.22b)$$

where V_{th} is the thermal voltage equal to 25.85 mV and N_S is the number of cells in the module [-]. The unknowns of this equation, which are the light-generated current I_L [A], the dark saturation current I_0 [A], and the ideality factor n , can be found by evaluating the curve in the three points reported by the module manufacturer: the open circuit, the short circuit, and the MPP points.

$$I_{SC} = I_L \quad (11.23a)$$

$$0 = I_L - I_0 \cdot (\exp v_{OC} - 1) \quad (11.23b)$$

$$I_{MPP} = I_L - I_0 \cdot (\exp v_{MPP} - 1) \quad (11.23c)$$

By rearranging these equations, one can express the voltage as a function of the current relying only on the known parameters, obtaining Equation 11.24a. The ideality factor n is found by iteratively solving Equation 11.24b. This curve is computed for each series-connected module, and the voltages are added up to obtain the system's IV curve. By finding the MPP point of this curve, one can estimate the DC power output of the system P_{sys} [W].

$$V = \frac{n \cdot N_S \cdot k_B \cdot T_{STC}}{q} \cdot \ln \left(\left(1 - \frac{I}{I_{SC}} \right) \cdot (\exp v_{OC} - 1) + 1 \right) \quad (11.24a)$$

$$I_{MPP} = I_{SC} \cdot \left(1 - \frac{\exp v_{MPP} - 1}{\exp v_{OC} - 1} \right) \quad (11.24b)$$

When each module operates at its MPP, like when connected to an optimizer with an MPP Tracker (MPPT), the system's DC power output is the sum of P_{MPP} of each module. Due to the high cost associated with this approach, it is common to connect the PV modules in series before pairing them with an inverter. Since the MPP conditions of each module will differ due to the varying incident irradiance, not all modules will operate at ideal conditions and the system DC power output will be lower than when using the MPPT strategy. The losses associated with this phenomenon are the power mismatch losses, which in the context of this research, are caused by the different tilts and orientations of the series-connected modules positioned along the structure.

By comparing the produced power output by the series-connected floating system $P_{floating}$ to the power produced by the system when each module is connected to an MPPT P_{MPPT} , the wave-induced mismatch power loss L_{mis} over a period of time T can be calculated using Equation 11.25.

$$L_{mis} = \left(1 - \frac{\int_0^T P_{floating}(t) dt}{\int_0^T P_{MPPT}(t) dt} \right) \cdot 100\% \quad (11.25)$$

Similarly, one can quantify the power loss due to the movement of the waves over a time period T as in Chapter 10, where the power produced by a system subjected to waves is compared to that produced by a horizontal one P_{hor} . P_{hor} is P_{MPPT} with all modules placed at a tilt of 0° . These three scenarios are graphically represented in Figure 11.3. The expression of the movement power loss is that of Equation 11.26.

$$L_{mov} = \left(1 - \frac{\int_0^T P_{MPPT}(t) dt}{\int_0^T P_{hor}(t) dt} \right) \cdot 100\% \quad (11.26)$$

The optoelectrical formulation was numerically implemented in Python using the PV library *pvlb* [178]. The library has been extensively applied in several works and validated under different conditions [445–447], ensuring the accuracy of the optoelectrical methodology.

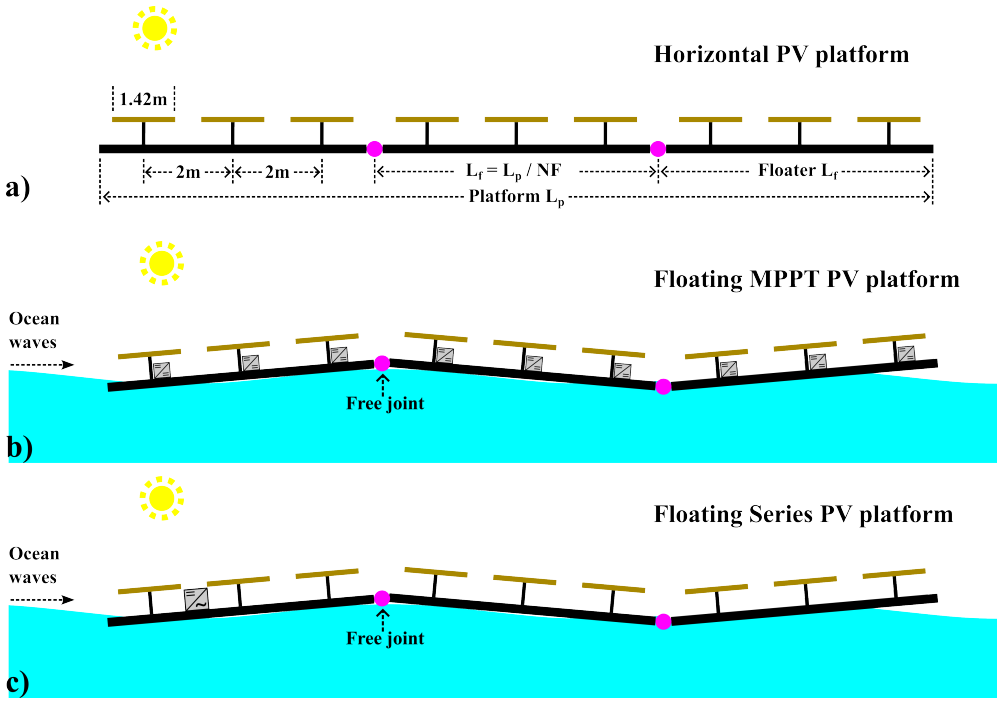


Figure 11.3: Schematic of the floating PV platform. a) Horizontal PV platform not subjected to the movement of waves. b) Floating PV platform, with individual optimizers containing an MPP tracker and a DC-DC converter for each PV module. c) Floating PV platform with all the modules connected in series and a single central inverter.

11.3. CASE STUDY DESCRIPTION

This section presents a hypothetical design of the floating PV platform. The proposed platform has a length of $L_p = 100\text{ m}$. It is constructed using High-density polyethylene (HDPE). For our analysis, we consider that HDPE has density $\rho_{HDPE} = 952\text{ kg m}^{-3}$, Young's modulus $E_{HDPE} = 500\text{ MPa}$, and tensile yield strength of 27 MPa . The platform is divided into individual identical floaters, interconnected with free joints, as shown in Figure 11.3. The weight of the auxiliary connectors is ignored in the present analysis.

GA-F200T c-Si PV modules [448] are selected due to their flexibility which facilitates their placement on a floating structure without impeding the structure's deformation or adding any resistance. The weight of the PV modules is negligible compared to that of the platform, even if they were not lightweight modules.

The PV modules are placed horizontally on top of the floating platform. This decision is made to ensure that the mechanical stability of the platform is not compromised by wind load, although optimal production would require the PV modules to be tilted. The modules are placed in a landscape orientation relative to the platform at a regular spacing of 2 m between the centers of the panels. Therefore, a total of 50 modules are connected in series. Since the modeling is done in two dimensions, the sea waves income parallel to the string of modules to emphasize the mismatch losses experienced by the modules.

We examine several scenarios, varying the number of floaters, the thickness of the floaters,

and the rigidity of the floaters. We first examine the case of floaters having a solid rectangular cross-section, assumed to be a beam thickness of $h_b = 0.20$ m, denoted as case PLA. The 100 m long platform is divided into a number of floaters, ranging from a single floater $NF = 1$ of length $L_f = 100$ m (PLA-NF01), to $NF = 50$ of length $L_f = 2$ m each (PLA-NF50). The platform characteristics for PLA are summarised in Table 11.1. The platform properties form the required input for the FE-FSI model.

Table 11.1: Properties for the base case PLA.

Property	Symbol	Value
Floater beam thickness	h_b	0.20 m
Cross-section Moment of Inertia per unit width	$I = \frac{1}{12} h_b^3$	$6.667 \times 10^4 \text{ m}^3$
Young's Modulus	$E = E_{HDPE}$	500 MPa
Mass per unit area (incl. PV modules)	m	$192.956 \text{ kg m}^{-2}$
Characteristic hydro-elastic length Equation 11.28	λ_c	15.076 m
Floater orientation	A_z	180° (South)

The analysis is conducted for the PV platform floating in an offshore environment with a water depth $d = 30$ m. The temperature and wind data were obtained for a location in the North Sea near the coast, with coordinates 53°24'42"N, and 6°11'57"E for the year 2017. The data was taken from the KNW KNMI North Sea Wind atlas [442], which is a public re-analysis dataset of meteorological measurements. Specifically, we have taken the dataset named *KNW-CSV-TS-UPDATE* at the ID 064-116. The GHI was not measured offshore thus it was taken from the nearest ground station, 160 km from the offshore location. The data was sourced from KNMI dataset of automatic weather stations in the Netherlands [449]. Specifically, the dataset *Actuele10mindataKNMIstations* at station-ID 277. The wind data is updated hourly and serves as the input for obtaining the hourly JONSWAP ocean-wave spectrum. The wave-spectrum input along with the complex-valued hydro-elastic response from the FE-FSI model will then provide the tilt of each panel at a resolution of 0.25 s, as summarised in Subsection 11.2.2. The temperature and irradiance data are updated every 10 minutes. These are the inputs for the optoelectrical model. These environmental parameters are interpolated to match the 0.25 s resolution of the tilt using cubic spline interpolation. Applying the BRL model [417], the components of DNI and DHI are derived from the GHI. Despite the weather data not originating from an offshore location, it is presumed that the irradiance profiles observed along the coast are comparable to those encountered in offshore settings, providing a reasonable representation.

11.4. NUMERICAL RESULTS: BASE CASE

11.4.1. SETUP OF THE NUMERICAL DOMAIN

The structural analysis of the floating PV platform is carried out using the monolithic FE-FSI model from Subsection 11.2.1. The presented FE model is capable of solving 2D and 3D problems, as demonstrated in [433]. However, in this work, we restrict ourselves to a 2D analysis in the vertical plane. We consider a rectangular numerical domain, with a constant water depth of d and total length of $5.5L_p$, spanning between $x \in [-4L_p, 1.5L_p]$, where L_p is the length of the floating platform. The floating PV platform is placed in $x \in [0, L_p]$. The wave input is prescribed using the inlet boundary Γ_{in} . We additionally place a wide damp-

ing zone between $x \in [-4.0L_p, -0.5L_p]$ for absorbing any waves reflected by the floating PV platform.

This numerical domain is discretized using 4th order rectangular elements, with constant $\Delta x = 0.01L_p$, while the vertical mesh size varies exponentially from $0.48d$ near the bottom to $0.04d$ near the free-surface, to capture the high-gradients near the free-surface. The choice of 4th order polynomial basis functions enables us to use coarser mesh size and also allows us to compute the bending moments from the beam deflection. As per the convergence analysis from [433], the selected mesh size is notably smaller than necessary for capturing the beam deflection. However, our choice is dictated by the necessity to accommodate the highest wave frequency with non-zero amplitude in the chosen wave spectrum.

11.4.2. PLATFORM RESPONSE TO MONOCHROMATIC WAVES

The above-mentioned numerical setup is used to calculate the complex-valued response of the floating PV platform. We first investigate the case PLA-NF01 (see Section 11.3), which is made of a single floater of length $L_f = L_p = 100$ m. We study the response of PLA-NF01 to wave-frequency $\omega = 1.5 \text{ rad s}^{-1}$ and wave-amplitude $\kappa_0 = 1$ m. Figure 11.6a-b present the solution obtained from the FE-FSI model for the beam deflection η , free-surface elevation κ and the velocity potential ϕ in the fluid domain. It should be noted that the FE model returns a complex-valued solution for all unknowns, thus containing information about both the amplitude of the response and the phase of the response.

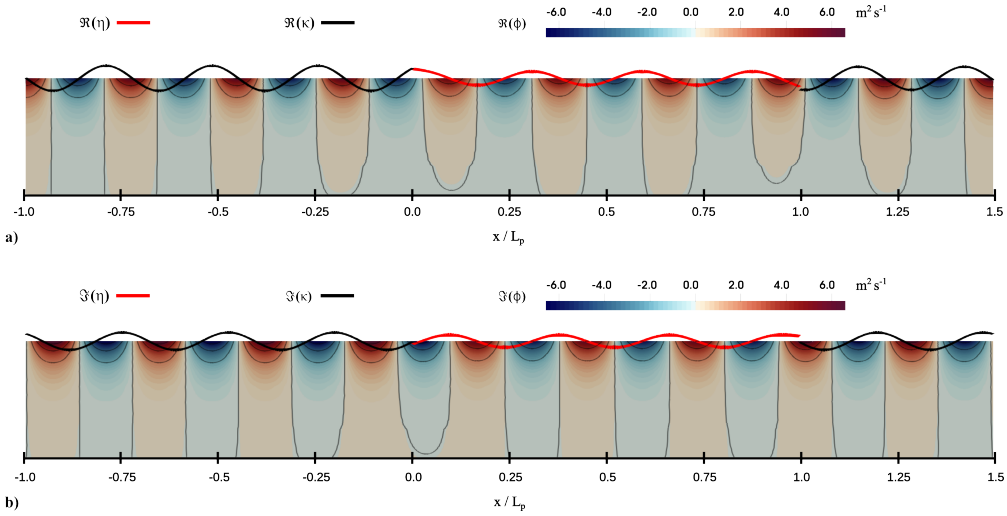


Figure 11.6: Contour plot of ϕ and line plots of κ and η , obtained from the FE-FSI model, due to the interaction of PLA-NF01 platform with ocean wave of frequency $\omega = 1.5 \text{ rad s}^{-1}$. a) Real values of the solution. b) Imaginary values of the solution.

For the purpose of the floating PV analysis, we are primarily interested in the deflection η of the beam. In Figure 11.4, we present the magnitude of the complex-valued solutions for deflection, slope, bending moment, and shear force, along the length of the platform. The results are presented for two wave excitation frequencies $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$. These frequencies were selected to demonstrate the significant impact that excitation fre-

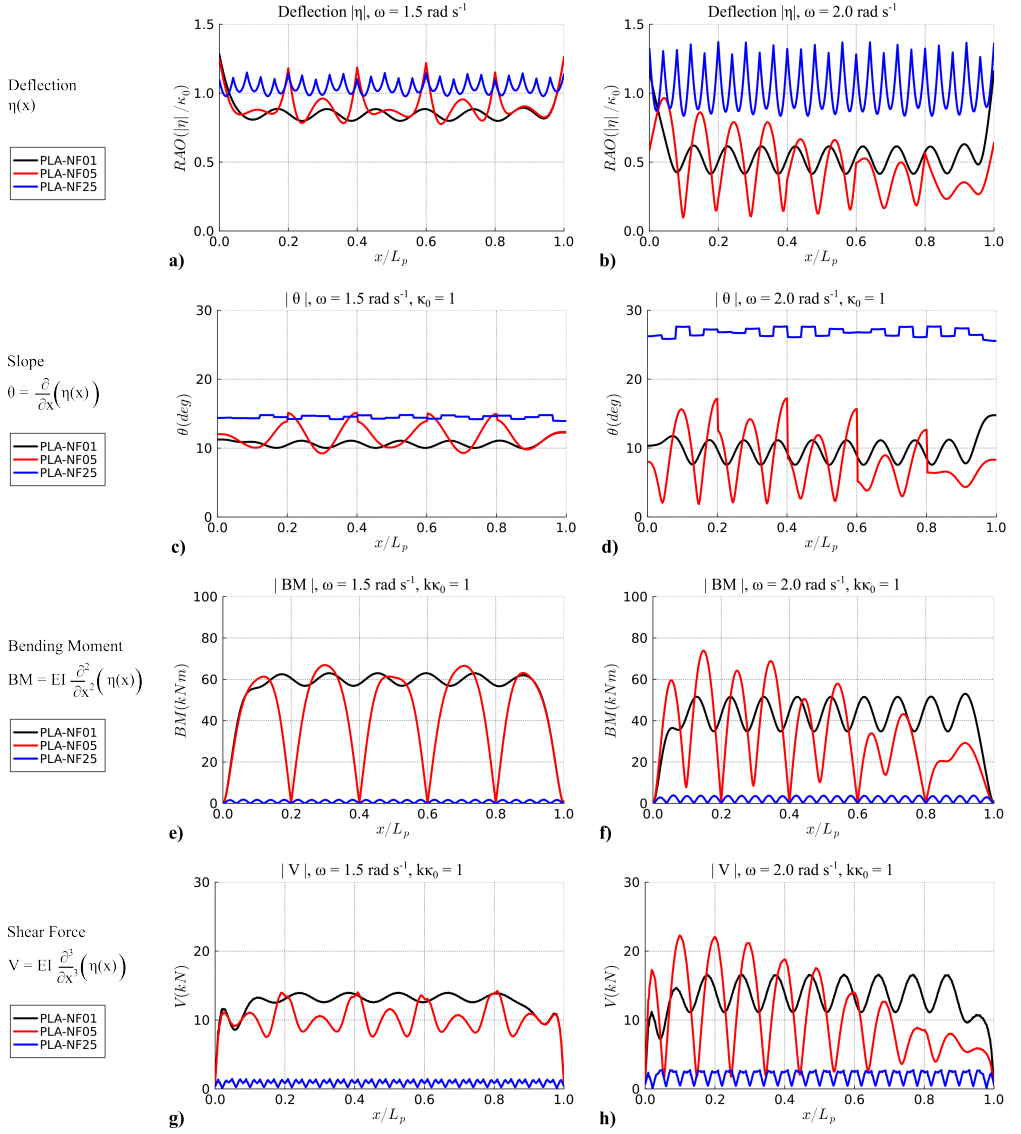


Figure 11.4: Structural response of PLA-NF01, PLA-NF05 and PLA-NF25 floating platforms to monochromatic ocean-wave excitation of $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$. a-b) Magnitude of beam deflection and c-d) magnitude of tilt, along the length of the platform. e-f) Magnitude of bending moment (BM) and g-h) magnitude of shear force in the cross-section, per unit beam-width for unit-slope wave excitation.

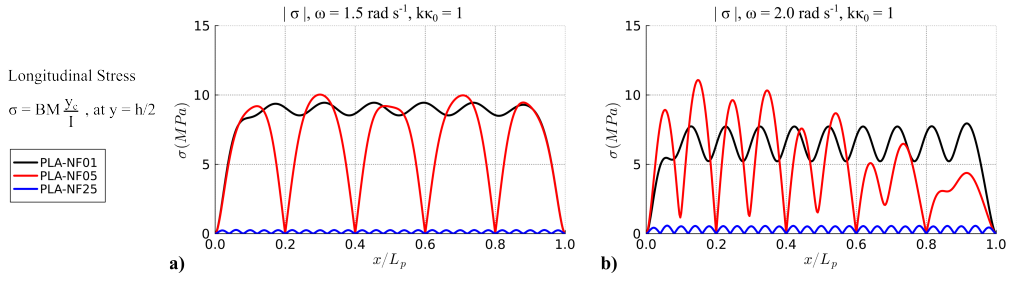


Figure 11.5: Structural response of PLA-NF01, PLA-NF05 and PLA-NF25 floating platforms to monochromatic ocean-wave excitation of $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$. a-b) Magnitude of longitudinal stress in the cross-section, at $y_c = h_b/2$. Here y_c is the normal distance between the rotation axis for bending and the location on the cross-section.

quency can have on the structural response. In each sub-figure of Figure 11.4, we compare the results for cases PLA-NF01, PLA-NF05, and PLA-NF25 (see Section 11.3), thus showing the influence of changing the number of floaters.

Let us first analyze the response for the single floater case PLA-NF01. The solution in Figure 11.4a-b highlights that the magnitude of deflection of the beam is maximum at the endpoints. Secondly, the magnitude of the motion is larger for the lower frequency $\omega = 1.5 \text{ rad s}^{-1}$ compared to $\omega = 2.0 \text{ rad s}^{-1}$. However, in both cases, the response in the leading and the trailing portion of the platform is quite similar.

Next, we study the response for the 5-floater case PLA-NF05. In this case, the free joints connecting the floaters are located at $x \in \{0.2L_p, 0.4L_p, 0.6L_p, 0.8L_p\}$. In Figure 11.4a-b, we observe distinct cusps at these locations, along with local peaks in the magnitude of the deflection. This highlights that the heave motion of the platform is maximum at the endpoints and the free joints. Once again, the overall response of the platform PLA-NF05 is lower for the frequency $\omega = 2.0 \text{ rad s}^{-1}$ compared to $\omega = 1.5 \text{ rad s}^{-1}$. However, for $\omega = 2.0 \text{ rad s}^{-1}$, we can notably observe that the magnitude of the response is larger in the upstream portion of the platform compared to the downstream portion. This is due to the partial reflection of the incoming waves by the platform, resulting in a gradual decrease in the wave energy propagating along the length of the platform.

Finally, we consider the case PLA-NF25 with 25 floaters. Figure 11.4a-b once again highlights cusps and local peaks at the joints, consistent with the previous observation. However, here the response for $\omega = 2.0 \text{ rad s}^{-1}$ is notably higher compared to $\omega = 1.5 \text{ rad s}^{-1}$. Finally, there is no notable decay in the response in the downstream section of the platform, an indication of little reflection of the incoming wave energy for this frequency.

TILT, BENDING MOMENT, SHEAR FORCE AND LONGITUDINAL STRESS

The complex-valued solution of deflection η can be used to calculate the tilt, bending moment, shear force, and longitudinal stress within the floater beams. The expressions for the local tilt (θ) of the floater beam, the bending moment (BM), the corresponding shear force (V) in the cross-section, and the longitudinal stress (σ) in the cross-section are expressed in Equation 11.27. Here y_c is the normal distance between the rotation axis for bending and the location on the cross-section.

$$\theta = \eta_x \quad (11.27a)$$

$$BM = EI \eta_{xx} \quad (11.27b)$$

$$V = EI \eta_{xxx} \quad (11.27c)$$

$$\sigma = BM \frac{y_c}{I} \quad (11.27d)$$

Figure 11.4c-d present the magnitude of the complex-valued solution for slope (or tilt) along the length of the platform, for $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$. Firstly, it is notable that for cases with joints, such as PLA-NF05 and PLA-NF25, a discontinuity in the slope is evident across the free joints. This behavior is expected, as the joints offer no bending resistance. Next, for PLA-NF01 and PLA-NF05, we note that the slope varies continuously within each floater, indicating the presence of an elastic response for the floater beam. In contrast, the plots for PLA-NF25 show a nearly constant slope within each platform. This indicates that the response of PLA-NF25 for both $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$ is not elastic.

In Figure 11.4e-f we report the magnitude of the complex-valued solution for bending moment along the length of the platform for $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$. The bending moment is reported assuming unit width platform and unit-slope wave input $k \kappa_0 = 1$.

Here k is the wave number for ocean-wave frequency ω , obtained from Equation 11.4. For the corresponding cases, Figure 11.4g-h report the magnitude of shear force $|V|$ and Figure 11.5 reports the magnitude of longitudinal stress σ at $y_c = h_b/2$ in the cross-section of the floater beam. We obtain zero bending moment at the end points of the platform because of the free-edge boundary condition. Also, given our choice of free joints, we obtain zero bending moment at the joints for cases PLA-N05 and PLA-N25. The structural response for PLA-N01 and PLA-N05 exhibit significant bending moment, shear force, and longitudinal stresses, indicating strong elastic behaviour. On the other hand, the 25-floater PLA-NF25 is observed to have a relatively small bending moment, indicating a largely rigid body response. Further, the bending moment and stresses for case PLA-NF05 once again highlight the gradual decay in the structural response along the length, owing to the partial reflection of the incoming waves by the platform.

Table 11.2: Mean of the magnitude of the shear-load on the joints.

Case	$\omega = 1.5 \text{ rad s}^{-1}$	$\omega = 2.0 \text{ rad s}^{-1}$
PLA-NF05	13.45 kN	15.74 kN
PLA-NF25	0.83 kN	2.39 kN

The shear force plot in Figure 11.4g-h also provides us with information about loads on the joints connecting the floaters in the platform. PLA-NF05 consists of 4 joints, at $x/L_p = \{0.2, 0.4, 0.6, 0.8\}$. The magnitude of shear load on these joints is given by the value of plot at these four locations. In Table 11.2, we report the mean of the load at these 4 joints. Similarly, we can obtain the magnitude of the loads at the 24 joints in PLA-NF25. The mean of these loads at the 24 joints of PLA-NF25 is also listed in Table 11.2 for the tested frequencies $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$.

11.4.3. RESPONSE FOR A WIDE RANGE OF MONOCHROMATIC WAVES

In the previous section, we restricted the analysis to just two excitation frequencies. In this section, we repeat the analysis for a wide range of excitation frequencies to better understand the differences in the total structural response for 1, 5, and 25 floater cases.

We repeat the analysis for $\omega \in [0.5, 5.0]$, which corresponds to ocean-wave length $\lambda \in [2.5, 190] \text{ m}$ for the assumed water-depth $d = 30 \text{ m}$. This results in the contour plot Figure 11.7a,c,e for PLA-NF01, PLA-NF05 and PLA-NF25, respectively, which are analogous to Figure 11.4a. In these contour plots, the x-axis is the position along the platform, the y-axis is the excitation frequency and the contour levels are for the magnitude of the complex-valued solution of beam deflection η . The arrows in the plots highlight $\omega = 1.5 \text{ rad s}^{-1}$ and $\omega = 2.0 \text{ rad s}^{-1}$, which were studied in the previous section. Similar contour plots are prepared for the calculated slope response, bending moment, and shear force. The contour plots Figure 11.7b,d,f present the magnitude of slope for unit-magnitude wave excitation. The contour plots Figure 11.8a,c,e present the magnitude of bending moment for unit-slope wave excitation. Finally, the contour plots Figure 11.8b,d,f present the magnitude of the shear force for unit-slope wave excitation. Note that the contour levels for PLA-NF25 are different compared to PLA-NF01 and PLA-NF05 across Figure 11.7 and Figure 11.8.

There are certain common observations across the plots, 1) the deflection magnitude is highest at the endpoints of the platform and at the free joints, 2) the slope is discontinuous

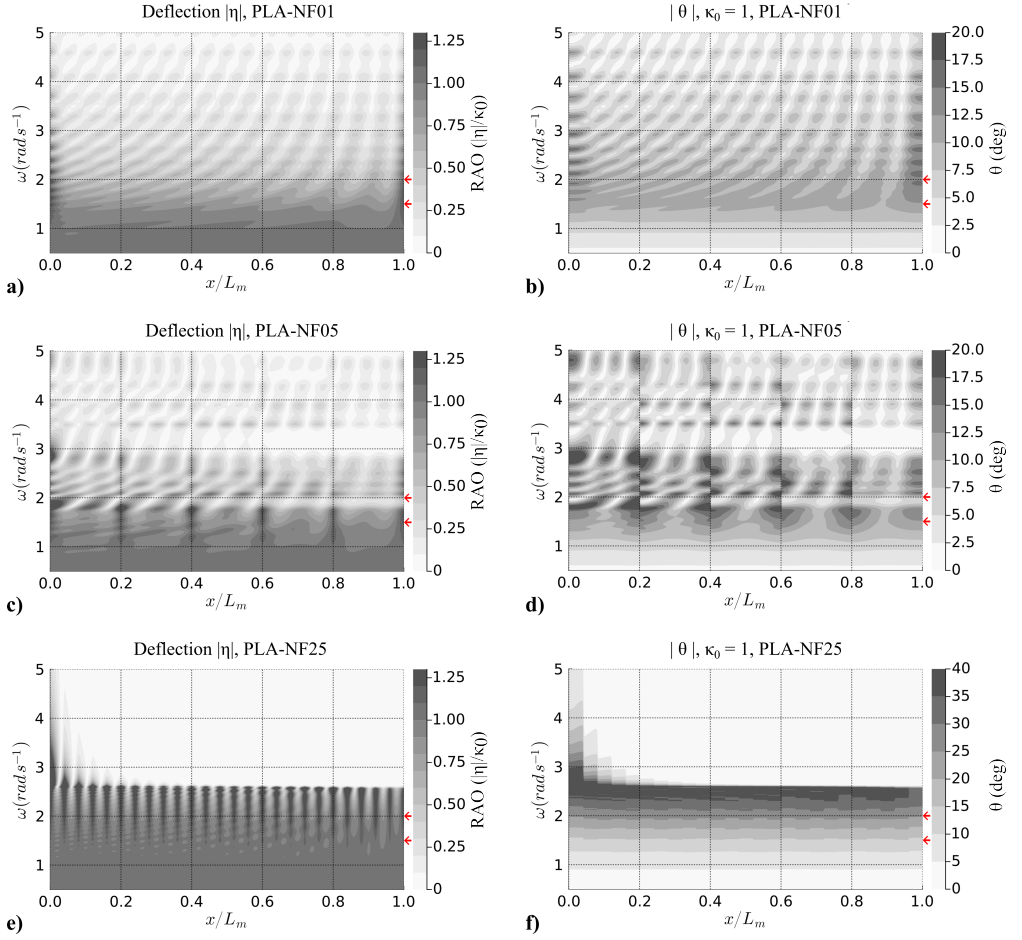


Figure 11.7: Structural response along the length of the platform for ocean-wave frequency in range $\omega \in [0.7, 5.0]$. a-b) Magnitude of deflection and slope for PLA-NF01. c-d) Magnitude of deflection and slope for PLA-NF05. e-f) Magnitude of deflection and slope for PLA-NF25. The arrows indicate the two frequencies that were plotted in Figure 11.5.

at the free joints, 3) the bending moment and longitudinal stress are zero at the endpoints and the free joints.

To explain the specific observations for each case, we recount the classification of the global response of floating structures in Figure 11.9, based on the works [450, 451]. The response can be classified using two ratios, length of floater vs ocean-wave length L_f/λ , and length of floater vs characteristic hydro-elastic length L_f/λ_c . Here the characteristic hydro-elastic length λ_c is defined as Equation 11.28 [450, 451]. For the structural properties of PLA, given in Table 11.1, the characteristic hydro-elastic length $\lambda_c = 15.076$ m.

$$\lambda_c = 2\pi \left(\frac{EI}{\rho g} \right)^{1/4} \quad (11.28)$$

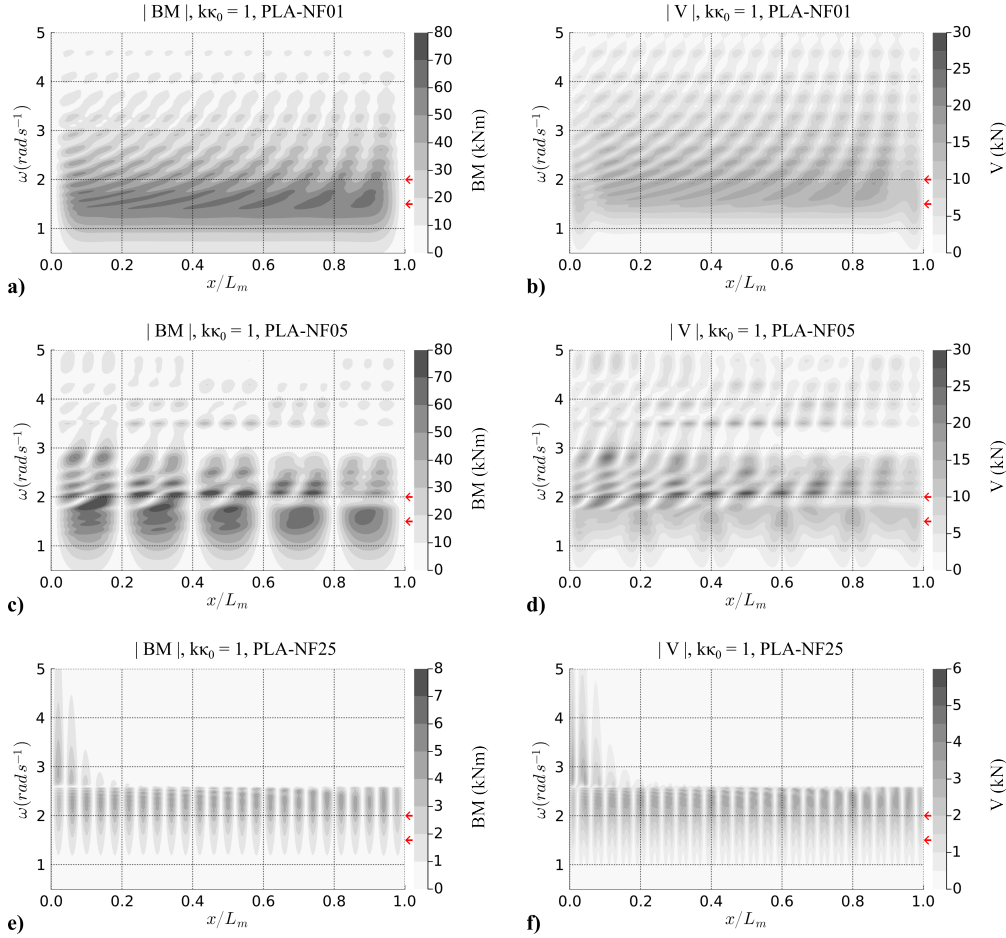


Figure 11.8: Structural response along the length of the platform for ocean-wave frequency in range $\omega \in [0.7, 5.0]$. a-b) Magnitude of bending moment and shear force for PLA-NF01. c-d) Magnitude of bending moment and shear force slope for PLA-NF05. e-f) Magnitude of bending moment and shear force for PLA-NF25. The arrows indicate the two frequencies that were plotted in Figure 11.5.

In Figure 11.9, we have indicated the response regime in which the 1, 5, and 25 floater cases are situated. For case PLA-NF01, the length of the floaters is the longest, i.e., both L_f/λ_c and L_f/λ are large. This indicates a very flexible floating structure response, which is evident in the results. Figure 11.8a shows a significant bending moment across the length of the platform and for all frequencies, thus highlighting the presence of the elastic response. This elastic behaviour however provides a stiffness against the wave excitation, thus resulting in smaller motions, as seen in lower deflection magnitudes in Figure 11.7a. Therefore, PLA-N01 is observed to have lower motions, but higher structural loads.

On the other hand, PLA-N25 with 25 floaters has the smallest values of L_f/λ_c and L_f/λ , suggesting a rigid body response. The deflection contour in Figure 11.7e highlights that the motion of PLA-N25 is larger compared to PLA-N01. This motion of PLA-N25 is partic-

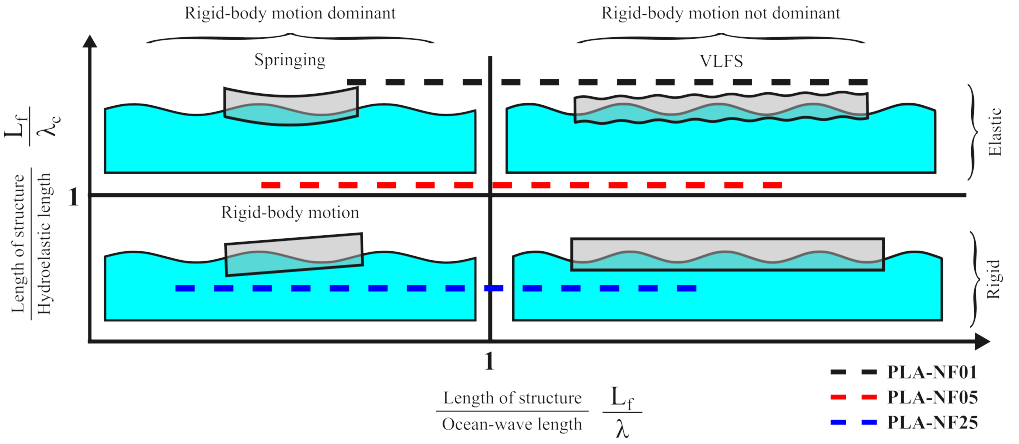


Figure 11.9: Classification of the global response of floating structures [450]. The lines indicate the regime for PLA-NF01, PLA-NF05 and PLA-NF25.

ularly large within the critical wave frequency range of $[0.5, 3] \text{ rad s}^{-1}$, which encompasses the majority of fully developed wind-generated sea-states. The contour plot in Figure 11.8e shows that the bending moment for PLA-N25 is at least one order of magnitude lower than PLA-N01 and PLA-N25. This indicates that a reduced elastic response leads to overall larger motions in the platform. Finally, we also observe that the leading floaters have higher magnitudes of motion, especially for high frequencies. This suggests that the leading floaters act as sacrificial barriers and reflect the incoming wave energy for the high frequencies. Overall, the PLA-N25 platform is expected to have large motion but smaller structural loads.

The case PLA-N05 has intermediate values of L_f/λ_c and L_f/λ . Figure 11.7c and Figure 11.8c show a larger magnitude of deflection and bending moment in PLA-NF05 compared to PLA-NF01. Hence this design choice would have worse motion and structural loads compared to PLA-NF01. Furthermore, the plots highlight a gradual decay of the response along the length of the platform due to successive partial reflection of the incoming wave energy.

Lastly, we examine the loads on the joints for cases PLA-NF05 and PLA-NF25. In Table 11.2, we provide the average magnitude of the loads across the joints, for two excitation frequencies. In Figure 11.10, we present the average magnitude of loads for excitation frequency $\omega \in [0.5, 5.0]$. We also present the maximum magnitude of loads, among the joints. In general, we note that the PLA-NF25 has significantly lower loads on joints compared to PLA-NF05. Secondly, for low frequencies, the mean and maximum of the loads across joints is quite similar. However, for higher frequencies, notable differences emerge in the mean and maximum joint loads. This disparity arises due to the increasing reflection of higher frequencies along the length of the platform, resulting in higher loads in the upstream portion and lower loads in the downstream portion. This phenomenon is particularly pronounced in the case of PLA-NF25 for $\omega > 2.7 \text{ rad s}^{-1}$, where the majority of the structural response is borne by the upstream floaters, leading to large loads on the upstream joints.

In summary, based on these observations, we can infer that increasing the number of floaters amplifies the platform motion. However, the structural loads exhibit a non-monotonic trend. This occurs because increasing the number of floaters results in a shorter

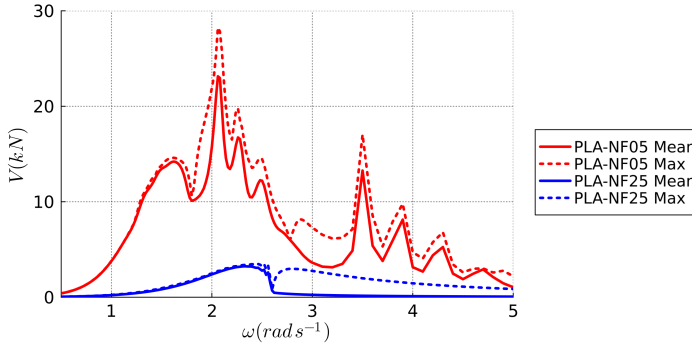


Figure 11.10: Maximum and average values of the magnitude of shear force $|V|$ at the joints of cases PLA-NF05 and PLA-NF25, for unit beam-width and unit-slope wave excitation.

length for each floater, causing the structural response to shift from elastic to rigid body behaviour. Since structural loads depend on both the extent of the elastic response and platform motion, a non-monotonic trend emerges. Therefore, there is a trade-off in selecting the optimal number of floaters.

Finally, we would like to use these results to highlight the significance of the hydro-elastic natural frequency of the floating system. In Figure 11.7, we can see that the response of PLA-NF25 peaks around $\omega = 2.8 \text{ rad s}^{-1}$, response of PLA-NF05 peaks around $\omega = 1.95 \text{ rad s}^{-1}$ and the response of PLA-NF01 peaks around $\omega = 1.5 \text{ rad s}^{-1}$. These frequencies correspond to the natural hydro-elastic frequencies of the floaters, which are inversely proportional to the length of the floater. In the interest of brevity, we refrain from delving into the natural frequencies of the floating elastic system in this work. However, the readers are directed to [452, 453] for the discussions on the natural frequencies of such a floating elastic system.

11.4.4. TILT DISTRIBUTION

As mentioned in Subsection 11.2.2, the tilt is the link between the structural and the opto-electrical models. Figure 11.11a shows the standard deviation of the tilt along the platform for PLA with 1, 5, and 25 floaters. As expected from the previously presented results, the variation in tilt is larger for a higher number of floaters. In the case of 5 floaters, the highest deviation is experienced at the hinges of the platform and the lowest one at the center of the platform. In contrast, for the case of 25 floaters, the location of the hinges is not appreciated. The highest deviation is experienced at the beginning and the lowest one at the end of the platform. This is due to the already-mentioned effect of wave reflection. For the single floater case, the highest deviation is experienced at the extremes but it undergoes in general the lowest deviation in tilt. The modules at such unstable positions should be replaced by fake ones or glass to observe the same mechanical behavior without compromising the power output.

After having explored the variation of the tilt along the platform, Figure 11.11b shows the distribution of tilt values over time at a specific position. One can observe again that the variation in tilt is larger for a higher number of floaters. In particular, while the tilt oscillates within a range of $\pm 10^\circ$ for a single floater, it doubles to $\pm 20^\circ$ when the platform is composed

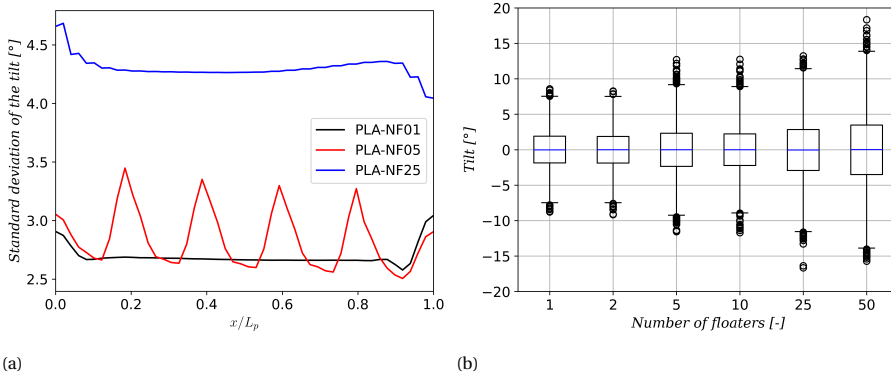


Figure 11.11: (a) Standard deviation of the tilt along the platform for PLA-NF01, PLA-NF05, and PLA-NF25. (b) Tilt values over time at $x = 0.19 \cdot L_p$ for an increasing number of floaters. Both figures considered the tilt from 9:00 to 10:00 on January 1st.

of 50 floaters. Along time, the tilt values oscillate around 0°; however, when fixing the time instant, the median values of the tilt along the platform may be positive or negative.

This greater variation in tilt for a higher number of floaters indicates that the power losses will also increase with the number of floaters. A wider variation in tilt translates into a greater range of irradiance values. This has a high influence on the power mismatch losses as the wider the range, the larger the losses.

11.4.5. POWER LOSSES

The power output of the floating PV system is computed following the approach developed in Subsection 11.2.3. To obtain a time series of the tilt and azimuth, the negative tilts are transformed into positive ones after considering that the azimuth during those instants of time would be the opposite of that which the system faces.

Table 11.3 shows the annual specific energy yield of the OFPV system with a single inverter (referred to as *floating*), the OFPV system with MPPT at the module level (*MPPT*), and the one laying horizontally not subjected to the effect of waves (*horizontal*). These systems are facing South hence they are denoted as PLS. The values have been normalized to the nominal power of the system. As expected, the floating system produces the least of the three scenarios. In particular, the energy produced is 0.7% to 2.6% lower than when all modules are laying horizontally. The MPPT scenario generates slightly less energy than the horizontal scenario, with a difference ranging from 0.1% to 0.2%, depending on the number of floaters. Production decreases as the number of floaters increases, with the floating system being more impacted than the MPPT system. The horizontal scenario is not affected by the number of floaters.

The results from Table 11.3 can be better analyzed by computing the mismatch and movement losses. Figure 11.12a shows how the mismatch losses experience an increase of 0.6% to 2.4% as the number of floaters increase, while the movement losses are almost negligible. The latter can be attributed to the fact that, on average, the oscillations are centered around 0°. Due to the wave effect, the modules will face the sun at one time instant, increasing productivity. However, at the next time instant, the modules will look away from

Table 11.3: Annual specific energy yield [kWh/W_p] for the floating, MPPT connected, and horizontal PV systems for different numbers of floaters for the PLS case.

Number of floaters [-]	1	2	5	10	25	50
Floating	1.001	1.001	1.000	0.999	0.991	0.982
MPPT	1.007	1.007	1.007	1.007	1.007	1.006
Horizontal	1.008	1.008	1.008	1.008	1.008	1.008

the sun, decreasing the production. If each module produces at its MPP, any extra power generated at some time points is offset by lower power generation at others. Similar results were reported in [Chapter 10](#).

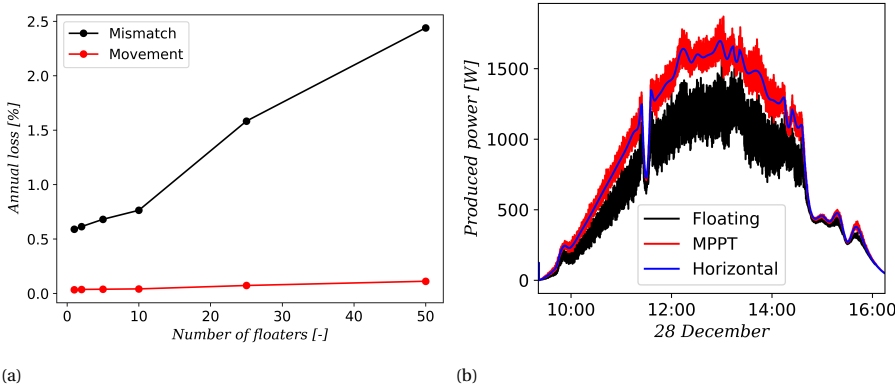


Figure 11.12: (a) Yearly mismatch and movement losses of PLS systems for an increasing number of floaters. (b) Power produced by the floating, MPPT, and horizontal PLS-NF25 systems on the day with the highest mismatch losses.

This phenomenon can be seen in [Figure 11.12b](#), which shows the power produced by the floating, MPPT, and horizontal systems on the day with the highest mismatch losses. The case of 25 floaters has been chosen as an example. The system with MPPT presents oscillations in power that are centered around the power produced by the horizontal system. With a higher magnitude of the oscillations and shifted down, one can encounter the power produced by the series-connected modules. In that same figure, the dip at around 11:30 shows the effect of a passing cloud on the power generation. The three systems produce a similar output during that cloudy instant since the mismatch losses occur mostly during sunny moments.

Three phenomena explain why December 28th is the day with the highest mismatch losses. First, it is a sunny day. Mismatch losses are higher on sunny days when there is a high portion of DNI than on cloudy ones. Out of the three irradiance components explained in [Subsection 11.2.3](#), the direct one G_{dir} , which is directly proportional to the DNI, is most affected by whether the module is facing the sun or not. Second, it is a windy day. The waves considered in this work are generated by wind. A stronger wind generates a higher tilt variation [Chapter 10](#). Third, it is a winter day. The mismatch is a relative value that depends on the power production. Since the production is higher in summer than in winter, the same losses will be relatively larger in winter than in summer.

For considerable oscillations to occur, all three phenomena must coincide simultaneously. It should be a sunny windy winter day. To showcase this, the power produced during three more days is plotted in Figure 11.13. December 10th is a windy winter day but with a low DNI component. January 22nd is a sunny winter day but with a low wind. July 9th is a sunny windy day but during the summer. For the three days, the generation by the three cases under study is equivalent.

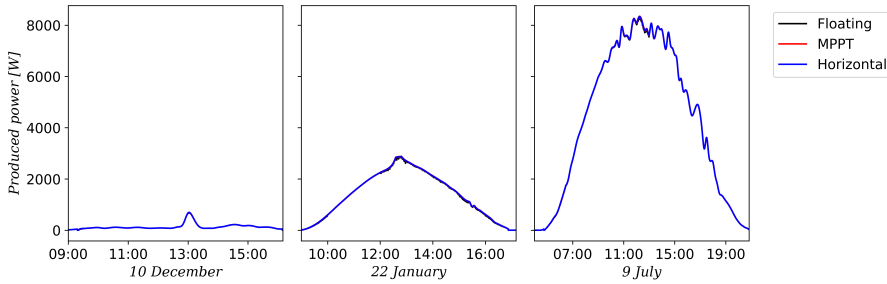


Figure 11.13: Power produced by the floating, MPPT, and horizontal PLS-NF25 systems on selected days. December 10th is a cloudy windy winter day; January 22nd is a sunny not windy winter day; and July 9th is a sunny windy summer day.

In terms of production, these findings are favorable as most of the losses will be reported during periods of low generation. These findings also add a benefit to the symbiosis between offshore photovoltaic and wind electricity, as the mismatch losses caused by a greater wind can be compensated by a higher production of wind power.

Before ending this section, we would like to comment on the tilt of the PV modules. To enhance the mechanical stability of the system, the modules have been placed horizontally on top of the platform. However, the optimal tilt for the North Sea is around 34° [414]. If the system under study were installed at this inclination, the mismatch losses would be relatively lower. The magnitude of the oscillations would remain similar, resulting in equivalent power loss. However, since the overall power production would be higher, the relative power loss would be smaller.

11.5. VARIATIONS TO THE BASE CASE

In this section, we systematically alter the characteristics of the base case PLA to analyze the impact of changing structural properties, such as Young's modulus, thickness and cross-section, and panel orientation on power mismatch losses.

11.5.1. ORIENTATION

This subsection explores how the power losses are affected when the PV system faces East instead of South. Therefore, the system is denoted as PLE.

The movement and power mismatch annual losses for the two orientations are shown in Figure 11.14a. The movement losses experience the same increase from 0.1% to 0.2% for the two orientations. This indicates that orientation appears to have no appreciable impact on the annual yield in the case of the MPPT connection. The power mismatch losses stay in the range 0.6% to 2.6% for the two orientations. The losses are slightly lower for the East-facing system for platforms with 10 floaters or less, but higher for 25 floaters or more.

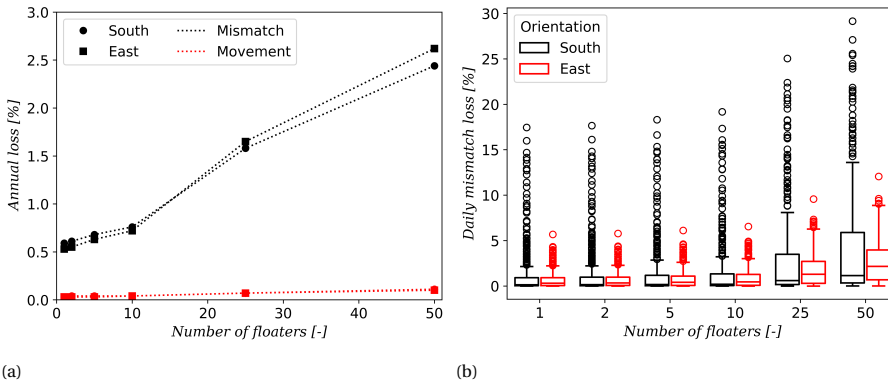


Figure 11.14: (a) Annual power losses and (b) daily mismatch for the systems facing South (PLS) and East (PLE) for an increasing number of floaters.

The distribution of mismatch loss for the two orientations, shown in the boxplots in Figure 11.14b for an increasing number of floaters, shines a light on this behavior. The South-oriented system experiences a wider range of daily mismatch values, reaching on some days higher losses than the East-oriented system. For instance, the maximum daily mismatch loss of 1 floater is close to 18% for the South-oriented system but lower than 6% for the East-oriented system. This behaviour occurs independently of the number of floaters. However, in terms of the median, the losses of the system facing East are higher, although the difference in median values between the two systems depends on the number of floaters. For a low number of floaters, when the oscillations are small, the median loss for the two orientations is similar. The outlying values of the South-oriented system make that, throughout the year, the system experiences slightly higher losses. On the other hand, when the number of floaters is 25 or 50 and the oscillations are higher, the considerable difference in median loss between the two orientations compensates for the outlying daily losses of the South-oriented system. This makes that the East-oriented system experiences a higher annual mismatch loss.

The plots in Figure 11.15 help understand this difference in the distribution of mismatch losses for the two orientations.

Figure 11.15a shows the power production on the day with the highest mismatch for the East-oriented case, January 14th. The power losses for this East-facing PV system occur mostly in the morning and afternoon instead of during noon as for the South-facing system. One can also note that this is again a sunny winter day. However, as opposed to December 28th, there is a drop in production near noon due to clouds passing.

The difference in orientation also affects the daily loss throughout the year, as shown in Figure 11.15b. A South-facing floater, which loses most of the power at noon, faces relatively high losses in the cold months when the days are short and most of the irradiance is centered around noon. An East-facing floater during those months is beneficial as most mismatch losses will occur when the irradiance is not so high. During the long summer days, the losses occurring at noon by the South-facing floater are not so detrimental because the irradiance is also high at noon. However, for the East-facing floater, the losses are relatively larger as the irradiance is still high in the hours around noon. However, the

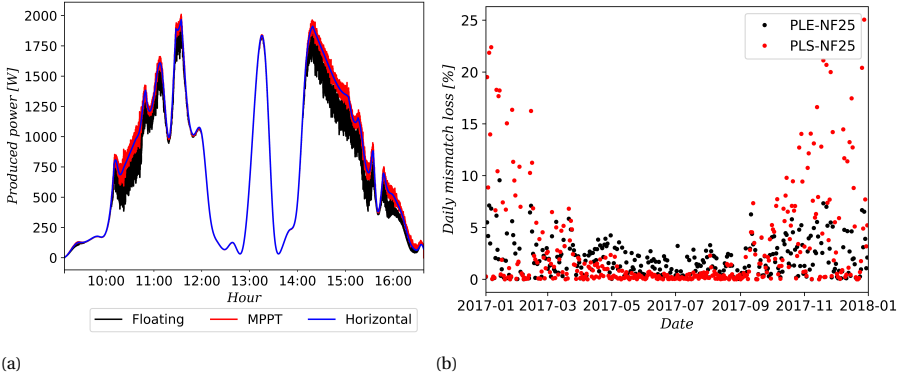


Figure 11.15: (a) Power produced by the floating, MPPT, and horizontal PLE-NF25 systems on January 14th, the day with the highest daily mismatch losses. (b) Daily mismatch throughout the year for PLE-NF25 and PLS-NF25.

highest daily mismatch losses still occur in winter due to the higher wind speeds and lower produced power.

11.5.2. YOUNG'S MODULUS

We start with the base case PLA, as defined in Table 11.1, and vary only the Young's modulus of the material. We explore three scenarios: $E = 0.5 E_{HDPE}$, $E = 1.0 E_{HDPE}$, and $E = 2.0 E_{HDPE}$. Figure 11.16a illustrates the annual mismatch and movement losses, for platforms with various number of floaters, for these three values of Young's modulus. Figure 11.16b presents the distribution of the daily mismatch losses.

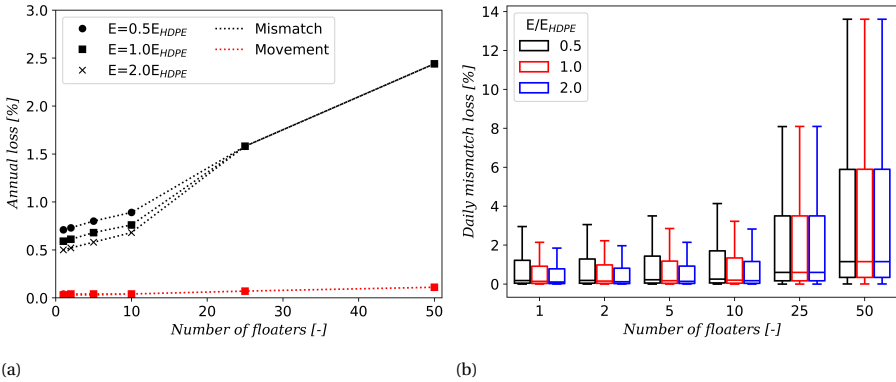


Figure 11.16: (a) Annual power losses and (b) daily mismatch for systems with varying Young's modulus for an increasing number of floaters.

From the plots, an interesting trend emerges, which can be explained, based on our analysis in Section 11.4. For 25 and 50 floaters, there is no discernible difference in the mismatch losses across the three Young's modulus values. In scenarios with a high number of floaters, the length of each floater L_f is small, resulting in small ratios L_f/λ and L_f/λ_c , indicative of a rigid body response. As observed earlier in Figure 11.8e-f, under the rigid body

response regime, the bending moment within the structure is minimal. This emphasises that the bending stiffness of the beam has little influence on the structure's response, hence altering the Young's modulus of the material yields little difference in the mismatch losses.

Conversely, scenarios with a low number of floaters will have higher values of L_f/λ and L_f/λ_c , implying an elastic response. As per our observation in Figure 11.8a-b, in these instances, the elastic response of the floater dominates. As a result, we observe distinct differences in the mismatch losses when the Young's modulus is altered for scenarios with a low number of floaters. Moreover, increasing the Young's modulus enhances the bending stiffness of the structure, thereby reducing the platform's motion and consequently diminishing the overall mismatch losses. This explains the observed reduction in losses with increasing Young's modulus, for cases with fewer number of floaters.

11.5.3. THICKNESS

In this section, we revisit the base case PLA, varying only the thickness of the beam. Case PLA has a thickness of $h_b = 0.2\text{m}$. We additionally test scenarios with thicknesses $h_b = 0.1\text{m}$ and $h_b = 0.4\text{m}$. Figure 11.17a presents the annual mismatch and movement losses for platforms with 1-50 floaters, for these three floater thickness values. Figure 11.17b presents the distribution of daily mismatch losses for the same thickness values. These plots reveal a curious trend. For scenarios with 1, 2, 5, and 10 floaters, an increase in platform thickness leads to a reduction in mismatch losses. However, for 25 and 50 floaters, an increase in thickness conversely results in increased mismatch losses.

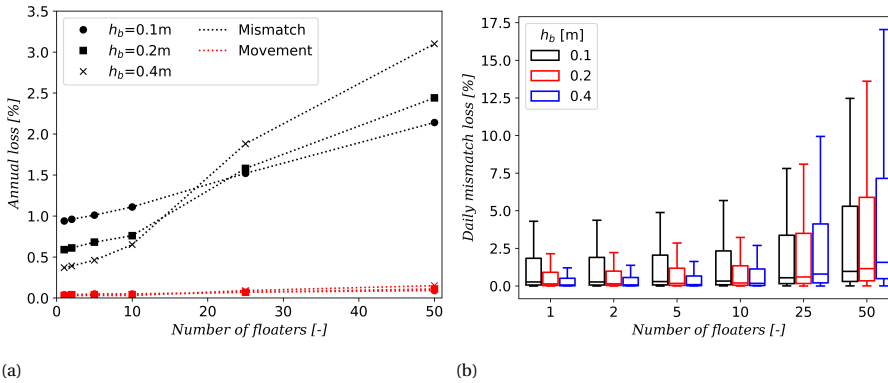


Figure 11.17: (a) Annual power losses and (b) daily mismatch for systems with varying thicknesses for an increasing number of floaters.

The observation for fewer floaters is easily explained. In cases with fewer floaters, the structure operates in the elastic regime due to high values of L_f/λ_c and L_f/λ . Therefore, the elastic response dominates. Since the bending stiffness $EI \propto h_b^3$, an increase in thickness significantly enhances the bending stiffness, resulting in substantially smaller motions for $h_b = 0.4\text{m}$ compared to $h_b = 0.1\text{m}$, thereby leading to lower mismatch losses.

The reversal of this trend for 25 and 50 floater scenarios can be explained by examining the influence of h_b on the characteristic hydro-elastic length of the structure, i.e., $\lambda_c \propto (h_b)^{3/4}$. For a given floater length L_f , changing the floater thickness alters the ratio $L_f/\lambda_c \propto (h_b)^{-3/4}$. Consequently, a thinner beam exhibits a higher L_f/λ_c , indicating an

elastic-dominant response, while a thicker beam results in a lower L_f/λ_c , leading to a rigid body response.

This distinction is further supported by the contour plots of longitudinal stress in the floater cross-section, as presented in Figure 11.18. In Figure 11.18a-c, we present the contour plots of the magnitude of stress for a platform composed of a single floater of $L_f = 100$ m (PLA-NF01) with floater thicknesses h_b of 0.1 m, 0.2 m, and 0.4 m, respectively. In Figure 11.18d-f, we present the contour plots of the magnitude of stress for a platform composed of 25 floaters of $L_f = 4$ m (PLA-NF25), with floater thicknesses h_b of 0.1 m, 0.2 m, and 0.4 m, respectively. In the PLA-NF01 scenario, there is significant stress across a wide range of excitation frequencies for all thicknesses, indicating elastic response. In the PLA-NF25 scenario, we observe nearly zero stress for the 0.4 m floater, indicating a rigid body response; while strong stresses are observed for the 0.1 m floater, indicating an elastic response.

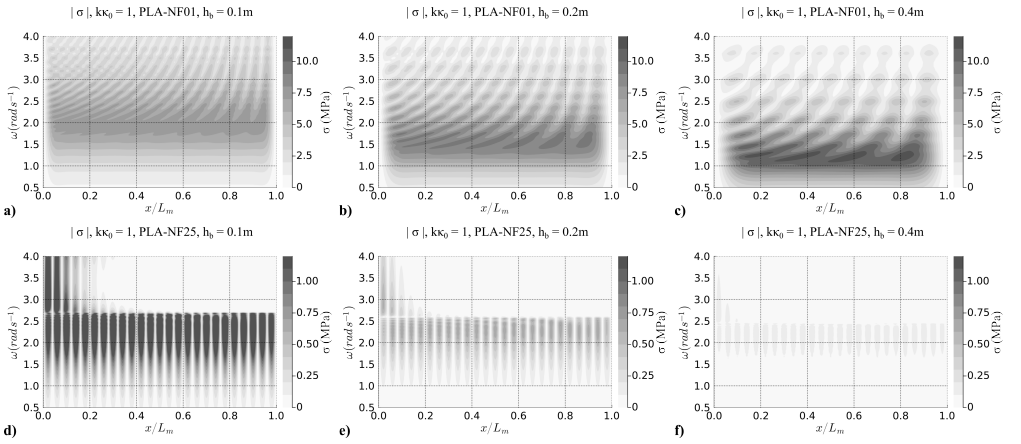


Figure 11.18: Longitudinal stress in the floater beam cross-section. a-c) PLA-NF01 with $h_b = 0.1$ m, 0.2 m, 0.4 m. d-f) PLA-NF25 with $h_b = 0.1$ m, 0.2 m, 0.4 m.

As $h_b = 0.4$ m for PLA-NF25 has very little elastic response, there is no additional stiffness against wave excitation, resulting in large motions and mismatch losses for 25 and 50 floater cases. Conversely, $h_b = 0.1$ m exhibits significant elastic response even for 25 and 50 floater scenarios, providing additional stiffness against wave excitation and resulting in smaller motions and mismatch losses.

11.5.4. FILL RATIO

In this test, we modify the cross-section of the floater. The base case PLA has a solid cross-section. We test scenarios, where the middle portion of the cross-section is hollowed out, to minimise weight with limited influence on the bending stiffness. We quantify this using a *fill ratio* γ . In addition to solid cross-section having $\gamma = 1$, we test cases with $\gamma = 0.5$ and the lightest $\gamma = 0.25$.

Figure 11.19a presents the annual mismatch and movement losses for platforms with 1-50 floaters, for these three fill-ratio values. Figure 11.19b presents the distribution of daily mismatch losses for the same. Similar to Subsection 11.5.3, the obtained trend is dependent on an interplay between rigid-body and elastic response, driven by the change in bending stiffness, mass and characteristic hydro-elastic length.

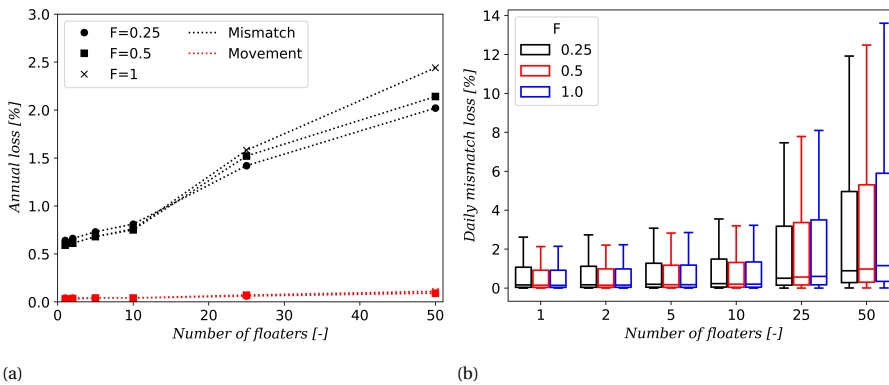


Figure 11.19: (a) Annual power losses and (b) daily mismatch for systems with varying fill factors for an increasing number of floaters.

11.6. CONCLUSIONS

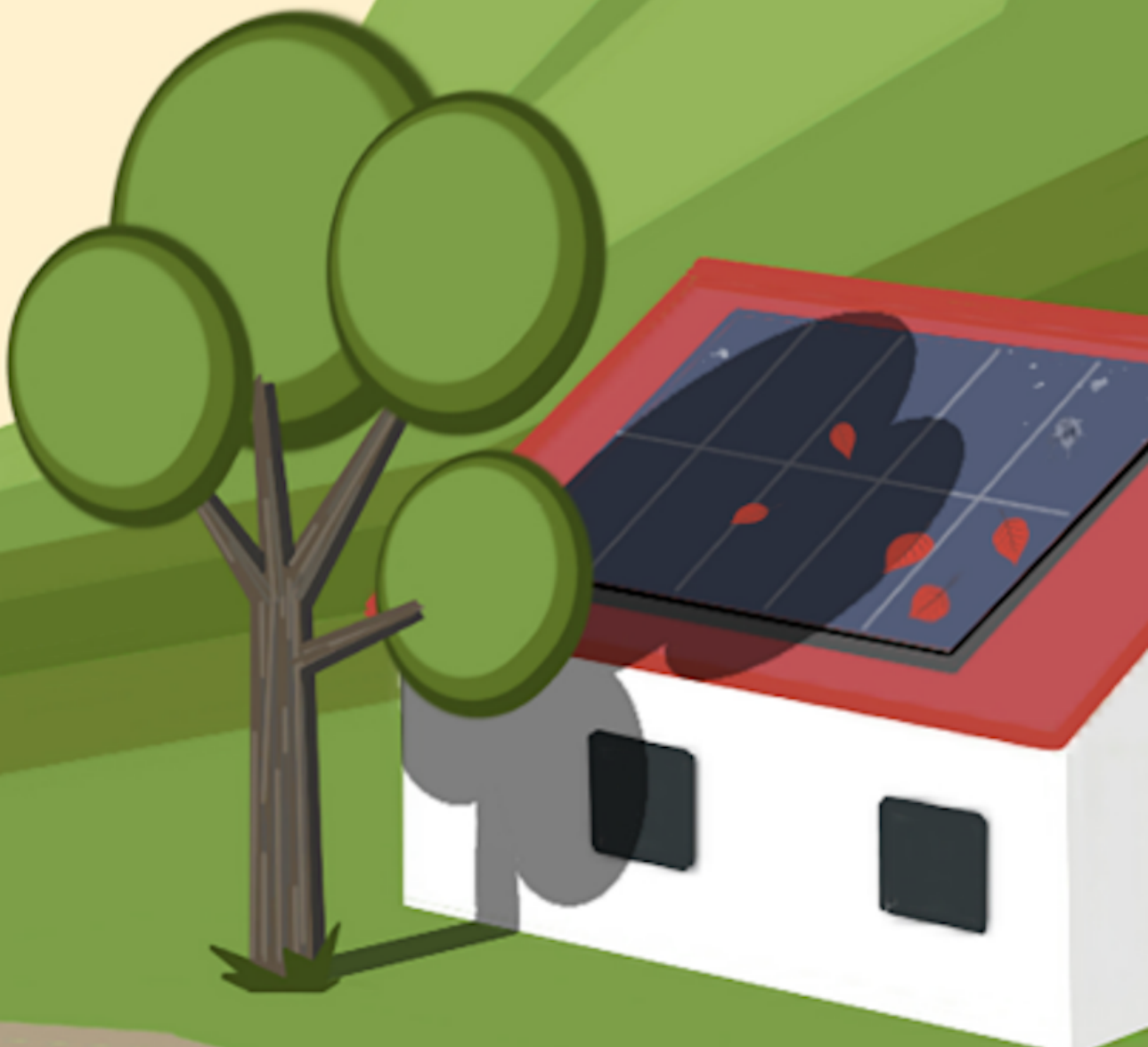
In this work, we have analyzed the interaction of an OFPV system with surface waves focusing on the structural loads of the platform and the mismatch losses experienced by the system. Using a combined framework to study this multi-physics problem, we have examined how the number of floaters, orientation, platform's Young modulus, thickness, and fill ratio affect the structural and optoelectrical response of the OFPV system. The main key findings are:

- There exists a trade-off in selecting the number of floaters. An increase in the number of floaters amplifies the platform motion. Optoelectric considerations favor smaller motions, making a design with a single or a few long floaters preferable to minimize the tilting of the PV panels. However, this choice of fewer and longer floaters introduces higher structural stresses and potential fatigue-related concerns. Opting for a large number of small floaters leads to a transition from elastic to rigid body response, resulting in minimal elastic stresses. Fortunately, the highest mismatch losses occur on sunny windy winter days, therefore periods of low generation.
- The system orientation shows a limited effect on the mismatch losses. This was expected considering the original horizontal position of the PV modules. However, North-South oscillations are preferred over East-West oscillations as the losses in the latter case are relatively higher for most of the year except for winter.
- The structural properties of the platform can have a significant influence on the overall power mismatch losses. A change in Young's modulus, thickness, or cross-section of the floater beam can alter the hydro-elastic response from a rigid body response to an elastic dominant response. This influence depends on the number of floaters of the platform.
- The scenarios with up to 10 floaters were observed to have an elastic dominant response. As a consequence of this, the power mismatch loss is observed to reduce with an increase in Young's modulus of the material, with an increase in the thickness of the floater beam, or with an increase in the fill ratio for the cross-section.

- For the 25 and 50 floaters scenarios, we observed a rigid body response for all tested values of Young's modulus. Therefore, Young's modulus did not have a significant impact on the power mismatch losses. On the other hand, the 25 and 50 floaters scenarios showed an elastic dominant response for low floater thickness and a rigid body response for high floater thickness. As a result of this, a thinner floater resulted in lower mismatch losses, due to added resilience from the hydro-elastic response. A similar trend was seen when varying the fill ratio, where a low fill ratio offered lower power mismatch losses due to the hydro-elastic response.

Overall, the work highlights the interplay between the hydro-elastic response and the design and properties of VLFS such as a floating PV platform. The work further presents the consequence of this hydro-elastic response on the potential optoelectric performance of the floating PV platform in offshore conditions. Our future work will focus on 3D analysis of this problem, allowing for irregularly shaped floating PV platforms and presenting the response in the presence of mooring lines. Further, the hydro-elastic model will be developed to account for the non-linearity in the ocean waves and the structural response. The exploration of alternative locations and different floating structures, such as membranes, is also worthwhile.

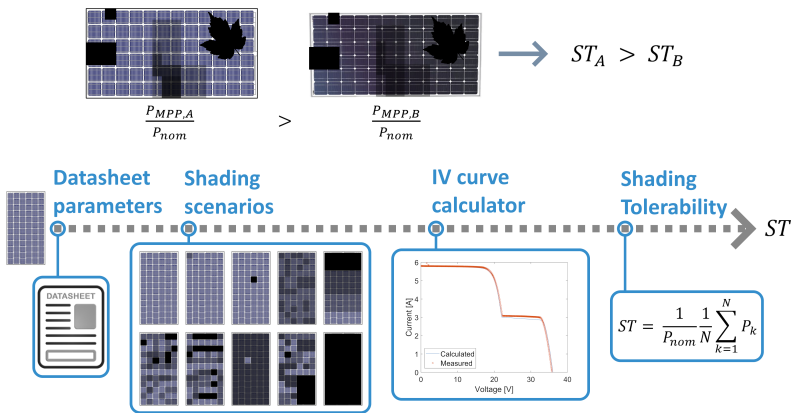
SHADING TOLERABILITY



12

SHADING TOLERABILITY

Partial shading significantly reduces the performance of PV modules, though its impact varies across different models. Assessing this tolerance is challenging due to the unpredictable nature of shading events. In this chapter, we develop a shading tolerability (ST) calculator, a tool that quantifies a module's ST as a numeric parameter for comparative analysis. ST is calculated based on a literature-defined methodology, which involves determining the power at the maximum power point under various shading scenarios with equal probability of occurrence. The tool was adapted to calculate ST for four irradiance levels efficiently. We compiled a database of ST values for 423 commercial modules with a conventional layout of series-connected cells, ranging from 18.4% to 21.4%.



This chapter has been adapted from the Trust-PV report A. Alcañiz, J.C. Ortiz Lizcano, H. Ziar, M. Dallapiccola, Y. Snow, I. Kaaya, "Residential-scale integrated energy yield simulation and reliability modelling framework" (2023) [234] and from the work A. Alcañiz, N. Rukhshi, R. Koutarapu, O. Isabella, and H. Ziar "Quantifying Shading Tolerability in Commercial Photovoltaic Modules" submitted to *Progress in Photovoltaics*.

12.1. INTRODUCTION

One important source of mismatch power losses affecting photovoltaic (PV) systems is partial shading. Partial shading can be caused by various phenomena such as passing clouds, dust, dirt, snow, or bird droppings affecting all types of PV systems [454, 455]. However, residential PV systems are especially affected by partial shading due to obstacles present in the urban environment [456]. This higher shading frequency can decrease the performance ratio up to 25% [457], although the exact value depends on the shading profile of the location and the PV module.

Partial shading can drastically reduce a PV system's energy yield and cause the cells to operate in reverse bias. This leads to rising temperatures resulting in hot spots, which can lead to local defects, permanent damage, and degradation of PV modules [458–461]. Active or passive bypass diodes can mitigate these issues but cannot eliminate them [462, 463]. Selecting a PV module with minimal shading losses is crucial.

Although hot spot endurance test procedures are available in standards [464], there is currently no straightforward method to compare the electrical performance under shading of different PV modules. It is hard to understand and quantify how a PV module reacts to shading compared to others. Currently, a module's resistance to shading is rarely listed on datasheets, and when present, it is expressed qualitatively with statements such as “Better performance under shade” [465], “excellent performance in low-light environments” [466], or “high performance under low light conditions” [467]. Such general qualitative statements make it difficult for a PV system designer to compare modules when trying to select the best one for a certain location in terms of shading.

With this issue in mind, Ziar et al. suggested the concept of shading tolerability [457]. Their objective was to develop a quantified numerical parameter that measures the ability of PV modules to withstand shading. This can be used to rank, compare, and choose the optimum module. Since shading is a random process, it is not practically feasible to test a PV module in all possible shading scenarios. For that reason, Ziar et al. employed a probabilistic approach to define shading tolerability (ST) as the “mathematical expectation of power production of a PV module under shading”. The mathematical expectation $E(x)$ of a discrete variable x is:

$$E(x) = \sum_k x_k \cdot p(x_k) \quad (12.1)$$

where $p(x)$ is the probability that x occurs.

In the problem at hand, k would be the shading scenarios, which are all possible ways the module can be shaded and are based on the assumption that the incident irradiance across the surface of a PV cell is homogeneous, and can be any value between 0 and 1000 W/m². Additionally, the probability of different scenarios to occur and of different cells within the module to be shaded is equal. Since the objective is to compute how much power a module would produce when subjected to shading, x_k would be the power output under the shading scenario k (P_k). For a PV module with c cells subjected to i different irradiance levels, there are a total of $N = i^c$ shading scenarios. Therefore, each scenario has an equal occurring chance of $\frac{1}{i^c}$. Adding all of these factors and normalizing the output power with the nominal power P_{nom} to compare modules with different rated powers, one obtains Equation 12.2 for the shading tolerability,

$$ST = \frac{1}{P_{nom}} \sum_{k=1}^{i^c} P_k \frac{1}{i^c} = \frac{1}{P_{nom}} \frac{1}{N} \sum_{k=1}^N P_k \quad (12.2)$$

where P_k is the maximum power point (MPP) power produced by the module under the shaded scenario k .

After defining the figure of merit, Ziar et al. measured the shading tolerability for several commercial PV modules. To make ST measurable, the irradiance levels were set to 2 (the minimum possible), and the modules were divided into 6 geometrically equal sections. They validated this approximation by showing that a module with a higher ST at $i = 2$ also has a higher ST as i approaches infinity. This resulted in a total of $2^6 = 64$ measurements to perform, which led to an average measurement time of around 6 hours per module [457, 468]. As the objective is to include ST as a figure of merit in the PV modules datasheet, such long measurement times may hinder its promotion. Moreover, the division of the PV module into 6 sections ignores several phenomena that affect when a few cells are shaded, such as the breakdown voltage of the cell.

A solution to these limitations would be to compute numerically the shading tolerability, which is what Klasen et al. performed in their work [469]. Instead of using a probabilistic approach, they considered Monte Carlo simulations for this purpose. They emulated a large number of rectangular and random scenarios to compute the “average normalized power for partial shading” $\overline{P_{ps}}$. They also studied the effect of different module layouts (conventional, butterfly, shingle string, and shingle matrix). Similarly, Fauzan et al. studied the tolerance to partial shading of several PV module layouts [470]. Rectangular, circular, and diagonal shading shapes of different sizes were emulated. They showed the superiority of their proposed layout, the small-area-high-voltage module, which showed a similar shading response as the Tessera concept [471]. These two studies focused only on the impact of module layout on shading tolerability, ignoring several other parameters that could be relevant as well, such as the number of bypass diodes and the cell electrical characteristics.

Another concern with the latter approach is the predefined shape of shading scenarios, which fails to capture the randomness of the problem. This limitation is also found in other studies, which either use fixed shading shapes or focus on the shading profile of a specific environment [456, 472–474].

Motivated by these limitations, this work aims to develop a tool for efficiently computing the shading tolerability of commercial PV modules. The tool is designed to quickly calculate ST as a numeric parameter using only the datasheet parameters of the module. Building on the methodology established by Ziar et al., this approach ensures both effectiveness and efficiency. The tool is utilized to generate a database of ST values for modules with a conventional layout of series-connected cells. We aim to explore the effect of various electrical parameters such as nominal power or breakdown voltage on ST. Additionally, we demonstrate the tool's application to commercial modules, paving the way for the inclusion of ST in module datasheets.

This work continues by developing the ST calculator in Section 12.2. Essential parameters for the calculator are determined in Section 12.3. The tool is improved in Section 12.4 by reducing its computational requirements. Finally, Section 12.5 presents the database of ST values and related discussion before concluding in Section 12.6.

12.2. SHADING TOLERABILITY CALCULATOR

The first step consists of developing a program that computes the ST. MATLAB was chosen as the programming language for this. Figure 12.1 shows the steps required to build the calculator. One feature of this calculator is that it only uses the datasheet parameters as input, facilitating its use. The shading scenarios are determined by knowing the module layout, number of cells, and bypass diodes. Since calculating the maximum power point of i^c scenarios is highly computationally intensive, an approach is developed to remove equivalent scenarios, defined as those that yield the same output power. The IV curve is calculated for the reduced scenarios, with which the ST can be computed using Equation 12.2.

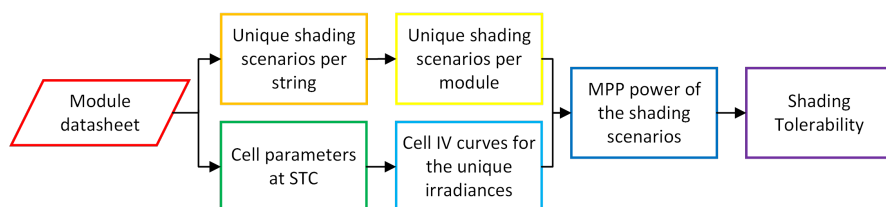


Figure 12.1: Flowchart of the shading tolerability calculator.

12.2.1. FINDING THE SHADING SCENARIOS

The shading scenarios are all the possible ways that a module can be shaded. For an 18-cell module and two irradiance levels, there are $N = 2^{18} = 2.6 \times 10^5$ scenarios. Some of these scenarios are shown in Figure 12.2. Examples (a) and (b) or (c) and (d) yield the same power output numerically. We call these *equivalent scenarios*, defined as those scenarios whose output IV curve is theoretically the same. This is based on the plausible assumption that all cells are identical within a PV module. To identify the equivalent scenarios, one needs to consider the number of shaded cells and the effect of bypass diodes. Examples (d), (e), and (f) yield a different power output despite having the same number of shaded cells because of their distribution among the bypass diode (BD) strings. The scenarios that are not equivalent to each other are referred to as *unique*. Therefore, scenarios (a), (c), (e), and (f) are unique. Instead of scenario (a) one could select its equivalent scenario (b) as unique, hence the set of unique scenarios would be (b), (c), (e), and (f).

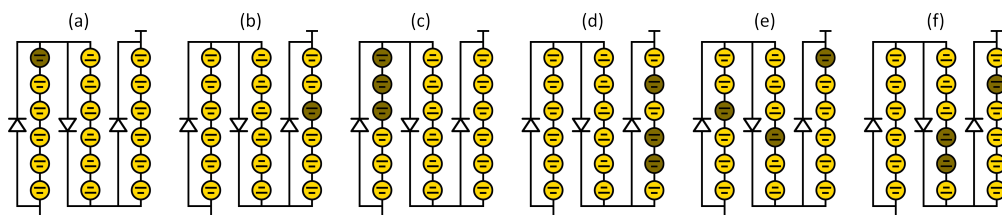


Figure 12.2: Examples of shading scenarios in an 18-cell series-connected module with two irradiance levels. Illuminated cells are represented in yellow while shaded ones are colored in brown.

One unique scenario can represent all of its equivalent scenarios. For instance, the shading scenario in Figure 12.2(a) can represent the power produced by 18 other scenarios,

since the shaded cell can be allocated in 18 different positions. Figure 12.2(b) will be one of these represented scenarios. We define *counts* as the number of equivalent scenarios represented by each unique scenario. Hence, the counts of the scenarios in Figure 12.2(a) and Figure 12.2(b) are 18.

Using equivalent scenarios reduces the number of shading scenarios to be considered for computing the shading tolerability. Instead of computing the power P_k for all the N shading scenarios and adding them up during the summation in Equation 12.2, one can identify the N_{uq} unique scenarios, calculate the power for each, and then multiply the power of each unique scenario by its counts.

To identify the unique scenarios and their counts, we start by calculating the unique shading scenarios on a string level, the *string scenarios*. A string is defined as a group of series-connected cells protected by a parallel bypass diode. We call this a *BD string*. Within a BD string, only the number of shaded cells is required to know the power output, not their position within that string. This simplifies the identification of the string scenarios.

The first step towards finding the string scenarios consists of calculating all the unique shading scenarios within a BD string. Consider a string with 6 cells and 2 irradiance levels, like those composing the modules from Figure 12.2. The unique scenarios are no cells shaded, one cell shaded, two cells shaded, and so on until all 6 cells are shaded. There are a total of $c_{str} + 1$ unique string shading scenarios for 2 irradiance levels, with c_{str} being the number of cells in the BD string.

The counts for each of these unique string scenarios n_{str} is expressed with the binomial coefficient presented in Equation 12.3, where c_{sh} is the number of shaded cells within the BD string. The binomial coefficient expresses combinations without repetition. In our case, the binomial coefficient expresses the number of possible combinations of c_{sh} shaded cells from a total of c_{str} cells in the string. Figure 12.3 illustrates all the unique string scenarios for the module shown in Figure 12.2, along with the counts for each scenario.

$$n_{str} = \binom{c_{str}}{c_{sh}} = \frac{c_{str}!}{c_{sh}! \cdot (c_{str} - c_{sh})!} \quad (12.3)$$

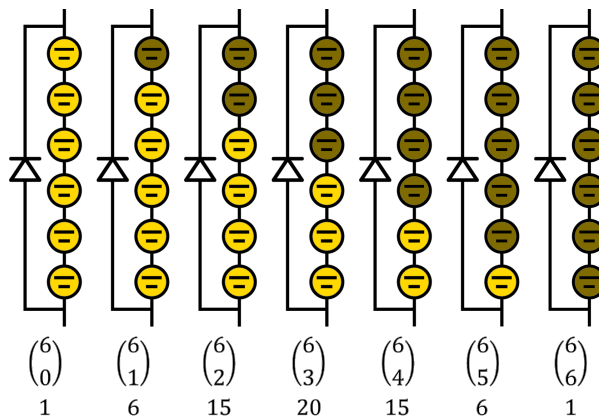


Figure 12.3: Unique string shading scenarios with two irradiance levels and counts n_{str} for each scenario for a BD string composed of 6 cells. Illuminated cells are represented in yellow while shaded ones are colored in brown.

The unique string scenarios are combined to obtain the unique shading scenarios on the module level. Consequently, the counts of each module-level scenario depend on the counts of the string scenarios that compose each scenario. To consider all the possible combinations for a specific scenario, the counts of the string scenarios that compose the shading scenario need to be multiplied. Additionally, depending on the position of the string scenarios, different shading scenarios are created corresponding to the same equivalent scenario. Figure 12.4 illustrates this phenomenon for the three possible cases occurring for the 18-cell module from Figure 12.2. When all the string shading scenarios are different, there is a total of 6 different possible combinations (case 1). When two of the three shading strings are equal, there are 3 possible combinations (case 2). If all the string scenarios are the same, there is only 1 possible combination (case 3). The number of equivalent scenarios that can be created by combining string scenarios is represented as n_{comb} . To obtain this value, one needs to consider the number of repetitions rep_{str} of the unique string scenarios $N_{uq, str}$ that compose the shading scenario. This is expressed in Equation 12.4, where s_{BD} is the number of BD strings. Figure 12.4 illustrates the usage of this equation for the possible cases in the 18-cell module with three BD strings.

$$n_{comb} = \frac{s_{BD}!}{\prod_{str=1}^{N_{uq, str}} rep_{str}!} \quad (12.4)$$

Combining all this information, one finds the counts for a shading scenario on the module level n_{sc} . First, one needs to consider the counts of each of the string scenarios that compose the shading scenario using Equation 12.3 and then the number of ways that these string scenarios can be combined using Equation 12.4. With those, one reaches Equation 12.5.

$$n_{sc} = n_{comb} \cdot \prod_{b=1}^{s_{BD}} n_{str_b} = \frac{s_{BD}!}{\prod_{str=1}^{N_{uq, str}} rep_{str}!} \cdot \prod_{b=1}^{s_{BD}} \binom{c_{str_b}}{c_{sh_b}} \quad (12.5)$$

Figure 12.5 exemplifies this process for a shading scenario of an 18-cell module with 3 bypass diodes. The shading scenario is formed by combining three unique string scenarios. Referring to Figure 12.3 one can find the counts for each string scenario. Then, one considers how many of these strings are repeated, if any. In the example, all strings are different hence $n_{comb} = 6$. After multiplying all the computed values, one obtains the counts of the exemplified shading scenario.

Equation 12.5 was obtained for the case of two irradiance levels, so it should be generalized to any number of irradiance levels i . The extension of the binomial coefficient for more than two classes of elements is the multinomial coefficient. To adapt it, the counts for the string scenarios n_{str} need to include the number of cells c_j with each irradiance value j . Therefore, Equation 12.3 transforms to Equation 12.6.

$$n_{str} = \binom{c_{str}}{c_1 c_2 c_3 \dots} = \frac{c_{str}!}{\prod_{j=1}^i c_j!} \quad \text{with} \quad c_{str} = \sum_{j=1}^i c_j \quad (12.6)$$

The combinations of the string scenarios n_{comb} as illustrated in Figure 12.4 are independent of the number of irradiance values. Therefore, the general form of Equation 12.5 for i irradiance levels transforms to Equation 12.7.

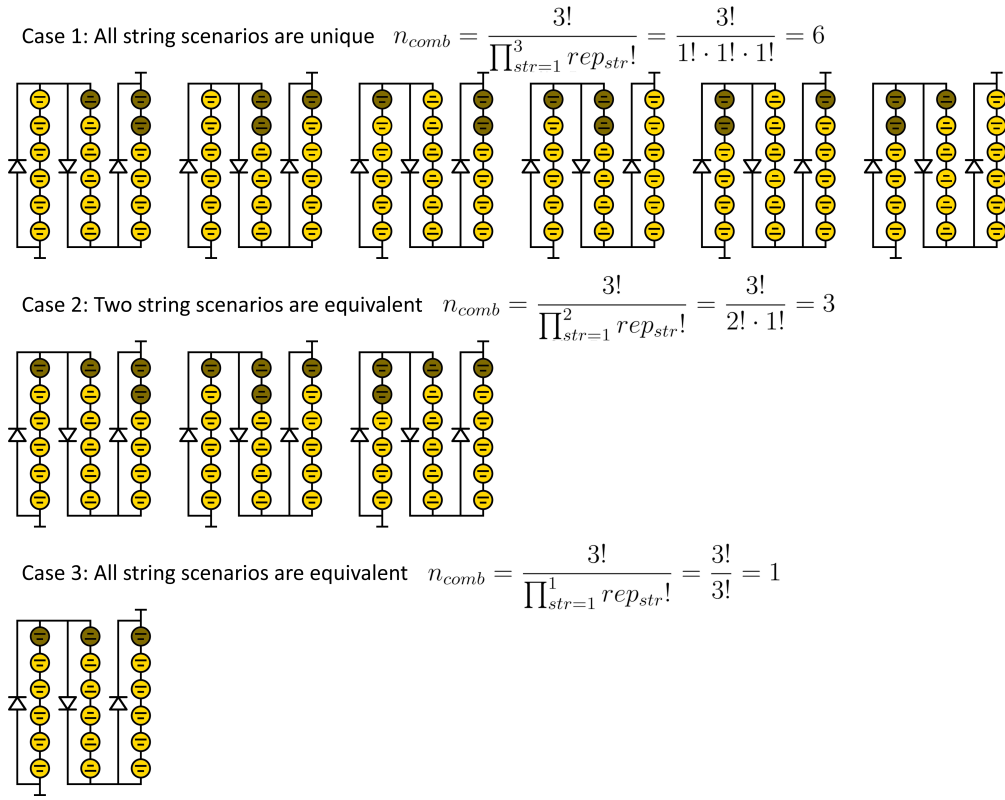


Figure 12.4: Possible case combinations of string scenarios to form shading scenarios on the module level. The examples correspond to an 18-cell module with 3 BD strings. Depending on the number of repeated string scenarios, a different number of combinations can be formed.

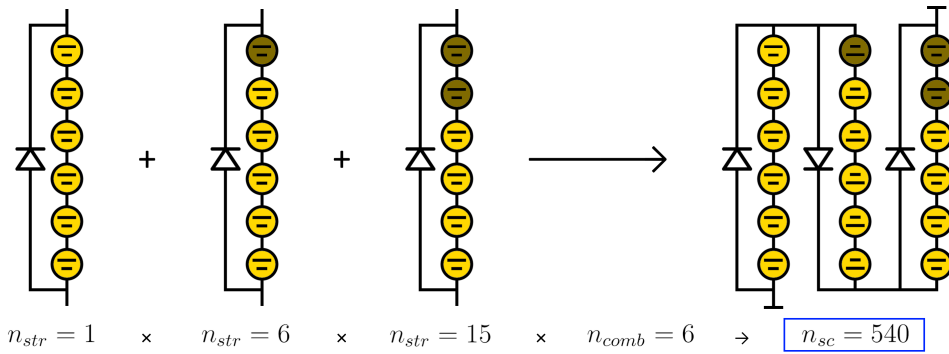


Figure 12.5: Graphical process to find the counts for a shading scenario. The example corresponds to an 18-cell module with 3 bypass diodes and considering 2 irradiance levels.

$$n_{sc} = n_{comb} \cdot \prod_{b=1}^{SBD} n_{str_b} = \frac{SBD!}{\prod_{str=1}^{N_{uq, str}} rep_{str}!} \cdot \prod_{b=1}^{SBD} \frac{c_{str_b}!}{\prod_{j=1}^i c_{j_b}!} \quad (12.7)$$

Using this equation, one finds that the counts of the scenario in Figure 12.2(a) are 18; 60 in (c); 216 in (e); and 540 in (f). With this approach, the power output for a commercial 60-cell PV module with 3 bypass diodes and considering 2 irradiance values needs to be calculated for only the N_{uq} 1771 unique scenarios instead of for all the $2^{60} = 1.2 \times 10^{18}$ shading scenarios.

12.2.2. DEVELOPING THE IV-CURVE CALCULATOR

The next step consists of developing an efficient and accurate IV curve calculator that provides P_{MPP} of a PV module under a specific shading scenario. The approach employed is based on the 2-diode model developed by Ishaque, Salam, et al. [475]. It consists of a parametric approach with several simplifications which considerably reduce the computational time and complexity, yet being able to model a two-diode model. The model was modified to consider i) the curves at the cell level, ii) the dependency with temperature and irradiance [476, Eq. 2, 4–6], and iii) the reverse bias region [477]. The implemented IV curve is shown in Equation 12.8 where

I is the output current [A],

I_{PV} is the photogenerated current at Standard Test Conditions (STC) [A] considered equal to the short circuit current at STC,

I_0 is the saturation current [A] equal for both diodes and found explicitly [478, Eq. 7],

V is the output voltage [V],

R_s is the series resistance [Ω],

V_{th} is the thermal voltage [V] equal to $\frac{k_B \cdot T_c}{q}$ with k_B being the Boltzmann constant 1.38×10^{-23} J/K, q being the electron charge 1.60×10^{-19} C, and T_c representing the cell temperature [$^{\circ}$ C],

p is the sum of the two diode ideality factors which can be chosen arbitrarily as long as it is higher than 2.2 (a value of 2.5 was set in this work),

R_p is the parallel or shunt resistance [Ω] found together with R_s by iteratively solving Equation 12.8 under MPP conditions,

V_{br} is the breakdown voltage [V] which depends on the cell material/technology, and

a_{br} and m_{br} are constants used to model the reverse bias region with values 2×10^{-3} and 3, respectively [479].

$$I = I_{PV} - I_0 \left[\exp \left(\frac{V + I \cdot R_s}{V_{th}} \right) + \exp \left(\frac{V + I \cdot R_s}{(p-1) \cdot V_{th}} \right) - 2 \right] - \frac{V + I \cdot R_s}{R_p} - a_{br} \cdot (V + I \cdot R_s) \left(1 - \frac{V + I \cdot R_s}{V_{br}} \right)^{-m_{br}} \quad (12.8)$$

The cell temperature T_c was calculated using the NOCT (Normal Operating Cell Temperature) model shown in Equation 12.9 [261, Eq. 20.2]. This model was selected for its simplicity and the presence of the NOCT temperature T_{NOCT} in the modules' datasheet.

$$T_c = T_a + \frac{T_{NOCT} - 20}{800} G_c \quad (12.9)$$

where T_a is the ambient temperature [$^{\circ}$ C] (set to 25° C in this work), and G_c is the incident irradiance on the cell [W/m^2].

The model was extended to the module level by considering the interconnection of cells in series or parallel and the implementation of bypass diodes. The bypass diode is activated when the string voltage drops below $-V_F$, the negative of the diode's forward voltage. The forward voltage changes as a function of the current in the range 0.1 to 10 A, modeled based on the p-n F1200A-Semikron bypass diode [480] at 25 °C.

The complete process for the IV calculator is as follows. First, the IV parameters under STC at the cell level are computed. Then, the calculator obtains the cell temperatures for the possible incident irradiance values. Using those values and the corresponding incident irradiances, the IV curves on the cell level can be calculated. Considering the module layout, the cell IV curves are combined in series or in parallel with the bypass diodes taken into account to obtain the module IV curves. From there, the P_{MPP} for each scenario is obtained. The benefits of this methodology are its low input requirements (datasheet parameters and the breakdown voltage, which is assumed based on the cell material/technology) and high computational speed (6 ms per shading scenario for a personal computer with Intel(R) Xeon(R) W-2223 CPU@3.60GHz and 8 GB of RAM).

The model was validated against measurements of a Canadian Solar CS6P 230P 60-cell module under various partial shading conditions, provided by EURAC Research [481]. Shading is done by covering a certain percentage of all cells in the first and/or third column and three different irradiance values. The P_{MPP} absolute errors for all measurements can be seen in Figure 12.6. The errors at 1000 W/m² are the lowest, while those at 200 W/m² are the highest. This indicates the higher difficulty of simulating IV curves at lower irradiance conditions. All errors lie below 4.5%, with most actually below 3%, considered low enough to develop the ST calculator.

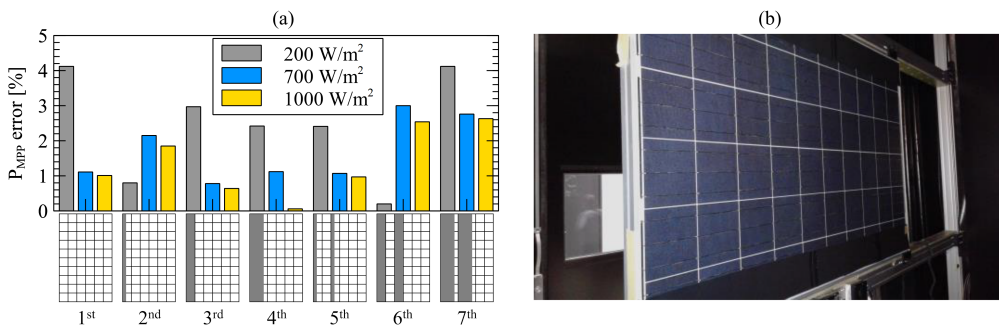


Figure 12.6: (a) Validation of the IV calculator by showing the maximum power point errors for measurements executed on Canadian Solar CS6P 230P modules. 1st to 7th indicate seven applied shading scenarios. (b) Photo of the module being tested under a shading scenario equivalent to the 6th.

12.2.3. VISUALIZING THE SHADING TOLERABILITY

Once the P_{MPP} for each unique shading scenario is obtained, the shading tolerability can be calculated. Since the power is calculated for only the unique shading scenarios instead of all of them, Equation 12.2 needs to be adapted. As mentioned in Subsection 12.2.1, a unique scenario sc yields the same power output as its n_{sc} equivalent scenarios. In Equation 12.2, the power of these equivalent scenarios would be added up. However, one obtains the same value by calculating the P_{MPP} of the unique scenario and multiplying it by

its counts n_{sc} . Thus, Equation 12.2 is modified to sum, not the power of all N shading scenarios, but rather the MPP power P_{sc} of the N_{uq} unique scenarios multiplied by its counts n_{sc} . Since all shading scenarios contribute to the ST calculation, the sum of all the counts for the unique scenarios $\sum_{sc=1}^{N_{uq}} n_{sc}$ should be equal to the total number of scenarios $N = i^c$. By incorporating these changes into Equation 12.2, one reaches Equation 12.10, which enables the calculation of ST by considering only the power produced by the unique shading scenarios.

$$ST = \frac{1}{P_{nom}} \frac{1}{N} \sum_{k=1}^N P_k = \frac{1}{P_{nom}} \frac{1}{\sum_{sc=1}^{N_{uq}} n_{sc}} \sum_{sc=1}^{N_{uq}} P_{sc} \cdot n_{sc} \quad (12.10)$$

To increase understanding of the shading tolerability concept, Ziar et al. introduced a relevant metric: the percentage ST $ST_{\%}$ [457]. This metric quantifies how shading-tolerant a module is relative to the theoretical maximum shading tolerance. Additionally, the metric allows comparing results obtained when the minimum irradiance is difference. $ST_{\%}$ is calculated by dividing the ST value by the maximum possible theoretical ST value ST_{max} . The maximum theoretical ST is obtained for a module with no additional losses when shaded, where the module produces power directly proportional to its unshaded area. A module that generates 50% of its rated power when 50% of its area is shaded would have this maximum ST. For this module, ST equals the average irradiance normalized by the irradiance under STC conditions (1000 W/m^2). For an irradiance range between 0 W/m^2 and 1000 W/m^2 , $ST_{max} = 0.5$.

Another way to express shading tolerability is through visualization. The goal is to observe the power output across all considered scenarios for the module. While one option would be to plot the power output for each shading scenario, it poses a challenge as the scenarios have no physical significance. To address this, the shaded area of each shading scenario is employed. We define the shaded area fraction (SAF) of a specific shading scenario as the area percentage that receives the lowest irradiance G_{min} assuming that the remaining area receives the maximum irradiance G_{max} . When two irradiance levels are considered, the SAF of a shading scenario is equal to the number of shaded cells divided by the total number of cells in the module c_{sh}/c . When more irradiance levels are considered, this approach no longer works but one can use the average module irradiance for the shading scenario G_{avg} to calculate the SAF, as in Equation 12.11.

$$SAF = \frac{G_{max} - G_{avg}}{G_{max} - G_{min}} \quad (12.11)$$

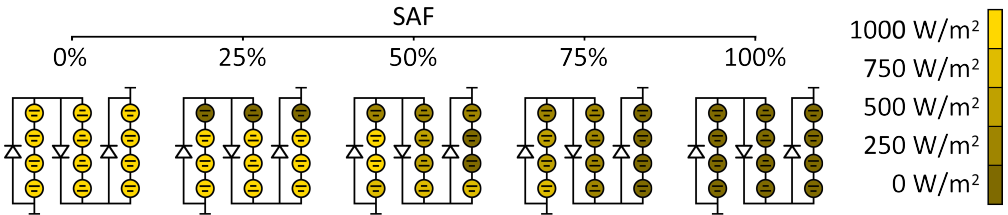


Figure 12.7: Shading scenarios with diverse shaded area fractions (SAF) to illustrate the SAF concept. The examples correspond to a 12-cell module with 3 bypass diodes and considering 5 irradiance levels.

The SAF takes values between 0 and 1 (or 0% and 100% when expressed in percentage terms). Figure 12.7 displays several shading scenarios with diverse SAFs to illustrate the concept. The first SAF value different than 0 corresponds to the scenario when only one cell receives the 2nd highest irradiance and the remaining cells receive the highest irradiance. For an 18-cell module with 5 irradiance levels (0, 250, 500, 750, and 1000 W/m²), the average irradiance of this scenario would be 986.1 W/m² corresponding to a SAF of 1.4%. This value is equal to the step between SAFs ΔSAF . The number of SAFs s_{SAF} depends on the number of irradiance levels and the number of cells in the module c , as shown in Equation 12.12.

$$s_{SAF} = 1 + c \cdot (i - 1) \quad (12.12)$$

Since the SAF takes s_{SAF} values between 0 and 1, the step between SAFs can also be calculated as $\Delta SAF = \frac{1}{s_{SAF}-1}$.

The shading scenarios are grouped based on their SAF, thus the shading scenarios receive a physical meaning, but several power points can be present per shaded area. To obtain a single point, a weighted average of the output power of the shading scenarios with the specific SAF is considered. The weights are determined based on the counts of each scenario. If this value is normalized by the nominal power of the PV module as in Equation 12.13, one obtains the normalized weighted average of power per SAF $\overline{P_{SAF}}$.

$$\overline{P_{SAF}} = \frac{1}{P_{nom}} \frac{\sum_{sc=1}^{N_{uq,SAF}} P_{sc} \cdot n_{sc}}{\sum_{sc=1}^{N_{uq,SAF}} n_{sc}} = \frac{1}{P_{nom}} \frac{\sum_{sc=1}^{N_{uq,SAF}} P_{sc} \cdot n_{sc}}{N_{SAF}} \quad (12.13)$$

In the equation, $N_{uq,SAF}$ corresponds to the number of unique scenarios with a specific SAF while N_{SAF} is the number of shading scenarios per SAF.

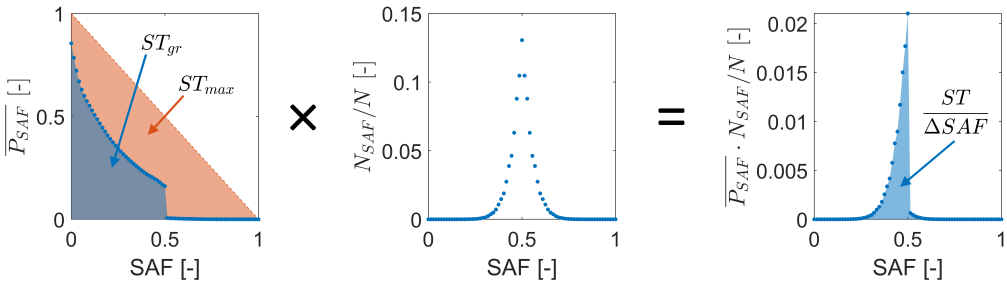


Figure 12.8: Differences between the shading tolerability-related concepts presented in this work. The dotted line in the left plot represents an example of how the shading tolerability is visualized. The area under that curve is ST_{gr} . In that same plot, the dashed red line visualizes the shading tolerability for a module with the maximum ST ST_{max} . The two remaining plots indicate the relation between these quantities and ST. Results were obtained for an 18-cell series-connected module and considering 5 irradiance levels.

The dotted line in the left plot of Figure 12.8 shows an example of how shading tolerability can be visualized. Only one point is present per SAF that represents the average weighted power for each shading fraction. The obtained trend is similar to the ones reported in Figures 6 and 7 of [469], but our method results in a continuous line instead of points of random scenarios. Additionally, the dashed red line indicates $\overline{P_{SAF}}$ for a module with the maximum ST ST_{max} . The two remaining plots indicate the relation between the visualization of the

ST and the definition of ST. In the ST definition, all shading scenarios have the same weight. However, when visualizing the ST, not all scenarios have the same importance since a single point is plotted per SAF and the number of scenarios per SAF differs. The middle graph in Figure 12.8 represents this concept. It shows the number of shading scenarios per SAF N_{SAF} normalized by the total number of shading scenarios N . This is a Gaussian-style curve whose maximum is at 0.5. When this graph is multiplied by the weighted averaged power and then integrated, one obtains the ST multiplied by the SAF step ΔSAF .

One can identify an additional parameter related to the shading tolerability in Figure 12.8: the integral of the shading tolerability visualization. This value is equivalent to the area under the curve and can be referred to as ST graph ST_{gr} . As opposed to ST which assigns the same weight to all scenarios, ST_{gr} assigns the same weight to all $\overline{P_{SAF}}$, therefore the weight of each shading scenario will depend on its SAF. This metric will not be employed in this work to avoid confusion with ST and $ST\%$. However, when visualizing the ST is not feasible, ST_{gr} can provide additional insights to ST .

12.2.4. VALIDATING THE CALCULATOR

Once the calculator is developed, it is essential to verify its accuracy before proceeding any further. This cannot be performed with measurements. However, one can use the theoretical limit for ST found by Ziar et al. [457]. In their manuscript, they showed that the general expression for ST is:

$$ST_{(i \rightarrow \infty, c_{ser})} = \lambda_{(i \rightarrow \infty, c_{ser})} \frac{1}{c_{ser} + 1} \quad (12.14)$$

where c_{ser} is the number of series-connected PV cells and λ is a coefficient that depends on the PV module's design and manufacturing.

To validate our calculator, we will calculate the shading tolerability of a hypothetical 12-cell ($c_{ser} = 12$) mono-crystalline silicon series-connected PV module with no bypass diodes. The irradiance limits are set between 0 and 1000 W/m² and the irradiance levels are increased progressively. If the approach is correctly implemented, ST should converge to $\frac{1}{12+1} = 0.077$. Since the value of λ for this module is unknown, we have assumed its minimum value, which is 1. Considering the layout of the hypothetical PV module, a low λ is expected rather than a high one since in this module the shaded cells affect the performance of sunny cells. This assumption could however cause a disagreement with the real value. The results are shown in Figure 12.9, where the dashed line represents the theoretical value. One can observe how the ST increases with the irradiance levels and its value approaches the theoretical one. Irradiance levels higher than 18 could not be simulated due to computational limitations, but the increasing trend indicates that ST has not converged yet and is approaching its theoretical limit.

12.3. SENSITIVITY ANALYSIS

With the ST calculator ready, only one parameter remains to be determined: the optimum irradiance levels. For that, one needs to fix the limits of irradiance. Since ST is meant as a figure of merit, the STC irradiance (1000 W/m²) is selected as the maximum irradiance incident on each cell. To determine the minimum irradiance, we try to emulate a shading condition in which the sun is blocked. In this case, the irradiance incident on the module would correspond to the diffuse component, which during clear sky conditions is about

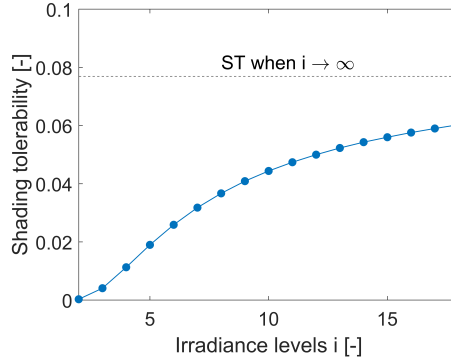


Figure 12.9: Shading tolerability for a 12-cell series-connected module without bypass diodes as the irradiance levels increase. The dashed line represents the theoretical limit of ST when the irradiance level is infinite.

10% of the global irradiance [261]. Overall, the shaded cells will receive irradiance values between 100 W/m^2 and 1000 W/m^2 . This sets ST_{max} to 0.55.

The next step consists of determining how many irradiance values are considered between these minimum and maximum. This is parameter i from Equation 12.2. Ideally, this value should be infinite, while computationally it should be as low as possible. A sensitivity analysis is performed to find the appropriate number of irradiance levels. To ease the computational constraints, the analysis is performed using hypothetical modules.

Figure 12.10a represents a hypothetical module of 12 monocrystalline silicon cells connected in series forming 3 BD strings due to the presence of one BD in parallel to every 4 cells. Figure 12.10b depicts the ST for this hypothetical module for an increasing number of irradiance levels between 100 and 1000 W/m^2 . The module produces less than its nominal power when not shaded because the ambient temperature is set to 25°C . This decision was taken while developing the ST calculator so that the response of the modules to temperature could be considered, which is also aligned with the previous works on ST [468]. Note as well that the module produces $\approx 10\%$ of its nominal power when it is fully shaded because shaded cells are receiving an irradiance of 100 W/m^2 .

As the number of irradiance levels increases, the curves gradually converge. These curves exhibit more significant variations for SAFs below ≈ 0.6 . At a shaded area of approximately 0.05, a kink is observed when considering 3 irradiance levels. This kink persists until 6 irradiance levels are considered.

Since the aim is to determine the optimum number of irradiance levels, we will study how the different parameters of a PV module affect the ST , altering only one at a time. The hypothetical module sketched in Figure 12.10a composed of 12 series-connected monocrystalline silicon cells with 3 BD strings will be employed as the default. The modified parameters are:

- Number of cells in series. The modules can have 12, 18, or 24 cells connected in series, being 12 the default value.
- Number of BD strings, determined by the number of BD. Bypass diodes in the module can be 0, 1, 2, 3, 4, and 12, being 3 the default value.
- Cell material/technology. Modules can be composed of mono-crystalline silicon cells

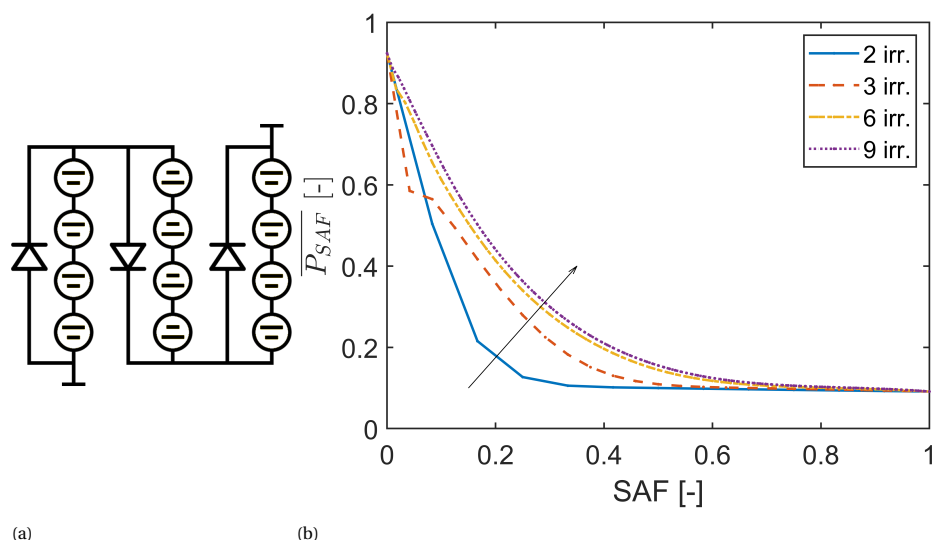


Figure 12.10: (a) Hypothetical PV module consisting of 12 series-connected cells and 3 BD strings. (b) Graphical shading tolerability for the hypothetical module in (a) for an increasing number of irradiance levels.

(mono c-Si) with Aluminium Back Surface Field (Al-BSF), poly-crystalline silicon cells (poly c-Si) with Al-BSF, mono c-Si IBC (Interdigitated Back Contact) cells, mono c-Si PERC (Passivated Emitter and Rear Contact) cells, amorphous silicon (a-Si) cells, CIS (Copper Indium Selenium) cells or tandem a-Si/ μ c-Si cells, mono c-Si with Al-BSF being the default material/technology. A commercial module per category is selected as a reference to determine the electrical values.

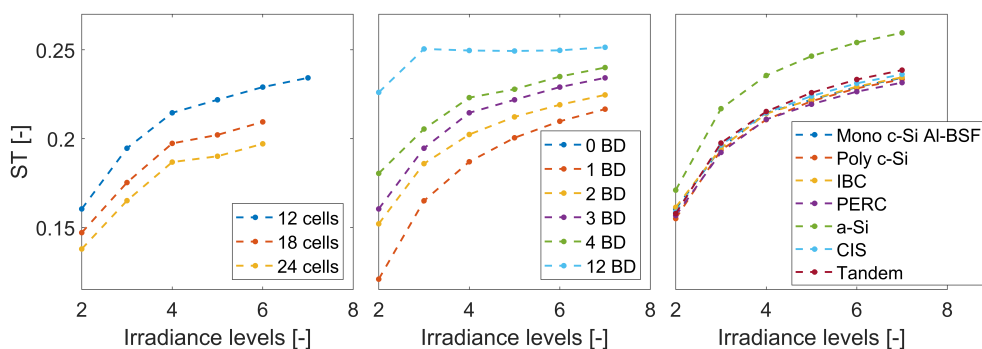


Figure 12.11: Effect of PV module parameters on the ST as a function of the irradiance levels: (a) number of cells, (b) number of bypass diodes, and (c) cell material/technology.

Figure 12.11 shows the main results of this analysis. Starting with the effect of the number of cells in Figure 12.11(a) and bypass diodes in Figure 12.11(b), the irradiance levels do not influence the trends. In both cases, the ranking of the STs when 2 or 5 irradiance levels

are considered is the same. In other words, there is no crossing of the lines that indicates a change in the ranking of the modules. Moreover, the variation of these parameters has the expected effect: a lower ST when increasing the number of cells in series, which is in line with the theoretical formulation, and a higher ST when increasing the number of bypass diodes. There is no difference at the module level between having one or no bypass diodes (there is an overlapping of the curves), although when connecting modules in series the bypass diode will be beneficial.

When examining the cell material/technology in Figure 12.11(c), the irradiance level does have an impact. Compare, for instance, the ST of the tandem module for 2 or 4 irradiance levels. A shift can be observed in the module rankings, except for a-Si which is scattered. This ranking remains consistent for irradiance levels higher than 5. It is important to acknowledge that these results may not be entirely robust. To conduct this analysis, we utilized cell characteristics from commercial modules to construct the series-connected hypothetical modules but did not consider the difference in module layout related to each cell material/technology. For instance, thin-film modules tend to have a higher number of cells connected in series than mono-crystalline modules. Nonetheless, we strived to present the best possible outcome.

So far, only the results from hypothetical modules with a reduced number of cells in series have been shown, but the calculator needs to be applied to commercial PV modules. From the previous analysis, a proper comparison of the ST of PV modules requires 5 irradiance levels or more. The ST calculation of a 72-cell series-connected module with 3 bypass diodes and 5 irradiance levels implies around 10^{12} unique scenarios to simulate, which require a computational memory of 32,000 GB.

Modifications to this approach need to be conducted to make the ST computation feasible. Inspired by [469], a solution could be to consider a random percentage of these scenarios. This, however, would increase the computational burden of the methodology. Therefore, alternative approaches that use the symmetry of the problem can be employed. These will be elaborated upon in the next section.

12.4. COMPUTATIONAL BURDEN REDUCTION

The previous section has shown that at least 5 irradiance levels are needed to compare PV modules accurately. This section explains the changes made to the methodology presented in Section 12.2 towards this objective.

Unless stated otherwise, the computational times reported in this section were completed in MATLAB R2021b and a personal computer with a Windows 10 Enterprise operating system, Intel(R) Xeon(R) W-2223 CPU@3.60GHz and 8 GB of RAM. The values were obtained for the Bisol BMU 260 W_p 60-cell conventional series-connected poly-crystalline module with three bypass diodes [482].

12.4.1. OBTAINING THE UNIQUE SCENARIOS PER SAF

The required computational memory is the main limitation when running the calculator for more than 2 irradiance levels. The bottleneck is the calculation of the unique shading scenarios of the module, which are all calculated at once. To lessen the memory burden, the scenarios were calculated per SAF. After obtaining the unique shading scenarios per string, all the SAFs were identified. The number of SAF depends on the number of cells in the module and the irradiance levels, as determined by Equation 12.12.

Knowing the values of SAF, one can now find the unique scenarios for each SAF. For that, the shaded area fraction of each of the unique shading scenarios per string is calculated. With that, one can identify which combinations of string scenarios result in a certain SAF. That avoids storing all the unique shading scenarios of the module at once. The order of the steps as depicted in Figure 12.1 had to be altered to implement this change. The flowchart in Figure 12.12 shows the new methodology.

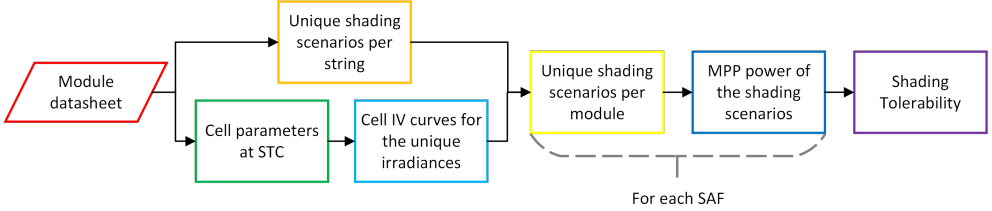


Figure 12.12: Computationally optimized flowchart of the shading tolerability calculator.

12.4.2. CONSIDERING ONLY THE MOST COMMON SCENARIOS

ST by definition considers that all shading scenarios have the same occurring chance. However, when simplifying the calculation using the unique scenarios and their counts in Equation 12.10, the unique scenarios with the highest counts will have the highest occurring chance and influence. Considering this, one can neglect the scenarios with the lowest counts as they will not influence the final output. Then the question arises of how many scenarios can be neglected. The number of ignored scenarios depends on the SAF, as the number of scenarios per SAF is unequal. For instance, when $SAF = 1$ (maximum power output), there is only one scenario with one count and that cannot be ignored as it is relevant for the graphical ST. Hence an approach is needed to find the number of ignored scenarios per SAF $N_{ign,SAF}$.

The objective is to obtain the same $\overline{P_{SAF}}$ despite ignoring the power produced by certain scenarios. Therefore, the summation in Equation 12.13 can be rewritten to distinguish between ignored and not-ignored (or considered) scenarios as in Equation 12.15.

$$\begin{aligned}
 \overline{P_{SAF}} &= \frac{1}{P_{nom}} \frac{\sum_{sc=1}^{N_{uq,SAF}} P_{sc} \cdot n_{sc}}{\sum_{sc=1}^{N_{uq,SAF}} n_{sc}} = \\
 &= \frac{1}{P_{nom}} \frac{\sum_{sc=1}^{N_{uq,con}} P_{sc} \cdot n_{sc} + \sum_{sc=1}^{N_{uq,ign}} P_{sc} \cdot n_{sc}}{N_{SAF}} = \\
 &= \frac{\overline{P_{con}} \cdot N_{con,SAF} + \overline{P_{ign}} \cdot N_{ign,SAF}}{N_{SAF}}
 \end{aligned} \tag{12.15}$$

In this equation, $N_{uq,con}$ is the number of unique considered scenarios per SAF, while $N_{uq,ign}$ is the number of unique ignored scenarios per SAF, and they add up to $N_{uq,SAF}$. $N_{con,SAF}$ is the number of considered scenarios per SAF and together with $N_{ign,SAF}$ add up to N_{SAF} as in Equation 12.16.

$$N_{SAF} = N_{con,SAF} + N_{ign,SAF} = \sum_{sc=1}^{N_{uq,con}} n_{sc} + \sum_{sc=1}^{N_{uq,ign}} n_{sc} = \sum_{sc=1}^{N_{uq,SAF}} n_{sc} \quad (12.16)$$

$\overline{P_{con}}$ and $\overline{P_{ign}}$ are the normalized weighted average of power of the considered and ignored scenarios per SAF, respectively. They are defined in Equation 12.17 by substituting xx with con or ign . This equation is equivalent to Equation 12.13 for $\overline{P_{SAF}}$ but includes only some of the shading scenarios per SAF instead of all of them.

$$\overline{P_{xx}} = \frac{1}{P_{nom}} \frac{\sum_{sc=1}^{N_{uq,xx}} P_{sc} \cdot n_{sc}}{\sum_{sc=1}^{N_{uq,xx}} n_{sc}} \quad (12.17)$$

$\overline{P_{con}}$ is the approximation of $\overline{P_{SAF}}$ that will be employed from now on. Therefore, it has to be within a margin ϵ_{max} of $\overline{P_{SAF}}$, as expressed in Equation 12.18. This margin can be interpreted as the maximum error that one can assume for $\overline{P_{SAF}}$.

$$\overline{P_{con}} = \overline{P_{SAF}} \pm \epsilon_{max} \quad (12.18)$$

Similarly, $\overline{P_{ign}}$ is deviated from $\overline{P_{SAF}}$ by a certain amount $\frac{\Delta P}{P_{nom}}$. ΔP is the deviation from $\overline{P_{SAF}}$ caused by the power produced by the ignored scenarios. In the worst-case scenario, the ignored scenarios produce the maximum power for that SAF, and $\overline{P_{SAF}}$ is the minimum power for that SAF. In that case, ΔP takes its maximum value equal to the difference between the maximum P_{max} and minimum P_{min} power produced at that SAF.

Assuming that only the scenarios with the lowest and highest power outputs will be ignored is quite conservative, thus a factor $f_{dP} \in [0, 1]$ can be included for a more realistic computation:

$$\overline{P_{ign}} = \overline{P_{SAF}} \mp \frac{\Delta P}{P_{nom}} = \overline{P_{SAF}} \mp \frac{f_{dP} \cdot (P_{max} - P_{min})}{P_{nom}}. \quad (12.19)$$

Note that this equation has a \mp sign instead of a \pm sign as in Equation 12.18. These signs are dependent on each other and indicate that if $\overline{P_{con}}$ is larger than $\overline{P_{SAF}}$, then $\overline{P_{ign}}$ has to be smaller, and the other way around. This ensures that the relation between these variables stated in Equation 12.15 holds.

Merging the previous equations into Equation 12.15, one reaches the following equation:

$$\overline{P_{SAF}} = \frac{(\overline{P_{SAF}} \pm \epsilon_{max}) \cdot (N_{SAF} - N_{ign,SAF}) + (\overline{P_{SAF}} \mp \frac{\Delta P}{P_{nom}}) \cdot N_{ign,SAF}}{N_{SAF}} \quad (12.20)$$

The number of ignored scenarios $N_{ign,SAF}$ can be isolated from this equation and expressed independently of $\overline{P_{SAF}}$ as in Equation 12.21.

$$N_{ign,SAF} = \frac{\epsilon_{max}}{\frac{\Delta P}{P_{nom}} + \epsilon_{max}} \cdot N_{SAF} \quad (12.21)$$

This approach allows us to identify the shading scenarios that can be ignored before computing the produced power.

N_{SAF} is obtained while computing the scenarios. A value of 0.0002 was selected for ϵ_{max} , corresponding to a maximum error of 0.05 W for a 260 W module. f_{dp} was set to 1 to consider the worst-case scenario. Determining P_{min} and P_{max} implies identifying the shading scenarios producing the least and the most power for a certain SAF without computing the power for all scenarios.

In series-connected modules with bypass diodes, the shading scenario that produces the least power occurs when all strings generate minimal power. This scenario has the highest number of low-irradiated cells possible for that SAF. Among such scenarios, the one where the cells within each string experience the most varying incident irradiances will produce the least power.

On the contrary, to identify the scenario producing the most one needs to identify first the scenarios where the cells within each string are more evenly distributed. Scenarios where the cells within each string have the most similar incident irradiance. That condition alone is not enough, so the scenarios are filtered by selecting the ones with the lowest number of low-irradiated cells.

With $N_{ign,SAF}$ determined, the $N_{uq,SAF}$ unique scenarios are sorted based on their counts. Those with the lowest counts are ignored until their cumulative count reaches $N_{ign,SAF}$. The remaining scenarios are the ones considered for the calculation.

Figure 12.13 displays the reduction in the number of computed scenarios for each SAF when applying this methodology. 75% of the shading scenarios were ignored without affecting the ST methodology. The number of ignored scenarios depends on the SAF, being the highest for $SAF = 50\%$. After implementing this approach, the computational time was reduced from 4 hours to 1.2 hours for series-connected modules and 3 irradiance levels.

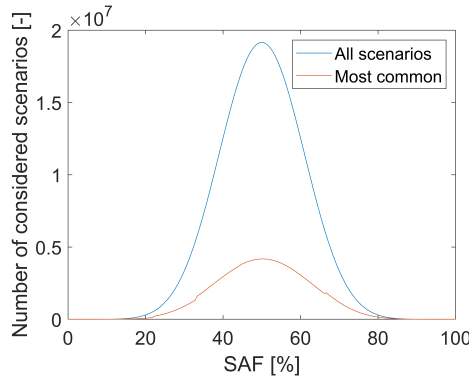


Figure 12.13: Number of scenarios for which the power has to be computed when considering all shading scenarios and when considering only the most common ones. Results are shown for 4 irradiance levels and a conventional 60-cell series-connected module with three BD strings.

12.4.3. OPTIMIZING THE CODE

To further reduce the computational time, three approaches were implemented which optimized the MATLAB code without having any effect on the results.

The construction of shading scenarios at the module level is achieved by combining the shading scenarios at the string level. For a module with three bypass diode strings, a triple for loop computes all the string combinations leading to a unique scenario and retaining

only those with the SAF of interest. This time-inefficient methodology was initially implemented in Subsection 12.4.1 to minimize memory requirements. Within a bypass diode string, and when considering more than two irradiance levels, multiple scenarios may exhibit the same SAF at the string level. Therefore, rather than looping through all string combinations, the code was optimized to loop through all unique SAF levels at the string level. This significantly reduced the computational time required to calculate the scenarios for four irradiance levels, from an estimated 4000 hours to less than 4 hours.

The result of every SAF iteration was the power produced by all the computed shading scenarios and their corresponding counts. Since it was not possible to know beforehand the number of scenarios that would be computed, storing that data increased the computational time. Considering the definition of $\overline{P_{SAF}}$ from Equation 12.13, only the sum of those counts and the power weighted by those counts are needed. By adapting the code and storing only those two values for each SAF, the computational time was reduced by almost 2 hours for four irradiance levels.

When computing the scenarios on the module level for each SAF, symmetry can be found for complementary SAF values. The identification and counts of the scenarios with a SAF of 20% are equivalent to those with a SAF of 80%. This is shown in the middle plot of Figure 12.8. As an example, Figure 12.14 represents two shading scenarios that exhibit duality. While one has only 3 shaded cells in one string, the other has only 3 unshaded cells in one string and the remaining shaded. By exploiting this symmetry, the computational time needed to identify the unique shading scenarios for four irradiance levels was reduced from less than 4 hours to half of that time.

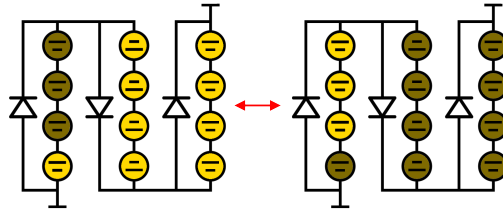


Figure 12.14: Example of two dual scenarios for a 12-cell conventional series-connected module with three bypass diodes.

12.4.4. REDUCING THE NUMBER OF SCENARIOS

Despite the implementation of these approaches and the achieved time reduction during the identification of the unique shading scenarios, more than 350 hours were needed to compute the power of the unique scenarios with 4 irradiance levels. On average, only 6 ms were needed to compute the power of one shading scenario, but there were still an excessive number of scenarios (2.2×10^8). Three different approaches were implemented to further reduce the number of scenarios.

The initially selected values for ϵ_{max} (0.0002) and f_{dP} (1) were overly conservative. The values were changed to 0.008 for ϵ_{max} and 0.25 for f_{dP} to reduce the computational burden. The approximation was suitable for SAFs above 40%, with errors below 10^{-3} in $\overline{P_{SAF}}$, but higher for SAFs below 40%, as seen in Figure 12.15a. The errors were calculated relative to the initially selected values for ϵ_{max} and f_{dP} . A sensitivity analysis was conducted to determine the values of ϵ_{max} and f_{dP} for each SAF that reduce the number of scenarios and

ensure that $\overline{P_{SAF}}$ does not change beyond the third decimal point. Figure 12.15a presents the error for the new values, with the two bottom plots showing the selected values of ϵ_{max} and f_{dP} for each SAF. This approach reduced the time needed to compute ST for 3 irradiance values from 1.2 hours to 0.35 hours. For 4 irradiance levels, the number of scenarios was reduced from 2.2×10^8 to 3.7×10^7 .

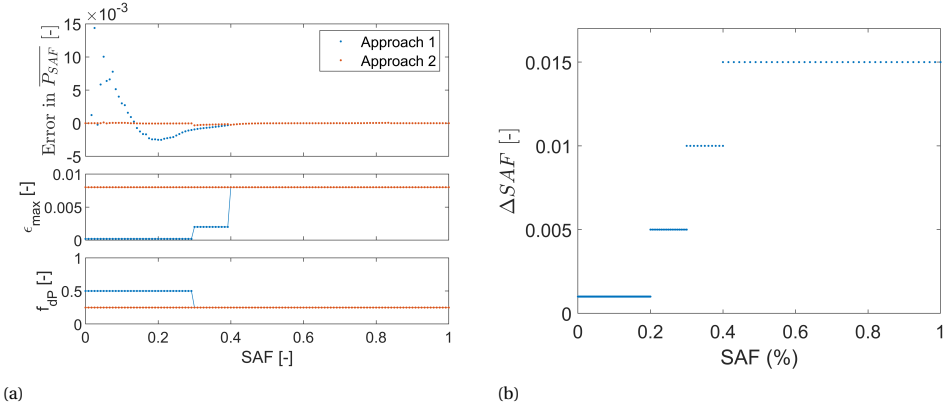


Figure 12.15: (a) Error in $\overline{P_{SAF}}$ for each SAF for two different approaches that affect the number of ignored scenarios. The number of ignored scenarios depends on the values of ϵ_{max} and f_{dP} , which are shown in the bottom plots. (b) Step in SAF values for different SAF.

The second approach consisted of considering fewer SAFs. The number of SAFs increases with the number of irradiance levels and cells (Equation 12.12). More SAFs are considered as the number of irradiance levels increases. While this is beneficial for SAFs below 40%, beyond that point there are virtually no changes in the ST graph. However, as the SAF approaches 50%, the number of scenarios increases significantly. Considering this, one can ignore certain SAFs by altering the step between SAFs for different SAF values. Figure 12.15b plots the values selected for the step as a function of the SAF. Since the trapezoidal method used to obtain ST_{gr} requires the same resolution for all points, the $\overline{P_{SAF}}$ results are interpolated for all SAFs. For 4 irradiance levels, this approach reduced the number of scenarios from 3.7×10^7 to 2.1×10^7 .

The last implemented approach consisted of identifying the scenarios that yield the same power output despite not being equivalent. Those scenarios fulfill two conditions: 1 AND (2 OR 3), being the conditions: 1) the scenarios have the same number of cells with each irradiance level but are differently distributed along the module; 2) the scenarios have more than one low-irradiated cell in all the strings; and 3) the scenarios have no low-irradiated cells. The first condition ensures that the cell-level IV curves are the same, although differently distributed. With at least one low-irradiance cell in each string, that cell will limit the current, limiting all the strings similarly. No steps will be present in the IV curve. We note that the number of low-irradiated cells needed for this condition to be fulfilled depends strongly on the breakdown voltage of the module. The lower the breakdown voltage (in absolute terms), the higher the minimum number of low-irradiated cells required in each string. On the other hand, if there are no low irradiated cells, the current is not highly limited in any of the strings, resulting in the same power output. For 4 irradiance levels, this approach reduces the number of scenarios from 2.1×10^7 to 7.9×10^6 .

12.4.5. OVERVIEW

After implementing these approaches, ST could be computed for 4 irradiance levels within a reasonable time of 16 hours and the memory requirements of a standard computer. Figure 12.16 provides an overview of the steps described in this section and their effect on the computational time. When running the method in the TU Delft supercomputer *delftblue* [483] using 4 nodes and 2 GB per node, the computational time was reduced to around 4 hours. For 3 irradiance levels, the computational time dropped from 4 hours to 4 minutes in a standard computer.

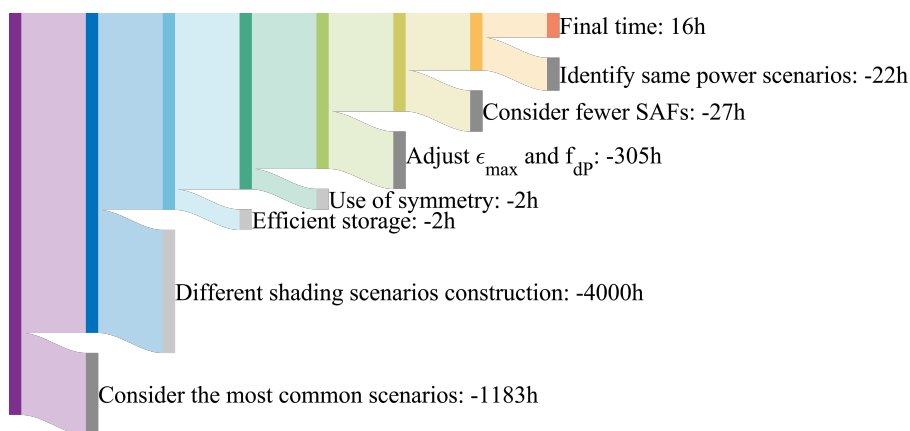


Figure 12.16: Impact on the computation times of the steps taken to reduce processing duration for ST calculation using 4 irradiance levels. The dark grey bars indicate the steps taken to reduce the number of shading scenarios. The light grey bars indicate the steps affecting the obtention of the unique shading scenarios. For aesthetic purposes, the width of the bars is not proportional to the actual time values. Graph performed using [484].

Despite all the efforts, applying the methodology for 5 irradiance values was out of reach. The computational resources required were too large. Obtaining the reduced form of the scenarios for a SAF of 50% (the most computationally intensive) takes more than 7 hours and more than 60 GB of memory.

Despite that, several points of $\overline{P_{SAF}}$ for a commercial module were computed for 5 irradiance levels. Figure 12.17 shows that plot together with the graphical ST for different irradiance values. Despite the incompleteness of the 5 irradiance levels curve, one can appreciate how the curves overlap after a certain SAF of around 40%. As already observed, most changes occur for small SAF values. The last computed point for 5 irradiance levels, corresponding to a SAF of 29.6%, took 18 GB of memory and more than 2 days to compute in the *delftblue* supercomputer.

12.5. RESULTS AND DISCUSSION

With the calculator ready to compute the ST for 4 irradiance levels, one can generate a database of ST values. The PV modules database from the California Energy Commission (CEC) is used for this purpose [485]. It contains the datasheet information of more than 17,000 PV modules. This allows us to identify the effect of various datasheet parameters on the ST.

One important parameter that is not available in the CEC database is the breakdown

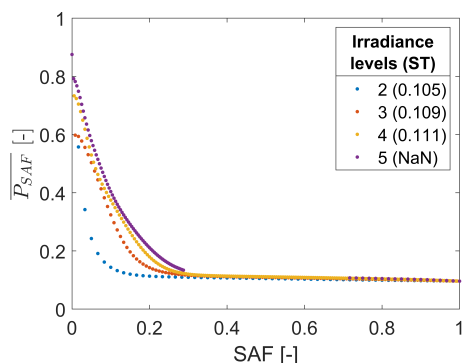


Figure 12.17: Graphical ST for the series-connected module under study for increasing number of irradiance levels.

voltage. The breakdown voltage V_{br} is essential to model the reverse bias region of the IV curves. This parameter depends on the cell material/technology, and has a strong effect on the ST, as shown in Figure 12.18. The effect is mostly visible for small SAF values since the breakdown voltage has a high impact on the IV curve when a few cells are shaded.

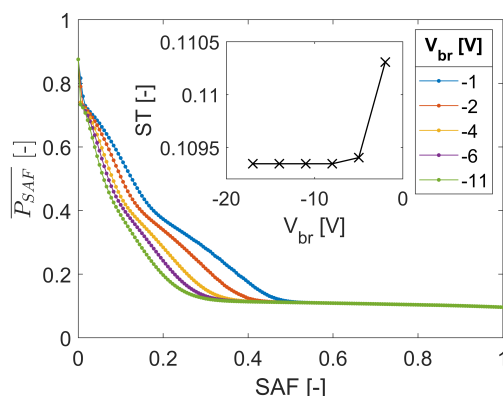


Figure 12.18: Effect on the variation of the breakdown voltage on the graphical ST and the ST value (in the inset). The results were obtained for the Bisol module employed in the previous section and 4 irradiance levels.

As seen in the inset of Figure 12.18, the ST is not affected for breakdown voltages smaller than -8 V. This is beneficial when assuming the breakdown voltage of crystalline silicon modules, as the range of values is as wide as -10 V to -30 V [474, 486–488]. A breakdown voltage of -10 V or -30 V does not impact the ST. It is reasonable to assume a value of -18 V for all non-IBC and non-PERC crystalline silicon modules considered in this work, with confidence that this assumption will not affect the results.

In fact, a different story is observed with other materials/technologies. IBC modules tend to have an earlier breakdown at smaller voltages, with the absolute value of breakdown voltage values generally falling below 6 V [487]. An incorrect breakdown voltage estimation affects the ST value. However, its value can only be known when measuring the module's IV curve. The same situation occurs with CIGS, and CdTe cells [489–491]. An average value of

-4.5 V is chosen for all these cases.

Ultimately, the ST values of 423 commercial PV modules with series-connected cell layouts and 3 BD strings from the CEC database were calculated. The database was filtered to avoid considering modules of the same model but with different powers. ST ranged between 0.10 and 0.12 and consequently $ST_{\%}$ was between 18.4% and 21.5%. This suggests that when a module with a series-connected cell layout is shaded, it generally produces 20% of the power that an ideal module (one with an ST of ST_{max}) would generate under the same shading conditions.

When comparing these values with the literature, Klasen et al. reported a \bar{P}_{ps} value of 0.207 for random shading of PV modules [469, Fig. 8.]. Since they obtained these results by setting the minimum irradiance that cells could receive to 0 W/m², $\bar{P}_{ps} = ST_{\%}$. Hence, we can directly compare our results with theirs because $ST_{\%}$ allows comparing ST that were obtained with a different minimum irradiance. Overall, there is an excellent agreement between the values reported by Klasen et al. and those presented in this work. There is also a strong alignment with the work performed by Fauzan et al. They reported values of average normalized power for partial shading for conventional modules in the range of 0.199-0.271 depending on the type of shading [470, Figure 4.].

Correlations between different module parameters were explored to see their impact, if any, on ST. The results are presented in Figure 12.19. As expected, a positive correlation is observed between ST and the breakdown voltage (0.4), which is in line with the literature [474]. The lower breakdown voltage reduces the current-limiting effect of the string resulting in a higher power output when a few cells are shaded.

The remaining input parameters show negligible correlation with the shading tolerability, even the ones related to temperature. As shown in Figure 12.19 the correlation between any of the two temperature coefficients (K_V for voltage and K_I for current) and the shading tolerability was negligible (-0.1 and -0.0). These results are opposed to those reported by Mishra et al. [468], where they found that the thermal features of a PV module can influence its ST by up to 12.25%. They reported that modules with a higher absolute temperature coefficient of power will have a higher ST. However, Figure 12.19 does not support such affirmation which could be because other parameters are affecting the ST or because the effect of the temperature coefficients on ST is limited.

Similarly, in [457], a theoretical relation was found between ST and the number of cells. The higher the number of cells in the module, the lower the ST. Although Figure 12.19 weakly supports this statement, the sensitivity analysis performed with hypothetical modules in Section 12.3 shows this relation. This could be again due to other parameters having a stronger influence on ST than the number of cells, as the analysis in Section 12.3 was conducted by altering only one parameter at a time. Alternatively, the relationship between ST and the number of cells may be nonlinear, with the number of cells having a more significant impact on ST when the cell count is low. Further investigation is needed in this aspect.

The database and related analysis have focused on modules with a series-connected cells layout. However, one parameter worth studying with a proven significant impact on shading tolerability [469, 470] is the module layout. As a next step, the shading tolerability of so-called *half-cell butterfly* modules could be studied. Similarly, the impact of parallel strings and the number of bypass diodes could be assessed.

One could also adjust the shading tolerability value to a specific location. ST assigns the same value to all the shading scenarios. However, one could obtain the shading profile of

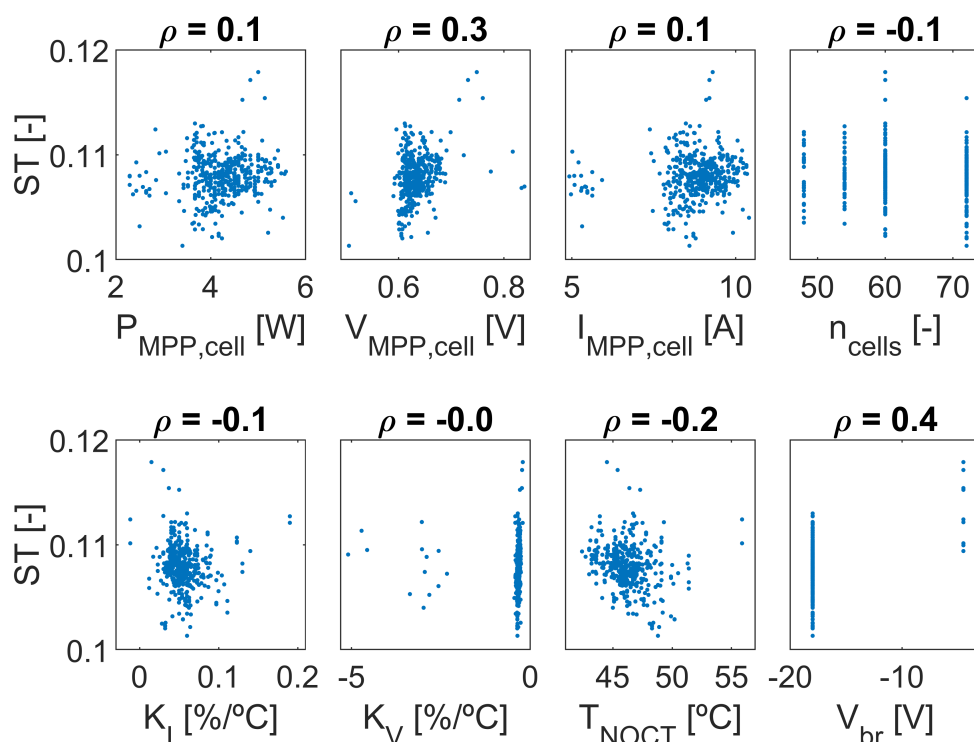


Figure 12.19: ST as a function of several input parameters for the 423 commercial PV modules tested with a series-connected cells layout and 3 BD strings. ρ is the correlation coefficient, K_I is the temperature coefficient for current, and K_V is the temperature coefficient for voltage.

the installation location of a PV module and obtain the probability distribution function of the SAE. When substituting that graph with the middle plot from Figure 12.8 and following the approach illustrated in that figure, one can obtain the ST for a specific location.

Additionally, the adaptation of the ST definition for bifacial modules could be considered. The easiest option could be to compute the ST independently for each face by assuming a constant irradiance value on the other face, although it may not be rigorous since the power is not independently produced by each face of the module. It remains to be explored how to best extend this concept to reflect best the real conditions.

Lastly, the relationship between ST and energy yield is unknown. No studies to date have correlated energy yield gains with increases in ST. Such research would extend the application of ST and demonstrate its benefits in real-world scenarios.

12.6. CONCLUSIONS

This work developed a shading tolerability (ST) tool to efficiently evaluate the shading tolerability of PV modules. The tool employs an IV curve calculator with an error below 4.5% for the power at the maximum power point and a probabilistic approach to consider all possible shading scenarios. A sensitivity analysis indicated that at least five irradiance levels are

necessary for accurate comparisons between PV modules, which require significant computational resources. To address this, several methods were developed to reduce computational demands by leveraging the probabilistic and symmetric nature of the problem, enabling the computation of ST for more than 400 commercial conventional series-connected modules. The obtained values were in line with the literature. A statistical analysis was conducted to find the relation between ST with the module datasheet parameters. This analysis revealed discrepancies with existing literature, particularly regarding the impact of thermal coefficients and the number of cells on ST. The calculator could be extended to other module layouts and technologies to contribute to and extend this discussion.

PERSPECTIVE: EXTENDING THE CALCULATOR TO OTHER LAYOUTS

A key factor that can potentially influence shading tolerability is the module layout [469, 470]. Among the possible options, the *half-cell butterfly layout* is increasing in popularity with some studies predicting a market share higher than 60% by 2030 [492].

A half-cell module, as the name suggests, is a PV module containing half-cells instead of full cells. A half-cell is essentially a solar cell cut in half. It generates half as much current as its full-cell counterpart, which reduces the resistive losses and the hotspot impact when partially shaded [493–495]. Half-cell modules are generally a series-parallel-series configuration, sometimes referred to as a *butterfly layout* [496]. With this layout, half-cell modules generate an output current and voltage comparable to conventional modules despite the halving of the cell current. Additionally, the series-parallel-series configuration of the half-cell butterfly module effectively creates six BD strings, even though it contains only three bypass diodes, increasing its shading response [495, 497].

Therefore, we aimed to calculate the shading tolerability for butterfly modules. The first step consisted of adapting the IV curve to consider the series-parallel-series interconnection. For implementing the parallel connection, the currents at each voltage point are added up since the voltage across parallel-connected cells is equal [261].

The IV model was validated for butterfly modules using measurements of a Trinasolar Duomax 144 half-cells module. The comparison between the measured and modeled IV curves for three different shading scenarios is presented in Figure 12.20. In the blue-colored scenario, all cells receive an irradiance of 1000 W/m^2 . In the red-colored scenario, one substring is shaded receiving 0 W/m^2 . In the black-colored scenario, two substrings are shaded.

A strong agreement can be observed between the modeled and measured IV curves. This agreement is reflected in the P_{MPP} errors, which are 0.0%, -0.2%, and -4.6% for the blue, red, and black scenarios, respectively. An underestimation in power is observed for all scenarios. As with the validation of the conventional modules, the error increases in value as the average module irradiance decreases. However, the error values are still within a reasonable range.

With the IV calculator ready, the next step consists of identifying the unique shading scenarios for butterfly modules. The new interconnection of the solar cells has to be considered in the process. Besides that, the same logic as explained in Subsection 12.2.1 can be applied to find the equivalent scenarios.

The tool as explained in Section 12.2 was implemented for butterfly modules. Figure 12.21b compares the graph of shading tolerability for 3 irradiance levels of a 12 full-cell module with a series-connection layout (Figure 12.10a) to that of a 24 half-cell one with a butter-

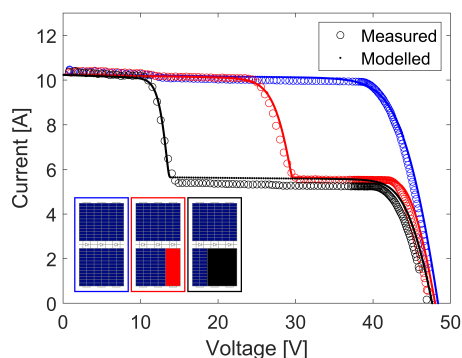


Figure 12.20: Validation of the IV model by showing the difference in IV curves for a butterfly module.

fly layout (Figure 12.21a). Both modules contain monocrystalline silicon cells and 3 bypass diodes.

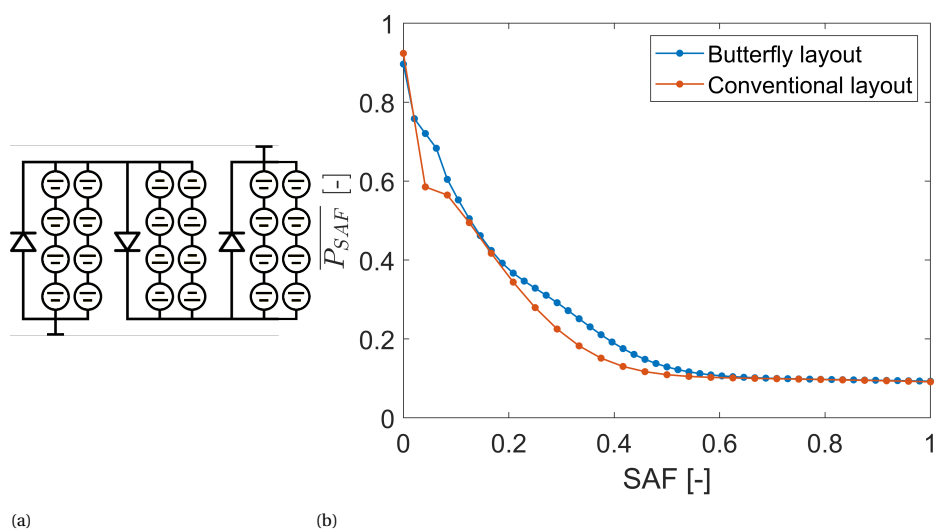


Figure 12.21: (a) Hypothetical 24-half-cell PV module with a butterfly layout and three bypass diodes. (b) Graphical shading tolerability for 3 irradiance levels of two hypothetical modules with different layouts: the butterfly module in (a) and the conventional module in Figure 12.10a.

From the graph, one determines that the ST of the half-cell module is higher than that of the equivalent series-connected one. This is in line with the literature which reports that butterfly modules experience fewer power losses when shaded [469, 470]. This benefit occurs mostly for smaller shaded area fractions (SAFs).

The ST for commercial butterfly PV modules could not be reported due to computational limitations. The memory constraints reported before implementing the changes detailed in Section 12.4 were particularly pronounced for butterfly modules. The number of

unique scenarios for a 120-half-cell butterfly module with 3 irradiance levels is $\approx 1.5 \times 10^{12}$, which requires around 72,000 GB of memory. The methodology explained in [Section 12.4](#) was not fully adapted yet for the butterfly layout hence the results with this layout could not be compared with the ones reported for the conventional layout.

In the future, when the ST concept will take hold, further optimizing the code, switching to low level coding language, and accessing to more powerful computational hardware (e.g. cloud computing) will enable the calculation of the ST metric of any type of PV module topology.

13

CONCLUSIONS AND OUTLOOK

13.1. CONCLUSIONS

This doctoral thesis contributed to the main objective of the Trust-PV project, which was to increase the performance and reliability of PV systems. Within the broad range of topics covered by this objective, this thesis focused on the energy yield modeling of PV systems. Depending on the tools used to create these models or the aspect to be modeled, four distinct blocks were identified: the machine learning block, the analytical models block, the floating photovoltaic block, and the shading tolerability block.

The block of machine learning forecasts started in [Chapter 2](#) by reviewing the literature that employed machine learning (ML) to forecast the power output of photovoltaic systems. This review provided an overview of the most common algorithms, methods, and system characteristics used when forecasting the power output. Based on the information learned, an outlook provided guidelines and recommendations for future research. For instance, the use of a very complex ML model - such as one consisting of a hybrid algorithm with various preprocessing steps - for PV power forecasting is not recommended. The added complexity and time investment does not justify the marginal gain in accuracy. Simpler models offered better explainability and control while maintaining acceptable accuracy.

One identified flaw during the review was that most studies employ at most two systems when developing their algorithms. Since ML algorithms require high-quality data to work, the limited data can be detrimental to the development of accurate models. However, since most devices automatically create and store recorded data, the main issue lies in data sharing rather than data recording. Motivated by this, an extensive search of open-source data was made and gathered in a developed website. [Chapter 3](#) explained this process. Any researcher can access the data from the website and employ it to create forecasting models with several systems.

Related to the previous point, because most researchers focus on systems within their own regions, most ML models forecast the performance of systems in similar climates. However, areas with high potential for photovoltaic energy are often ignored. [Chapter 4](#) explored how climate affects the accuracy of the ML models. The results show that models located in dry areas are easier to forecast than those located in tropical ones, which are the hardest. This behaviour can be explained by the weather variability of each climate.

Another work related to the previously mentioned flaw of focusing on a few number of systems is the uncertainty about whether the developed models will maintain the accuracy for different systems. Moreover, when dealing with a large number of PV systems, developing a machine learning model for each system can be highly time-consuming. [Chapter 5](#) elaborated upon this point by creating a single ML model able to forecast the power output of a fleet of systems. The algorithm achieved higher accuracy than the reference analytical model, but it had a significant drawback. The results were dependent on the magnitude of the production, with smaller errors observed when the power output was lower. This was detrimental especially for the smaller systems and during low-production periods.

To overcome this problem, a different alternative was explored in [Chapter 6](#) based on peer-to-peer communication. The methodology relied on the production of neighboring systems to create the predictions. The power output of analogous PV systems was compared to identify any malfunctions during production. A machine learning model was employed to identify the best parameters that increased the accuracy of the predictions. The main advantage of this method was the independence from weather data since all systems experienced the same local climate.

Sticking with the importance of weather data, we moved to the next block of analytical models. This block initiated in [Chapter 7](#) by presenting the PVMD toolbox, a state-of-the-art simulation framework to predict the PV power of photovoltaic systems. The accuracy of this simulation framework was assessed for residential PV systems. The key finding of this study was the importance of accurate irradiance data, as errors in irradiance were amplified in the PV power predictions.

Although the previous conclusion was derived for residential-scale systems, the importance of accurate irradiance measurements is as large for utility-scale systems. Those systems generally have irradiance sensors allocated throughout the PV plant. However, deciding the amount and location of these sensors is a challenging task for which there are no clear guidelines. [Chapter 8](#) aims to fill this gap by creating an irradiance sensor allocation software tool that suggests the optimal number of irradiance sensors and their best locations for any PV plant considering its topology. The methodology behind the algorithm lies in finding where the irradiance profile is most similar to the average to ensure the best results when those sensors are employed in the predictions.

The following block focuses again on utility-scale PV systems but located in a different environment: water. The topic is first introduced in [Chapter 9](#) where the effect of waves movement, dust, and module temperature is assessed for a floating PV system located in a quarry lake in France. The results showed that this system is loosely influenced by the surrounding water for the three aspects explored.

That is not the case for the offshore floating PV system considered in [Chapter 10](#). That chapter details the effect of waves movement on the power production of a hypothetical floating PV system located on the North Sea. By developing a 3D mechanical model to represent the impact of waves on the floating platform, the effect on both the DC and the AC yield is assessed. The main conclusion is that the losses are the highest when the sea is the roughest.

The previous study was extended in [Chapter 11](#) by using an improved model which allowed to explore the impact of the floater characteristics on the DC losses. The results highlighted the existence of a trade-off between mechanical stability and power mismatch losses. Additionally, depending on the floater characteristics, the response of the floater

experienced an elastic or a rigid response.

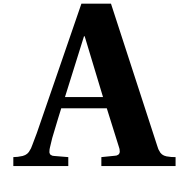
The fourth block on shading tolerability is covered in the last chapter, which focuses again on one aspect affecting especially residential-scale systems: the losses due to partial shading in PV modules. Partial shading decreases the power produced by PV modules and eventually their lifetime. However, there is currently no factor in the PV modules' datasheet that quantifies the power lost due to partial shading. That is not a straightforward task since the shading conditions that a module is subjected to depend on its location. That can be solved by considering all the possible shading conditions using a probabilistic approach, as done by the shading tolerability metric. However, measuring the power lost by all the possible shading conditions is unfeasible. Therefore, in [Chapter 12](#) we developed a tool that provides the shading tolerability of a PV module given its datasheet information. The tool was employed to create a database of shading tolerability of commercial PV modules and to promote its usability.

13.2. OUTLOOK

This thesis presented findings and insights into the energy yield modeling for PV systems with the aim to increase the performance and reliability of PV systems. The results reported in this doctoral thesis demonstrate progress in the field, although there remains considerable scope for further research. Further possibilities for development are:

- One important issue with machine learning models for PV power forecasts lies in the machine learning nature of the algorithms that focus mostly on reducing the larger errors. Since the magnitude of the errors in PV power forecasts depends on the PV power production, the errors are relatively higher during low-production periods. However, accurate forecasts are needed also on winter and cloudy days. This issue could be solved by developing a set of metrics and loss functions that take into consideration the diurnal and seasonal profile of PV power production.
- Both machine learning and analytical models for PV power predictions are sensitive to the input irradiance data. While methodologies based on the peer-to-peer communication can avoid using this data, these methodologies are not always a solution. An alternative could be to obtain irradiance data from satellite images. Machine learning models can predict the cloud position and irradiance in the near future using satellite images. This method can provide irradiance with higher temporal and spatial resolution than that measured by regional meteorological stations.
- The irradiance sensor allocation software tool developed in [Chapter 8](#) needs an improved user interface and added capabilities before commercialization. The most important addition considering the photovoltaic projects under development is the extension to bifacial systems and systems with an East-West configuration.
- In [Chapter 11](#), the impact of waves on the power production was assessed on an offshore system composed of floaters. However, several pilot offshore floating PV projects are testing the use of membranes instead of floaters to place PV modules on. It would be interesting to extend the study to these kind of platforms and assess the impact of mismatch power losses for floating systems composed of membranes.
- Lastly, the shading tolerability calculator could be extended to other layouts such as *butterfly* and *shingle* before the inclusion of the shading tolerability metric in the

datasheet of PV modules. The main limitation in this process is the computational resources needed. Therefore, the code optimization developed in [Chapter 12](#) for conventional modules should be adjusted to these layouts.



REVIEWED ML FORECASTING LITERATURE

Table A.1: Summary of the main characteristics of the reviewed literature. Below the table, one can find the meaning of the abbreviations.

First author	Year	Ref.	Country (Köppen)	Features	Algorithms
Abdel-Nasser	2019	[498]	Egypt (B)	pow	LR dt FFNN LSTM
Abuella	2017	[55]	Australia (B)	pow irr ta h cc r ws wd p t	LR SVR MLP dr
Agoua	2019	[131]	France (C)	pow irr ta h wd	bay regu ELM
Al-Dahidi	2020	[129]	Jordan (B)	pow irr ta h ws t	FFNN ELM
Alessandrini	2015	[123]	Italy (C)	pow irr ang ta cc	regr FFNN ens
Alfadda	2017	[58]	USA (C)	pow ta h cc ws wd t	LR regu SVR
Alkandari	2020	[88]	-	irr ta h r ws wd	LSTM DL hyb
Almeida	2015	[104]	Spain (C)	irr ta h cc ws wd p	dt dr
Alomari	2019	[499]	Jordan (B)	pow irr ta t	BPNN
Anagnostos	2019	[146]	Germany (C)	irr ang tpv cc	MLP
Asrari	2017	[126]	USA (C)	pow	FFNN GA opt
Baharin	2016	[105]	Malaysia (A)	irr ta h ws	SVR dr
Carrera	2020	[81]	South Korea (C)	irr ang ta h cc r ws wd p	LR regu k-NN SVR dt ens
Cervone	2017	[94]	Italy (C)	pow irr ang ta cc	BPNN ens hyb

Table A.1: Summary of the main characteristics of the reviewed literature. Below the table, one can find the meaning of the abbreviations.

First author	Year	Ref.	Country (Köppen)	Features	Algorithms
Chang	2020	[141]	Taiwan (C)	pow	SVR BPNN ANN DL
Chen	2020	[95]	Australia (-)	pow	LR SVR dt FFNN LSTM ens hyb
Chu	2015	[100]	USA (B)	pow	k-NN FFNN GA
Das	2017	[59]	Malaysia (A)	pow irr ta ws	SVR BPNN
De Felice	2015	[76]	Italy (C)	irr ta	SVR
De Giorgi	2016	[93]	Italy (C)	pow irr ta tpv	LS-SVR DL hyb
Do	2016	[67]	-	pow ta cc	regr FFNN
Eseye	2018	[120]	China (D)	pow irr ta h cc ws p	SVR BPNN pre GA PSO
Ferlito	2017	[64]	Italy (C)	pow irr ta cc	LR MARS k-NN SVR dt ANN ELM ens
Gensler	2017	[89]	Germany (C)		MLP LSTM DL hyb
Golestaneh	2016	[130]	-	pow	ELM
Graditi	2016	[138]	Italy (C)	irr tpv	MLP LSTM pre
Gulin	2017	[500]	Croatia (C)	irr ta h ws wd p	MLP
Han	2019	[107]	-	pow ta h p	SVR LSTM dr
Hossain	2017	[35]	Malaysia (A)	pow irr ta tpv ws	SVR FFNN ELM
Huang	2016	[501]	USA (D)	irr ang tpv	bay SVR FFNN
Huang	2019	[60]	Taiwan (C)	pow irr ta	SVR dt MLP LSTM CNN ens
Isaksson	2018	[110]	Sweden (D)	pow irr ta h cc r ws wd p	regu k-NN BPNN ens dr
Jung	2020	[92]	South Korea (-)	irr ta h cc rwsr ws t	LSTM DL
Kazem	2017	[502]	Oman (B)	irr ta	SVR FFNN MLP DL
Konstantinou	2021	[503]	Cyprus (B)	pow	LSTM
Kraemer	2020	[132]	Norway (D)	ang ta h cc r p	ANN DL ens
Kumar	2018	[504]	India (A)	pow irr ta h rwsr ws p	FFNN fuz
Kuzmiakova	2017	[69]	USA (D)	ta h cc ws p	regr LSTM ens dr
Lee	2019	[111]	South Korea (D)	ang ta tpv h cc rwsr ws wd	SVR dt FFNN DL ens dr
Lee	2018	[65]	South Korea (-)	pow irr ta h rwsr ws	LR regu regu SVR ens hyb

Table A.1: Summary of the main characteristics of the reviewed literature. Below the table, one can find the meaning of the abbreviations.

First author	Year	Ref.	Country (Köppen)	Features	Algorithms
Lee	2019	[505]	South Korea (D)	irr ta h cc t	BPNN LSTM
Li	2019	[127]	Australia (B)	irr ta h	SVR ANN PSO opt
Li	2016	[61]	China (C)	ta h r ws p	LR MARS k-NN SVR dt FFNN
Li	2015	[116]	China (C)	pow	BPNN ELM
Li	2020	[119]	Australia (B)	pow irr ta h ws	MLP DL LSTM DL pre
Li	2018	[140]	South Africa (B)	pow ta h	BPNN DL
Li	2016	[142]	USA (C)	pow ang ta ws wd	SVR FFNN
Li	2019	[143]	Belgium (C)	pow	SVR BPNN ANN DL LSTM
Lin	2016	[99]	Taiwan (C)	pow	LS-SVR ANN GA
Lin	2018	[98]	Australia (B)	pow irr ta h ws	LS-SVR BPNN ANN DL hyb pre
Lin	2020	[97]	Australia (-)	pow irr ta r ws	ANN hyb
Maitanova	2019	[506]	Germany (C)	pow ta h cc r	LSTM opt
Majumder	2020	[507]	USA (-)	pow	SVR ANN ELM opt
Massaoudi	2019	[109]	Australia (B)	pow irr ta h ws wd t	bay regu k-NN ens dr
Massidda	2017	[72]	Germany (C)	irr ta cc ws p	MARS
Mishra	2020	[62]	USA (D)	pow ang ta h cc ws p	LR regu LSTM pre opt
Nageem	2017	[56]	India (A)	irr ta h ws p t	SVR
Nayak	2020	[96]	India (B)	irr	SVR dt ANN ELM hyb
Ni	2017	[124]	Singapore (A)	pow irr ta t pv ws wd	ELM
Ogliari	2018	[508]	Italy (C)	irr ta cc rwsr ws wd p t	hyb
Pan	2020	[78]	Australia (B)	irr ta h wd t	SVR opt
Panamtash	2021	[122]	USA (A)	pow	BPNN
Paulescu	2017	[137]	Italy (C)	irr ta cc	LR regr fuz
Pawar	2020	[82]	Australia (C)	irr	SVR
Pierro	2017	[108]	Italy (C)	pow irr ang ta h ws	k-NN MLP pre dr
Pujić	2020	[192]	Denmark (C)	irr ta h cc ws wd p	LR k-NN SVR FFNN ens

Table A.1: Summary of the main characteristics of the reviewed literature. Below the table, one can find the meaning of the abbreviations.

First author	Year	Ref.	Country (Köppen)	Features	Algorithms
Ramsami	2015	[68]	UK (C)	irr ta h rwsr ws wd p	LR FFNN ANN dr
Rana	2015	[77]	Australia (C)	pow irr ta h ws	SVR MLP
Raza	2018	[118]	Australia (C)	pow irr ta h ws	FFNN ANN DL pre
Rosato	2017	[70]	Italy (C)	pow	LR ANN fuz
Rosiek	2018	[103]	Spain (C)	irr ang ta h ws wd p	BPNN pre dr
Semero	2018	[63]	China (D)	irr ta tpv h cc p	LR BPNN fuz dr GA PSO
Shang	2018	[113]	USA (B)		MLP BPNN ANN fuz pre dr PSO
Sharadga	2020	[84]	China (C)	pow	MLP ANN LSTM DL fuz pre
Sheng	2020	[80]	-	irr ta h ws t	bay SVR ANN ELM opt
Soufiane	2020	[121]	-	pow	ANN pre
Sun	2019	[145]	USA (C)	pow cc	CNN
Suresh	2020	[86]	Poland (C)	irr ta tpv ws	LR CNN hyb
Takeda	2017	[125]	Japan (C)	pow irr ta ws	ens
Tang	2016	[509]	USA (C)	irr ta h ws	ANN ANN ELM
Theocharides	2018	[147]	Cyprus (B)	irr ang ta h ws wd	SVR dt FFNN
Torabi	2017	[510]	Portugal (C)	pow irr	BPNN
Touati	2017	[66]	Qatar (B)	irr ta tpv h ws	LR dt dr
Van Tai	2019	[106]	Russia (D)	irr tpv ws	LR FFNN fuz dr
Wang	2019	[87]	Australia (B)	irr ta h ws wd	LSTM CNN hyb
Wang	2018	[79]	Australia (B)	irr ta h cc r ws wd p	SVR BPNN LSTM DL hyb dr
Wang	2020	[115]	USA (B)	irr ta	k-NN BPNN LSTM
Wang	2017	[117]	Belgium (C)	pow	regr SVR BPNN CNN pre
Wang	2019	[135]	USA (C)	pow	CNN
Wang	2016	[511]	China (C)	irr ta h rwsr ws p	LR regr MLP ANN
Wen	2019	[91]	Australia (B)	irr ta h ws t	SVR MLP hyb
Wolff	2016	[114]	Germany (C)	pow irr ta cc	SVR

Table A.1: Summary of the main characteristics of the reviewed literature. Below the table, one can find the meaning of the abbreviations.

First author	Year	Ref.	Country (Köppen)	Features	Algorithms
Yadav	2018	[512]	India (A)	irr tpv	ANN
Yang	2020	[513]	Netherlands (C)	pow	ANN opt
Yin	2020	[101]	China (D)	irr ta h ws	SVR BPNN ELM
Yousif	2017	[514]	Oman (B)	irr ta	SVR MLP DL
Zang	2018	[90]	China (C)	pow irr ta ws	bay SVR BPNN CNN ens hyb pre
Zhang	2015	[134]	-	irr ang ta tpv cc ws	k-NN LS-SVR ANN
Zhang	2018	[85]	Japan (C)	pow cc	MLP LSTM CNN
Zhou	2018	[57]	China (C)	pow tpv t	SVR pre

The five Köppen categories are: A - tropical; B - dry; C - mild temperate; D - continental; and E - polar.

The abbreviations for the features are: pow – PV power; irr – irradiance; ang – solar angles; ta – air temperature; tpv – PV temperature; h – humidity; cc – cloud coverage; r – rainfall; ws – wind speed; wd – wind direction; p – pressure; t – time. Be aware that only the most relevant features have been mentioned, but some authors have included more.

The abbreviations for the ML algorithms have been mentioned in the text. If one or more algorithms from a certain family have been employed and are not one of the popular algorithms, the abbreviation of the family is mentioned. The abbreviations correspond to: regr – regression; regu – regularization; bay – Bayesian algorithms; dt – decision trees; ens – ensemble algorithms; ANN – Artificial Neural Networks; DL – Deep Learning; fuz – fuzzy algorithms; hyb – hybrid algorithms. In the next appendix, the main characteristics of each family are briefly mentioned.

B

MACHINE LEARNING FAMILIES

This appendix outlines the key characteristics of each of the 10 ML families considered, aiming to highlight differences rather than provide detailed explanations.

REGRESSION

With their origin in statistics, regression methods have been the starting point of machine learning thus they have been adopted as part of the discipline [128]. This group is composed of the linear regression method and its variations. This is different than auto-regressive methods, mainly in the way how the linear coefficients are obtained.

Equation B.1 exemplifies the estimation of the PV power \hat{P} at time step t , using a simple linear regression model. Only the irradiance I and the temperature T at the previous time step are considered as model features.

$$\hat{P}(t, \vec{\alpha}) = \alpha_0 + \alpha_1 \cdot I(t-1) + \alpha_2 \cdot T(t-1) \quad (\text{B.1})$$

The linear regression coefficients $\vec{\alpha} = [\alpha_0, \alpha_1, \alpha_2]$ are found by minimizing the loss function \mathcal{L} between estimated and real PV power P , Equation B.2. The loss function can take several forms, with quadratic loss being one of the most popular ones.

$$\min_{\vec{\alpha}} \left(\sum_t \mathcal{L}(\hat{P}(t, \vec{\alpha}) - P(t)) \right) \quad (\text{B.2})$$

REGULARIZATION

A common issue when solving ML problems is *data overfitting*. This occurs when the algorithms fit the testing set so much that they are unable to fit unseen data. To solve this issue, regularization techniques were developed to limit the size of the parameters to be optimized. Regularization of the linear regression from Equation B.1 can be applied by limiting the size of the coefficients $\vec{\alpha}$. This is imposed by adding a penalization term $\mathcal{R}(\vec{\alpha})$ to the function to be minimized. Hence, the objective function in Equation B.2 becomes Equation B.3.

$$\min_{\vec{\alpha}} \left(\sum_t \mathcal{L}(\hat{P}(t, \vec{\alpha}) - P(t)) + \mathcal{R}(\vec{\alpha}) \right) \quad (\text{B.3})$$

B

Depending on how the penalization of the parameters is defined, one encounters several algorithms: Least Absolute Shrinkage and Selection Operator (LASSO) when 1-norm (absolute value) penalization is imposed; Ridge Regression for 2-norm (root squared value) penalization; and Elastic Net for 1-norm and 2-norm penalizations.

BAYESIAN ALGORITHMS

Bayesian algorithms use Bayes' theorem as a basis to develop their models [54]. Bayes' theorem relates the probability of a certain event to the knowledge a priori of conditions related to that event. For example, Bayes' theorem finds $p(P | I, T)$, the probability of producing the PV power P given a certain irradiance I and temperature T (time dependencies were removed for simplicity). $p(P | I, T)$ is related to the prior, the distribution of the PV power $p(P)$ before knowing the meteorological conditions, and to the likelihood, the probability to measure certain weather conditions knowing the produced PV power $p(I, T | P)$. Equation B.4 depicts this relation, where $p(I, T)$ is the distribution of meteorological conditions, obtained via normalization of the product of probabilities.

$$p(P | I, T) = \frac{p(P) \cdot p(I, T | P)}{p(I, T)} \quad (\text{B.4})$$

Different Bayesian algorithms arise based on the assumed distribution in Bayes' theorem. Assuming a Gaussian distribution leads to Gaussian process regression [515], while assuming feature independence results in Gaussian Naïve Bayes [54].

INSTANCE-BASED

These algorithms find important points within the training data set that describe the problem at hand [54]. One of the most popular algorithms inside this block is the k-nearest neighbour (k-NN). In the presence of a new input, k-NN looks at the k nearest neighbors of the input to decide the output.

The main actor in this family is the Support Vector Regression (SVR). Figure B.1 exemplifies the use of SVR to predict the PV power \hat{P} as a function of the kernel of the irradiance $\phi(I)$. The kernel is a transformation made to the input so that a linear relationship can be found between the transformed input and output. This linear relationship is found by minimizing the margin of tolerance ϵ so that all points fall within the margin. The points within the borders of the margin determine the parameters of the linear relationship and are called support vectors.

DECISION TREES

These algorithms predict using a tree-like model of decisions. The data space is recursively partitioned and a simple prediction model is then fitted within each partition [182], as shown in Figure B.2. The differences between models in this family stem from their underlying algorithms, which dictate, for example, the type of prediction model fitted to each partition [182].

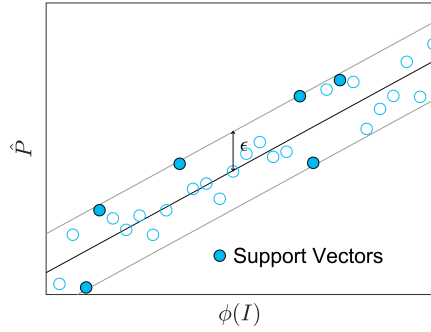


Figure B.1: Schematic example of Support Vector Regression applied for the prediction of the PV power as a function of the transformed irradiance.

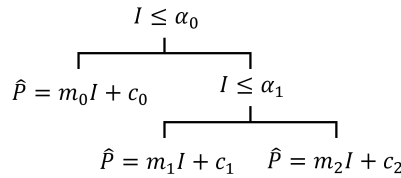


Figure B.2: Example of a decision tree for PV power prediction.

ENSEMBLE ALGORITHMS

Ensemble algorithms combine weak learners (also called predictors) into one strong learner [54]. A weak learner is defined as one whose performance is slightly better than random chance. For instance, random forests and gradient boosts regression trees are made of a combination of decision trees. An example of their structure can be seen in Figure B.3, where the result of three different decision trees is combined to improve the estimation of the output.

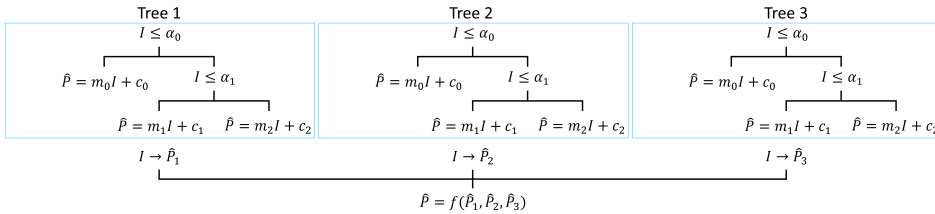


Figure B.3: Schematic of random forest, where the output of three different decision trees is combined to improve the prediction.

One can also combine deterministic forecasts to create a probabilistic one. This is the working method of the Analog Ensemble. This algorithm starts by producing a short-term deterministic forecast with training data. When new data is provided, a deterministic forecast is computed, and its result is compared to the most similar past forecasts. These are then combined to generate a bias-adjusted probabilistic forecast.

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) were inspired by the synaptic connections in the nervous system, hence their name [516]. An ANN consists of an input layer of nodes (neurons), several hidden layers, and a final layer of output neurons [83]. The nodes are interconnected via a set of weights such that signals can travel through them [517]. Each node's output is determined by applying an activation function to the aggregated weighted inputs [516, 518]. The activation function introduces nonlinearity which increases the flexibility, making this configuration very powerful [83]. ANN's main drawback is that the optimal configuration and learning algorithm are hard to obtain a priori.

If the ANN nodes never form a cycle, the ANN is called a feed-forward neural network (FFNN) or multilayer perceptron (MLP) [519]. This is the first and simplest type of ANN [516]. Figure B.4 shows graphically its structure for a PV power forecasting problem. All the nodes are fully connected between consequent layers and the information flows from the input to the hidden and finally the output. The weights assigned to each node are modified by experience while training. The most common method to train these structures is using back-propagation [519], hence these ANNs are referred to as back-propagation neural networks (BPNN).

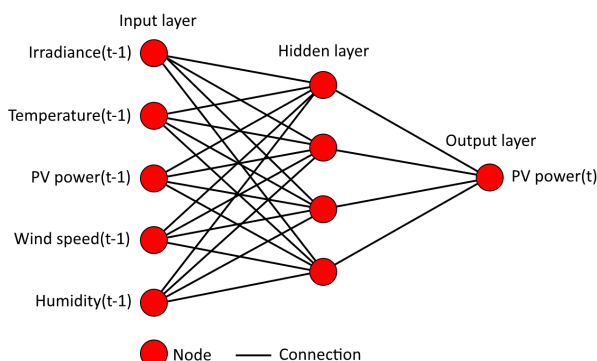


Figure B.4: MLP or FFNN structure for a PV power forecasting problem.

Several characteristics can be tuned to create different types of ANN [520], such as applying a different activation function in each node, implemented by the Radial Basis Function, and generalized regression ANNs.

DEEP LEARNING

Deep Learning (DL) is a subset of ANNs. These ANNs have a deeper number of processing layers so their problem-solving capacity is increased, and have higher extraction capabilities [136, 521]. Famous examples of the use of DL are face recognition [522] and autonomous driving [523]. This increased complexity requires vast amounts of data, leading to high computational time and hard, often impossible, interpretability [524].

One example is the Recurrent Neural Network (RNN). These structures are characterized by having at least one loop [521], for instance, by making that each hidden cell receives its output with a fixed delay [520], vide Figure B.5. The recurrent cells can be upgraded by including *memory* cells, leading to long-short-term memory (LSTM).

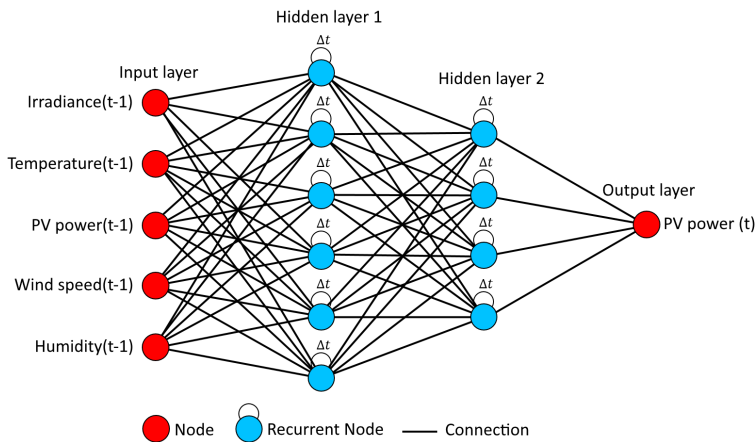


Figure B.5: Recurrent deep learning neural network structure.

An issue with FFNNs is their time-consuming learning process due to slow training algorithms repeatedly applied [525]. This problem was solved with the Extreme Learning Machine (ELM) [526]. ELM reduces the computational burden by selecting the hidden nodes randomly and calculating the output weights only once.

FUZZY ALGORITHMS

The algorithms in this section use fuzzy logic methods, an extension of binary logic where values range from 0 (False) or 1 (True). Fuzzy code allows components to take values between 0 and 1, interpreted as probabilities. Takagi and Sugeno developed a model using this logic [527] which has been employed to modify several ML algorithms.

For instance, when fuzzy logic is applied to k-means clustering, the elements can belong simultaneously to more than one cluster up to some degree [128], as represented in Figure B.6. Fuzzy logic can also be integrated into ANN yielding to the Adaptive Neuro Fuzzy Inference System [528], making the ANN more flexible.

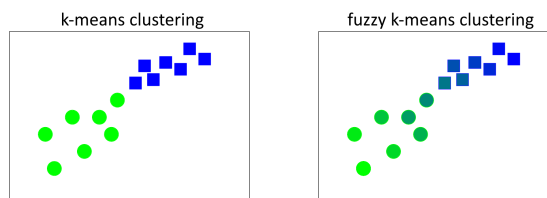


Figure B.6: Comparison between k-means and fuzzy k-means clustering, where the association to a group is defined by a probability rather than a certainty.

C

METRICS

Metrics are employed to measure how close predicted values are to real ones. The most basic metric is the error, defined as the difference between real and forecasted values. This measure can be made unit-free through normalization by the nominal capacity of the power plant, the real value (percentage error), or the average of the real values, among others.

Time series are not practical to work with, since they do not provide a single value and hence are hard to compare. Some transformations can be applied to this error to express it as one figure, such as:

- Averaging, giving place to the Mean Bias Error (MBE) or simply Bias
- Taking the absolute and averaging, Mean Absolute Error (MAE)
- Squaring and averaging, Mean Squared Error (MSE)
- Squaring, averaging, and taking the square root, Root Mean Squared Error (RMSE)

All these metrics can be further normalized, giving place to the Normalized Mean Bias Error (NMBE) and equivalent measures (NMAE, NMSE, and NRMSE). When the MAE is normalized by the real value of the data series, it gives the Mean Absolute Percentage Error (MAPE), which is a special case of NMAE.

Normalization provides unit-free errors and allows for the comparison of the performance of PV power plants of different nominal power. However, one needs to be careful with the way that this normalization is performed, especially when comparing metrics of different works. Normalization can be achieved with the average value of the series, with the nominal power of the plant, or with the difference between maximum and minimum values, among other options. These normalizations are correct and widely accepted, but they can result in very different NRMSE ranges.

Time resolution also impacts the metrics. If daily resolution is used instead of hourly, night values are excluded, bringing the average closer to the difference between the extremes. Therefore, time resolution should always be considered when comparing metrics of different works.

Each metric provides different information about the type of error performed. MBE indicates whether the predictions lean towards underestimation or overestimation, but posi-

tive errors are canceled by negative ones. MAE describes the quantitative error of the forecasts, giving the same importance to all values [103]. This last fact is sometimes undesirable since larger errors are the most costly. MSE and RMSE solve this issue at the cost of giving higher importance to outliers. Between the two, RMSE is preferred since the scale remains intact.

Other measures that are not so straightforward to derive but also extensively employed for PV power forecasting are the coefficient of determination or R-squared (R^2), the skill score (SS), and the correlation. R^2 indicates the variability between real and predicted values; it represents how well future outcomes are likely to be predicted [103]. The best value it can take is 1, hence the higher the score, the better. The skill score compares the error of the proposed method relative to the error of a standard method (such as persistence). Correlation determines the forecast performance even if any systematic correction or re-scaling has occurred.

There is a group of metrics specific to probabilistic forecasting. These measure the error made on prediction intervals, not on single points. The most employed metric in this category is the Continuous Ranked Probability Score (CRPS), which computes the difference between the real value and the prediction interval [529]. It combines deterministic outcomes with probabilistic predictions, and it is equivalent to MAE for deterministic forecasts. An alternative to CRPS is the measure of kurtosis and skewness, which are usually computed together. They represent how different the shape of the prediction is relative to a normal distribution. While the skewness measures the asymmetry of the distribution, the kurtosis computes its sharpness [128].

Table C.1 shows each of the mentioned metrics with their equation and main characteristics. Here, x_i are the true values, \hat{x}_i are the predicted values, \bar{x} is the average true prediction value, N represents the number of samples, $F_i^f(x)$ is the cumulative distribution function (CDF) of the i^{th} probabilistic forecast, $F_i^0(x)$ is the CDF of the i^{th} observation, and σ is the variance of the prediction distribution.

Table C.1: Most employed metrics for PV power forecasting with their equations and main characteristics.

Metric	Equation	Main characteristics
Error	$x_i - \hat{x}_i$	Time series.
NMBE	$\frac{1}{\bar{x}} \frac{\sum_i (x_i - \hat{x}_i)}{N}$	Measure of central tendency. Indicates over and under estimations.
NMAE	$\frac{1}{\bar{x}} \frac{\sum_i x_i - \hat{x}_i }{N}$	Measure of average error. Same importance to all values.
NMSE	$\frac{1}{\bar{x}^2} \frac{\sum_i (x_i - \hat{x}_i)^2}{N}$	Measure of average squared distance. Higher importance to outliers.
NRMSE	$\frac{1}{\bar{x}} \sqrt{\frac{\sum_i (x_i - \hat{x}_i)^2}{N}}$	Measure of average error. Higher importance to outliers.
MAPE	$\frac{1}{N} \sum_i \left \frac{x_i - \hat{x}_i}{x_i} \right $	Measure of prediction accuracy as a percentage of the error.
R ²	$1 - \frac{\sum_i (x_i - \hat{x}_i)^2}{\sum_i (\bar{x} - x_i)^2}$	Measure of variability. Indicates model's goodness of fit.
Skill score	$1 - \frac{RMSE_{\text{proposed}}}{RMSE_{\text{reference}}}$	Represents an improvement over the reference model (e.g. persistence).
Correlation	$\frac{\sum_i (x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (\hat{x}_i - \bar{\hat{x}})^2}}$	Measure of linear relationship between variables.
CRPS	$\frac{1}{N} \sum_i \int_{-\infty}^{\infty} \left(F_i^f(x) - F_i^0(x) \right)^2 dx$	Measure of difference between the real value and the prediction interval.
Kurtosis	$\frac{1}{N} \frac{\sum_i (\hat{x}_i - \bar{\hat{x}})^4}{\sigma^4}$	Measure of asymmetry of the distribution.
Skewness	$\frac{1}{N} \frac{\sum_i (\hat{x}_i - \bar{\hat{x}})^3}{\sigma^3}$	Measure of sharpness of the distribution.

D

DECOMPOSITION MODELS COMPARISON

Since no decomposition model stands out in the literature, a comparison was conducted to determine the most suitable one. The analysis utilized all 12 European meteorological stations from the Baseline Surface Radiation Network (BSRN) [329] (see [Figure 8.2a](#) for their location). The stations provide ground-based measurements of GHI, DHI, and DNI, enabling the comparison of modeled values with measured ones. BSRN combines data from numerous previously independent radiation measurement stations globally. Given the diverse sources, no standardized quality checks are applied. Thus, preliminary quality assessments, including the conversion to hourly resolution, are conducted on all irradiance data. All hourly data points where the solar zenith exceeds 85° are discarded to prevent anomalies that occur at low solar altitude angles. This zenith limit was used for preparing the BSRN data and the irradiance data of the software tool.

This study focuses on six widely used models: Erbs [530], DISC [531], Reindl2 [532, 533], BRL [280], Every [534], and Engerer [535, 536]. For every station, the hourly time series $\widehat{\text{DHI}}$ was calculated for each model using the GHI of that station, and compared to the BSRN-measured DHI. All these decomposition models use the clearness index to estimate the diffuse fraction, which is related to the time of day. Some models also included other time-dependent variables such as the sun altitude in Reindl2, the zenith angle in DISC, and apparent solar time in BRL, Every, and Engerer. These latter three furthermore included continuity in the model by including daily clearness index and persistence. The performance indicators are nRMSE and nMBE.

Where H is the number of hourly time instants. The results were analyzed to identify the best decomposition model for hourly irradiance data in Europe, see [Figure D.1](#).

The nRMSE is notably higher for the BUD and SON stations compared to the others. This can be attributed to the relatively poor data quality of these stations. The BRL model outperforms all other models in all stations except for the SON station, which had shown data limitations due to the high altitude. The simple Erbs model outperforms the bivariate models Reindl2 and DISC, the two least accurate models. Therefore, increasing the complexity of Erbs, resulting in Reindl2 and DISC, did not improve its performance. The Every

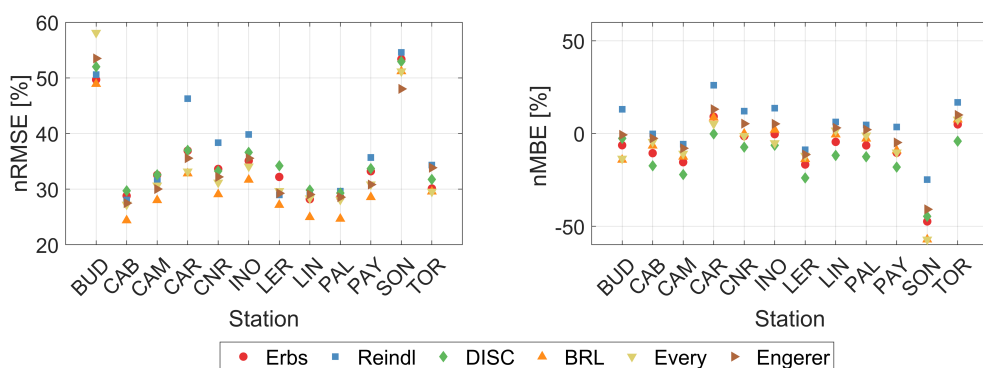


Figure D.1: nRMSE and nMBE of six decomposition models, shown for the twelve different datasets retrieved using the BSRN database. Refer to [Figure 8.2a](#) for the location of the BSRN station employed.

D

and Engerer models are adaptations of the BRL model. However, their nRMSE is higher than the original model. For the Every model, the decreased performance could be due to the model being parameterized in Australia. The Engerer model also showed decreased performance compared to the BRL model, possibly due to the model being designed for minute data originally.

Regarding the nMBE of the analysis, no model outperforms the others at all stations. However, on average, the Engerer model performs best, with an average absolute nMBE of 8.9% compared to the worst-performing 14% for the DISC model and close to 11% for the remaining models. Nevertheless, Engerer outperforms BRL in only seven out of twelve stations. In contrast to the nRMSE results, Every and Engerer models slightly surpass the preceding BRL model. Furthermore, the Erbs outperforms its succeeding models DISC and Reindl2. All models highly underestimate the diffuse fraction at the SON station, which could be again due to the extremely high altitude of the station. This raises a general concern for using decomposition models at high altitudes.

The performance analysis concludes that the Engerer model performs best based on average nMBE. However, the BRL model is considered the overall best model, consequently outperforming other models based on nRMSE and performing averagely based on the nMBE. In this study, these results are assumed to apply to satellite irradiance data. However, a note should be made that this conclusion is only robust when hourly ground-based input irradiance is used.

In addition to examining annual metrics, seasonal variations were investigated. The daily errors of all stations are averaged, leading to an average nRMSE or nMBE for each day. When exploring the nRMSE results (of which no figure is included), most models have their highest accuracy in late autumn and early winter. However, no clear pattern has been found since the Reindl2 and DISC models show different seasonal behavior, and the similarities in the BRL, Every, and Engerer model patterns probably emerge from their relation.

More interesting results are found when exploring the seasonal pattern in nMBE shown in [Figure D.2](#). All models overpredict the diffuse component in autumn and underpredict it in spring. This bias pattern also emerges when the same analysis is done at a station level (not shown). It is hard to explain this behavior since most seasonal fluctuations have extremes in winter and summer. Despite conducting a comprehensive analysis of the sea-

sonality of variables, the observed nMBE pattern could not be explained.

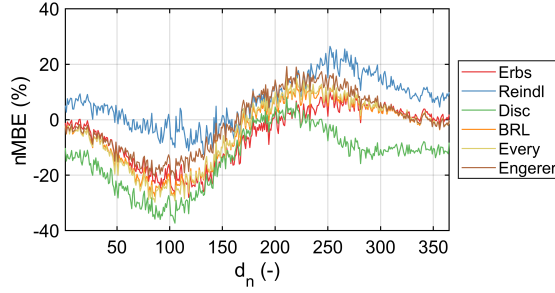


Figure D.2: Average daily nMBE of six decomposition models for a dataset containing irradiance data of twelve European BSRN stations. A seasonal bias pattern occurs with negative bias in spring and positive bias in autumn.

D

While the origin of this seasonal bias remains unknown, compensating for it can improve existing models. A correction factor can be applied to a model with high nMBE to create a much better fit. However, this approach does not apply to a model with a high nRMSE. Based on the results, opting for the BRL model is recommended. To compensate for the seasonal bias, a multiplication factor is introduced to the BRL model:

$$\widehat{DHI}_{improved} = \widehat{DHI} \frac{1}{nMBE_{fit}} \quad (D.1)$$

By simply including the multiplication factor, the diffuse fraction k_d (ratio of irradiance that is diffusely reflected) may exceed unity. This is mitigated within the new model by capping all values of k_d at 1.

The daily average nMBE curve for the BRL model was used to determine the multiplication factor. A Fourier expansion with two terms and a period of 365 days was used.

$$nMBE_{fit} = a_0 + a_1 \cos(D_n) + b_1 \sin(D_n) + a_2 \cos(2 \cdot D_n) + b_2 \sin(2 \cdot D_n) \quad (D.2)$$

$$D_n = \frac{d_n \cdot 2\pi}{365}$$

With $a_0 = 1.037$, $a_1 = 0.02052$, $a_2 = -0.02688$, $b_1 = 0.1069$, $b_2 = -0.002202$, and d_n the day number in a calendar year. Using Equation D.2, the seasonal nMBE pattern is mitigated using the compensating factor as seen in Figure D.3. The new bias-compensated model is named the BRL-MvK model.

The improvement of the model is also observed with the annual nMBE of each station. The annual nMBE is shifted towards a more positive bias for all stations (of which no figure is included), decreasing the inter-station average of the absolute value of the annual nMBE values from 11.2% to 10.6%. This shift results in improved model performance for most (formerly) negative-biased stations and a decreased performance for all formerly positive-biased stations. Kindly note that Figure D.3 refers to the daily biased error averaged for all BSRN stations while 10.6% refers to the mean absolute of the nMBE of the 12 stations. The nRMSE of the BRL-MvK model outperforms the BRL model in all stations and the mean is improved from 31.8% to 30.5%.

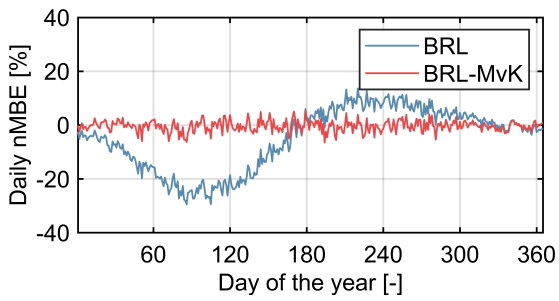


Figure D.3: Daily nMBE averaged for all the BSRN stations of the BRL and BRL-MvK model throughout a year.

E

SENSORS ALLOCATION SOFTWARE TOOL OVERVIEW

The software tool initiates by prompting the user to input location parameters:

- What is the central longitude [°] and latitude [°] of the PV plant (with 4 decimal points of accuracy)?
- A square image is downloaded. How many km should the sides be to include the full PV plant area?
- What is the capacity of the PV plant?

After which a prompt is generated confirming the input location of the user using a map tile retrieved via Geoapify [537].

When the location is deemed correct, the perimeters of the PV plant and the PV group area are defined. The user is asked to draw polygons, as seen in Figure E.1. The blue area indicates the PV plant area and the yellow areas are the PV groups area, which is provided per group.

Before being prompted to draw the areas, the user is asked whether all PV modules in the plant are uniformly tilted and oriented. If so, the following questions are asked:

- What is the azimuth orientation (deg) of the PV panels (E=90°, S=180°, W=270°)?
- What is the tilt (deg) of the PV panels (flat=0°)?

If the modules are not uniformly tilted and oriented, the two latter questions are asked for every group after drawing it.

Before defining the yellow group areas, the user is asked if all groups have an equal number of solar panels. If so, the N_{gp} of each group is set to equal values. Otherwise, they are asked to input the number of modules after each drawn PV group.

Afterward, the user is prompted to input the data purposes using the GUI element shown in Figure E.2. If more than one goal is checked in the GUI element, the user is asked to input the relative priority of each purpose.

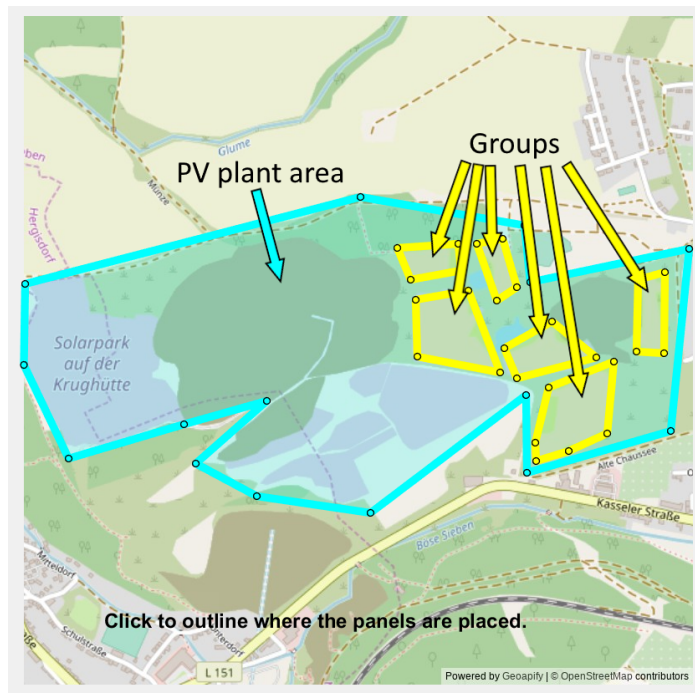


Figure E.1: GUI where the user is asked to draw the PV group areas (yellow) for the Solar Park Krughütte in Eisleben, Germany. The already drawn blue area is the PV plant area.

How will the sensor measurements be used?

- ☐ Yearly calculation of performance ratio
- ☐ Power loss analysis on a daily base
- ☐ Connection with satellite data
- ☐ Forecasting models (using *Ghor*)
- ☐ Forecasting models (using *Gpoa*)
- ☐ Power loss analysis in real time

Confirm choice

Figure E.2: GUI element where the user inputs the purpose of the data for future use.

The last question is about the number of sensors to be allocated. If the user does not input a predefined number of sensors, advice is given.

During the calculations, the user is updated on the progress and the undertaken steps. The code is profiled to give insight into how time-intensive the software tool is and what the speed-determining modules are. [Table E.1](#) provides the time profiling for the two case studies presented in [Subsection 8.4.1](#).

Table E.1: Runtime of the different steps of the software tool for the two case studies.

Runtime [min]	Kolindros	Eisleben
Distribute sensors	0.0	0.0
Import data	0.4	0.4
Calculate SVF	4.5	19.9
Create irradiance maps	0.0	0.0
Create error maps	0.1	0.5
Suggest locations	0.0	0.0
Create report	0.6	1.0
Total	5.6	21.8

For the two case studies presented, the software takes around 5.6 minutes for the 6 MW_p solar park in Kolindros (Greece), and 21.8 minutes for the 29.1 MW_p solar park in Eisleben (Germany). Although in general the higher the capacity, the longer the computational time, the capacity is not the sole indicator of computational time demand. The total runtime of the software tool mostly depends on the SVF calculations, for which the higher the local height differences, the more time-intensive the computations become. The creation of the output report and the data import are also time-intensive steps.

F

SENSORS ALLOCATION ALGORITHM SIMPLIFICATION

The allocation algorithm is based on the minimization problem in [Equation 8.1](#). Independently of the time resolution of the error, the measurement deviations take the absolute difference between $G_{POA,l}$ and G_{ref} . For clarity, the definition of $G_{POA,l}$ is shown in [Equation E1](#).

$$G_{POA,l} = SVF_l \cdot R_d \cdot DHI + (1 - SVF_l) \cdot \alpha \cdot GHI + DNI \cdot \cos(\gamma) \quad (E1)$$

Where γ is the angle of incidence onto the PV modules. $G_{POA,l}$, R_d , DHI, DNI, GHI, and α are time series but the time dependency has been removed from this equation for ease of understanding.

The reference irradiance is the weighted average of the G_{POA} of a certain area, as shown in [Equation 8.3](#). Since the reference irradiance is location-independent, it can be rewritten including the definition of $G_{POA,l}$:

$$G_{ref} = \widetilde{SVF} \cdot R_d \cdot DHI + (1 - \widetilde{SVF}) \cdot \alpha \cdot GHI + DNI \cdot \cos(\gamma) \quad (E2)$$

Where \widetilde{SVF} is the weighted average of the SVF of all grid points inside the PV group for which the sensor measures irradiance, [Equation E3](#).

$$\widetilde{SVF} = \frac{1}{\sum_{n=1}^{N_{gp}} N_n \cdot \sum_{n=1}^{N_{gp}} L_n} \sum_{n=1}^{N_{gp}} \sum_{l=1}^{L_n} SVF_l \cdot N_n \quad (E3)$$

The simplification of G_{ref} only holds because DHI, DNI, GHI, R_d , α , and γ are all (assumed to be) spatially independent. Using this new notation of G_{ref} , the following derivation can be done:

$$|G_{POA,l} - G_{ref}| = |SVF_l \cdot R_d \cdot DHI + (1 - SVF_l) \cdot \alpha \cdot GHI - \bar{SVF} \cdot R_d \cdot DHI - (1 - \bar{SVF}) \cdot \alpha \cdot GHI| \quad (F4)$$

$$= |(SVF_l - \bar{SVF}) \cdot (R_d \cdot DHI - \alpha \cdot GHI)| \quad (F5)$$

Equation F5 holds for sums over all data points of in-plane and reference irradiance time series of length T . The minimization problem for the three time resolutions of the measurement deviation can be rewritten using $\Delta SVF_l = SVF_l - \bar{SVF}$:

$$\arg \min_{l \in L} (\epsilon_{h,l}) = \arg \min_{l \in L} \sum_{h=1}^H |\Delta SVF_l \cdot [R_d(h) \cdot DHI(h) - \alpha(h) \cdot GHI(h)]| \cdot \frac{H'}{H} \quad (F6)$$

$$\arg \min_{l \in L} (\epsilon_{d,l}) = \arg \min_{l \in L} \sum_{d=1}^D |\Delta SVF_l \sum_{hd=1}^{24} [R_d(hd) \cdot DHI(hd) - \alpha(hd) \cdot GHI(hd)]| \cdot \frac{H'}{24 \cdot D} \quad (F7)$$

$$\arg \min_{l \in L} (\epsilon_{y,l}) = \arg \min_{l \in L} \sum_{y=1}^Y |\Delta SVF_l \sum_{hy=1}^{8760} [R_d(hy) \cdot DHI(hy) - \alpha(hy) \cdot GHI(hy)]| \cdot \frac{1}{Y} \quad (F8)$$

Where $h \in H$, $d \in D$, and $y \in Y$ are the hour, day, and year of each time step, respectively. The last term normalizes to a yearly value, with $H' = 8760$ being the hours per year.

All terms that are both inside the absolute signs and location independent cancel out in the minimization problem. The fractions at the end of Equation F6, F7 and F8 cancel out because they are necessarily a positive constant. Therefore, all minimization problems are reduced to Equation 8.6.

This simplification does not imply that the error maps are equal. Differences occur when the time series inside the absolute brackets from Equation F6-F8 have both negative and positive terms. This leads to measurement deviations canceling out when different periods are chosen to minimize bias. However, that condition rarely occurs, being met in only 0.2% and 0.14% of the data points in the Kolindros and Eisleben case studies, respectively. This simplification has been implemented for these two case studies, and the difference in results is negligibly small.

BIBLIOGRAPHY

- [1] A.M. van Valkengoed, L. Steg, and G. Perlaviciute. “Development and validation of a climate change perceptions scale”. In: *Journal of Environmental Psychology* 76 (2021), p. 101652. ISSN: 0272-4944. DOI: <https://doi.org/10.1016/j.jenvp.2021.101652>. URL: <https://www.sciencedirect.com/science/article/pii/S0272494421001055>.
- [2] Jimmy Gallegos et al. “Sustainable electrification—advances and challenges in electrical-distribution networks: a review”. In: *Sustainability* 16.2 (2024), p. 698.
- [3] IEA. *Tracking Clean Energy Progress 2023*. Tech. rep. Licence: CC BY 4.0. 2023.
- [4] F Nijse et al. “Is a solar future inevitable?” In: (2022).
- [5] Photovoltaic Power Systems Program. *Snapshot of Global PV Markets*. Tech. rep. International Energy Agency, 2021. URL: www.iea-pvps.org.
- [6] Vasilis M Fthenakis, Hyung Chul Kim, and Erik Alsema. “Emissions from photovoltaic life cycles”. In: *Environmental science & technology* 42.6 (2008), pp. 2168–2174.
- [7] SolarPower Europe. *Global Market Outlook for Solar Power 2023-2027*. Tech. rep. 2023.
- [8] Christopher D O’Connor, Kaitlin Fredericks, and Kaylee Kosoralo. “People’s perceptions of energy technologies in an era of rapid transformation”. In: *Environmental Innovation and Societal Transitions* 43 (2022), pp. 331–342.
- [9] Julia Cousse. “Still in love with solar energy? Installation size, affect, and the social acceptance of renewable energy technologies”. In: *Renewable and Sustainable Energy Reviews* 145 (2021), p. 111107.
- [10] Juliet E Carlisle et al. “Utility-scale solar and public attitudes toward siting: A critical examination of proximity”. In: *Land Use Policy* 58 (2016), pp. 491–501.
- [11] Kantar Belgium. *Special Eurobarometer 492: Europeans’ attitudes on EU energy policy*. Tech. rep. European Union, 2019.
- [12] International Energy Agency. *Trends in PV Applications*. Tech. rep. Accessed: Dec 14, 2023. 2022.
- [13] Sandra George O’Neil. “Community obstacles to large scale solar: NIMBY and renewables”. In: *Journal of Environmental Studies and Sciences* 11.1 (2021), pp. 85–92.
- [14] Roberta S Nilson and Richard C Stedman. “Are big and small solar separate things?: The importance of scale in public support for solar energy development in upstate New York”. In: *Energy Research & Social Science* 86 (2022), p. 102449.
- [15] Wouter Schram. *Collective Self-Consumption: The Next Step in PV Prosuming Policy to Foster Citizen Engagement?* EUPVSEC-40. 2023.

- [16] Goda Perlaviciute et al. "Values in the backyard: the relationship between people's values and their evaluations of a real, nearby energy project". In: *Environmental Research Communications* 3.10 (2021), p. 105004.
- [17] SocialRES. *Together we make energy*. <https://socialres.eu/>. Accessed: Dec 14, 2023.
- [18] Damián Copena and Xavier Simón. "Wind farms and payments to landowners: Opportunities for rural development for the case of Galicia". In: *Renewable and Sustainable Energy Reviews* 95 (2018), pp. 38–47.
- [19] Lawrence Gottschamer and Qiong Zhang. "The dynamics of political power: The socio-technical transition of California's electricity system to renewable energy". In: *Energy Research & Social Science* 70 (2020), p. 101618.
- [20] Inga Ydersbond. "Multi-level lobbying in the EU: The case of the Renewables Directive and the German energy industry". In: (2012).
- [21] Trevor Culhane, Galen Hall, and J Timmons Roberts. "Who delays climate action? Interest groups and coalitions in state legislative struggles in the United States". In: *Energy Research & Social Science* 79 (2021), p. 102114.
- [22] Aitong Li, Yuan Xu, and Hideaki Shiroyama. "Solar lobby and energy transition in Japan". In: *Energy Policy* 134 (2019), p. 110950.
- [23] Isabelle Cadoret and Fabio Padovano. "The political drivers of renewable energies policies". In: *Energy Economics* 56 (2016), pp. 261–269.
- [24] Gilles Notton et al. "Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting". In: *Renewable and sustainable energy reviews* 87 (2018), pp. 96–105.
- [25] K. Kawabe and K. Tanaka. "Impact of Dynamic Behavior of Photovoltaic Power Generation Systems on Short-Term Voltage Stability". In: *IEEE Trans. Power Syst.* 30.6 (Nov. 2015), pp. 3416–3424. DOI: [10.1109/TPWRS.2015.2390649](https://doi.org/10.1109/TPWRS.2015.2390649).
- [26] Abdullah Alshahrani et al. *The technical challenges facing the integration of small-scale and large-scale PV systems into the grid: A critical review*. 2019.
- [27] Shady S Refaat et al. "Impact of grid-tied large-scale photovoltaic system on dynamic voltage stability of electric power grids". In: *IET Renewable Power Generation* 12.2 (2018), pp. 157–164.
- [28] NetBeheer Nederland. *Landelijk Actieprogramma - Netcongestie*. Tech. rep. 2022.
- [29] International Energy Agency. *Renewables 2023: Analysis and forecast to 2028*. Tech. rep. 2023.
- [30] NetBeheer Nederland. *Capaciteitskaart invoeding elektriciteitsnet*. <https://capaciteitskaart.netbeheernederland.nl/>. Accessed: Feb 2, 2024.
- [31] Can Wan et al. "Photovoltaic and solar power forecasting for smart grid energy management". In: *CSEE Journal of Power and Energy Systems* 1.4 (2015), pp. 38–46.
- [32] Alba Alcañiz et al. "Trends and gaps in photovoltaic power forecasting with machine learning". In: *Energy Reports* 9 (2023), pp. 447–471.

- [33] R. Ahmed et al. *A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization*. May 2020. DOI: [10.1016/j.rser.2020.109792](https://doi.org/10.1016/j.rser.2020.109792). URL: <https://doi.org/10.1016/j.rser.2020.109792>.
- [34] Javier Antonanzas et al. "Review of photovoltaic power forecasting". In: *Solar energy* 136 (2016), pp. 78–111.
- [35] M. Hossain et al. "Application of Extreme Learning Machine for Short Term Output Power Forecasting of Three Grid-connected PV Systems". In: *Journal of Cleaner Production* 167 (Nov. 2017), pp. 395–405. DOI: [10.1016/j.jclepro.2017.08.081](https://doi.org/10.1016/j.jclepro.2017.08.081).
- [36] Clarivate Analytics. *Web of Science*. Accessed Aug. 08, 2022. URL: <https://www.webofscience.com/wos/woscc/basic-search>.
- [37] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra. "Solar forecasting methods for renewable energy integration". In: *Progress in Energy and Combustion Science* 39.6 (Dec. 2013), pp. 535–576. DOI: [10.1016/j.pecs.2013.06.002](https://doi.org/10.1016/j.pecs.2013.06.002).
- [38] Robert Ulbricht et al. "First steps towards a systematical optimized strategy for solar energy supply forecasting". In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLP-KDD 2013), Prague, Czech Republic*. Vol. 2327. 2013.
- [39] Y. Ren, P. N. Suganthan, and N. Srikanth. "Ensemble methods for wind and solar power forecasting - A state-of-the-art review". In: *Renewable and Sustainable Energy Reviews* 50 (Oct. 2015), pp. 82–91. DOI: [10.1016/j.rser.2015.04.081](https://doi.org/10.1016/j.rser.2015.04.081).
- [40] F. H. Gandoman, F. Raeisi, and A. Ahmadi. "A literature review on estimating of PV-array hourly power under cloudy weather conditions". In: *Renewable and Sustainable Energy Reviews* 63 (Sept. 2016), pp. 579–592. DOI: [10.1016/j.rser.2016.05.027](https://doi.org/10.1016/j.rser.2016.05.027).
- [41] M. Q. Raza, M. Nadarajah, and C. Ekanayake. "On recent advances in PV output power forecast". In: *Solar Energy* 136 (Oct. 2016), pp. 125–144. DOI: [10.1016/j.solener.2016.06.073](https://doi.org/10.1016/j.solener.2016.06.073).
- [42] F. Barbieri, S. Rajakaruna, and A. Ghosh. "Very short-term photovoltaic power forecasting with cloud modeling: A review". In: *Renewable and Sustainable Energy Reviews* 75 (Aug. 2017), pp. 242–263. DOI: [10.1016/j.rser.2016.10.068](https://doi.org/10.1016/j.rser.2016.10.068).
- [43] Utpal Kumar Das et al. "Forecasting of photovoltaic power generation and model optimization: A review". In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 912–928.
- [44] Sobrina Sobri, Sam Koohi-Kamali, and Nasrudin Abd Rahim. "Solar photovoltaic generation forecasting methods: A review". In: *Energy Conversion and Management* 156 (2018), pp. 459–497.
- [45] D. W. van der Meer, J. Widén, and J. Munkhammar. "Review on probabilistic forecasting of photovoltaic power production and electricity consumption". In: *Renewable and Sustainable Energy Reviews* 81 (Jan. 2018), pp. 1484–1512. DOI: [10.1016/j.rser.2017.05.212](https://doi.org/10.1016/j.rser.2017.05.212).
- [46] Amir Mosavi et al. "State of the art of machine learning models in energy systems, a systematic review". In: *Energies* 12.7 (2019), p. 1301.

- [47] Dazhi Yang. "A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES)". In: *Journal of Renewable and Sustainable Energy* 11.2 (Mar. 2019), p. 22701. ISSN: 19417012. DOI: [10.1063/1.5087462](https://doi.org/10.1063/1.5087462). URL: <https://doi.org/10.1063/1.5087462>.
- [48] Mohamed Massaoudi et al. "Convergence of photovoltaic power forecasting and deep learning: State-of-art review". In: *IEEE Access* 9 (2021), pp. 136593–136615.
- [49] Adel Mellit et al. "Advanced methods for photovoltaic output power forecasting: A review". In: *Applied Sciences* 10.2 (2020), p. 487.
- [50] Abdul Rahim Pazikadin et al. "Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend". In: *Science of The Total Environment* 715 (2020), p. 136848.
- [51] R. A. Rajagukguk, R. A. A. Ramadhan, and H.-J. Lee. "A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power". In: *Energies* 13.24 (2020), p. 6623. DOI: [10.3390/EN13246623](https://doi.org/10.3390/EN13246623).
- [52] Cong Feng, Yuanzhi Liu, and Jie Zhang. "A taxonomical review on recent artificial intelligence applications to PV integration into power grids". In: *International Journal of Electrical Power & Energy Systems* 132 (2021), p. 107176.
- [53] Priya Gupta and Rhythm Singh. "PV power forecasting based on data-driven models: a review". In: *International Journal of Sustainable Engineering* 14.6 (2021), pp. 1733–1755. DOI: [10.1080/19397038.2021.1986590](https://doi.org/10.1080/19397038.2021.1986590). URL: <https://doi.org/10.1080/19397038.2021.1986590>.
- [54] Jason Brownlee. *Machine Learning Mastery*. Accessed Aug. 30, 2021. URL: <https://machinelearningmastery.com/>.
- [55] M. Abuella and B. Chowdhury. "Solar Power Forecasting Using Support Vector Regression". In: (Mar. 2017).
- [56] Ruby Nageem and R Jayabarathi. "Predicting the power output of a grid-connected solar panel using multi-input support vector regression". In: *Procedia computer science* 115 (2017), pp. 723–730.
- [57] Hangxia Zhou et al. "Short-term photovoltaic power forecasting based on stacking-SVM". In: *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE. 2018, pp. 994–998.
- [58] Abdullah Alfadda et al. "Hour-ahead solar PV power forecasting using SVR based approach". In: *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE. 2017, pp. 1–5.
- [59] Utpal Kumar Das et al. "SVR-based model to forecast PV power generation under different weather conditions". In: *Energies* 10.7 (2017), p. 876.
- [60] C. J. Huang and P. H. Kuo. "Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting". In: *IEEE Access* 7 (2019), pp. 74822–74834. DOI: [10.1109/ACCESS.2019.2921238](https://doi.org/10.1109/ACCESS.2019.2921238).
- [61] Yanting Li et al. "Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines". In: *Applied Energy* 180 (2016), pp. 392–401.

- [62] Manohar Mishra et al. “Deep learning and wavelet transform integrated approach for short-term solar PV power prediction”. In: *Measurement* 166 (2020), p. 108250.
- [63] Yordanos Kassa Semero, Jianhua Zhang, and Dehua Zheng. “PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy”. In: *CSEE Journal of Power and Energy Systems* 4.2 (2018), pp. 210–218.
- [64] S Ferlito, G Adinolfi, and G Graditi. “Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production”. In: *Applied Energy* 205 (2017), pp. 116–129.
- [65] W. Lee et al. “Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks”. In: *IEEE Access* 6 (2018), pp. 73068–73080. DOI: [10.1109/ACCESS.2018.2883330](https://doi.org/10.1109/ACCESS.2018.2883330).
- [66] Farid Touati et al. “Long-term performance analysis and power prediction of PV technology in the State of Qatar”. In: *Renewable Energy* 113 (2017), pp. 952–965.
- [67] Minh-Thang Do, Ted Soubdhan, and Benoît Robyns. “A study on the minimum duration of training data to provide a high accuracy forecast for PV generation between two different climatic zones”. In: *Renewable Energy* 85 (2016), pp. 959–964.
- [68] P. Ramsami and V. Oree. “A Hybrid Method for Forecasting the Energy Output of Photovoltaic Systems”. In: *Energy Conversion and Management* 95 (May 2015), pp. 406–413. DOI: [10.1016/j.enconman.2015.02.052](https://doi.org/10.1016/j.enconman.2015.02.052).
- [69] A. Kuzmiakova, G. Colas, and A. McKeehan. *Short-term Memory Solar Energy Forecasting at University of Illinois*. 2017.
- [70] A. Rosato et al. “Embedding of Time Series for the Prediction in Photovoltaic Power Plants”. In: *Energies* 10.1003 (2017). DOI: [10.3390/en10071003](https://doi.org/10.3390/en10071003).
- [71] J. H. Friedman. “Multivariate Adaptive Regression Splines”. In: *The Annals of Statistics* 19.1 (1991), pp. 1–141.
- [72] L. Massidda and M. Marrocu. “Use of Multilinear Adaptive Regression Splines and Numerical Weather Prediction to Forecast the Power Output of a PV Plant in Borkum, Germany”. In: *Solar Energy* 146 (Apr. 2017), pp. 141–149. DOI: [10.1016/j.solener.2017.02.007](https://doi.org/10.1016/j.solener.2017.02.007).
- [73] A. Gosiewska, A. Kozak, and P. Biecek. “Simpler is Better: Lifting Interpretability-Performance Trade-off via Automated Feature Engineering”. In: *Decision Support Systems* 150 (Nov. 2021), p. 113556. DOI: [10.1016/J.DSS.2021.113556](https://doi.org/10.1016/J.DSS.2021.113556).
- [74] C. Rudin. “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead”. In: *Nature Machine Intelligence* 1.5 (May 2019), pp. 206–215. DOI: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x).
- [75] Hilarie Sit. *Quick Start to Gaussian Process Regression*. 2019. URL: <https://towardsdatascience.com/quick-start-to-gaussian-process-regression-36d838810319> (visited on 11/27/2020).
- [76] M. De Felice, M. Petitta, and P. M. Ruti. “Short-term Predictability of Photovoltaic Production over Italy”. In: *Renewable Energy* 80 (Aug. 2015), pp. 197–204. DOI: [10.1016/j.renene.2015.02.010](https://doi.org/10.1016/j.renene.2015.02.010).

- [77] M. Rana, I. Koprinska, and V. G. Agelidis. "2D-Interval Forecasts for Solar Power Production". In: *Solar Energy* 122 (Dec. 2015), pp. 191–203. DOI: [10.1016/j.solener.2015.08.018](https://doi.org/10.1016/j.solener.2015.08.018).
- [78] Mingzhang Pan et al. "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization". In: *Journal of Cleaner Production* 277 (2020), p. 123948.
- [79] Y. Wang, W. Liao, and Y. Chang. "Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting". In: *Energies* 11.8 (Aug. 2018), p. 2163. DOI: [10.3390/EN11082163](https://doi.org/10.3390/EN11082163).
- [80] H. Sheng et al. "Solar Power Forecasting Based on Domain Adaptive Learning". In: *IEEE Access* 8 (2020). DOI: [10.1109/ACCESS.2020.3034100](https://doi.org/10.1109/ACCESS.2020.3034100).
- [81] B. Carrera and K. Kim. "Comparison Analysis of Machine Learning Techniques for Photovoltaic Prediction Using Weather Sensor Data". In: *Sensors* 20.3129 (2020). DOI: [10.3390/s20113129](https://doi.org/10.3390/s20113129).
- [82] P. Pawar, N. Mithulananthan, and M. Q. Raza. "Solar PV Power Forecasting Using Modified SVR with Gauss-Newton Method". In: *2020 IEEE 2nd Global Power, Energy and Communication Conference (GPECOM)*. Oct. 2020, pp. 226–231. DOI: [10.1109/GPECOM49333.2020.9247935](https://doi.org/10.1109/GPECOM49333.2020.9247935).
- [83] S. S. Cross, R. F. Harrison, and R. L. Kennedy. "Introduction to Neural Networks". In: *The Lancet* 346 (1995), pp. 1075–1079. DOI: [10.1016/S0140-6736\(95\)91746-2](https://doi.org/10.1016/S0140-6736(95)91746-2).
- [84] H. Sharadga, S. Hajimirza, and R. S. Balog. "Time Series Forecasting of Solar Power Generation for Large-Scale Photovoltaic Plants". In: *Renewable Energy* 150 (May 2020), pp. 797–807. DOI: [10.1016/j.renene.2019.12.131](https://doi.org/10.1016/j.renene.2019.12.131).
- [85] J. Zhang et al. "Deep Photovoltaic Nowcasting". In: *Solar Energy* 176 (Dec. 2018), pp. 267–276. DOI: [10.1016/j.solener.2018.10.024](https://doi.org/10.1016/j.solener.2018.10.024).
- [86] V. Suresh et al. "Forecasting Solar PV Output Using Convolutional Neural Networks with a Sliding Window Algorithm". In: *Energies* 13.3 (Feb. 2020), p. 723. DOI: [10.3390/EN13030723](https://doi.org/10.3390/EN13030723).
- [87] K. Wang, X. Qi, and H. Liu. "Photovoltaic Power Forecasting based LSTM-Convolutional Network". In: *Energy* 189 (Dec. 2019), p. 116225. DOI: [10.1016/J.ENERGY.2019.116225](https://doi.org/10.1016/J.ENERGY.2019.116225).
- [88] M. Alkandari and I. Ahmad. "Solar Power Generation Forecasting Using Ensemble Approach Based on Deep Learning and Statistical Methods". In: *Applied Computing and Informatics* (2020). DOI: [10.1016/j.aci..](https://doi.org/10.1016/j.aci..)
- [89] Andre Gensler et al. "Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks". In: *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings* (2017), pp. 2858–2865. DOI: [10.1109/SMC.2016.7844673](https://doi.org/10.1109/SMC.2016.7844673).
- [90] Haixiang Zang et al. "Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network". In: *IET Generation, Transmission & Distribution* 12.20 (2018), pp. 4557–4567.

- [91] L. Wen et al. "Optimal Load Dispatch of Community Microgrid with Deep Learning Based Solar Power and Load Forecasting". In: *Energy* 171 (Mar. 2019), pp. 1053–1065. DOI: [10.1016/J.ENERGY.2019.01.075](https://doi.org/10.1016/J.ENERGY.2019.01.075).
- [92] Y. Jung et al. "Long Short-Term Memory Recurrent Neural Network for Modeling Temporal Patterns in Long-Term Power Forecasting for Solar PV Facilities: Case Study of South Korea". In: *Journal of Cleaner Production* 250 (Mar. 2020), p. 119476. DOI: [10.1016/J.JCLEPRO.2019.119476](https://doi.org/10.1016/J.JCLEPRO.2019.119476).
- [93] M. G. De Giorgi, M. Malvoni, and P. M. Congedo. "Comparison of Strategies for Multi-step Ahead Photovoltaic Power Forecasting Models Based on Hybrid Group Method of Data Handling Networks and Least Square Support Vector Machine". In: *Energy* 107 (July 2016), pp. 360–373. DOI: [10.1016/j.energy.2016.04.020](https://doi.org/10.1016/j.energy.2016.04.020).
- [94] G. Cervone et al. "Short-term Photovoltaic Power Forecasting Using Artificial Neural Networks and an Analog Ensemble". In: *Renewable Energy* 108 (Aug. 2017), pp. 274–286. DOI: [10.1016/j.renene.2017.02.052](https://doi.org/10.1016/j.renene.2017.02.052).
- [95] Z. Chen and I. Koprinska. *Ensemble Methods for Solar Power Forecasting*. 2020.
- [96] A. Nayak and L. Heistrene. "Hybrid Machine Learning Model for Forecasting Solar Power Generation". In: *2020 International Conference on Smart Grids and Energy Systems (SGES)*. 2020. DOI: [10.1109/SGES51519.2020.00167](https://doi.org/10.1109/SGES51519.2020.00167).
- [97] Yang Lin et al. "Solar power forecasting based on pattern sequence similarity and meta-learning". In: *Artificial Neural Networks and Machine Learning–ICANN 2020: 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part I* 29. Springer. 2020, pp. 271–283.
- [98] P. Lin et al. "Short-term Power Prediction for Photovoltaic Power Plants Using a Hybrid Improved Kmeans-GRA-Elman Model Based on Multivariate Meteorological Factors and Historical Power Datasets". In: *Energy Conversion and Management* 177 (Dec. 2018), pp. 704–717. DOI: [10.1016/j.enconman.2018.10.015](https://doi.org/10.1016/j.enconman.2018.10.015).
- [99] K. P. Lin and P. F. Pai. "Solar Power Output Forecasting Using Evolutionary Seasonal Decomposition Least-Square Support Vector Regression". In: *Journal of Cleaner Production* 134.Part B (Oct. 2016), pp. 456–462. DOI: [10.1016/j.jclepro.2015.08.099](https://doi.org/10.1016/j.jclepro.2015.08.099).
- [100] Y. Chu et al. "Short-term Reforecasting of Power Output from a 48 MWe Solar PV Plant". In: *Solar Energy* 112 (Feb. 2015), pp. 68–77. DOI: [10.1016/j.solener.2014.11.017](https://doi.org/10.1016/j.solener.2014.11.017).
- [101] W. Yin et al. "A Novel Non-iterative Correction Method for Short-term Photovoltaic Power Forecasting". In: *Renewable Energy* 159 (Oct. 2020), pp. 23–32. DOI: [10.1016/J.RENENE.2020.05.134](https://doi.org/10.1016/J.RENENE.2020.05.134).
- [102] I. Guyon and A. Elisseeff. "An Introduction to Variable and Feature Selection". In: *Journal of Machine Learning Research* 3.3 (2003), pp. 1157–1182.
- [103] S. Rosiek, J. Alonso-Montesinos, and F. J. Batlles. "Online 3-h Forecasting of the Power Output from a BIPV System using Satellite Observations and ANN". In: *International Journal of Electrical Power & Energy Systems* 99 (July 2018), pp. 261–272. DOI: [10.1016/j.ijepes.2018.01.025](https://doi.org/10.1016/j.ijepes.2018.01.025).

- [104] M. P. Almeida, O. Perpiñán, and L. Narvarte. "PV Power Forecast using a Nonparametric PV Model". In: *Solar Energy* 115 (May 2015), pp. 354–368. DOI: [10.1016/j.solener.2015.03.006](https://doi.org/10.1016/j.solener.2015.03.006).
- [105] K. A. Baharin et al. "Short-term Forecasting of Solar Photovoltaic Output Power for Tropical Climate using Ground-based Measurement Data". In: *Journal of Renewable and Sustainable Energy* 8.5 (Sept. 2016), p. 053701. DOI: [10.1063/1.4962412](https://doi.org/10.1063/1.4962412).
- [106] D. Van Tai. "Solar Photovoltaic Power Output Forecasting using Machine Learning Technique". In: *Journal of Physics: Conference Series*. Vol. 1327. 1. Oct. 2019, p. 012051. DOI: [10.1088/1742-6596/1327/1/012051](https://doi.org/10.1088/1742-6596/1327/1/012051).
- [107] S. Han et al. "Mid-to-long term Wind and Photovoltaic Power Generation Prediction based on Copula Function and Long Short Term Memory Network". In: *Applied Energy* 239 (Apr. 2019), pp. 181–191. DOI: [10.1016/J.APENERGY.2019.01.193](https://doi.org/10.1016/J.APENERGY.2019.01.193).
- [108] Marco Pierro et al. "Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data". In: *Solar Energy* 158 (Dec. 2017), pp. 1026–1038. ISSN: 0038092X. DOI: [10.1016/j.solener.2017.09.068](https://doi.org/10.1016/j.solener.2017.09.068).
- [109] Mohamed Massaoudi et al. "PV Power Forecasting Using Weighted Features for Enhanced Ensemble Method". In: *IEEE Access* (Oct. 2019). arXiv: [1910.09404](https://arxiv.org/abs/1910.09404). URL: <http://arxiv.org/abs/1910.09404>.
- [110] E. Isaksson and M. K. Conde. "Solar Power Forecasting with Machine Learning Techniques". MA thesis. KTH Royal Institute of Technology, 2018.
- [111] D. Lee et al. "Improvement of Short-Term BIPV Power Predictions Using Feature Engineering and a Recurrent Neural Network". In: *Energies* 12.17 (Aug. 2019), p. 3247. DOI: [10.3390/EN12173247](https://doi.org/10.3390/EN12173247).
- [112] C. Ratanamahatana and D. Gunopulos. "Feature Selection for the Naive Bayesian Classifier using Decision Trees". In: *Machine Learning* 17.5-6 (May 2010), pp. 475–487. DOI: [10.1080/713827175](https://doi.org/10.1080/713827175).
- [113] C. Shang and P. Wei. "Enhanced Support Vector Regression based Forecast Engine to Predict Solar Power Output". In: *Renewable Energy* 127 (Nov. 2018), pp. 269–283. DOI: [10.1016/j.renene.2018.04.067](https://doi.org/10.1016/j.renene.2018.04.067).
- [114] Björn Wolff et al. "Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data". In: *Solar Energy* 135 (2016), pp. 197–208.
- [115] Fei Wang et al. "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework". In: *Energy Conversion and Management* 212 (2020), p. 112766.
- [116] Zhongwen Li et al. "Day-ahead hourly photovoltaic generation forecasting using extreme learning machine". In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE. 2015, pp. 779–783.
- [117] Huaizhi Wang et al. "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network". In: *Energy conversion and management* 153 (2017), pp. 409–422.

- [118] M. Q. Raza, N. Mithulananthan, and A. Summerfield. "Solar Output Power Forecast using an Ensemble Framework with Neural Predictors and Bayesian Adaptive Combination". In: *Solar Energy* 166 (May 2018), pp. 226–241. DOI: [10.1016/j.solener.2018.03.066](https://doi.org/10.1016/j.solener.2018.03.066).
- [119] P. Li et al. "A Hybrid Deep Learning Model for Short-term PV Power Forecasting". In: *Applied Energy* 259 (Feb. 2020), p. 114216. DOI: [10.1016/J.APENERGY.2019.114216](https://doi.org/10.1016/J.APENERGY.2019.114216).
- [120] A. T. Eseye, J. Zhang, and D. Zheng. "Short-term Photovoltaic Solar Power Forecasting using a Hybrid Wavelet-PSO-SVM Model based on SCADA and Meteorological Information". In: *Renewable Energy* 118 (Apr. 2018), pp. 357–367. DOI: [10.1016/J.RENENE.2017.11.011](https://doi.org/10.1016/J.RENENE.2017.11.011).
- [121] Soufiane Gaizen, Ouafia Fadi, and Ahmed Abbou. "Solar power time series prediction using wavelet analysis". In: *International Journal of Renewable Energy Research (IJRER)* 10.4 (2020), pp. 1764–1773.
- [122] H. Panamtash, S. Mahdavi, and Q. Zhou. "Probabilistic Solar Power Forecasting: A Review and Comparison". In: *2020 52nd North American Power Symposium (NAPS)*. Apr. 2021. DOI: [10.1109/NAPS50074.2021.9449746](https://doi.org/10.1109/NAPS50074.2021.9449746).
- [123] S. Alessandrini et al. "An Analog Ensemble for Short-term Probabilistic Solar Power Forecast". In: *Applied Energy* 157 (Nov. 2015), pp. 95–110. DOI: [10.1016/j.apenergy.2015.08.011](https://doi.org/10.1016/j.apenergy.2015.08.011).
- [124] Q. Ni et al. "An Ensemble Prediction Intervals Approach for Short-term PV Power Forecasting". In: *Solar Energy* 155 (Oct. 2017), pp. 1072–1083. DOI: [10.1016/j.solener.2017.07.052](https://doi.org/10.1016/j.solener.2017.07.052).
- [125] H. Takeda. "Short-term Ensemble Forecast for Purchased Photovoltaic Generation". In: *Solar Energy* 149 (June 2017), pp. 176–187. DOI: [10.1016/j.solener.2017.03.088](https://doi.org/10.1016/j.solener.2017.03.088).
- [126] A. Asrari, T. X. Wu, and B. Ramos. "A Hybrid Algorithm for Short-Term Solar Power Prediction - Sunshine State Case Study". In: *IEEE Transactions on Sustainable Energy* 8.2 (Apr. 2017), pp. 582–591. DOI: [10.1109/TSTE.2016.2613962](https://doi.org/10.1109/TSTE.2016.2613962).
- [127] Ling-Ling Li et al. "Renewable energy prediction: A novel short-term prediction model of photovoltaic output power". In: *Journal of Cleaner Production* 228 (2019), pp. 359–375.
- [128] Sergios Theodoridis Konstantinos Koutrombas. *Pattern recognition*. Elsevier, 2008.
- [129] S. Al-Dahidi, M. Louzazni, and N. Omran. "A Local Training Strategy-based Artificial Neural Network for Predicting the Power Production of Solar Photovoltaic Systems". In: *IEEE Access* 8 (2020), pp. 150262–150281. DOI: [10.1109/ACCESS.2020.3016165](https://doi.org/10.1109/ACCESS.2020.3016165).
- [130] F. Golestaneh, P. Pinson, and H. B. Gooi. "Very Short-term Nonparametric Probabilistic Forecasting of Renewable Energy Generation - With Application to Solar Energy". In: *IEEE Transactions on Power Systems* 31.5 (2016), pp. 3850–3863. DOI: [10.1109/TPWRS.2015.2502423](https://doi.org/10.1109/TPWRS.2015.2502423).
- [131] G. Agoua, R. Girard, and G. Kariniotakis. "Probabilistic Models for Spatio-Temporal Forecasting of Photovoltaic Production". In: *IEEE Transactions on Sustainable Energy* 10.2 (2019), pp. 780–789.

- [132] F. A. Kraemer et al. "Operationalizing Solar Energy Predictions for Sustainable, Autonomous IoT Device Management". In: *IEEE Internet of Things Journal* 7.12 (2020). DOI: [10.1109/JIOT.2020.3002330](https://doi.org/10.1109/JIOT.2020.3002330).
- [133] D. Chen and H. W. Chen. "Using the Köppen Classification to Quantify Climate Variation and Change: An Example for 1901-2010". In: *Environment, Development and Sustainability* 6.1 (Apr. 2013), pp. 69–79. DOI: [10.1016/j.envdev.2013.03.007](https://doi.org/10.1016/j.envdev.2013.03.007).
- [134] Yue Zhang et al. "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators". In: *IEEE Transactions on Smart Grid* 6.5 (2015), pp. 2253–2262.
- [135] Kang Wang et al. "Multiple convolutional neural networks for multivariate time series prediction". In: *Neurocomputing* 360 (2019), pp. 107–119.
- [136] K. Wang, X. Qi, and H. Liu. "A Comparison of Day-ahead Photovoltaic Power Forecasting Models based on Deep Learning Neural Network". In: *Applied Energy* 251 (Oct. 2019), p. 113315. DOI: [10.1016/J.APENERGY.2019.113315](https://doi.org/10.1016/J.APENERGY.2019.113315).
- [137] M. Paulescu et al. "Structured, Physically Inspired (Gray Box) Models versus Black Box Modeling for Forecasting the Output Power of Photovoltaic Plants". In: *Energy* 121 (Feb. 2017), pp. 792–802. DOI: [10.1016/j.energy.2017.01.015](https://doi.org/10.1016/j.energy.2017.01.015).
- [138] G. Graditi, S. Ferlito, and G. Adinolfi. "Comparison of Photovoltaic Plant Power Production Prediction Methods using a Large Measured Dataset". In: *Renewable Energy* 90 (May 2016), pp. 513–519. DOI: [10.1016/j.renene.2016.01.027](https://doi.org/10.1016/j.renene.2016.01.027).
- [139] J. Antonanzas et al. "Influence of Electricity Market Structures on Deterministic Solar Forecasting Verification". In: *Solar Energy* 210 (Nov. 2020), pp. 44–46. DOI: [10.1016/J.SOLENER.2020.04.017](https://doi.org/10.1016/J.SOLENER.2020.04.017).
- [140] Qian Li, Zhou Wu, and Xiaohua Xia. "Estimate and characterize PV power at demand-side hybrid system". In: *Applied Energy* 218 (2018), pp. 66–77.
- [141] Gary W Chang and Heng-Jiu Lu. "Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast". In: *IEEE Transactions on Sustainable Energy* 11.1 (2018), pp. 185–194.
- [142] Z. Li et al. "A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting". In: *Energies* 9.1 (Jan. 2016), p. 55. DOI: [10.3390/EN9010055](https://doi.org/10.3390/EN9010055).
- [143] G. Li et al. "Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach". In: *Energies* 12.13 (July 2019), p. 2538. DOI: [10.3390/EN12132538](https://doi.org/10.3390/EN12132538).
- [144] Soumyabrata Dev et al. "Ground-based image analysis: A tutorial on machine-learning techniques and applications". In: *IEEE Geoscience and Remote Sensing Magazine* 4.2 (2016), pp. 79–93.
- [145] Y. Sun, V. Venugopal, and A. R. Brandt. "Short-term solar power forecast with deep learning: Exploring optimal input and output configuration". In: *Solar Energy* 188 (Aug. 2019), pp. 730–741. DOI: [10.1016/j.solener.2019.06.041](https://doi.org/10.1016/j.solener.2019.06.041).
- [146] Dimitrios Anagnostos et al. "A method for detailed, short-term energy yield forecasting of photovoltaic installations". In: *Renewable Energy* 130 (2019), pp. 122–129.

- [147] S. Theocharides et al. "Machine Learning Algorithms for Photovoltaic System Power Output Prediction". In: *2018 IEEE International Energy Conference (ENERGYCON)*. 2018, pp. 1–6.
- [148] H. J. Escalante. *A Comparison of Outlier Detection Algorithms for Machine Learning*. 2005.
- [149] L. L. Li et al. "Short-term output power forecasting of photovoltaic systems based on the deep belief net". In: *Journal of Green Building* 9.9 (Sept. 2017), pp. 1–13. DOI: [10.1177/1687814017715983](https://doi.org/10.1177/1687814017715983).
- [150] Jie Zhang et al. "A suite of metrics for assessing the performance of solar power forecasting". In: *Solar Energy* 111 (2015), pp. 157–175.
- [151] J. S. Armstrong and F. Collopy. "Error measures for generalizing about forecasting methods: Empirical comparisons". In: *International Journal of Forecasting* 8.1 (June 1992), pp. 69–80. DOI: [10.1016/0169-2070\(92\)90008-W](https://doi.org/10.1016/0169-2070(92)90008-W).
- [152] Taraneh Saadati and Burak Barutcu. "Forecasting Solar Energy: Leveraging Artificial Intelligence and Machine Learning for Sustainable Energy Solutions". In: *Journal of Economic Surveys* (2025).
- [153] Nurul Jannah et al. "Recent Advances and Future Challenges of Solar Power Generation Forecasting". In: *IEEE Access* (2024).
- [154] Gabriele Piantadosi et al. "Photovoltaic power forecasting: A Transformer based framework". In: *Energy and AI* 18 (2024), p. 100444.
- [155] Rafiq Asghar et al. "Artificial neural networks for photovoltaic power forecasting: a review of five promising models". In: *IEEE Access* (2024).
- [156] Dheeraj Kumar Dhaked et al. "Exploring the landscape of deep learning for solar photovoltaic power output forecasting: A review". In: *Renewable Energy Focus* (2025), p. 100682.
- [157] Qiushi Zhang et al. "A Review of Photovoltaic Power Generation Forecasting Techniques and Deep Learning Models". In: *2024 The 9th International Conference on Power and Renewable Energy (ICPRE)*. IEEE. 2024, pp. 1342–1347.
- [158] Luke Taylor and Geoff Nitschke. "Improving deep learning with generic data augmentation". In: *2018 IEEE symposium series on computational intelligence (SSCI)*. IEEE. 2018, pp. 1542–1547.
- [159] Alba Alcañiz et al. "Effect of Climate on Photovoltaic Yield Prediction Using Machine Learning Models". In: *Global Challenges* 7.1 (2023), p. 2200166.
- [160] Tao Hong et al. "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond". In: *International Journal of Forecasting* 32.3 (July 2016), pp. 896–913. ISSN: 01692070. DOI: [10.1016/j.ijforecast.2016.02.001](https://doi.org/10.1016/j.ijforecast.2016.02.001).
- [161] Bill Marion et al. "New data set for validating PV module performance models". In: *2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014*. Institute of Electrical and Electronics Engineers Inc., Oct. 2014, pp. 1362–1366. ISBN: 9781479943982. DOI: [10.1109/PVSC.2014.6925171](https://doi.org/10.1109/PVSC.2014.6925171).

- [162] Christil Pasion et al. "Machine learning modeling of horizontal photovoltaics using weather and location data". In: *Energies* 13.10 (May 2020), p. 2570. ISSN: 19961073. DOI: [10.3390/en13102570](https://doi.org/10.3390/en13102570). URL: www.mdpi.com/journal/energies.
- [163] NREL: Developer Network. *PVDAQ (PV Data Acquisition) API*. URL: <https://developer.nrel.gov/docs/solar/pvdaq-v3/> (visited on 02/21/2022).
- [164] Desert Knowledge Australia. *DKA Solar Centre*. URL: <http://dkasolarcentre.com.au/> (visited on 06/15/2021).
- [165] SMA Solar Technology AG. *PV System List - Sunny Portal*. URL: <https://www.sunnyportal.com/Plants> (visited on 06/15/2021).
- [166] NREL. *Datasets - DuraMAT Data Hub*. URL: <https://datahub.duramat.org/dataset> (visited on 02/21/2022).
- [167] *Photovoltaic (PV) Solar Panel Energy Generation data - London Datastore*. URL: <https://data.london.gov.uk/dataset/photovoltaic--pv--solar-panel-energy-generation-data> (visited on 02/21/2022).
- [168] Google. *Kaggle: Your Machine Learning and Data Science Community*. URL: <https://www.kaggle.com/> (visited on 02/21/2022).
- [169] Roger H French et al. *Assessment of Performance Loss Rate of PV Power Systems*. 2021.
- [170] Varun. Mehra. *Solar Home System Jharkhand India*. Aug. 2016. URL: <https://data.openei.org/submissions/681>.
- [171] Paul Norton et al. *Hathaway"Solar Patriot"House: A Case Study in Efficiency and Renewable Energy*. Tech. rep. National Renewable Energy Lab., Golden, CO (US), 2005.
- [172] Elia Group. *Elia: Belgian's Electricity System Operator*. URL: <https://www.elia.be/> (visited on 06/20/2022).
- [173] Bundesnetzagentur. *SMARD- Market data*. URL: <https://www.smard.de/en/> (visited on 06/20/2022).
- [174] Korea Power Exchange. *Electric Power Statistics Information System (EPSIS)*. URL: <http://epsis.kpx.or.kr/epsisnew/selectMain.do?locale=eng> (visited on 06/20/2022).
- [175] Paul-Frederik Bach. URL: <http://www.pfbach.dk/> (visited on 02/21/2022).
- [176] Neon Neue Energieökonomik, Technical University of Berlin, and ETH Zürich. *Open Power System Data*. URL: <https://open-power-system-data.org/> (visited on 06/20/2022).
- [177] Sandia Energy. *Photovoltaic Collaborative to Advance Multi-climate Performance and Energy Research (PV CAMPER)*. URL: <https://energy.sandia.gov/programs/renewable-energy/photovoltaic-solar-energy/projects/photovoltaic-collaborative-to-advance-multi-climate-performance-and-energy-research-pv-camper/> (visited on 06/08/2022).
- [178] William F Holmgren, Clifford W Hansen, and Mark A Mikofski. "pvlib python: A python package for modeling solar energy systems". In: *Journal of Open Source Software* 3.29 (2018), p. 884. DOI: [10.21105/joss.00884](https://doi.org/10.21105/joss.00884).

- [179] Deutsche Wetterdienst. *Wetter und Klima aus einer Hand*. URL: <https://www.dwd.de/DE/Home/home%5C%5Fnode.html> (visited on 05/25/2022).
- [180] Jan Kleissl. *Solar energy forecasting and resource assessment*. Academic Press, 2013.
- [181] Christopher M Bishop. *Pattern recognition and machine learning*. Information Science and Statistics. Springer, 2006.
- [182] Wei Yin Loh. “Classification and regression trees”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1.1 (2011), pp. 14–23. ISSN: 19424787. DOI: [10.1002/widm.8](https://doi.org/10.1002/widm.8).
- [183] Tianqi Chen and Carlos Guestrin. “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.
- [184] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [185] Julián Ascencio-Vásquez, Kristijan Brecl, and Marko Topič. “Methodology of Köppen-Geiger-Photovoltaic climate classification and implications to worldwide mapping of PV system performance”. In: *Solar Energy* 191 (2019), pp. 672–685. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2019.08.072>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X19308527>.
- [186] Wilberth Herrera. *Climate of Costa Rica*. 2016.
- [187] Markus Kottek et al. “World Map of the Köppen-Geiger climate classification updated”. In: *Meteorologische Zeitschrift* 15.3 (July 2006), pp. 259–263. ISSN: 09412948. DOI: [10.1127/0941-2948/2006/0130](https://doi.org/10.1127/0941-2948/2006/0130).
- [188] Tareq Salamah et al. “Effect of dust and methods of cleaning on the performance of solar PV module for different climate regions: Comprehensive review”. In: *Science of The Total Environment* 827 (June 2022), p. 154050. ISSN: 0048-9697. DOI: [10.1016/J.SCITOTENV.2022.154050](https://doi.org/10.1016/J.SCITOTENV.2022.154050).
- [189] Piya Narkwatchara, Chavalit Ratanatamskul, and Achara Chandrachai. “Effects of particulate matters and climate condition on photovoltaic system efficiency in tropical climate region”. In: *Energy Reports* 6 (Nov. 2020), pp. 2577–2586. ISSN: 2352-4847. DOI: [10.1016/J.EGYR.2020.09.016](https://doi.org/10.1016/J.EGYR.2020.09.016).
- [190] Hadia Awad et al. “Predicting the energy production by solar photovoltaic systems in cold-climate regions”. In: <https://doi.org/10.1080/14786451.2017.1408622> 37.10 (Nov. 2017), pp. 978–998. ISSN: 1478646X. DOI: [10.1080/14786451.2017.1408622](https://doi.org/10.1080/14786451.2017.1408622). URL: <https://www.tandfonline.com/doi/abs/10.1080/14786451.2017.1408622>.
- [191] Herman Böök et al. “Photovoltaic system modeling: A validation study at high latitudes with implementation of a novel DNI quality control method”. In: *Solar Energy* 204 (July 2020), pp. 316–329. ISSN: 0038-092X. DOI: [10.1016/J.SOLENER.2020.04.068](https://doi.org/10.1016/J.SOLENER.2020.04.068).
- [192] D. Pujić, M. Jelić, and N. Tomašević. “Comparison between different ML approaches for PV and STC production forecasting using real world data”. In: *Proceedings of the International Conference on Information Science and Technology*. 2020.

- [193] Daniel Grzebyk et al. "Individual yield nowcasting for residential PV systems". In: *Solar Energy* 251 (2023), pp. 325–336.
- [194] Wenyong Zhang, Qingwei Li, and Qifeng He. "Application of machine learning methods in photovoltaic output power prediction: A review". In: *Journal of Renewable and Sustainable Energy* 14.2 (Apr. 2022), p. 022701. ISSN: 19417012. DOI: [10.1063/5.0082629](https://doi.org/10.1063/5.0082629). URL: <https://aip.scitation.org/doi/abs/10.1063/5.0082629>.
- [195] Boudewijn Elsinga and Wilfried GJHM van Sark. "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems". In: *Applied Energy* 206 (2017), pp. 1464–1483.
- [196] KNMI - Koninklijk Nederlands Meteorologisch Instituut. <https://www.knmi.nl/home>. Accessed: 2021-08-06.
- [197] Nailya Maitanova et al. "Machine Learning Approach to a Low-cost Day-ahead Photovoltaic Power Prediction Based on Publicly available Weather Reports". In: *Energies* 13.3 (2020). DOI: [10.3390/en13030735](https://doi.org/10.3390/en13030735).
- [198] Solar Monkey. <https://solarmonkey.nl/>. Accessed: 2021-08-06.
- [199] Isabelle Guyon et al. "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46.1 (2002), pp. 389–422.
- [200] Pablo M Granitto et al. "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products". In: *Chemometrics and intelligent laboratory systems* 83.2 (2006), pp. 83–90.
- [201] Jason Brownlee. *Feature Importance and Feature Selection With XGBoost in Python*. <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/>, accessed: 25.11.2021. 2016.
- [202] Tim N.C. de Vries et al. "A quick-scan method to assess photovoltaic rooftop potential based on aerial imagery and LiDAR". In: *Solar Energy* 209.2 (2020), pp. 96–107. DOI: [10.1016/j.solener.2020.07.035](https://doi.org/10.1016/j.solener.2020.07.035). URL: <https://doi.org/10.1016/j.solener.2020.07.035>.
- [203] William Vorhies. *Want to Win Competitions? Pay Attention to Your Ensembles*. May 2016. URL: <https://www.datasciencecentral.com/profiles/blogs/want-to-win-at-kaggle-pay-attention-to-your-ensembles>.
- [204] Qi Wang et al. "A comprehensive survey of loss functions in machine learning". In: *Annals of Data Science* (2020), pp. 1–26.
- [205] Eneco. *Our exits: peeks*. <https://www.eneco.com/wat-we-doen/Ventures/onze-exits/>. Accessed: 2021-08-09.
- [206] Alba Alcañiz et al. "Photovoltaic system monitoring and fault detection using peer systems". In: *Progress in Photovoltaics: research and applications* 30.9 (2022), pp. 1072–1086.
- [207] Dhivya Sampath Kumar et al. "Solar irradiance resource and forecasting: a comprehensive review". In: *IET Renewable Power Generation* 14.10 (2020), pp. 1641–1656.
- [208] Vincent P.A. Lonij et al. "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors". In: *Solar Energy* 97 (Nov. 2013), pp. 58–66. ISSN: 0038092X. DOI: [10.1016/j.solener.2013.08.002](https://doi.org/10.1016/j.solener.2013.08.002).

- [209] Boudewijn Elsinga and Wilfried van Sark. "Spatial power fluctuation correlations in urban rooftop photovoltaic systems". In: *Progress in Photovoltaics: Research and Applications* 23.10 (2015), pp. 1390–1397.
- [210] Anastasios Golnas et al. "Performance assessment without pyranometers: Predicting energy output based on historical correlation". In: *Conference Record of the IEEE Photovoltaic Specialists Conference* 6 (2011), pp. 002006–002010. ISSN: 01608371. DOI: [10.1109/PVSC.2011.6186347](https://doi.org/10.1109/PVSC.2011.6186347).
- [211] Odysseas Tsafarakis, Kostas Sinapis, and Wilfried G.J.H.M. Van Sark. "PV system performance evaluation by clustering production data to normal and non-normal operation". In: *Energies* 11.4 (2018). ISSN: 19961073. DOI: [10.3390/en11040977](https://doi.org/10.3390/en11040977). URL: www.mdpi.com/journal/energies.
- [212] Ivan Popovic and Ilija Radovanovic. "Methodology for detection of photovoltaic systems underperformance operation based on the correlation of irradiance estimates of neighboring systems". In: *Journal of Renewable and Sustainable Energy* 10.5 (Sept. 2018), p. 053701. ISSN: 19417012. DOI: [10.1063/1.5042579](https://doi.org/10.1063/1.5042579). URL: <https://aip.scitation.org/doi/abs/10.1063/1.5042579>.
- [213] Jonathan Leloux et al. "Automatic fault detection on BIPV systems without solar irradiation data". In: *29th European Photovoltaic Solar Energy Conference and Exhibition* (2014).
- [214] Jonathan Leloux et al. "Performance to Peers (P2P): A benchmark approach to fault detections applied to photovoltaic system fleets". In: *Solar Energy* 202 (May 2020), pp. 522–539. ISSN: 0038092X. DOI: [10.1016/j.solener.2020.03.015](https://doi.org/10.1016/j.solener.2020.03.015).
- [215] Granino A. Korn and Theresa M. Korn. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. Vol. 15. 76. Dover Publications, 1961, p. 421. ISBN: 9780486411477. DOI: [10.2307/2003035](https://doi.org/10.2307/2003035).
- [216] Mirko Stojiljkovic. "Stochastic Gradient Descent Algorithm With Python and NumPy". In: *Real Python* (2021). URL: <https://realpython.com/gradient-descent-algorithm-python/>.
- [217] John A Nelder and Roger Mead. "A simplex method for function minimization". In: *The computer journal* 7.4 (1965), pp. 308–313.
- [218] Sidhartha Panda and Narayana Prasad Padhy. "Comparison of particle swarm optimization and genetic algorithm for FACTS-based controller design". In: *Applied Soft Computing Journal* 8.4 (Sept. 2008), pp. 1418–1427. ISSN: 15684946. DOI: [10.1016/j.asoc.2007.10.009](https://doi.org/10.1016/j.asoc.2007.10.009).
- [219] Ajith Abraham et al. "Multi-objective peer-to-peer neighbor-selection strategy using genetic algorithm". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 4873 LNCS. Springer, Berlin, Heidelberg, Dec. 2007, pp. 443–451. ISBN: 9783540772194. DOI: [10.1007/978-3-540-77220-0_41](https://doi.org/10.1007/978-3-540-77220-0_41). URL: <https://link.springer.com/chapter/10.1007/978-3-540-77220-0%7B%5C%5F%7D41>.
- [220] Sana Rehman et al. "Bi-Directional Mutual Energy Trade between Smart Grid and Energy Districts Using Renewable Energy Credits". In: *Sensors* 21.9 (2021), p. 3088.

- [221] Shijie Sun et al. "Device-to-device resource allocation in LTE-advanced networks by hybrid particle swarm optimization and genetic algorithm". In: *Peer-to-Peer Networking and Applications* 9.5 (2016), pp. 945–954.
- [222] John Henry Holland et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [223] Melanie Mitchell. "Genetic algorithms: An overview." In: *Complex*. Vol. 1. 1. Citeseer. 1995, pp. 31–39.
- [224] Ofer M. Shir. "Niching in Evolutionary Algorithms". In: *Handbook of Natural Computing*. Ed. by Grzegorz Rozenberg, Thomas Bäck, and Joost N. Kok. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1035–1069. ISBN: 978-3-540-92910-9. DOI: [10.1007/978-3-540-92910-9_32](https://doi.org/10.1007/978-3-540-92910-9_32).
- [225] Tobias Blickle and Lothar Thiele. "A comparison of selection schemes used in evolutionary algorithms". In: *Evolutionary Computation* 4.4 (1996), pp. 361–394.
- [226] Thomas Back. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.
- [227] V Berdugo et al. "Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems. Next Gener. Data Min. Summit Ubiquitous Knowledge Discov". In: *Energy Manage. Smart Grids Intell. Mach. to-Machine Telemat* (2011).
- [228] Chen Yang and Le Xie. "A novel ARX-based multi-scale spatio-temporal solar power forecast model". In: *2012 North American Power Symposium (NAPS)*. IEEE. 2012, pp. 1–6.
- [229] A. G.R. Vaz et al. "An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands". In: *Renewable Energy* 85 (Jan. 2016), pp. 631–641. ISSN: 18790682. DOI: [10.1016/j.renene.2015.06.061](https://doi.org/10.1016/j.renene.2015.06.061).
- [230] Charles Spearman. "The proof and measurement of association between two things". In: *The American journal of psychology* 100.3/4 (1987), pp. 441–471.
- [231] Yitzi Snow. *Peer to peer (P2P) optimization methodology for medium size commercial-residential PV systems*. Tech. rep. private report. Trust-PV, 2023.
- [232] MR Vogt et al. "Introducing a comprehensive physics-based modelling framework for tandem and other PV systems". In: *Solar Energy Materials and Solar Cells* 247 (2022), p. 111944.
- [233] I.T. Horvath et al. *Automated PV digital twin-based yield simulation framework*. Tech. rep. Trust-PV, 2022.
- [234] A. Alcañiz et al. *Residential-scale integrated energy yield simulation and reliability modelling framework*. Tech. rep. Trust-PV, 2023.
- [235] David Faiman. "Assessing the outdoor operating temperature of photovoltaic modules". In: *Progress in Photovoltaics: Research and Applications* 16.4 (2008), pp. 307–315.

- [236] Martin K Fuentes. *A simplified thermal model for flat-plate photovoltaic arrays*. Tech. rep. Sandia National Labs., Albuquerque, NM (USA), 1987.
- [237] E. Garcia Goma. “Development of Cell to System Annual Energy Yield Toolbox for Bifacial Modules”. MA thesis. Delft University of Technology, 2018.
- [238] Hesani Ziar et al. “Innovative floating bifacial photovoltaic solutions for inland water areas”. In: *Progress in Photovoltaics: research and applications* 29.7 (2021), pp. 725–743.
- [239] Julen A. Garro Etxebarria. “Toolbox for the design and simulation of a floating bifacial PV plant with reflectors”. MA thesis. Delft University of Technology, 2018.
- [240] I. Narváez Alavez. “Monitoring of bifacial floating PV systems and validation of a toolbox for its simulation”. MA thesis. Delft University of Technology, 2019.
- [241] Z. Wang. “Improvements and Experimental Validation of the PVMD Toolbox: An energy yield prediction model for PV systems”. MA thesis. Delft University of Technology, 2019.
- [242] Tim Stark. “Modeling and Monitoring of a Floating Photovoltaic Pilot System”. MA thesis. Delft University of Technology, 2020.
- [243] A. Nour El Din. “Improved Electrical Model and Experimental Validation of the PVMD Toolbox”. MA thesis. Delft University of Technology, 2020.
- [244] SISIFO. URL: <https://www.sisifo.info> (visited on 03/22/2024).
- [245] NREL. *PVWatts Calculator*. URL: <https://pvwatts.nrel.gov/> (visited on 03/22/2024).
- [246] *Solar Pathfinder*. URL: <https://www.solarpathfinder.com/> (visited on 03/22/2024).
- [247] RatedPower. *pvDesign*. URL: <https://ratedpower.com/pvdesign/> (visited on 03/25/2024).
- [248] TU Delft. *The Dutch PV Portal 3.0*. URL: <https://pvportal-3.ewi.tudelft.nl/PVP3.1NL/Mod%5C%5FDesignTool/Design.php> (visited on 03/22/2024).
- [249] First solar. *Plant predict*. URL: <https://www.plantpredict.com/> (visited on 10/07/2023).
- [250] NREL. *System Advisor Model*. <https://sam.nrel.gov/>. Accessed: 2023-06-15.
- [251] Valentin software GmbH. *PV*Sol*. URL: <https://www.valentinsoftware.com/en/products/photovoltaics/57/pvsol-premium> (visited on 10/07/2023).
- [252] EnerBIM. *BIMSolar*. URL: <https://www.bim-solar.com/> (visited on 03/25/2024).
- [253] DNV. *SolarFarmer*. URL: <https://www.dnv.com/software/services/solarfarmer/> (visited on 03/25/2024).
- [254] *eurac research*. URL: <https://www.eurac.edu/> (visited on 03/22/2024).
- [255] Trace Software. *Archelios Pro*. URL: <https://www.trace-software.com/archelios-pro/solar-pv-design-software/> (visited on 03/22/2024).
- [256] Aurora Solar. *HelioScope*. <https://helioscope.aurorasolar.com/>. Accessed: 2023-06-15.
- [257] PVsyst S.A. *PVsyst - Photovoltaic Software*. <https://www.pvsyst.com/>. Version 7.2. Accessed: 2023-10-23. 2023.

- [258] Imec. *Precision model*. URL: <https://www.imec-int.com/en/articles/predicting-energy-yields-photovoltaic-systems-imecs-precision-model> (visited on 03/26/2024).
- [259] PVcase. *Yield*. URL: <https://pvcase.com/yield/> (visited on 03/26/2024).
- [260] Ismail Kaaya et al. "Modeling Outdoor Service Lifetime Prediction of PV Modules: Effects of Combined Climatic Stressors on PV Module Power Degradation". In: *IEEE Journal of Photovoltaics* 9.4 (2019), pp. 1105–1112. DOI: [10.1109/JPHOTOV.2019.2916197](https://doi.org/10.1109/JPHOTOV.2019.2916197).
- [261] Arno Smets et al. *Solar Energy: The physics and engineering of photovoltaic conversion, technologies and systems*. Bloomsbury Publishing, 2016.
- [262] Rudi Santbergen et al. "GenPro4 optical model for solar cell simulation and its application to multijunction solar cells". In: *IEEE journal of photovoltaics* 7.3 (2017), pp. 919–926.
- [263] Robert Siegel. "Net radiation method for transmission through partially transparent plates". In: *Solar Energy* 15 (1973).
- [264] Rudi Santbergen and RJ Ch van Zolingen. "The absorption factor of crystalline silicon PV cells: A numerical and experimental study". In: *Solar energy materials and solar cells* 92.4 (2008), pp. 432–444.
- [265] R Santbergen et al. "The AM1. 5 absorption factor of thin-film solar cells". In: *Solar energy materials and solar cells* 94.5 (2010), pp. 715–723.
- [266] K Jäger et al. "A scattering model for nano-textured interfaces and its application in opto-electrical simulations of thin-film silicon solar cells". In: *Journal of Applied Physics* 111.8 (2012).
- [267] M Zeman et al. "Modelling of thin-film silicon solar cells". In: *Solar Energy Materials and Solar Cells* 119 (2013), pp. 94–111.
- [268] PVMD group. *ASA web simulator*. URL: <https://asa.ewi.tudelft.nl/> (visited on 11/10/2023).
- [269] R Santbergen et al. "Calculation of irradiance distribution on PV modules by combining sky and sensitivity maps". In: *Solar Energy* 150 (2017), pp. 49–54.
- [270] National Ocean Service - National Oceanic and Atmospheric Administration. *What is lidar?* URL: <https://oceanservice.noaa.gov/facts/lidar.html> (visited on 03/28/2024).
- [271] Martijn Keijzer. "A multi-surface reflected irradiance model for pyranometer corrections and PV yield calculations in complex urban geometries". MA thesis. the Netherlands: Delft University of Technology, 2019. URL: <http://resolver.tudelft.nl/uuid:d1a6baf0-7721-441e-bc2c-3a05bdaea89d>.
- [272] Furkan Fatih Sönmez et al. "Fast and accurate ray-casting-based view factor estimation method for complex geometries". In: *Solar Energy Materials and Solar Cells* 200 (2019), p. 109934.
- [273] DG Steyn. "The calculation of view factors from fisheye-lens photographs: Research note". In: (1980).

- [274] Richard Perez, Robert Seals, and Joseph Michalsky. "All-weather model for sky luminance distribution—preliminary configuration and validation". In: *Solar energy* 50.3 (1993), pp. 235–245.
- [275] Christian Gueymard. "Simple model of the atmospheric radiative transfer of sunshine (SMARTS2): algorithms and performance assessment". In: *Florida: Solar Energy Center* (1995).
- [276] Amit Jain and Avinashi Kapoor. "Exact analytical solutions of the parameters of real solar cells using Lambert W-function". In: *Solar Energy Materials and Solar Cells* 81.2 (2004), pp. 269–277.
- [277] William Earl Boyson et al. *Performance model for grid-connected photovoltaic inverters*. Tech. rep. Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA ..., 2007.
- [278] Meteonorm. *Horicatcher*. URL: <https://meteonorm.com/en/product/horicatcher> (visited on 03/28/2024).
- [279] S. Mishra. "Selection Map for PV Module Installation Based on Shading Tolerability and Temperature Coefficient". MA thesis. the Netherlands: PVMD, TU Delft, 2018.
- [280] Barbara Ridley, John Boland, and Philippe Lauret. "Modelling of diffuse solar fraction with multiple predictors". In: *Renewable Energy* 35.2 (2010), pp. 478–483. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2009.07.018>. URL: <http://www.sciencedirect.com/science/article/pii/S0960148109003012>.
- [281] Ricardo Marquez, Hugo TC Pedro, and Carlos FM Coimbra. "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs". In: *Solar Energy* 92 (2013), pp. 176–188.
- [282] Zhiyuan Si et al. "Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks". In: *IEEE Transactions on Industry Applications* 57.1 (2020), pp. 5–16.
- [283] Paulo AC Rocha and Victor Oliveira Santos. "Global horizontal and direct normal solar irradiance modeling by the machine learning methods XGBoost and deep neural networks with CNN-LSTM layers: a case study using the GOES-16 satellite imagery". In: *International Journal of Energy and Environmental Engineering* 13.4 (2022), pp. 1271–1286.
- [284] Raimondo Gallo et al. "Solar radiation forecasting with deep learning techniques integrating geostationary satellite images". In: *Engineering Applications of Artificial Intelligence* 116 (2022), p. 105493.
- [285] Jong-Min Yeom et al. "Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional LSTM networks for South Korea". In: *Environmental Research Letters* 15.9 (2020), p. 094025.
- [286] Andreas H Nielsen, Alexandros Iosifidis, and Henrik Karstoft. "IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting". In: *Solar Energy* 228 (2021), pp. 659–669.
- [287] A.J. Wieringa. "Irradiance Map forecasting using satellite images and deep learning". MA thesis. Delft University of Technology, 2024.

- [288] Alba Alcañiz et al. “Wide-area sky view factor analysis and Fourier-based decomposition model for optimizing irradiance sensors allocation in European solar photovoltaic farms: A software tool”. In: *Solar Energy* 286 (2025), p. 113139.
- [289] K. Lambers. Private communication. Asset manager at Everyday. Mar. 2023.
- [290] J. Hommel. Private communication. Bakker en Co. Mar. 2023.
- [291] H. Buikema. Private communication. Powerfield. Mar. 2023.
- [292] Nils H Reich et al. “Performance ratio revisited: is PR> 90% realistic?” In: *Progress in Photovoltaics: Research and Applications* 20.6 (2012), pp. 717–726.
- [293] IEC. *Photovoltaic system performance - Part 1: Monitoring (IEC 61724-1:2021)*. Standard. CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels: International Electrotechnical Commission, 2021. URL: <https://webstore.iec.ch/publication/65561>.
- [294] Marco Pau, Nicola Locci, and Carlo Muscas. “A tool to define the position and the number of irradiance sensors in large PV plants”. In: *2014 IEEE International Energy Conference (ENERGYCON)*. IEEE. 2014, pp. 374–379.
- [295] G. Oviedo. Private communication. BayWa r.e. 2023.
- [296] Marc Korevaar and Damon Nitzel. “Simulation of POA front irradiance sensor mounting position”. In: *European PVPMC* 8. 2023.
- [297] Lian Lian Jiang, R Srivatsan, and Douglas L Maskell. “Optimal Irradiance Sensor Placement for Photovoltaic Systems Using Mutual Information Based Greedy Algorithm in Gaussian Process”. In: *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*. IEEE. 2017, pp. 120–126.
- [298] Annalisa Di Piazza, Maria Carmela Di Piazza, and Gianpaolo Vitale. “A kriging-based partial shading analysis in a large photovoltaic field for energy forecast”. In: *International Conference on Renewable Energies and Power Quality (ICREPQ'09) Valencia, Spain*. 2009.
- [299] Marco Ernst et al. “Comparison of ground-based and satellite-based irradiance data for photovoltaic yield estimation”. In: *Energy Procedia* 92 (2016), pp. 546–553.
- [300] A Harsarapama et al. “Accuracy of irradiance measurement for a PV park versus the number of sensors”. In: *Optics for Solar Energy*. Optica Publishing Group. 2017, RTh2B–2.
- [301] M. Alasdair and L. Ben. *Utility scale solar power plants: A guide for developers and investors*. Tech. rep. South asia Department of International Finance Corporation World Bank Group, 2012.
- [302] R. Perez et al. “Modeling daylight availability and irradiance components from direct and global irradiance”. In: *Solar energy* 44 (1990), pp. 271–289. ISSN: 5. DOI: [https://doi.org/10.1016/0038-092X\(90\)90055-H](https://doi.org/10.1016/0038-092X(90)90055-H). URL: <https://www.sciencedirect.com/science/article/pii/0038092X9090055H>.
- [303] Dora de Jong and Hesam Ziar. “Photovoltaic Potential of the Dutch Inland Shipping Fleet: An Experimentally Validated Method to Simulate the Power Series from Vessel-Integrated Photovoltaics”. In: *Solar RRL* 7.8 (2023), p. 2200642.

- [304] Global Energy Observatory. Google. KTH Royal Institute of Technology in Stockholm. Enipedia. World Resources Institute. *Global Energy Observatory*. Published on Resource Watch and Google Earth Engine; <http://resourcewatch.org/> <https://earthengine.google.com/>. 2018.
- [305] Sinergise Laboratory for geographical information systems, Ltd. *Sentinel-Hub*. <http://www.sentinel-hub.com/>.
- [306] Dazhi Yang. “Solar radiation on inclined surfaces: Corrections and benchmarks”. In: *Solar Energy* 136 (2016), pp. 288–302. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2016.06.062>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X16302432>.
- [307] Dazhi Yang. “Estimating 1-min beam and diffuse irradiance from the global irradiance: A review and an extensive worldwide comparison of latest separation models at 126 stations”. In: *Renewable and Sustainable Energy Reviews* 159 (2022), p. 112195. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2022.112195>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032122001174>.
- [308] Lavinia Laiti et al. “Estimating Hourly Beam and Diffuse Solar Radiation in an Alpine Valley: A Critical Assessment of Decomposition Models”. In: *Atmosphere* 9.4 (2018). ISSN: 2073-4433. DOI: [10.3390/atmos9040117](https://doi.org/10.3390/atmos9040117). URL: <https://www.mdpi.com/2073-4433/9/4/117>.
- [309] Cédric Bertrand, Gilles Vanderveken, and Michel Journée. “Evaluation of decomposition models of various complexity to estimate the direct solar irradiance over Belgium”. In: *Renewable Energy* 74 (2015), pp. 618–626. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2014.08.042>. URL: <https://www.sciencedirect.com/science/article/pii/S0960148114005011>.
- [310] Sokol Dervishi and Ardeshtir Mahdavi. “Comparison of models for the derivation of diffuse fraction of global irradiance data for vienna, austria”. In: *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association* (Jan. 2011).
- [311] Jayanta Deb Mondol, Yigzaw G. Yohanis, and Brian Norton. “Solar radiation modelling for the simulation of photovoltaic systems”. In: *Renewable Energy* 33.5 (2008), pp. 1109–1120. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2007.06.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0960148107002157>.
- [312] F. Lanini. “Division of global radiation into direct radiation and diffuse radiation”. MA thesis. Bern, Switzerland: Institute of Geography, University of Bern, 2010.
- [313] Ana Pérez-Burgos et al. “Retrieval of monthly average hourly values of direct and diffuse solar irradiance from measurements of global radiation in Spain”. In: *Journal of Renewable and Sustainable Energy* 10 (Mar. 2018), p. 023707. DOI: [10.1063/1.5016926](https://doi.org/10.1063/1.5016926).
- [314] Paulescu Eugenia and Robert Blaga. “Regression models for hourly diffuse solar radiation”. In: *Solar Energy* 125 (Nov. 2015), pp. 111–124. DOI: [10.1016/j.solener.2015.11.044](https://doi.org/10.1016/j.solener.2015.11.044).

- [315] D et al. Martinsen. "Poster: evaluation of irradiance and transposition models for a Nordic climate". In: Oct. 2022.
- [316] Colienne Demain, Michel Journée, and Cédric Bertrand. "Evaluation of different models to estimate the global solar radiation on inclined surfaces". In: *Renewable Energy* 50 (2013), pp. 710–721. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2012.07.031>. URL: <https://www.sciencedirect.com/science/article/pii/S0960148112004570>.
- [317] G. Notton, P. Poggi, and C. Cristofari. "Predicting hourly solar irradiations on inclined surfaces based on the horizontal measurements: Performances of the association of well-known mathematical models". In: *Energy Conversion and Management* 47.13 (2006), pp. 1816–1829. ISSN: 0196-8904. DOI: <https://doi.org/10.1016/j.enconman.2005.10.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0196890405002645>.
- [318] Miklós Horváth and Tamás Csoknyai. "Evaluation of Solar Energy Calculation Methods for 45° Inclined, South Facing Surface". In: *Energy Procedia* 78 (2015). 6th International Building Physics Conference, IBPC 2015, pp. 465–470. ISSN: 1876-6102. DOI: <https://doi.org/10.1016/j.egypro.2015.11.700>. URL: <https://www.sciencedirect.com/science/article/pii/S1876610215024327>.
- [319] M. Cucumo et al. "Experimental testing of models for the estimation of hourly solar radiation on vertical surfaces at Arcavacata di Rende". In: *Solar Energy* 81.5 (2007), pp. 692–695. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2006.09.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X06002258>.
- [320] Andrea Padovan and Davide Del Col. "Measurement and modeling of solar irradiance components on horizontal and tilted planes". In: *Solar Energy* 84.12 (2010), pp. 2068–2084. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2010.09.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X10003014>.
- [321] Ajay Jamodkar. "Energy Yield Prediction of Solar Powered E-bike Charging Station". MA thesis. the Netherlands: Delft University of Technology, 2017. URL: <https://repository.tudelft.nl/islandora/object/uuid:2abcd917-7298-4ae5-8c0a-392bb31a1ef2/datastream/OBJ/download>.
- [322] C Vasar, O Prostean, and G Prostean. "Evaluating solar radiation on a tilted surfaces—a study case in Timis (Romania)". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 106. 1. IOP Publishing. 2016, p. 012026.
- [323] R. Posadillo and R. López Luque. "Evaluation of the performance of three diffuse hourly irradiation models on tilted surfaces according to the utilizability concept". In: *Energy Conversion and Management* 50.9 (2009), pp. 2324–2330. ISSN: 0196-8904. DOI: <https://doi.org/10.1016/j.enconman.2009.05.014>. URL: <https://www.sciencedirect.com/science/article/pii/S0196890409001885>.

- [324] P.G. Loutzenhiser et al. “Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation”. In: *Solar Energy* 81.2 (2007), pp. 254–267. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2006.03.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X06000879>.
- [325] Pierre Ineichen. *Global irradiance on tilted and oriented planes: model validations*. Tech. rep. Univeristy of Geneva, 2011.
- [326] R. Perez et al. “A new simplified version of the perez diffuse irradiance model for tilted surfaces”. In: *Solar energy* 39 (1987), pp. 221–231. DOI: [https://doi.org/10.1016/S0038-092X\(87\)80031-2](https://doi.org/10.1016/S0038-092X(87)80031-2).
- [327] S. Lloyd. “Least squares quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137. DOI: [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).
- [328] Thomas Huld, Richard Müller, and Attilio Gambardella. “A new solar radiation database for estimating PV performance in Europe and Africa”. In: *Solar energy* 86.6 (2012), pp. 1803–1815.
- [329] A. Driemel et al. “Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017)”. In: *Earth System Science Data* 10.3 (2018), pp. 1491–1501. DOI: [10.5194/essd-10-1491-2018](https://doi.org/10.5194/essd-10-1491-2018). URL: <https://essd.copernicus.org/articles/10/1491/2018/>.
- [330] Fanglin Yang et al. “Dependence of land surface albedo on solar zenith angle: Observations and model parameterization”. In: *Journal of Applied Meteorology and Climatology* 47.11 (2008), pp. 2963–2982.
- [331] Robert E Dickinson. “Land surface processes and climate—Surface albedos and energy balance”. In: *Advances in geophysics*. Vol. 25. Elsevier, 1983, pp. 305–353.
- [332] Hesam Ziar et al. “A comprehensive albedo model for solar energy applications: Geometric spectral albedo”. In: *Applied Energy* 255 (2019), p. 113867. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2019.113867>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261919315545>.
- [333] W. Lucht, C.B. Schaaf, and A.H. Strahler. “An algorithm for the retrieval of albedo from space using semiempirical BRDF models”. In: *IEEE Transactions on Geoscience and Remote Sensing* 38.2 (2000), pp. 977–998. DOI: [10.1109/36.841980](https://doi.org/10.1109/36.841980).
- [334] Ying Qu et al. “Direct-Estimation Algorithm for Mapping Daily Land-Surface Broadband Albedo From MODIS Data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.2 (2014), pp. 907–919. DOI: [10.1109/TGRS.2013.2245670](https://doi.org/10.1109/TGRS.2013.2245670).
- [335] Shreedatta Marathe. “Effect of urban development on local albedo”. MA thesis. the Netherlands: Delft University of Technology, 2022. URL: <https://repository.tu-delft.nl/islandora/object/uuid%5C%3A7c13032f-1758-4c11-b9df-ec823ab5b3f7>.
- [336] Adam H. Sparks. “nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R”. In: *The Journal of Open Source Software* 3.30 (Oct. 2018), p. 1035. DOI: [10.21105/joss.01035](https://doi.org/10.21105/joss.01035).

- [337] Xingwen Lin et al. "Spatiotemporal Variability of Land Surface Albedo over the Tibet Plateau from 2001 to 2019". In: *Remote Sensing* 12 (Apr. 2020), p. 1188. DOI: [10.3390/rs12071188](https://doi.org/10.3390/rs12071188).
- [338] E. Palle et al. "Earth's albedo variations 1998-2014 as measured from ground-based earthshine observations". In: *Geophysical Research Letters* 43.9 (May 2016), pp. 4531–4538. DOI: [10.1002/2016gl068025](https://doi.org/10.1002/2016gl068025). URL: <https://doi.org/10.1002/2016gl068025>.
- [339] Yves-Marie Saint-Drenan et al. "An approach for the estimation of the aggregated photovoltaic power generated in several European countries from meteorological data". In: *Advances in Science and Research* 15 (May 2018), pp. 51–62. DOI: [10.5194/asr-15-51-2018](https://doi.org/10.5194/asr-15-51-2018).
- [340] Google Maps. Web-based mapping service. Accessed: April 22, 2023. URL: <https://www.google.com/maps>.
- [341] Kipp & Zonen. *Suncertainty: the Kipp & Zonen pyranometer uncertainty calculator*. Visited on 4/7/2024. 2016. URL: <https://www.kippzonen.com/News/743/Suncertainty-the-Kipp-Zonen-pyranometer-uncertainty-calculator>.
- [342] C Schaaf and Z Wang. *MCD43C3 MODIS/Terra+ Aqua BRDF/Albedo Albedo Daily L3 Global 0.05 Deg CMG. NASA EOSDIS Land Processes DAAC*. 2015.
- [343] S. Tonnel et al. *Guidelines for design, procurement and O&M friendly concepts for floating PV*. Tech. rep. Trust-PV, 2023.
- [344] S. Tonnel et al. *Demonstration of solution and challenges on floating systems*. Tech. rep. Trust-PV, 2024.
- [345] Richard J. A. M. Stevens et al. "Combining economic and fluid dynamic models to determine the optimal spacing in very large wind farms". In: *Wind Energy* 20.3 (2017), pp. 465–477. ISSN: 10954244. DOI: [10.1002/we.2016](https://doi.org/10.1002/we.2016).
- [346] David Firnando Silalahi and Andrew Blakers. "Global Atlas of Marine Floating Solar PV Potential". In: *Solar*. Vol. 3. MDPI. 2023, pp. 416–433.
- [347] World Bank Group. *Where Sun Meets Water: Floating Solar Market Report*. World Bank, 2019.
- [348] Haohui Liu et al. "Field experience and performance analysis of floating PV technologies in the tropics". In: *Progress in Photovoltaics: Research and Applications* 26.12 (2018), pp. 957–967.
- [349] Anik Goswami and Pradip Kumar Sadhu. "Degradation analysis and the impacts on feasibility study of floating solar photovoltaic systems". In: *Sustainable Energy, Grids and Networks* 26 (2021), p. 100425.
- [350] Maarten Dörenkämper et al. "The cooling effect of floating PV in two different climate zones: A comparison of field test data from the Netherlands and Singapore". In: *Solar Energy* 219 (2021), pp. 15–23.
- [351] Hesam Ziar. "Floating solar stations". In: *10 breakthrough ideas in energy for the next 10 years* (2021), pp. 30–43.

- [352] Manish Kumar, Humaid Mohammed Niyaz, and Rajesh Gupta. “Challenges and opportunities towards the development of floating photovoltaic systems”. In: *Solar Energy Materials and Solar Cells* 233 (2021), p. 111408.
- [353] Sherine El Baradei and Mai Al Sadeq. “Effect of solar canals on evaporation, water quality, and power production: An optimization study”. In: *Water* 12.8 (2020), p. 2103.
- [354] Marco Rosa-Clot and Giuseppe Marco Tina. “Levelized cost of energy (LCOE) analysis”. In: *Floating PV plants*. Elsevier, 2020, pp. 119–127.
- [355] Rafael M Almeida et al. “Floating solar power could help fight climate change—let’s get it right”. In: *Nature* 606.7913 (2022), pp. 246–249.
- [356] Centre National de la Recherche Scientifique (CNRS). *Project Solake*. URL: <https://www.inee.cnrs.fr/fr/dipee-occitanie-ouest> (visited on 08/22/2024).
- [357] S Zahra Golroodbari and Wilfried van Sark. “Simulation of performance differences between offshore and land-based photovoltaic systems”. In: *Progress in Photovoltaics: Research and Applications* 28.9 (2020), pp. 873–886.
- [358] Robert K. Lane. *Seiches*. Britannica. URL: <https://www.britannica.com/science/lake/Surface-waves> (visited on 04/16/2024).
- [359] Christian Schill et al. “Soiling Losses—Impact on the Performance of Photovoltaic Power Plants”. In: *International Energy Agency* (2022).
- [360] Wanxiang Yao et al. “New models for the influence of rainwater on the performance of photovoltaic modules under different rainfall conditions”. In: *Renewable and Sustainable Energy Reviews* 173 (2023), p. 113119.
- [361] J Zorrilla-Casanova et al. “Analysis of dust losses in photovoltaic modules”. In: *World renewable energy congress*. Linköping University Electronic Press Linköping, Sweden. 2011, pp. 2985–2992.
- [362] Hussein A Kazem et al. “A review of dust accumulation and cleaning methods for solar photovoltaic systems”. In: *Journal of Cleaner Production* 276 (2020), p. 123187.
- [363] Seyfi Şevik and Ahmet Aktaş. “Performance enhancing and improvement studies in a 600 kW solar photovoltaic (PV) power plant; manual and natural cleaning, rainwater harvesting and the snow load removal on the PV arrays”. In: *Renewable Energy* 181 (2022), pp. 490–503.
- [364] Julius Tanesab et al. “Dust effect and its economic analysis on PV modules deployed in a temperate climate zone”. In: *Energy Procedia* 100 (2016), pp. 65–68.
- [365] Global Modeling and Assimilation Office (GMAO). *MERRA-2 tavg1_2d_aer_Nx GES DISC*. 2015. DOI: <https://doi.org/10.5067/KLICTZ8EM9D>. (Visited on 04/22/2024).
- [366] Eau France. *HydroPortail v3.1.4.3*. URL: <https://hydro.eaufrance.fr/> (visited on 04/22/2024).
- [367] M Grech et al. “Floating photovoltaic installations in Maltese sea waters”. In: *Proceedings of the 32nd European Photovoltaic Solar Energy Conference and Exhibition, Munich, Germany*. 2016, pp. 20–24.
- [368] R Cazzaniga et al. “Floating photovoltaic plants: Performance analysis and design solutions”. In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 1730–1741.

- [369] Sara Oliveira-Pinto and Jasper Stokkermans. "Assessment of the potential of different floating solar technologies—Overview and analysis of different case studies". In: *Energy Conversion and Management* 211 (2020), p. 112747.
- [370] Leonardo Micheli. "The temperature of floating photovoltaics: Case studies, models and recent findings". In: *Solar Energy* 242 (2022), pp. 234–245. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2022.06.039>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X22004650>.
- [371] John A Duffie and William A Beckman. *Solar engineering of thermal processes*. Wiley New York, 1980.
- [372] Giuseppe Marco Tina et al. "Comparative analysis of monofacial and bifacial photovoltaic modules for floating power plants". In: *Applied Energy* 281 (2021), p. 116084.
- [373] Torunn Kjeldstad et al. "Cooling of floating photovoltaics and the importance of water temperature". In: *Solar Energy* 218 (2021), pp. 544–551.
- [374] David L King, Jay A Kratochvil, and William Earl Boyson. *Photovoltaic array performance model*. Vol. 8. Citeseer, 2004.
- [375] Waithiru Charles Lawrence Kamuyu et al. "Prediction model of photovoltaic module temperature for power performance of floating PVs". In: *Energies* 11.2 (2018), p. 447.
- [376] A Alcañiz et al. "Offshore floating PV-DC and AC yield analysis considering wave effects". In: *Energy Conversion and Management* 300 (2024), p. 117897.
- [377] Ana-Maria Chiroasca, Liliana Rusu, and Anca Bleoju. "Study on wind farms in the North Sea area". In: *Energy Reports* 8 (2022), pp. 162–168. ISSN: 23524847. DOI: [10.1016/j.egy.2022.10.244](https://doi.org/10.1016/j.egy.2022.10.244).
- [378] Oceans of Energy. *A world's first: offshore floating solar farm installed at the Dutch North Sea*. 2019. URL: <https://oceansofenergy.blue/2019/12/11/a-worlds-first-offshore-floating-solar-farm-installed-at-the-dutch-north-sea/>.
- [379] Tara Hooper, Alona Armstrong, and Brigitte Vlaswinkel. "Environmental impacts and benefits of marine floating solar". In: *Solar Energy* 219 (2021). Special Issue on Floating Solar: beyond the state of the art technology, pp. 11–14. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2020.10.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X2031063X>.
- [380] Manish Kumar and Arun Kumar. "Experimental characterization of the performance of different photovoltaic technologies on water bodies". In: *Progress in Photovoltaics: Research and Applications* 28.1 (2020), pp. 25–48.
- [381] Laura Essak and Aritra Ghosh. "Floating photovoltaics: A review". In: *Clean Technologies* 4.3 (2022), pp. 752–769.
- [382] Amandeep Singh Makhija and S Bohra. "Performance and degradation analysis for different solar photovoltaic technologies under hot and humid environment: A review". In: *Progress in Energy* (2023).
- [383] Shengnan Zhao et al. "Potential Root Causes for Failures in Floating PV Systems". In: *ASME 2023 42nd International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection. 2023, p. 11.

- [384] Md Atiqur Rahaman et al. "Floating photovoltaic module temperature estimation: Modeling and comparison". In: *Renewable Energy* 208 (2023), pp. 162–180.
- [385] Dag Lindholm et al. "CFD modelling to derive U-values for floating PV technologies with large water footprint". In: *Solar Energy* 238 (2022), pp. 238–247.
- [386] Anik Goswami et al. "Floating solar power plant for sustainable development: A techno-economic analysis". In: *Environmental Progress & Sustainable Energy* 38.6 (2019), e13268.
- [387] Jangwon Suh, Yonghae Jang, and Yosoon Choi. "Comparison of electric power output observed and estimated from floating photovoltaic systems: A case study on the hapcheon dam, Korea". In: *Sustainability* 12.1 (2019), p. 276.
- [388] Adimas Pradityo Sukarso and Kyung Nam Kim. "Cooling effect on the floating solar PV: Performance and economic analysis on the case of west Java province in Indonesia". In: *Energies* 13.9 (2020), p. 2126.
- [389] Alberto Ghigo et al. "Design and analysis of a floating photovoltaic system for off-shore installation: The case study of Lampedusa". In: *Energies* 15.23 (2022), p. 8804.
- [390] Kelley Ruehl et al. *WEC-Sim v5.0.1*. Version v5.0.1. 2022. DOI: [10.5281/zenodo.7121186](https://doi.org/10.5281/zenodo.7121186). URL: <https://zenodo.org/badge/latestdoi/20451353>.
- [391] Nallapaneni Manoj Kumar et al. "Advancing simulation tools specific to floating solar photovoltaic systems—Comparative analysis of field-measured and simulated energy performance". In: *Sustainable Energy Technologies and Assessments* 52 (2022), p. 102168.
- [392] Nagananthini Ravichandran, Nagavinothini Ravichandran, and Balamurugan Panneerselvam. "Comparative assessment of offshore floating photovoltaic systems using thin film modules for Maldives islands". In: *Sustainable Energy Technologies and Assessments* 53 (2022), p. 102490.
- [393] Young-Kwan Choi. "A study on power generation analysis of floating PV system considering environmental impact". In: *International journal of software engineering and its applications* 8.1 (2014), pp. 75–84.
- [394] Kim Trapani and Dean L. Millar. "Hydrodynamic Overview of Flexible Floating Thin Film PV Arrays". In: *Proceedings of the 3rd Offshore Energy and Storage Symposium*. Malta, 2016, pp. 13–15.
- [395] AM Al-Yacoub, Emiruddin Redza Bin Abdul Halim, and MS Liew. "Hydrodynamic analysis of floating offshore solar farms subjected to regular waves". In: *Advances in Manufacturing Engineering: Selected articles from ICMMPE 2019*. Springer. 2020, pp. 375–390.
- [396] Gyu-Han Lee et al. "Comparative study of effect of wind and wave load on floating PV: Computational simulation and design method". In: *Journal of the Korean Society for Precision Engineering* 18.11 (2019), pp. 9–17.
- [397] C Diendorfer, M Haider, and M Lauermann. "Performance analysis of offshore solar power plants". In: *Energy Procedia* 49 (2014), pp. 2462–2471.
- [398] Alessandro Toffoli and Elzbieta M Bitner-Gregersen. "Types of ocean surface waves, wave classification". In: *Encyclopedia of maritime and offshore engineering* (2017), pp. 1–8.

- [399] R Grotmaack. "Small rigid floating bodies under the influence of water waves". In: *Research Letters in the Information and Mathematical Sciences* 5 (2003), pp. 143–157.
- [400] LH Holthuijsen. *Waves in Coastal and Oceanic Waters*. 2007.
- [401] T. H. Dawson. *Offshore structural engineering*. Prentice-Hall, 1983.
- [402] Arthur Pecher and Jens Peter Kofoed. *Handbook of ocean wave energy*. Springer Nature, 2017.
- [403] Klaus Hasselmann et al. "Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)". In: *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A* (1973).
- [404] Yang Yu, Hailong Pei, and Chengzhong Xu. "Parameter identification of JONSWAP spectrum acquired by airborne LIDAR". In: *Journal of Ocean University of China* 16.6 (2017), pp. 998–1002.
- [405] MJ Tucker, Peter G Challenor, and DJT Carter. "Numerical simulation of a random sea: a common error and its effect upon wave group statistics". In: *Applied ocean research* 6.2 (1984), pp. 118–122.
- [406] S. M. Vazirizade. "An Intelligent Integrated Method for Reliability Estimation of Off-shore Structures Wave Loading Applied in Time Domain". Available at https://repository.arizona.edu/bitstream/handle/10150/636592/azu_etd_17617_sip1_m.pdf?sequence=1. PhD thesis. University of Arizona, 2019.
- [407] Edward H Owens. "Sea conditions". In: *Beaches and coastal geology. Encyclopedia of Earth Sciences Series*. Springer US, Boston (1982), pp. 722–804.
- [408] Clarence O Collins III. *Typhoon generated surface gravity waves measured by NOMAD-type buoys*. University of Miami, 2014.
- [409] J.S. Mani. *Coastal hydrodynamics*. PHI Learning Pvt. Ltd., 2012.
- [410] Paul R Pinet. *Essential invitation to oceanography*. Jones & Bartlett Publishers, 2014.
- [411] Murray R. Spiegel. *Manual de Fórmulas y Tablas Matemáticas*. U.S.A.: McGraw-Hill Book Co, 1986.
- [412] Robert H Stewart. *Introduction to physical oceanography*. Robert H. Stewart, 2008.
- [413] Pongpat Thavornpattanapong et al. "Mathematical analysis of added-mass instability in fluid-structure interaction". In: *International Journal of Mathematics and Statistics* 10.W11 (2011), pp. 43–51.
- [414] Mark Z Jacobson and Vijaysinh Jadhav. "World estimates of PV optimal tilt angles and ratios of sunlight incident upon tilted and tracked PV panels relative to horizontal panels". In: *Solar energy* 169 (2018), pp. 55–66.
- [415] SMA. *Sunny Boy 1.5/2/2.5 - Datasheet*. 2021. URL: <https://files.sma.de/downloads/SBxx-1VL-40-DS-en-51.pdf>.
- [416] F. N. Fritsch and R. E. Carlson. "Monotone Piecewise Cubic Interpolation". In: *SIAM Journal on Numerical Analysis* 17.2 (1980), pp. 238–246. DOI: [10.1137/0717021](https://doi.org/10.1137/0717021). URL: <https://doi.org/10.1137/0717021>.

- [417] John Boland, Jing Huang, and Barbara Ridley. “Decomposing global solar radiation into its direct and diffuse components”. In: *Renewable and Sustainable Energy Reviews* 28 (2013), pp. 749–756.
- [418] Sara Golroodbari and Wilfried van Sark. “On the effect of dynamic albedo on performance modelling of offshore floating photovoltaic systems”. In: *Solar Energy Advances* 2 (2022), p. 100016.
- [419] Gregory Sinnett and Falk Feddersen. “Observations and parameterizations of surf-zone albedo”. In: *Methods in Oceanography* 17 (2016), pp. 319–334.
- [420] HuiZhi Liu, Gang Tu, and WenJie Dong. “Three-year changes of surface albedo of degraded grassland and cropland surfaces in a semiarid area”. In: *Chinese Science Bulletin* 53.8 (2008), pp. 1246–1254.
- [421] Dock Marine. *Floating Cubes SunnyDock*. 2020. URL: <https://uploads-ssl.webflow.com/5e2394bfb966d96dd886cadf/5e2cd9f18c2b174d9dbee2ae%5C%5FSunnydock.pdf>.
- [422] LG. *LG400N2W-A5 Datasheet*. 2017. URL: <https://www.lg.com/us/business/download/resources/BT00002151/LG400N2W-A5.pdf>.
- [423] Álvaro de Gruijter. “Design and integration of an offshore off-grid system and an on-land system for comparison on photovoltaic performance”. MA thesis. Oceans of Energy and Delft University of Technology, 2023.
- [424] Manish Kumar et al. “Impact of Water Waves on the Output Power of Floating PV Systems on Water Bodies”. 8th world conference on photovoltaic energy conversion. 2022.
- [425] Sudhakar Babu Thanikanti et al. “A dynamic mismatch loss mitigation algorithm with dual input dual output converter for solar PV systems”. In: *Solar Energy Materials and Solar Cells* 251 (2023), p. 112163. ISSN: 0927-0248. DOI: <https://doi.org/10.1016/j.solmat.2022.112163>. URL: <https://www.sciencedirect.com/science/article/pii/S0927024822005803>.
- [426] Leonardo Micheli et al. “Techno-economic potential and perspectives of floating photovoltaics in Europe”. In: *Solar Energy* 243 (2022), pp. 203–214. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2022.07.042>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X22005217>.
- [427] Kim Trapani and Dean L. Millar. “Floating photovoltaic arrays to power the mining industry: A case study for the McFaulds lake (Ring of Fire)”. In: *Environmental Progress & Sustainable Energy* 35.3 (2016). _eprint: <https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/ep.12275>. pp. 898–905. DOI: <https://doi.org/10.1002/ep.12275>.
- [428] A. Berstad and P. Grøn. “Model basin testing of a floating solar PV plant compared to analysis”. In: *10th Conference on Computational Methods in Marine Engineering*. CIMNE, 2023. DOI: [10.23967/marine.2023.094](https://doi.org/10.23967/marine.2023.094). (Visited on 09/08/2023).
- [429] Dharma K.K. Sree et al. “Fluid-structural analysis of modular floating solar farms under wave motion”. In: *Solar Energy* 233 (2022), pp. 161–181. ISSN: 0038092X. DOI: [10.1016/j.solener.2022.01.017](https://doi.org/10.1016/j.solener.2022.01.017).

- [430] Maria Ikhennicheu et al. “Analytical method for loads determination on floating solar farms in three typical environments”. In: *Solar Energy* 219 (2021), pp. 34–41. ISSN: 0038092X. DOI: [10.1016/j.solener.2020.11.078](https://doi.org/10.1016/j.solener.2020.11.078).
- [431] Maarten Dörenkämper et al. “Influence of wave induced movements on the performance of floating PV systems”. In: *Proceedings of 36th European Photovoltaic Solar Energy Conference (EUPVSEC)*. Marseille, France. 2019, pp. 1759–1762.
- [432] R. Claus and M. López. “Key issues in the design of floating photovoltaic structures for the marine environment”. en. In: *Renewable and Sustainable Energy Reviews* 164 (Aug. 2022), p. 112502. ISSN: 13640321. DOI: [10.1016/j.rser.2022.112502](https://doi.org/10.1016/j.rser.2022.112502). (Visited on 08/03/2023).
- [433] Oriol Colomé, Francesc Verdugo, and Ido Akkerman. “A monolithic finite element formulation for the hydroelastic analysis of very large floating structures”. en. In: *International Journal for Numerical Methods in Engineering* 124.3 (Feb. 2023), pp. 714–751. ISSN: 0029-5981, 1097-0207. DOI: [10.1002/nme.7140](https://doi.org/10.1002/nme.7140).
- [434] Robert G. Dean and Robert A. Dalrymple. *Water wave mechanics for engineers and scientists*. Advanced series on ocean engineering v. 2. Singapore ; Teaneck, NJ: World Scientific, 1991. 353 pp. ISBN: 978-981-02-0420-4.
- [435] Jeff Bezanson et al. “Julia: A Fresh Approach to Numerical Computing”. en. In: *SIAM Review* 59.1 (Jan. 2017), pp. 65–98. ISSN: 0036-1445, 1095-7200. DOI: [10.1137/141000671](https://doi.org/10.1137/141000671).
- [436] Santiago Badia and Francesc Verdugo. “Gridap: An extensible Finite Element toolbox in Julia”. In: *Journal of Open Source Software* 5.52 (2020), p. 2520. ISSN: 2475-9066. DOI: [10.21105/joss.02520](https://doi.org/10.21105/joss.02520).
- [437] Francesc Verdugo and Santiago Badia. “The software design of Gridap: A Finite Element package based on the Julia JIT compiler”. en. In: *Computer Physics Communications* 276 (July 2022), p. 108341. ISSN: 00104655. DOI: [10.1016/j.cpc.2022.108341](https://doi.org/10.1016/j.cpc.2022.108341).
- [438] Oriol Colomes Gene. *Code used to generate the data underlying the publication: A monolithic Finite Element formulation for the hydroelastic analysis of Very Large Floating Structures*. 2022. DOI: [10.4121/19601419](https://doi.org/10.4121/19601419).
- [439] T.I. Khabakhpasheva and A.A. Korobkin. “Hydroelastic behaviour of compound floating plate in waves”. In: *Journal of Engineering Mathematics* 44.1 (2002), pp. 21–40. ISSN: 00220833. DOI: [10.1023/A:1020592414338](https://doi.org/10.1023/A:1020592414338). (Visited on 07/12/2024).
- [440] Xiaolei Liu, Xuefeng Wang, and Shengwen Xu. “A DMM-EMM-RSM hybrid technique on two-dimensional frequency-domain hydroelasticity of floating structures over variable bathymetry”. en. In: *Ocean Engineering* 201 (Apr. 2020), p. 107135. ISSN: 00298018. DOI: [10.1016/j.oceaneng.2020.107135](https://doi.org/10.1016/j.oceaneng.2020.107135). (Visited on 07/12/2024).
- [441] Kiyokazu Yago and Hisayoshi Endo. “On the Hydroelastic Response of Box-Shaped Floating Structure with Shallow Draft”. In: *Journal of the Society of Naval Architects of Japan* 1996.180 (1996), pp. 341–352. ISSN: 1884-2070, 0514-8499. DOI: [10.2534/j.jasnaoe1968.1996.180_341](https://doi.org/10.2534/j.jasnaoe1968.1996.180_341). (Visited on 07/12/2024).
- [442] KNMI. *KNW-CSV-TS-UPDATE*. 2023. URL: <https://dataplatfom.knmi.nl/dataset/knw-csv-ts-update-1-0>.

- [443] Subrata Kumar Chakrabarti. *Hydrodynamics of offshore structures*. Southampton: Computational Mechanics Publications, 1987. ISBN: 0-905451-66-X. URL: <https://www.witpress.com/books/978-0-90545-166-4>.
- [444] Arnas Majumder et al. "Performance analysis of a floating photovoltaic system and estimation of the evaporation losses reduction". In: *Energies* 14.24 (2021), p. 8336.
- [445] Silvana Ayala Pelaez et al. "Model and validation of single-axis tracking with bifacial PV". In: *IEEE Journal of Photovoltaics* 9.3 (2019), pp. 715–721.
- [446] Ezio Soares de Moura, Henrique Silveira Alves Marques, and Douglas Bressan Riffel. "Assessment of photovoltaic I–V curve translation methods in real-world situations". In: *Solar Energy* 276 (2024), p. 112654.
- [447] Borislav Stankov et al. "Influence of Wind and Rainfall on the Performance of a Photovoltaic Module in a Dusty Environment". In: *Energies* 17.14 (2024), p. 3394.
- [448] BosswerkGmbH. *GreenAkku*. 2023. URL: <https://greenakku.de/Solarmodule/Flexible-Solarmodule/Bosswerk-Sphere-Hoch-Flexibles-Solarmodule-ETFE-200Wp/:4499.html>.
- [449] KNMI. *Actuele10mindataKNMIstations*. 2023. URL: <https://dataplatfom.knmi.nl/dataset/actuele10mindataknmistations-2>.
- [450] Min Zhang and Sebastian Schreier. "Review of wave interaction with continuous flexible floating structures". en. In: *Ocean Engineering* 264 (Nov. 2022), p. 112404. ISSN: 00298018. DOI: [10.1016/j.oceaneng.2022.112404](https://doi.org/10.1016/j.oceaneng.2022.112404). (Visited on 12/02/2022).
- [451] H. Suzuki et al. "Very Large Floating Structures". In: *Volume 2: Structures, Safety and Reliability; Petroleum Technology Symposium*. San Diego, California, USA: ASMEDC, 2007, pp. 597–608. ISBN: 978-0-7918-4268-3 978-0-7918-3799-3. DOI: [10.1115/OMAE2007-29758](https://doi.org/10.1115/OMAE2007-29758). (Visited on 02/21/2024).
- [452] Eva Loukogeorgaki, Constantine Michailides, and Demos C. Angelides. "Hydroelastic analysis of a flexible mat-shaped floating breakwater under oblique wave action". en. In: *Journal of Fluids and Structures* 31 (2012), pp. 103–124. ISSN: 08899746. DOI: [10.1016/j.jfluidstructs.2012.02.011](https://doi.org/10.1016/j.jfluidstructs.2012.02.011).
- [453] Shagun Agarwal, Oriol Colomés, and Andrei Metrikine. *Dynamic analysis of viscoelastic floating membranes using monolithic Finite Element method*. 2023. DOI: [10.2139/ssrn.4689746](https://doi.org/10.2139/ssrn.4689746).
- [454] Romênia G Vieira et al. "A comprehensive review on bypass diode application on photovoltaic modules". In: *Energies* 13.10 (2020), p. 2472.
- [455] F Belhachat and C Larbes. "Modeling, analysis and comparison of solar photovoltaic array configurations under partial shading conditions". In: *Solar energy* 120 (2015), pp. 399–418.
- [456] K Sinapis et al. "A comprehensive study on partial shading response of c-Si modules and yield modeling of string inverter and module level power electronics". In: *Solar Energy* 135 (2016), pp. 731–741.
- [457] Hesam Ziar et al. "Quantification of shading tolerability for photovoltaic modules". In: *IEEE Journal of Photovoltaics* 7.5 (2017), pp. 1390–1399.

- [458] Soubhagya K Das et al. "Shading mitigation techniques: State-of-the-art in photovoltaic applications". In: *Renewable and Sustainable Energy Reviews* 78 (2017), pp. 369–390.
- [459] Jianbo Bai et al. "Characteristic output of PV systems under partial shading or mismatch conditions". In: *Solar Energy* 112 (2015), pp. 41–54.
- [460] Yaw-Juen Wang and Po-Chun Hsu. "An investigation on partial shading of PV modules with different connection configurations of PV cells". In: *Energy* 36.5 (2011), pp. 3069–3078.
- [461] Zuo Wang et al. "Quantitative estimation of mismatch losses in photovoltaic arrays under partial shading conditions". In: *Optik* 203 (2020), p. 163950.
- [462] Youssef El Basri et al. "A proposed graphical electrical signatures supervision method to study PV module failures". In: *Solar energy* 116 (2015), pp. 247–256.
- [463] John A Tsanakas, Long Ha, and Claudia Buerhop. "Faults and infrared thermographic diagnosis in operating c-Si photovoltaic modules: A review of research and future challenges". In: *Renewable and sustainable energy reviews* 62 (2016), pp. 695–709.
- [464] International Electrotechnical Commission (IEC). *IEC 61215-2:2016 Terrestrial photovoltaic (PV) modules - Design qualification and type approval - Part 2: Test procedures*. 2016.
- [465] Seraphim Energy. *Eclipse SRP-E01B Datasheet*. 2018.
- [466] Jinko Solar. *Eagle HC 60M-V 300-320 Watt Datasheet*.
- [467] DMEGC Solar. *DM330G1-66BU Datasheet*. 2020.
- [468] Sandeep Mishra et al. "Selection map for PV module installation based on shading tolerability and temperature coefficient". In: *IEEE Journal of Photovoltaics* 9.3 (2019), pp. 872–880.
- [469] Nils Klasen et al. "A comprehensive study of module layouts for silicon solar cells under partial shading". In: *IEEE Journal of Photovoltaics* 12.2 (2022), pp. 546–556.
- [470] Luthfan Fauzan et al. "Small area high voltage photovoltaic module for high tolerance to partial shading". In: *Iscience* 26.6 (2023).
- [471] LH Slooff et al. "New module concept for aesthetic PV integration with better shadow performance". In: *Proceedings Advanced Building Skins* (2017).
- [472] F.P. Baumgartner and C. Allenspach. "Performance Gain of Shading Tolerant PV Modules in Different Electrical PV System Setups". In: *40th EUPVSEC*. 2023.
- [473] J. Moereke et al. "A Measure for PV Module Performance under Partial Shading and Its Application to CIGS and cSi Technologies for Realistic Shading Scenarios". In: *40th EUPVSEC*. 2023.
- [474] Andres Calcabrini et al. "The relevance of the cell's breakdown voltage in the dc yield of partially shaded pv modules". In: *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*. IEEE. 2021, pp. 0092–0094.
- [475] Kashif Ishaque, Zainal Salam, et al. "A comprehensive MATLAB Simulink PV system simulator with partial shading capability based on two-diode model". In: *Solar energy* 85.9 (2011), pp. 2217–2227.

- [476] Juan David Bastidas-Rodríguez et al. “Model-Based Degradation Analysis of Photovoltaic Modules Through Series Resistance Estimation”. In: *IEEE Transactions on Industrial Electronics* 62.11 (2015), pp. 7256–7265. DOI: [10.1109/TIE.2015.2459380](https://doi.org/10.1109/TIE.2015.2459380).
- [477] Volker Quaschnig and Rolf Hanitsch. “Numerical simulation of current-voltage characteristics of photovoltaic systems with shaded solar cells”. In: *Solar energy* 56.6 (1996), pp. 513–520.
- [478] Marcelo Gradella Villalva, Jonas Rafael Gazoli, and Ernesto Ruppert Filho. “Comprehensive approach to modeling and simulation of photovoltaic arrays”. In: *IEEE Transactions on power electronics* 24.5 (2009), pp. 1198–1208.
- [479] Hajime Kawamura et al. “Simulation of I–V characteristics of a PV module with shaded PV cells”. In: *Solar Energy Materials and Solar Cells* 75.3-4 (2003), pp. 613–621.
- [480] SEMIKRON. *F 1200A Datasheet*. Axial leaded high-efficiency rectifier diode. 2020.
- [481] Mattia Dallapiccola et al. “Simplified method for partial shading losses calculation for series connected PV modules with experimental validation”. In: *35th European Photovoltaic Solar Energy Conference and Exhibition*. 2018.
- [482] Bisol. *Bisol Premium Series BMU 245-265 Wp*. 2014.
- [483] Delft High Performance Computing Centre (DHPC). *DelftBlue Supercomputer (Phase 2)*. <https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2>. 2024.
- [484] Zhaoxu Liu / slandarer. *Sankey Plot*. Accessed: 20/08/2024. 2024. URL: <https://www.mathworks.com/matlabcentral/fileexchange/128679-sankey-plot>.
- [485] California Energy Commission. *PV Modules Database*. Accessed: 21/10/2021. URL: <https://www.energy.ca.gov/data-reports/energy-almanac/renewables-and-clean-energy/solar-equipment-lists>.
- [486] W Herrmann, W Wiesner, and W Vaassen. “Hot spot investigations on PV modules-new concepts for a test standard and consequences for module design with respect to bypass diodes”. In: *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference-1997*. IEEE. 1997, pp. 1129–1132.
- [487] Haifeng Chu et al. “Shade performance of a back-contact module assembled with cells featuring soft breakdown characteristic”. In: *Energy Procedia* 92 (2016), pp. 540–545.
- [488] Otwin Breitenstein et al. “Understanding junction breakdown in multicrystalline solar cells”. In: *Journal of Applied Physics* 109.7 (2011).
- [489] Xingshu Sun et al. “A physics-based compact model for CIGS and CdTe solar cells: From voltage-dependent carrier collection to light-enhanced reverse breakdown”. In: *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*. IEEE. 2015, pp. 1–6.
- [490] Piotr Szaniawski et al. “Light-enhanced reverse breakdown in Cu (In, Ga) Se₂ solar cells”. In: *Thin Solid Films* 535 (2013), pp. 326–330.
- [491] Timothy J Silverman et al. “Thermal and electrical effects of partial shade in monolithic thin-film photovoltaic modules”. In: *IEEE Journal of Photovoltaics* 5.6 (2015), pp. 1742–1747.

- [492] Kristijan Brecl, Matevž Bokalič, and Marko Topič. “Annual energy losses due to partial shading in PV modules with cut wafer-based Si solar cells”. In: *Renewable Energy* 168 (2021), pp. 195–203.
- [493] S Guo et al. “A quantitative analysis of photovoltaic modules using halved cells”. In: *International Journal of Photoenergy* 2013.1 (2013), p. 739374.
- [494] Tianyu Tang et al. “A quantitative comparison between double glass photovoltaic modules using half-size cells and quarter-size cells”. In: *IEEE Journal of Photovoltaics* 7.5 (2017), pp. 1298–1303.
- [495] Hamed Hanifi, Jens Schneider, and Joerg Bagdahn. “Reduced shading effect on half-cell modules—measurement and simulation”. In: *31st European Photovoltaic Solar Energy Conference and Exhibition*. EU PVSEC. 2015, pp. 2529–2533.
- [496] Jiadong Qian et al. “Comparison of half-cell and full-cell module hotspot-induced temperature by simulation”. In: *IEEE Journal of Photovoltaics* 8.3 (2018), pp. 834–839.
- [497] Daniel de B Mesquita et al. “A review and analysis of technologies applied in PV modules”. In: *2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*. IEEE. 2019, pp. 1–6.
- [498] M. Abdel-Nasser and K. Mahmoud. “Accurate photovoltaic power forecasting models using deep LSTM-RNN”. In: *Neural Computing and Applications* 31.7 (July 2019), pp. 2727–2740. DOI: [10.1007/s00521-017-3225-z](https://doi.org/10.1007/s00521-017-3225-z).
- [499] M. H. Alomari, J. Adeeb, and O. Younis. “PVPF tool: An automated web application for real-time photovoltaic power forecasting”. In: *International Journal of Electrical and Computer Engineering* 9.1 (2019), pp. 34–41. DOI: [10.11591/ijece.v9i1.pp34-41](https://doi.org/10.11591/ijece.v9i1.pp34-41).
- [500] M. Gulin, T. Pavlović, and M. Vašak. “A one-day-ahead photovoltaic array power production prediction with combined static and dynamic on-line correction”. In: *Solar Energy* 142 (Jan. 2017), pp. 49–60. DOI: [10.1016/j.solener.2016.12.008](https://doi.org/10.1016/j.solener.2016.12.008).
- [501] C. Huang et al. “Improvement in artificial neural network-based estimation of grid connected photovoltaic power output”. In: *Renewable Energy* 97 (Nov. 2016), pp. 838–848. DOI: [10.1016/j.renene.2016.06.043](https://doi.org/10.1016/j.renene.2016.06.043).
- [502] H. A. Kazem and J. H. Yousif. “Comparison of prediction methods of photovoltaic power system production using a measured dataset”. In: *Energy Conversion and Management* 148 (Sept. 2017), pp. 1070–1081. DOI: [10.1016/j.enconman.2017.06.058](https://doi.org/10.1016/j.enconman.2017.06.058).
- [503] M. Konstantinou, S. Peratikou, and A. G. Charalambides. “Solar Photovoltaic Forecasting of Power Output Using LSTM Networks”. In: *Atmosphere* 12.1 (Jan. 2021), p. 124. DOI: [10.3390/ATMOS12010124](https://doi.org/10.3390/ATMOS12010124).
- [504] K. R. Kumar and M. S. Kalavathi. “Artificial intelligence based forecast models for predicting solar power generation”. In: *Materials Today: Proceedings* 5.1 (Jan. 2018), pp. 796–802. DOI: [10.1016/J.MATPR.2017.11.149](https://doi.org/10.1016/J.MATPR.2017.11.149).
- [505] D. Lee and K. Kim. “Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information”. In: *Energies* 12.2 (Jan. 2019), p. 215. DOI: [10.3390/EN12020215](https://doi.org/10.3390/EN12020215).

- [506] Nailya Maitanova et al. "Machine learning approach to a low-cost day-ahead photovoltaic power prediction based on publicly available weather reports for automated energy management systems". In: *Proc. 36th Eur. Photovolt. Sol. Energy Conf. Exhib.*, no. September. 2019, pp. 1711–1718.
- [507] I. Majumder, P. K. Dash, and R. Bisoi. "Short-term solar power prediction using multi-kernel-based random vector functional link with water cycle algorithm-based parameter optimization". In: *Neural Computing and Applications* 32 (2020). DOI: [10.1007/s00521-019-04290-x](https://doi.org/10.1007/s00521-019-04290-x).
- [508] E. Ogluari et al. "Computational Intelligence Techniques Applied to the Day Ahead PV Output Power Forecast: PHANN, SNO and Mixed". In: *Energies* 11.6 (June 2018), p. 1487. DOI: [10.3390/en11061487](https://doi.org/10.3390/en11061487).
- [509] P. Tang, D. Chen, and Y. Hou. "Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting". In: *Chaos, Solitons & Fractals* 89 (Aug. 2016), pp. 243–248. DOI: [10.1016/J.CHAOS.2015.11.008](https://doi.org/10.1016/J.CHAOS.2015.11.008).
- [510] Roham Torabi et al. "Development of an ANN model to predict the electricity produced by small scale roof-top PV systems in Madeira Island". In: *2017 International Conference in Energy and Sustainability in Small Developing Economies (ES2DE)*. IEEE. 2017, pp. 1–6.
- [511] G. Wang, Y. Su, and L. Shu. "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models". In: *Renewable Energy* 96 (Oct. 2016), pp. 469–478. DOI: [10.1016/J.RENENE.2016.04.089](https://doi.org/10.1016/J.RENENE.2016.04.089).
- [512] A. K. Yadav et al. "Daily array yield prediction of grid-interactive photovoltaic plant using relief attribute evaluator based Radial Basis Function Neural Network". In: *Renewable and Sustainable Energy Reviews* 81 (Jan. 2018), pp. 2115–2127. DOI: [10.1016/j.rser.2017.06.023](https://doi.org/10.1016/j.rser.2017.06.023).
- [513] Z. Yang et al. "A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting". In: *Neurocomputing* 397 (July 2020), pp. 415–421. DOI: [10.1016/J.NEUCOM.2019.09.110](https://doi.org/10.1016/J.NEUCOM.2019.09.110).
- [514] J. H. Yousif, H. A. Kazem, and J. Boland. "Predictive Models for Photovoltaic Electricity Production in Hot Weather Conditions". In: *Energies* 10.7 (July 2017), p. 971. DOI: [10.3390/EN10070971](https://doi.org/10.3390/EN10070971).
- [515] E. Schulz, M. Speekenbrink, and A. Krause. *A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions*. 2018.
- [516] Sun-Chong Wang. *Interdisciplinary Computing In Java Programming Language*. 2003.
- [517] L. V Fausett. *Fundamentals of neural networks: architectures, algorithms and applications*. 2006.
- [518] AD Dongare, RR Kharde, and Amit D Kachare. "Introduction to artificial neural network". In: *International Journal of Engineering and Innovative Technology (IJEIT)* 2.1 (2012), pp. 189–194.
- [519] Matt W Gardner and SR Dorling. "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences". In: *Atmospheric environment* 32.14–15 (1998), pp. 2627–2636.

- [520] Andrew Tch. *The Mostly Complete Chart of Neural Networks, Explained*. Accessed: Sep. 02, 2021. 2017. URL: <https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464>.
- [521] Ajay Shrestha and Ausif Mahmood. "Review of deep learning algorithms and architectures". In: *IEEE access* 7 (2019), pp. 53040–53065.
- [522] N. Jamil, S. Lqbal, and N. Iqbal. "Face recognition using neural networks". In: *Proceedings - IEEE International Multi Topic Conference 2001: Technology for the 21st Century, IEEE INMIC 2001*. 2001, pp. 277–281. DOI: [10.1109/INMIC.2001.995351](https://doi.org/10.1109/INMIC.2001.995351).
- [523] J. Kocić, N. Jovičić, and V. Drndarević. "An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms". In: *Sensors (Switzerland)* 19.9 (May 2019). DOI: [10.3390/s19092064](https://doi.org/10.3390/s19092064).
- [524] K. Chen and X. Meng. "Interpretation and Understanding in Machine Learning". In: *Journal of Computer Research and Development* 57.9 (Sept. 2020), p. 1971. DOI: [10.7544/ISSN1000-1239.2020.20190456](https://doi.org/10.7544/ISSN1000-1239.2020.20190456).
- [525] S. Ding et al. "Extreme learning machine: algorithm, theory and applications". In: *Artificial Intelligence Review* 44.1 (June 2015), pp. 103–115. DOI: [10.1007/s10462-013-9405-z](https://doi.org/10.1007/s10462-013-9405-z).
- [526] G. Bin Huang, Q. Y. Zhu, and C. K. Siew. "Extreme learning machine: Theory and applications". In: *Neurocomputing* 70.1-3 (Dec. 2006), pp. 489–501. DOI: [10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126).
- [527] Tomohiro Takagi and Michio Sugeno. "Fuzzy identification of systems and its applications to modeling and control". In: *IEEE transactions on systems, man, and cybernetics* 1 (1985), pp. 116–132.
- [528] J-SR Jang. "ANFIS: adaptive-network-based fuzzy inference system". In: *IEEE transactions on systems, man, and cybernetics* 23.3 (1993), pp. 665–685.
- [529] Michael Carney and Pádraig Cunningham. "Evaluating Density Forecasting Models". In: 2006. URL: <https://api.semanticscholar.org/CorpusID:14418294>.
- [530] D.G. Erbs, S.A. Klein, and J.A. Duffie. "Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation". In: *Solar Energy* 28.4 (1982), pp. 293–302. ISSN: 0038-092X. DOI: [https://doi.org/10.1016/0038-092X\(82\)90302-4](https://doi.org/10.1016/0038-092X(82)90302-4). URL: <https://www.sciencedirect.com/science/article/pii/0038092X82903024>.
- [531] E L Maxwell. "A quasi-physical model for converting hourly global horizontal to direct normal insolation". In: (Aug. 1987). URL: <https://www.osti.gov/biblio/5987868>.
- [532] D.T. Reindl, W.A. Beckman, and J.A. Duffie. "Diffuse fraction correlations". In: *Solar Energy* 45.1 (1990), pp. 1–7. ISSN: 0038-092X. DOI: [https://doi.org/10.1016/0038-092X\(90\)90060-P](https://doi.org/10.1016/0038-092X(90)90060-P). URL: <https://www.sciencedirect.com/science/article/pii/0038092X9090060P>.
- [533] D.T. Reindl, W.A. Beckman, and J.A. Duffie. "Evaluation of hourly tilted surface radiation models". In: *Solar Energy* 45.1 (1990), pp. 9–17. ISSN: 0038-092X. DOI: [https://doi.org/10.1016/0038-092X\(90\)90061-G](https://doi.org/10.1016/0038-092X(90)90061-G). URL: <https://www.sciencedirect.com/science/article/pii/0038092X9090061G>.

- [534] Jeremy P. Every, Li Li, and David G. Dorrell. “Köppen-Geiger climate classification adjustment of the BRL diffuse irradiation model for Australian locations”. In: *Renewable Energy* 147 (2020), pp. 2453–2469. ISSN: 0960-1481. DOI: <https://doi.org/10.1016/j.renene.2019.09.114>. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119314521>.
- [535] N.A. Engerer. “Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia”. In: *Solar Energy* 116 (2015), pp. 215–237. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2015.04.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X15001905>.
- [536] Jamie Bright and Nicholas Engerer. “Engerer2: Global re-parameterisation, update, and validation of an irradiance separation model at different temporal resolutions”. In: *Journal of Renewable and Sustainable Energy* 11 (May 2019), p. 033701. DOI: [10.1063/1.5097014](https://doi.org/10.1063/1.5097014).
- [537] Geoapify GmbH. *Geoapify*. URL: <https://www.geoapify.com/>.

ACKNOWLEDGEMENTS

“I don’t want to do a PhD” I said in 2020. Luckily, I’m sometimes wrong, as time has once more proved, since I’m very thankful of the PhD journey that I’ve experienced.

I’d first like to thank my promotor, Prof.dr. Olindo Isabella and Prof.dr. Miro Zeman, and co-promotor, Dr. Hesán Ziar, for the opportunity to pursue this PhD. I still remember Olindo’s PhD offer right after my MSc thesis defense, an evidence of the trust he placed in me from the very beginning. Trust, support, and independence who were also extended by Hesán during these four years.

I’d like to acknowledge, one last time, the funding from the Horizon 2020 Program, under Grant Agreement 952957, Trust-PV project, as well as all the people involved in it. I really enjoyed being part of such a collaborative project and discuss, get to know, and have lots of fun with the amazing people that composed it.

I’d also like to thank the conference organizers of the EUPVSEC for selecting me as a committee member after just two years in the PhD. I loved the recognition and the opportunity to chair a session, but especially spending one of the greatest nights with top-notch solar researchers at the committee dinner.

Many results wouldn’t have been achieved either without the collaboration outside of the PVMD group and the Trust-PV. Therefore, I’d like to thank once more all of those who shared their knowledge or data and contributed to this work.

And of course, I’m also grateful to the PVMD group members who helped with critical discussions, measurement equipment, and enjoyable sunny lunches. With a special mention to my students who did an amazing job and who taught me a lot during these four years.

I’d also like to thank my friends and family, for being next to my side in this journey. I especially want to recognize my parents for their constant support, critical view, and weekly Skype meetings. *Os habré enseñado a ser padres, pero vosotros me enseñasteis a ser persona.*

Last and most importantly, the person whose presence is explicitly nowhere but in reality everywhere in this book: Miguel. I can’t quantify how much he helped me along the journey, not only in constant support and encouragement, but also in actual help when coding, discussing solutions, and expressing ideas.

PS: Thanks in advance to the tiny human inside of me for the company during the defense and the support in form of kicking. I’m really looking forward to getting to know you, but know that I already love you.

LIST OF PUBLICATIONS

PEER-REVIEWED PUBLICATIONS

PART OF THIS THESIS

1. **A. Alcañiz**, M. M. Nikam, Y. Snow, O. Isabella, H. Ziar “Photovoltaic system monitoring and fault detection using peer systems” *Progress in Photovoltaics: research and applications*, 30(9), 1072-1086 (2022) *
2. M. R. Vogt, C. Ruiz Tobon, **A. Alcañiz**, P. Procel, Y. Blom, A. Nour El Din, T. Stark, Z. Wang, E. Garcia Goma, J. G. Etxebarria, H. Ziar, M. Zeman, R. Santbergen, O. Isabella “Introducing a comprehensive physics-based modelling framework for tandem and other PV systems” *Solar Energy Materials and Solar Cells*, 247, 111944 (2022)
3. **A. Alcañiz**, D. Grzebyk, H. Ziar, O. Isabella “Trends and gaps in photovoltaic power forecasting with machine learning” *Energy Reports*, 9, 447-471 (2023)
4. **A. Alcañiz**, A. V. Lindfors, M. Zeman, H. Ziar, O. Isabella, “Effect of Climate on Photovoltaic Yield Prediction Using Machine Learning Models” *Global Challenges*, 7(1), 2200166 (2023)
5. D. Grzebyk, **A. Alcañiz**, J. C. Donker, M. Zeman, H. Ziar, O. Isabella, “Individual yield nowcasting for residential PV systems” *Solar Energy*, 251, 325-336 (2023)
6. **A. Alcañiz**, N. Monaco, O. Isabella, H. Ziar, “Offshore floating PV-DC and AC yield analysis considering wave effects” *Energy Conversion and Management*, **300** 117897 (2024)
7. **A. Alcañiz**, M. I. van Kouwen, O. Isabella, H. Ziar “Wide-area sky view factor analysis and Fourier-based decomposition model for optimizing irradiance sensors allocation in European solar photovoltaic farms: a software tool” *Solar Energy*, 286, 113139 (2025)
8. **A. Alcañiz**, S. Agarwal, P. Tiwald, O. Isabella, H. Ziar, O. Colomès “Assessment of power losses and structural response of offshore floating solar platform” under review in *Energy*
9. **A. Alcañiz**, N. Rukhshi, R. Koutarapu, O. Isabella, H. Ziar “Quantifying Shading Tolerability in Commercial Photovoltaic Modules” submitted to *Progress in Photovoltaics*

*Featured on the cover of *Progress in Photovoltaics: research and applications*, volume 30, number 9, September 2022

OUTSIDE THE SCOPE OF THIS THESIS

1. **A. Alcañiz**, G. López, I. Martín, A. Jiménez, A. Datas, E. Calle, E. Ros, L. G. Gerling, C. Voz, C. del Cañizo, R. Alcubilla “Germanium photovoltaic cells with MoOx hole-selective contacts” *Solar Energy*, 181, 357-360 (2019)
2. I. Martín, **A. Alcañiz**, A. Jiménez, G. López, C. del Cañizo, A. Datas “Application of quasi-steady-state photoconductance technique to lifetime measurements on crystalline germanium substrates” *IEEE Journal of Photovoltaics*, 10(4), 1068-1075 (2020)
3. L. Mazzarella, **A. Alcañiz**, P. Procel, E. Kawa, Y. Zhao, U. Tiringier, C. Han, G. Yang, P. Taheri, M. Zeman, O. Isabella “Strategy to mitigate the dipole interfacial states in (i)a-Si:H/MoOx passivating contacts solar cells” *Progress in Photovoltaics: research and applications*, 29(3), 391-400 (2021)
4. L. Cao, P. Procel, **A. Alcañiz**, J. Yan, F. Tichelaar, E. Özkol, Y. Zhao, C. Han, G. Yang, Z. Yao, M. Zeman, R. Santbergen, L. Mazzarella, O. Isabella “Achieving 23.83% conversion efficiency in silicon heterojunction solar cell with ultra-thin MoOx hole collector layer via tailoring (i) a-Si: H/MoOx interface” *Progress in Photovoltaics: research and applications*, 31(12), 1245-1254 (2023)
5. Y. Zhao, P. Procel, C. Han, L. Cao, G. Yang, E. Özkol, **A. Alcañiz**, K. Kovačević, G. Limodio, R. Santbergen, A. Smets, A. Weeber, M. Zeman, L. Mazzarella, O. Isabella “Strategies for realizing high-efficiency silicon heterojunction solar cells” *Solar Energy Materials and Solar Cells*, 258, 112413 (2023)

TRUST-PV TECHNICAL REPORTS

1. I.T. Horvath, **A. Alcañiz**, F. Segata, J. Lemmens, T. Hall, W. Vanheusden, “Automated PV digital twin-based yield simulation framework” (2022)
2. **A. Alcañiz**, J. C. Ortiz Lizcano, H. Ziar, M. Dallapiccola, Y. Snow, I. Kaaya, “Residential-scale integrated energy yield simulation and reliability modelling framework” (2023)
3. A. Bakovasilis, O. Dupon, I. Kaaya, M. Turpin, **A. Alcañiz**, M. Pierro, “Forecasting for advanced operational stability” (2023)
4. S. Tonnel, M. Ikhennicheu, R. Le Failler, **A. Alcañiz**, H. Ziar “Guidelines for design, procurement and O&M friendly concepts for floating PV” (2023)
5. S. Tonnel, F. Huntingford, M. Ikhennicheu, R. Le Failler, F. Gorintin, **A. Alcañiz**, H. Ziar “Demonstration of solution and challenges on floating systems” (2024)

CONFERENCE CONTRIBUTIONS

1. **A. Alcañiz**, M. M. Nikam, Y. Snow, O. Isabella, H. Ziar “Oral: Photovoltaic system monitoring using peer systems” in *8th World Conference on Photovoltaic Energy Conversion*, Milan, Italy (2022)
2. **A. Alcañiz**, A. V. Lindfors, M. Zeman, H. Ziar, O. Isabella, “Oral: Climatic effect on Photovoltaic Yield Prediction with Machine Learning Models” in *8th World Conference on Photovoltaic Energy Conversion*, Milan, Italy (2022)

3. D. Grzebyk, **A. Alcañiz**, J. C. Donker, M. Zeman, H. Ziar, O. Isabella, “Visual: Photovoltaic Yield monitoring for residential scale systems” in *8th World Conference on Photovoltaic Energy Conversion*, Milan, Italy (2022)
4. **A. Alcañiz**, A. V. Lindfors, M. Zeman, H. Ziar, O. Isabella, “Oral: Climatic effect on Photovoltaic Yield Prediction with Machine Learning Models” in *33rd International Photovoltaic Science and Engineering Conference*, Nagoya, Japan (2022)
5. **A. Alcañiz**, N. Rukhshi, O. Isabella, H. Ziar “Visual: Calculation tool to determine the Shading Tolerability of Photovoltaic Modules” in *13th International Conference on Crystalline Silicon Photovoltaics*, Delft, the Netherlands (2023)
6. **A. Alcañiz**, N. Rukhshi, R. Koutarapu, O. Isabella, H. Ziar “Oral: Calculation tool to determine the Shading Tolerability of Photovoltaic Modules” in *40th EUPVSEC*, Lisbon, Portugal (2023)
7. **A. Alcañiz**, N. Monaco, O. Isabella, H. Ziar, “Oral: Offshore floating PV-DC and AC yield analysis considering wave effects” in *40th EUPVSEC*, Lisbon, Portugal (2023)
8. **A. Alcañiz**, N. Monaco, O. Isabella, H. Ziar, “Oral: Offshore floating PV-DC and AC yield analysis considering wave effects” in *European PV Performance Modeling Collaborative Workshop*, Mendrisio, Switzerland (2023)
9. **A. Alcañiz** “Panel discussion: Preventive mitigation measures for more accurate PV plant design” in *Solar Quality Summit Europe*, Barcelona, Spain (2024)

INTERVIEWS

1. “From Waves to Shadows: Optimizing Solar Systems by Modelling Their Energy Output” (English, Dutch) *Delft University of Technology*, Published: 05-09-2024
2. “New model for offshore floating PV system design” (English, Spanish) *PV magazine*, Published: 19-09-2024
3. “Kunstmatige intelligentie gaat wereld zonne-energie radicaal veranderen” (Dutch) *Solar Magazine*, Published: 25-09-2024

CURRICULUM VITÆ

Alba ALCANIZ MOYA

Alba Alcañiz Moya was born in Barcelona, Spain on the 21st of November 1996. She studied Engineering Physics at the Universitat Politècnica de Catalunya from which she graduated in 2018. Her interest to learn more about photovoltaic technology led her to pursue a Master's degree at Delft University of Technology in Sustainable Energy Technology. Her thesis in the Photovoltaic Materials and Devices entitled "Numerical simulation of c-Si solar cells based on transition metal oxide as carrier selective contact" contributed to fabricating a record efficiency solar cell. She continued her academic career within the same research group pursuing a PhD in photovoltaic yield modeling within a European Project. She is now working as a Project Engineer at TPSolar Nederland B.V.