

**From InSAR Time-Series to Crop Growth
Machine Learning and Physics-Guided Models for Radar-Based Vegetation Analysis**

Nikaein, T.

DOI

[10.4233/uuid:d5c77c32-2f3e-455b-9187-06c76f9abc22](https://doi.org/10.4233/uuid:d5c77c32-2f3e-455b-9187-06c76f9abc22)

Publication date

2025

Document Version

Final published version

Citation (APA)

Nikaein, T. (2025). *From InSAR Time-Series to Crop Growth: Machine Learning and Physics-Guided Models for Radar-Based Vegetation Analysis*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:d5c77c32-2f3e-455b-9187-06c76f9abc22>

Important note

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

Copyright

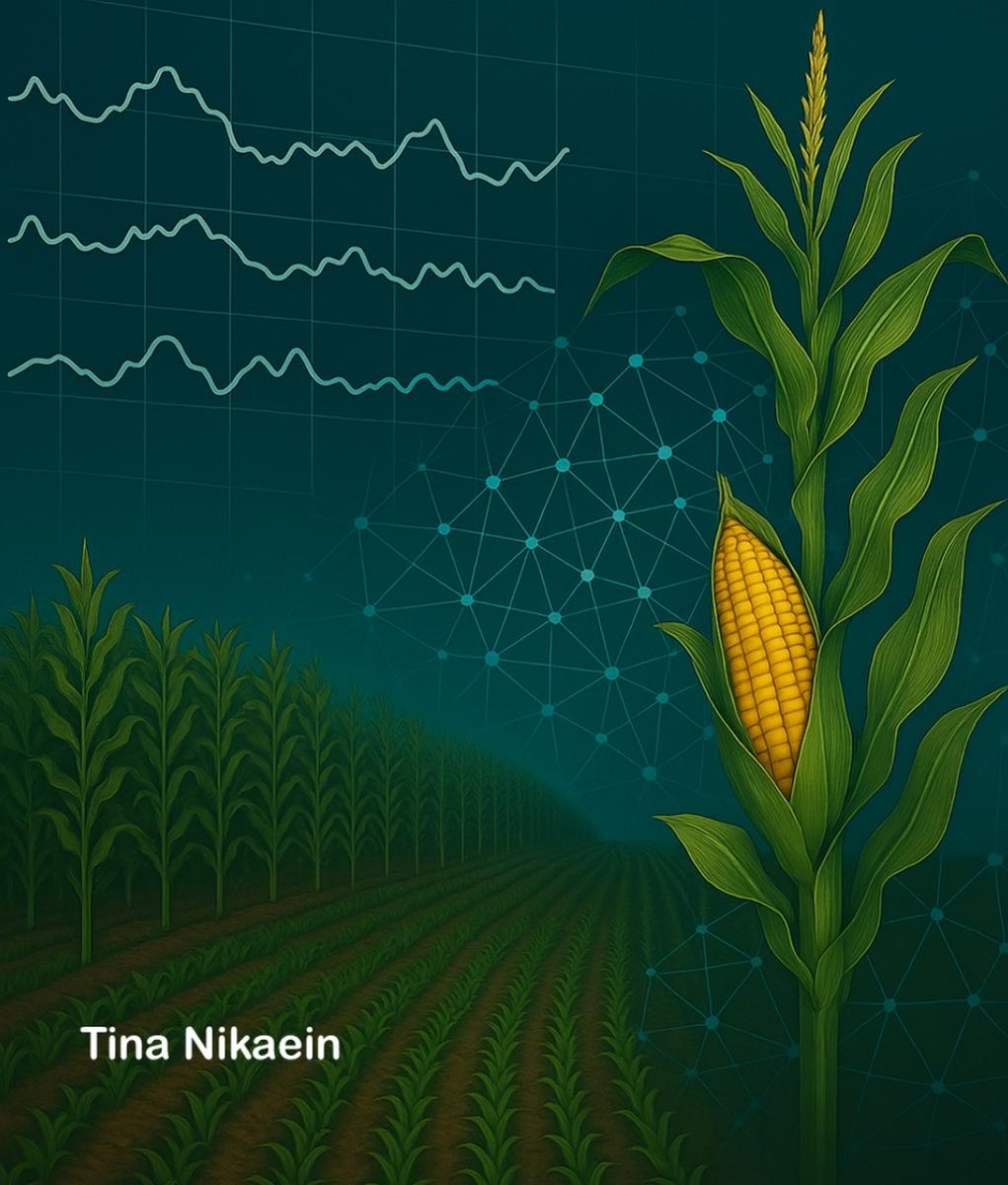
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

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 InSAR Time-Series to Crop Growth

Machine Learning and Physics-Guided Models
for Radar-Based Vegetation Analysis



Tina Nikaein

**FROM INSAR TIME-SERIES TO CROP GROWTH:
MACHINE LEARNING AND PHYSICS-GUIDED
MODELS FOR RADAR-BASED VEGETATION
ANALYSIS**

**FROM INSAR TIME-SERIES TO CROP GROWTH: MACHINE
LEARNING AND PHYSICS-GUIDED MODELS FOR
RADAR-BASED VEGETATION ANALYSIS**

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
Wednesday 11 June 2025 at 17:30 o'clock

by

Tina NIKAEIN

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. R.F. Hanssen	Delft University of Technology, <i>promotor</i>
Prof. dr. ir. P. López-Dekker	Delft University of Technology, <i>promotor</i>

Independent members:

Prof. dr. ir. S.C. Steele-Dunne	Delft University of Technology
Prof. dr. J.M. López-Sanchez	Universidad de Alicante, Spain
Prof. dr. J. Judge	University of Florida, USA
Dr. A. Marino	University of Stirling, UK
Dr. R. Taormina	Delft University of Technology
Prof. dr. ir. H.W.J. Russchenberg	Delft University of Technology, reserve member



Keywords: SAR, Sentinel-1, Land cover mapping, Crop growth modeling, Machine learning, Vegetation, Agriculture.

Printed by: Ridderprint || www.ridderprint.nl

Cover by: Tina Nikaein

Copyright © 2025 by T. Nikaein

ISBN 978-94-6384-790-2

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

To my brother

CONTENTS

Summary	ix
Samenvatting	xi
List of Abbreviations	xiii
1. Introduction	1
1.1. Background and Motivation	2
1.1.1. Land Cover Mapping and Monitoring	2
1.1.2. Crop Growth Monitoring	3
1.1.3. Machine Learning in SAR-based Vegetation Monitoring	5
1.1.4. Physics-guided machine learning	6
1.2. Objectives and Research Questions	7
1.3. Outline	7
2. On the Value of Sentinel-1 InSAR Coherence Time-Series for Vegetation Classification	19
2.1. Introduction	20
2.2. Study Area and Data	21
2.3. Methods	22
2.3.1. Pre-Processing	22
2.3.2. Interferometric Coherence	25
2.3.3. Land Cover Classification	27
2.3.4. Feature Relevance	29
2.4. Results and Discussion	29
2.4.1. Quantitative Accuracy	29
2.4.2. Spatial Analysis	32
2.4.3. Feature Relevance Analysis	34
2.5. Conclusions	42
3. Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning	47
3.1. Introduction	48
3.2. Study Area	50
3.3. Data and Methodology	50
3.3.1. DSSAT	50
3.3.2. Input Data for DSSAT	52
3.3.3. SAR and Optical Data	55
3.3.4. Support Vector Regression (SVR)	55

3.3.5.	Surface Roughness	56
3.3.6.	Evaluation of model performance	57
3.3.7.	Feature Analysis	57
3.4.	Results and Discussion	58
3.4.1.	DSSAT Calibration and Performance	58
3.4.2.	Modeled vs Measured Radar Observables	60
3.4.3.	Feature analysis	66
3.5.	Conclusions	67
4.	Physics-Guided Machine Learning Based Forward-Modeling of Radar Observables: a Case Study on Sentinel-1 Observations of Corn-Fields	79
4.1.	Introduction	80
4.2.	Study Area and Data	82
4.3.	Methodology	83
4.3.1.	Physics-based loss function	83
4.3.2.	Physics guided network topology	84
4.3.3.	NN Implementation	86
4.4.	Experiment Workflow	86
4.5.	Results	88
4.5.1.	Synthetic data	88
4.5.2.	Sentinel-1 data	88
4.6.	Conclusion	92
5.	Conclusions and Recommendations	101
5.1.	Conclusions and Main Contributions	101
5.2.	Further Research	105
A.	APPENDIX-A: SUPPLEMENTARY MATERIAL	109
A.1.	Supplementary Material for Chapter 3	109
	Acknowledgements	121
	Curriculum Vitæ	123
	List of Publications	125

SUMMARY

Spaceborne sensors, particularly Synthetic Aperture Radar (SAR), provide valuable tools for monitoring agricultural resources, improving yield predictions, and ensuring sustainable farming practices. In this research, we explore several venues to advance the use of SAR observation time series for agricultural and vegetation monitoring applications.

The first part of this research evaluates the added value of Sentinel-1 InSAR coherence time series for land cover classification, using an agricultural region in São Paulo, Brazil, as a case study. This region is characterized by a mixture of crops, pastures, and sugarcane plantations, all managed asynchronously. The findings demonstrate that incorporating InSAR coherence alongside SAR backscatter improves classification accuracy, particularly during the dry season when the distinctions between vegetation and bare soil are more pronounced. The research employed machine learning approaches to analyze pixel-level and field-level classifications using different sampling schemes. It highlights how multi-looking strategies can be adjusted to improve the accuracy of the classification outcomes in agricultural settings. This research highlights the usefulness of coherence data for the detection of events such as harvesting, offering valuable insights for more dynamic agricultural monitoring. The sensitivity of the coherence to agricultural changes leads to the observed improvement in Land Use Land Cover (LULC) mapping.

Forward models, or observation operators, are essential for the interpretation of radar observations and for the development of assimilation frameworks. In particular, in this research, we are interested in forward modeling the relation between crop bio-geophysical parameters, such as Leaf Area Index (LAI), Above Ground Biomass (AGB), and soil moisture, the inputs to our data-driven model, and radar observables, the outputs.

In the second part of this research, we integrate an existing crop growth model, the Decision Support System for Agrotechnology Transfer (DSSAT), with machine learning techniques to train a forward model to predict SAR observables over silage maize fields in The Netherlands across multiple years. Using crop growth models circumvents the dependency on limitedly available field measurements. When we use training and validation data from the same growth season, we obtain accurate predictions, with a mean absolute error (MAE) of less than 1.23 dB. Some of the field-to-field variability is accounted for by including the mean backscatter intensity during a few acquisitions before crop emergence. The obtained performance suggests the potential of using this approach to generate observation operators for data assimilation frameworks or for anomaly detection, supporting large-scale agricultural monitoring. However, the results also highlight one of the main challenges: the resulting data-driven model fails to generalize when presented with input bio-geophysical parameters that fall outside the regions of the parameter space spanned by the training dataset, as can happen, for example, during a

drought period.

In the final part, we develop a physics-guided machine learning approach to address the limitations of data-driven models: lack of generalizability, tendency to overfitting, and reliance on extensive training data sets. We introduce physical constraints in an artificial neural network (ANN) in two ways. First, by modifying the loss function, used to train the ANN, by including a penalty for unphysical behavior, in particular by penalizing negative values of the partial derivative of the predicted backscatter intensity with respect to the surface soil moisture, since we assume this should always be positive. Second, by mirroring the architecture of the widely used Water Cloud Model (WCM) in the network topology. The added physical term to the loss function improves the ANN performance in all cases considered, with an R^2 increase of 3 percentage points (p.p). The WCM-inspired model performs slightly worse when trained and tested with data from the same year, but it generalizes better, producing significantly better results for unseen conditions. In addition, the WCM-inspired model also produces individual contributions to the observed intensity, such as the surface-scattering component and the vegetation backscatter component.

SAMENVATTING

Aardobservatiesatellieten, in bijzonder radar satellieten, zijn waardevol voor het monitoren van landbouw, het verbeteren van opbrengstvoorspellingen en het waarborgen van duurzame landbouwpraktijken. In dit onderzoek verkennen we verschillende mogelijkheden om SAR-tijdreeksen in te zetten voor het verbeteren van toepassingen in landbouw- en vegetatiemonitoring.

Het eerste deel van dit onderzoek evalueert de toegevoegde waarde van Sentinel-1 InSAR-coherentietijdreeksen voor landbedekkingsclassificatie, waarbij een landbouwregio in São Paulo, Brazilië, als casestudy wordt gebruikt. Deze regio wordt gekenmerkt door een mix van gewassen, weilanden en suikerrietplantages, die allemaal asynchroon worden beheerd. De bevindingen tonen aan dat het combineren van InSAR-coherentie met SAR-weerkaatsing de classificatie nauwkeurigheid verbetert, vooral tijdens het droge seizoen, wanneer de verschillen tussen vegetatie en kale grond beter te onderscheiden zijn. Het onderzoek maakte gebruik van machine learning-technieken om pixel- en veldniveauclassificaties te analyseren met verschillende bemonsteringsstrategieën. Het benadrukt hoe strategieën zoals multi-looking kunnen worden aangepast om de nauwkeurigheid van classificatieresultaten in landbouwcontexten te verbeteren. Dit onderzoek toont het nut aan van coherentiegegevens voor het detecteren van gebeurtenissen zoals oogsten, wat waardevolle inzichten biedt voor dynamischer landbouwmonitoring. De gevoeligheid van de coherentie voor veranderingen in de landbouw draagt bij aan de waargenomen verbeteringen in Land Use Land Cover (LULC)-kaarten.

Forward models, oftewel observation operators, zijn essentieel voor de interpretatie van radarwaarnemingen en voor de ontwikkeling van assimilatiekaders. In dit werk zijn we in het bijzonder geïnteresseerd in het forward modelleren van de relatie tussen biofysische parameters van gewassen, zoals het bladoppervlakte-index (Leaf Area Index, LAI), bovengrondse biomassa (Above Ground Biomass, AGB) en bodemvocht, de input voor ons data-gedreven model, en radarwaarnemingen, de output.

In het tweede deel van dit onderzoek integreren we een gewasgroei-model, het Decision Support System for Agrotechnology Transfer (DSSAT), met machine learning-technieken om een forward-model te trainen dat SAR-metingen voorspelt voor maïsvelden in Nederland over meerdere jaren. Het gebruik van gewasgroei-modellen lost de afhankelijkheid van beperkt beschikbare veldmetingen op. Wanneer we trainings- en validatiegegevens uit hetzelfde groeiseizoen gebruiken, verkrijgen we nauwkeurige voorspellingen met een mean absolute error (MAE) van minder dan 1,23 dB. Een deel van de variabiliteit tussen velden is uit te leggen door het gemiddelde weerkaatsingsniveau tijdens enkele waarnemingen vóór de opkomst van het gewas toe te voegen. De behaalde resultaten suggereren dat deze aanpak potentieel heeft om observation operators te genereren voor data-assimilatiekaders of voor anomaliedetectie, ter ondersteuning van grootschalige landbouwmonitoring. De resultaten benadrukken echter ook een belangrijke

uitdaging: het datagedreven model generaliseert niet goed wanneer het wordt geconfronteerd met invoerparameters buiten het bereik van de trainingsdataset, zoals tijdens een droogteperiode.

In het laatste deel ontwikkelen we een door fysieke kennis gestuurde machine learning-aanpak om de beperkingen van datagedreven modellen te trotseren: gebrek aan generaliseerbaarheid, neiging tot overfitting en afhankelijkheid van omvangrijke trainingsdatasets. We introduceren inherente fysieke beperkingen in een artificial neural network (ANN) op twee manieren. Ten eerste door de loss function, die wordt gebruikt om het ANN te trainen, te wijzigen door een strafterm op te nemen voor fysisch onrealistisch gedrag, in het bijzonder door negatieve waarden van de partiële afgeleide van de voorspelde werkaatsingsintensiteit met betrekking tot het bodemvocht te bestraffen, omdat we aannemen dat deze altijd positief moet zijn. Ten tweede door de architectuur van het veelgebruikte Water Cloud Model (WCM) te spiegelen in de netwerktopologie. De aangepaste loss function verbetert de prestaties van het ANN in alle beschouwde gevallen met een R^2 -toename van 3 procentpunten (p.p). Het WCM-geïnspireerde model presteert iets slechter bij testen en trainen met gegevens uit hetzelfde jaar, maar generaliseert beter en levert aanzienlijk betere resultaten voor onbekende omstandigheden. Bovendien produceert het WCM-geïnspireerde model ook afzonderlijke bijdragen aan de waargenomen intensiteit, zoals de oppervlakverspreidingscomponent en de vegetatietierugstrooiingscomponent.

LIST OF ABBREVIATIONS

SAR	Synthetic Aperture Radar
InSAR	Interferometric Synthetic Aperture Radar
NRCS	Normalized Radar Cross Section
GLC	Global Land Cover
LULC	Land Use Land Cover
VV	Vertical transmit Vertical receive
VH	Vertical transmit Horizontal receive
CR	Cross-Ratio
p-p	Percentage Points
DSSAT	Decision Support System for Agrotechnology Transfer
AGB	Above Ground Biomass
DB	Dry Biomass
LAI	Leaf Area Index
RTM	Radiative Transfer Models
VWC	Vegetation Water Content
MIMICS	Michigan Microwave Canopy Scattering
CGM	Crop Growth Model
DA	Data Assimilation
EO	Earth Observation
ML	Machine Learning
SVR	Support Vector Regression
PGNN	Physic Guided Neural Network
ROSE-L	Radar Observing System for Europe - L-Band

ASCAT	Advanced SCATterometer
ISBA	Interaction-Soil-Biosphere-Atmosphere

1

INTRODUCTION

*Perhaps you are overvaluing what you don't have
and undervaluing what you do.*

Jordan B. Peterson

1.1. BACKGROUND AND MOTIVATION

Space-borne remote sensing images play a key role in vegetation monitoring by providing repeated observations over large areas. These observations are essential for understanding vegetation dynamics, detecting changes in land cover, and supporting agricultural and environmental management.

Synthetic Aperture Radar (SAR) products complement optical images by offering sensitivity to bio-geophysical parameters such as soil moisture, dry and wet biomass, or structural parameters of the canopy, while being independent of weather conditions and solar illumination. By acquiring data over the same area at regular time intervals, SAR enables the study of vegetation dynamics over time, such as seasonal growth patterns, and responses to environmental changes or agricultural practices. Radar backscatter from vegetation is a superposition of the electromagnetic signal reflected by the vegetation, the underlying soil, as well as multiple reflections involving both [1]. This sensitivity allows SAR data to be used for agricultural applications such as crop monitoring [2, 3], crop mapping [4], drought monitoring [5] and detecting waterlogging [6].

1.1.1. LAND COVER MAPPING AND MONITORING

Cropland and pastureland maps are important for a range of land monitoring applications, including food security monitoring and early warning [7, 8], water use efficiency [9], soil erosion prevention [10] and fraud detection in agricultural practices [11]. One of the primary challenges in these scenarios is managing the high spatial complexity and dynamic nature of these environments, which evolve due to various land uses and management practices over time.

Over the past twenty years, advancements in space-based remote sensing technology and the increasing computational resources available for processing large datasets have considerably improved the quality of land cover products [12, 13]. These improvements are pronounced in the spatial resolution, which has shifted from scales of tens of kilometers to tens of meters, and in update frequency, with some products updated annually and others as often as every few days [14–20]. Such progress has allowed for more precise and timely land cover monitoring, supporting applications in environmental management and agricultural monitoring. Previous studies provide a summary of the current state of available global land covers including their challenges [21–24]. Despite these advancements, most current land cover products rely on optical remote sensing data. This reliance presents challenges, particularly when addressing the temporal dynamics of land. The acreage and conditions of cropland and pastureland are continuously affected by climate variations and anthropogenic management. In cloud-prone regions, such as tropical areas experiencing frequent land-use and land-cover (LULC) changes [25], reliance on optical imagery alone can limit temporal coverage. This makes the land cover maps less effective and highlights the need to incorporate SAR sensors to overcome these limitations. As a result, there is growing interest in LULC products based on SAR [26] or fusion of SAR with optical data [27].

Long and dense time series are important for LULC mapping and monitoring as they allow for the detection of dissimilarities in the signature of different crop

types during specific days of the year. Time series of SAR images have been proven effective in discriminating different crop types and producing classification maps of good quality [2, 28]. The two satellites belonging to the Sentinel-1 (S1) constellation provide C-band intensity and coherence time series in dual-polarization (VV and VH) with revisit intervals of 6 or 12 days for consecutive acquisitions in the same geometry. These short revisit times make it possible to monitor crops and pasture conditions in near-real time and keep track of land cover changes. Time series analysis is invaluable for tracking phenological events, assessing crop health, and predicting yields. Furthermore, dual-polarized SAR enriches the information gathered, as different polarizations can reflect different aspects of vegetation structure and moisture content. This dual perspective enhances the differentiation between various types of vegetation and land cover, allowing for more accurate classification and monitoring.

In the past, most agricultural mapping studies using Sentinel-1 relied exclusively on backscatter intensities [29–32]. Coherence received less attention, primarily due to its more time-consuming processing and its noisy nature, and the fact that it represents the difference between epochs, unlike intensity or amplitude, which correspond to a single epoch. However, recent studies have highlighted its potential for identifying specific events such as mowing or grazing events [33–35], monitoring crop growth stages [36–38].

As land use monitoring techniques continue to advance, the impact of multitemporal coherence data should be considered in vegetation mapping methods. Developing robust models that effectively monitor and characterize crops and pastures using multitemporal Sentinel-1 data represents an essential next step.

1.1.2. CROP GROWTH MONITORING

Crop monitoring is essential for assessing crop development and productivity of agricultural systems. It involves tracking key indicators such as leaf area index (LAI), above ground biomass (AGB), and other indicators of vegetation health to ensure that crops are growing optimally. Monitoring these variables helps identify growth anomalies or stressors, enabling timely interventions to optimize yields. Remote sensing technologies, such as SAR, are valuable tools for monitoring these parameters [2, 39].

Radar backscattered intensity and coherence are highly sensitive to changes in canopy structures, variations in soil moisture, and vegetation water content (VWC) [40]. This high sensitivity to multiple bio-geophysical variables complicates the accurate retrieval of soil moisture and vegetation parameters, particularly in areas with dense vegetation. Converting SAR microwave data into measurable crop and/or soil-related parameters such as soil moisture and VWC, requires inversion models. However, these models often struggle to encapsulate the complex interactions of the electromagnetic wave with the soil and vegetation. Integrating SAR observations with modeled soil moisture and vegetation can be approached in two ways: i) converting the observed SAR signal directly into retrieval products such as soil moisture or VWC; or ii) mapping modeled bio-geophysical parameters to satellite observables using a forward operator. The first approach often uses algorithms that rely on

pattern changes between consecutive SAR images [41, 42] or other techniques [43]. The second approach involves mapping the crop model outputs to observations using a forward model. This model predicts the signal that would be received by the sensor, based on the simulated states of soil moisture and vegetation [44, 45]. This capability is important for applications such as data assimilation, where integrating SAR observations into crop models can improve predictions, anomaly detection for identifying deviations from normal conditions, and mission design for optimizing sensor specifications and data acquisition strategies.

Crop growth models (CGMs) predict the life cycle of crops, from planting to harvest, by simulating key growth processes and environmental interactions. These models provide quantitative forecasts of crop attributes including LAI, AGB, soil moisture, height, and yield. When field measurement data is scarce, the value of CGMs becomes particularly evident. They can use available data efficiently to mitigate the impact of data gaps and provide reliable predictions, making them invaluable where comprehensive data collection is challenging. These models rely on inputs such as meteorological data (such as solar radiation, temperature, and precipitation), soil characteristics, crop genetic traits, and field management practices, all of which are critical for accurate simulations of crop [46]. Figure. 1.1 provides an overview of the most relevant inputs and outputs for CGMs, specifically for the Decision Support System for Agrotechnology Transfer (DSSAT) crop-growth model [47, 48]. While CGMs perform well in theoretical studies and small-scale applications, they face difficulties when applied to larger regional scales. These models are typically tuned to individual plots or fields, with simplified parameters and structures, making it challenging to account for the complex interactions of weather, soil, and farming practices at a broader scale. One way to address these limitations is through data assimilation (DA), which involves updating model variables with observations [49–51]. DA techniques integrate multisource remote sensing data with CGM simulation results, aiming to enhance the accuracy of crop growth analysis [52, 53]. DA can involve either satellite observation (e.g. backscatter) directly or retrieval products (e.g. soil moisture). Previous studies have explored the direct assimilation of microwave data [54, 55]. The assimilation of directly observed signals limits the inconsistencies and cross-correlation errors between retrievals and model simulations [56].

Analytical, numerical models or data-driven ones can address this direct mapping between crop model variables and observed signals. Forward models, also known as observation operators, play a crucial role in analyzing and interpreting remote sensing data, designing observational frameworks, and integrating measurements into numerical models. Traditionally, these models rely on simple empirical approaches, basic physical models, or combinations of the two (e.g., physical models with empirically adjusted parameters). The water cloud model (WCM) [57] is one of the most commonly used forward operators. Although these methods are practical, they often fall short of capturing the full complexity of the physical system and the various phenomena it encompasses. There are more sophisticated physics-based models e.g. Tor Vergata [58] but the difficulties in their parametrization limit their usage. In practice, parameters in physical models are adjusted to minimize the

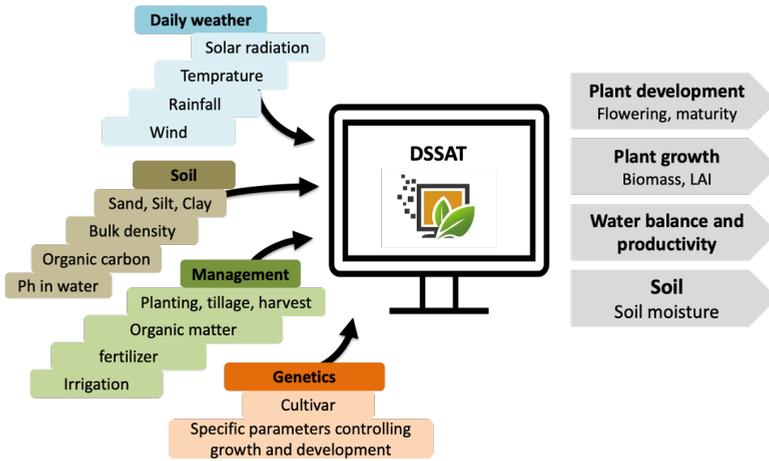


Figure 1.1.: Depiction of key inputs and outputs of the DSSAT [47, 48] crop-growth model.

difference between model predictions and observations. As physical models become more sophisticated, it becomes harder to determine if this parameter adjustment can be interpreted as a tuning of the targeted bio-geophysical parameter or a general model tuning that may account for missing contributions. Considering this general tuning exercise, a natural next step is to consider fully data-driven approaches.

During the last decades, ML algorithms have been widely adopted as an approach to modeling complex systems when a theoretical understanding of the systems is incomplete but data capturing the relations between inputs and outputs are abundant. Previous studies showed the potential of ML algorithms as a forward observation [55, 59, 60].

1.1.3. MACHINE LEARNING IN SAR-BASED VEGETATION MONITORING

The use of ML in remote sensing has advanced agricultural mapping and monitoring. The integration of ML with remotely sensed data has emerged as a powerful tool, providing insight into biomass estimation [61, 62], crop health [63, 64], crop mapping [65] and yield prediction [66]. ML in agricultural applications has been used for various purposes, from classification [67–69] to functioning as a regression model between vegetation parameters and earth observation (EO) datasets [55, 60].

Recently, ML models have been employed as forward models to map model variables to SAR observations [70]. ML methods can learn patterns and relationships from large datasets without requiring explicit governing equations. However, it is essential to use ML models with caution due to potential issues such as overfitting and the risk of losing important physical insights. Purely data-driven approaches may disregard fundamental physical laws, resulting in predictions that lack physical plausibility and interpretability. To address these challenges, physics-informed ML models are emerging as a valuable solution [71].

1.1.4. PHYSICS-GUIDED MACHINE LEARNING

In recent years, the integration of ML with physics-based knowledge has shown considerable promise in enhancing the accuracy and interpretability of model outputs in different applications [72]. This approach leverages the robust theoretical foundations of physics to guide and constrain ML models, resulting in predictions that are both accurate and physically plausible [73, 74].

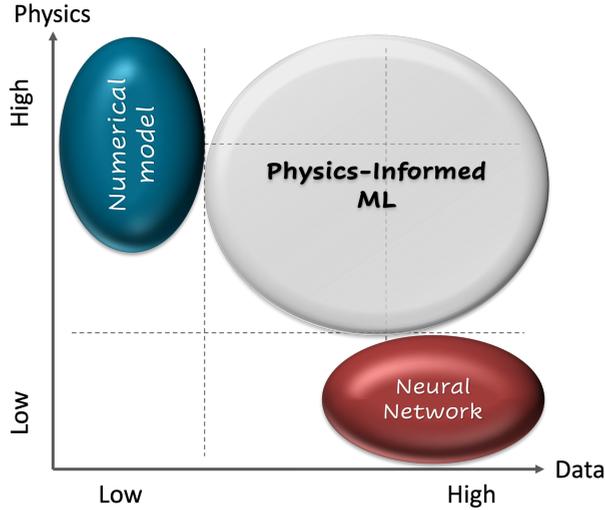


Figure 1.2.: Depiction of the relationship between a neural network, numerical model, and physics-informed ML ([72, 75, 76]).

While ML algorithms, particularly deep learning networks, excel at detecting complex patterns, they often require large amounts of data to train effectively and can produce results that, while statistically accurate, may not align with physical principles [77]. As ML models are typically training on limited data, they may struggle to generalize beyond their training domain, a challenge that is especially prominent in remote sensing where in-situ measurements are often scarce.

On the other hand, physical models represent reality using a set of equations and require extensive calibration to fine-tune parameters. It may happen that they cannot capture the full complexity of real-world phenomena due to either an incomplete understanding of the underlying physics or omitted variables. Integrating physical principles with ML can help mitigate these shortcomings by leveraging the strengths of both approaches. This integration enhances the transferability of models across different domains, ensuring that they remain robust and relevant under varying conditions. Figure. 1.2 represents a depiction of the relationship between different types of models.

1.2. OBJECTIVES AND RESEARCH QUESTIONS

This research explores the use of SAR data in agricultural monitoring by integrating the potential of machine learning and crop growth models. The main goal of this thesis is:

To advance Synthetic Aperture Radar-based vegetation monitoring capabilities exploiting the full range of radar observables combined with machine learning techniques.

Driven by the previous motivation and general objective, the research presented in the remainder of this dissertation is articulated around the following research questions:

1. *To what extent does the inclusion of short-term interferometric coherence time series improve the accuracy of vegetation mapping?*
2. *How can we address the unavoidable scarcity of field measurement data to train a data-driven forward model that relates crop bio-geophysical parameters to radar observables?*
3. *How can we guide neural network architectures with physical principles to improve the prediction of SAR observables for vegetation?*

1.3. OUTLINE

To address the first research question, **Chapter 2** examines the potential of Sentinel-1 InSAR coherence time-series for vegetation classification in a complex agricultural area in São Paulo, Brazil. It evaluates the added value resulting from combining dual-polarized (VV and VH) InSAR coherence with backscatter intensity time series to improve land-cover classification accuracy. Machine learning techniques provide us with an ideal tool to explore this added value, as well as a framework in which to analyze the importance of the different radar observables considered. We examine the impact of different classification methods and data preparation processes on the quality of land cover classification results and the effect of sampling strategies on the validity of the analysis. We investigate how multi-looking and field-based coherence estimation enhances the quality of coherence as an observable, leading to more accurate classification.

Chapter 3 addresses the second question. We train a machine learning model to predict the observed radar observables using the outputs of a crop-growth model as input. Insofar as the outputs of the crop-growth model are reliable, the result is a data-driven forward model. We obtain a close fit between modeled and observed C-band SAR observables.

In **Chapter 4** we turn our attention to the last question, exploring two approaches to combine physical understanding of electromagnetic scattering over vegetated

surfaces with artificial neural networks to yield a more accurate and/or robust forward model of radar observables. The chapter is a natural continuation of the work presented in Chapter 3, and it addresses some of the shortcomings of purely data-driven models.

Chapter 5 gives an overview of the main findings and recommendations for further research. Figure 1.3 presents a schematic overview that illustrates the various contributions made in this thesis.

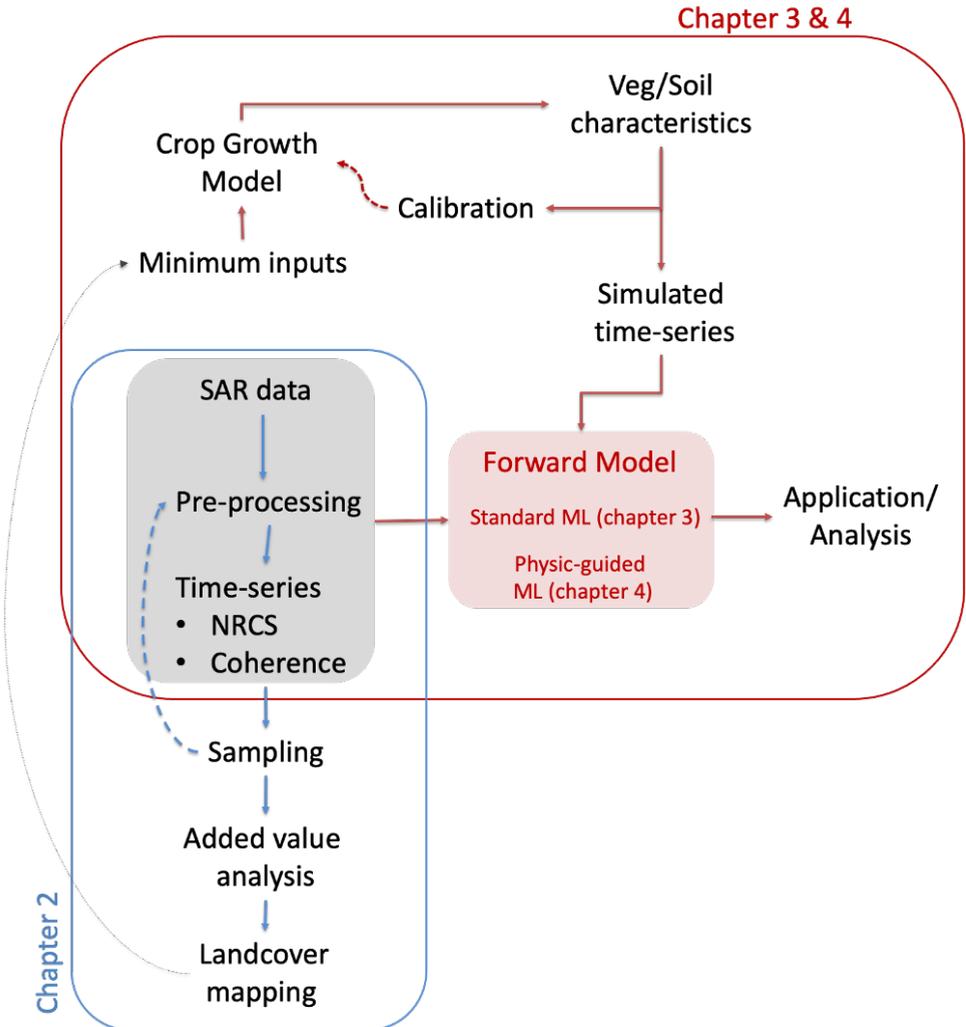


Figure 1.3.: Overview of study methods and chapters.

REFERENCES

- [1] F. T. Ulaby, P. C. Dubois, and J. van Zyl. “Radar Mapping of Surface Soil Moisture”. In: *Journal of Hydrology*. Soil Moisture Theories and Observations 184.1 (Oct. 1996), pp. 57–84. ISSN: 0022-1694. DOI: [10.1016/0022-1694\(95\)02968-0](https://doi.org/10.1016/0022-1694(95)02968-0).
- [2] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. Liu, and K. Papathanassiou. “Radar Remote Sensing of Agricultural Canopies: A Review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2249–2273. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2016.2639043](https://doi.org/10.1109/JSTARS.2016.2639043).
- [3] M. Vreugdenhil, W. Wagner, B. Bauer-Marschallinger, I. Pfeil, I. Teubner, C. Rüdiger, and P. Strauss. “Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study”. In: *Remote Sensing* 10.9 (Sept. 2018), p. 1396. ISSN: 2072-4292. DOI: [10.3390/rs10091396](https://doi.org/10.3390/rs10091396).
- [4] L. Dingle Robertson, A. M. Davidson, H. McNairn, M. Hosseini, S. Mitchell, D. de Abelleira, S. Verón, G. le Maire, M. Plannells, S. Valero, N. Ahmadian, A. Coffin, D. Bosch, M. H. Cosh, B. Basso, and N. Saliendra. “C-Band Synthetic Aperture Radar (SAR) Imagery for the Classification of Diverse Cropping Systems”. In: *International Journal of Remote Sensing* 41.24 (Dec. 2020), pp. 9628–9649. ISSN: 0143-1161. DOI: [10.1080/01431161.2020.1805136](https://doi.org/10.1080/01431161.2020.1805136).
- [5] M. Shorachi, V. Kumar, and S. C. Steele-Dunne. “Sentinel-1 SAR Backscatter Response to Agricultural Drought in The Netherlands”. In: *Remote Sensing* 14.10 (Jan. 2022), p. 2435. ISSN: 2072-4292. DOI: [10.3390/rs14102435](https://doi.org/10.3390/rs14102435).
- [6] N. den Besten, S. Steele Dunne, A. Mahmud, D. Jackson, B. Aouizerats, R. de Jeu, R. Burger, R. Houborg, M. McGlinchey, and P. van der Zaag. “Understanding Sentinel-1 Backscatter Response to Sugarcane Yield Variability and Waterlogging”. In: *Remote Sensing of Environment* 290 (May 2023), p. 113555. ISSN: 0034-4257. DOI: [10.1016/j.rse.2023.113555](https://doi.org/10.1016/j.rse.2023.113555).
- [7] B. Wu, R. Gommès, M. Zhang, H. Zeng, N. Yan, W. Zou, Y. Zheng, N. Zhang, S. Chang, Q. Xing, and A. Van Heijden. “Global Crop Monitoring: A Satellite-Based Hierarchical Approach”. In: *Remote Sensing* 7.4 (Apr. 2015), pp. 3907–3933. DOI: [10.3390/rs70403907](https://doi.org/10.3390/rs70403907).
- [8] C. Vancutsem, E. Marinho, F. Kayitakire, L. See, and S. Fritz. “Harmonizing and Combining Existing Land Cover/Land Use Datasets for Cropland Area Monitoring at the African Continental Scale”. In: *Remote Sensing* 5.1 (Jan. 2013), pp. 19–41. DOI: [10.3390/rs5010019](https://doi.org/10.3390/rs5010019).

- [9] A. Calera, I. Campos, A. Osann, G. D’Urso, and M. Menenti. “Remote Sensing for Crop Water Management: From ET Modelling to Services for the End Users”. In: *Sensors* 17.5 (May 2017), p. 1104. ISSN: 1424-8220. DOI: [10.3390/s17051104](https://doi.org/10.3390/s17051104).
- [10] M. M. Alkharabsheh, T. K. Alexandridis, G. Bilas, N. Misopolinos, and N. Silleos. “Impact of Land Cover Change on Soil Erosion Hazard in Northern Jordan Using Remote Sensing and GIS”. In: *Procedia Environmental Sciences. Four Decades of Progress in Monitoring and Modeling of Processes in the Soil-Plant-Atmosphere System: Applications and Challenges* 19 (Jan. 2013), pp. 912–921. ISSN: 1878-0296. DOI: [10.1016/j.proenv.2013.06.101](https://doi.org/10.1016/j.proenv.2013.06.101).
- [11] J. Schmedtmann and M. L. Campagnolo. “Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control”. In: *Remote Sensing* 7.7 (July 2015), pp. 9325–9346. ISSN: 2072-4292. DOI: [10.3390/rs70709325](https://doi.org/10.3390/rs70709325).
- [12] T. Kuemmerle, K. Erb, P. Meyfroidt, D. Müller, P. H. Verburg, S. Estel, H. Haberl, P. Hostert, M. R. Jepsen, T. Kastner, C. Levers, M. Lindner, C. Plutzer, P. J. Verkerk, E. H. van der Zanden, and A. Reenberg. “Challenges and Opportunities in Mapping Land Use Intensity Globally”. In: *Current Opinion in Environmental Sustainability. Human Settlements and Industrial Systems* 5.5 (Oct. 2013), pp. 484–493. ISSN: 1877-3435. DOI: [10.1016/j.cosust.2013.06.002](https://doi.org/10.1016/j.cosust.2013.06.002).
- [13] F. Gilić, M. Gašparović, and M. Baučić. “Current State and Challenges in Producing Large-Scale Land Cover Maps: Review Based on Recent Land Cover Products”. In: *Geocarto International* 38.1 (Dec. 2023), p. 2242693. ISSN: 1010-6049. DOI: [10.1080/10106049.2023.2242693](https://doi.org/10.1080/10106049.2023.2242693).
- [14] E. Bartholomé and A. S. Belward. “GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data”. In: *International Journal of Remote Sensing* 26.9 (May 2005), pp. 1959–1977. ISSN: 0143-1161. DOI: [10.1080/01431160412331291297](https://doi.org/10.1080/01431160412331291297).
- [15] P. Defourny, L. Schouten, S. Bartalev, S. Bontemps, P. Cacetta, A. De Wit, C. Di Bella, B. Gerard, C. Giri, V. Gond, G. Hazeu, A. Heinemann, M. Herold, J. Koozebani, G. Jaffrain, R. Latifovic, H. Lin, P. Mayaux, S. Muecher, A. Nonguierma, H.-J. Stibig, E. Van Bogaert, C. Vancutsem, P. Bicheron, M. Leroy, and O. Arino. “Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience”. In: *33rd International Symposium on Remote Sensing of Environment, Sustaining the Millennium Development Goals - ISBN 978-0-932913-13-5* (2009).
- [16] M. A. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. “MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets”. In: *Remote Sensing of Environment* 114.1 (Jan. 2010), pp. 168–182. ISSN: 0034-4257. DOI: [10.1016/j.rse.2009.08.016](https://doi.org/10.1016/j.rse.2009.08.016).

- [17] M. C. Hansen and T. R. Loveland. "A Review of Large Area Monitoring of Land Cover Change Using Landsat Data". In: *Remote Sensing of Environment*. Landsat Legacy Special Issue 122 (July 2012), pp. 66–74. ISSN: 0034-4257. DOI: [10.1016/j.rse.2011.08.024](https://doi.org/10.1016/j.rse.2011.08.024).
- [18] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, and J. Mills. "Global Land Cover Mapping at 30m Resolution: A POK-Based Operational Approach". In: *ISPRS Journal of Photogrammetry and Remote Sensing*. Global Land Cover Mapping and Monitoring 103 (May 2015), pp. 7–27. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2014.09.002](https://doi.org/10.1016/j.isprsjprs.2014.09.002).
- [19] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby. "Global Land Use / Land Cover with Sentinel 2 and Deep Learning". In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. July 2021, pp. 4704–4707. DOI: [10.1109/IGARSS47720.2021.9553499](https://doi.org/10.1109/IGARSS47720.2021.9553499).
- [20] C. Brown, S. Brumby, B. Guzder-Williams, T. Birch, S. Hyde, J. Mazzariello, W. Czerwinski, V. Pasquarella, R. Haertel, S. Ilyushchenko, K. Schwehr, M. Weisse, F. Stolle, C. Hanson, O. Guinan, R. Moore, and A. Tait. "Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping". In: *Scientific Data* 9.1 (2022). DOI: [10.1038/s41597-022-01307-4](https://doi.org/10.1038/s41597-022-01307-4).
- [21] G. Grekousis, G. Mountrakis, and M. Kavouras. "An Overview of 21 Global and 43 Regional Land-Cover Mapping Products". In: *International Journal of Remote Sensing* 36.21 (Nov. 2015), pp. 5309–5335. ISSN: 0143-1161. DOI: [10.1080/01431161.2015.1093195](https://doi.org/10.1080/01431161.2015.1093195).
- [22] N. Joshi, M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. R. Jepsen, T. Kuemmerle, P. Meyfroidt, E. T. A. Mitchard, J. Reiche, C. M. Ryan, and B. Waske. "A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring". In: *Remote Sensing* 8.1 (Jan. 2016), p. 70. ISSN: 2072-4292. DOI: [10.3390/rs8010070](https://doi.org/10.3390/rs8010070).
- [23] Y. Wang, Y. Sun, X. Cao, Y. Wang, W. Zhang, and X. Cheng. "A Review of Regional and Global Scale Land Use/Land Cover (LULC) Mapping Products Generated from Satellite Remote Sensing". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 206 (Dec. 2023), pp. 311–334. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2023.11.014](https://doi.org/10.1016/j.isprsjprs.2023.11.014).
- [24] P. Xu, N.-E. Tsendbazar, M. Herold, S. de Bruin, M. Koopmans, T. Birch, S. Carter, S. Fritz, M. Lesiv, E. Mazur, A. Pickens, P. Potapov, F. Stolle, A. Tyukavina, R. Van De Kerchove, and D. Zanaga. "Comparative Validation of Recent 10 M-Resolution Global Land Cover Maps". In: *Remote Sensing of Environment* 311 (Sept. 2024), p. 114316. ISSN: 0034-4257. DOI: [10.1016/j.rse.2024.114316](https://doi.org/10.1016/j.rse.2024.114316).
- [25] N. M. Kayitesi, A. C. Guzha, and G. Mariethoz. "Impacts of Land Use Land Cover Change and Climate Change on River Hydro-Morphology- a Review of Research Studies in Tropical Regions". In: *Journal of Hydrology* 615 (Dec. 2022), p. 128702. ISSN: 0022-1694. DOI: [10.1016/j.jhydrol.2022.128702](https://doi.org/10.1016/j.jhydrol.2022.128702).

- [26] R. d'Andrimont, A. Verhegghen, G. Lemoine, P. Kempeneers, M. Meroni, and M. van der Velde. "From Parcel to Continental Scale – A First European Crop Type Map Based on Sentinel-1 and LUCAS Copernicus in-Situ Observations". In: *Remote Sensing of Environment* 266 (Dec. 2021), p. 112708. ISSN: 0034-4257. DOI: [10.1016/j.rse.2021.112708](https://doi.org/10.1016/j.rse.2021.112708).
- [27] D. Zanaga, R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N.-E. Tsendbazar, P. Xu, F. Ramoino, and O. Arino. *ESA WorldCover 10 m 2021 V200*. Oct. 2022. DOI: [10.5281/zenodo.7254221](https://doi.org/10.5281/zenodo.7254221).
- [28] C. Hütt, W. Koppe, Y. Miao, and G. Bareth. "Best Accuracy Land Use/Land Cover (LULC) Classification to Derive Crop Types Using Multitemporal, Multisensor, and Multi-Polarization SAR Satellite Images". In: *Remote Sensing* 8.8 (Aug. 2016), p. 684. DOI: [10.3390/rs8080684](https://doi.org/10.3390/rs8080684).
- [29] N. Clerici, C. A. V. Calderón, and J. M. Posada. "Fusion of Sentinel-1A and Sentinel-2A Data for Land Cover Mapping: A Case Study in the Lower Magdalena Region, Colombia". In: *Journal of Maps* 13.2 (Nov. 2017), pp. 718–726. ISSN: null. DOI: [10.1080/17445647.2017.1372316](https://doi.org/10.1080/17445647.2017.1372316).
- [30] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data". In: *IEEE Geoscience and Remote Sensing Letters* 14.5 (May 2017), pp. 778–782. ISSN: 1545-598X. DOI: [10.1109/LGRS.2017.2681128](https://doi.org/10.1109/LGRS.2017.2681128).
- [31] E. Ndikumana, D. Ho Tong Minh, N. Baghdadi, D. Courault, and L. Hossard. "Deep Recurrent Neural Network for Agricultural Classification Using Multitemporal SAR Sentinel-1 for Camargue, France". In: *Remote Sensing* 10.8 (Aug. 2018), p. 1217. DOI: [10.3390/rs10081217](https://doi.org/10.3390/rs10081217).
- [32] R. A. Molijn, L. Iannini, P. López Dekker, P. S. G. Magalhães, and R. F. Hanssen. "Vegetation Characterization through the Use of Precipitation-Affected SAR Signals". In: *Remote Sensing* 10.10 (Oct. 2018), p. 1647. DOI: [10.3390/rs10101647](https://doi.org/10.3390/rs10101647).
- [33] T. Tamm, K. Zalite, K. Voormansik, and L. Talgre. "Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands". In: *Remote Sensing* 8.10 (Oct. 2016), p. 802. DOI: [10.3390/rs8100802](https://doi.org/10.3390/rs8100802).
- [34] K. Zalite, O. Antropov, J. Praks, K. Voormansik, and M. Noorma. "Monitoring of Agricultural Grasslands With Time Series of X-Band Repeat-Pass Interferometric SAR". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.8 (Aug. 2016), pp. 3687–3697. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2015.2478120](https://doi.org/10.1109/JSTARS.2015.2478120).
- [35] O. Chiboub, A. Kallel, P.-L. Frison, and M. Lopes. "Monitoring of Grasslands Management Practices Using Interferometric Products Sentinel-1". In: *Advances in Remote Sensing and Geo Informatics Applications*. Ed. by H. M. El-Askary, S. Lee, E. Heggy, and B. Pradhan. Advances in Science, Technology & Innovation. Springer International Publishing, 2019, pp. 239–242. ISBN: 978-3-030-01440-7.

- [36] R. Nasirzadehdizaji, Z. Cakir, F. Balik Sanli, S. Abdikan, A. Pepe, and F. Calò. “Sentinel-1 Interferometric Coherence and Backscattering Analysis for Crop Monitoring”. In: *Computers and Electronics in Agriculture* 185 (June 2021), p. 106118. ISSN: 0168-1699. DOI: [10.1016/j.compag.2021.106118](https://doi.org/10.1016/j.compag.2021.106118).
- [37] A. Villarroya-Carpio, J. M. Lopez-Sanchez, and M. E. Engdahl. “Sentinel-1 Interferometric Coherence as a Vegetation Index for Agriculture”. In: *Remote Sensing of Environment* 280 (Oct. 2022), p. 113208. ISSN: 0034-4257. DOI: [10.1016/j.rse.2022.113208](https://doi.org/10.1016/j.rse.2022.113208).
- [38] A. Villarroya-Carpio and J. M. Lopez-Sanchez. “Multi-Annual Evaluation of Time Series of Sentinel-1 Interferometric Coherence as a Tool for Crop Monitoring”. In: *Sensors* 23.4 (Jan. 2023), p. 1833. ISSN: 1424-8220. DOI: [10.3390/s23041833](https://doi.org/10.3390/s23041833).
- [39] H. McNairn and B. Brisco. “The Application of C-band Polarimetric SAR for Agriculture: A Review”. In: *Canadian Journal of Remote Sensing* (Jan. 2004). DOI: [10.5589/m03-069](https://doi.org/10.5589/m03-069).
- [40] H. McNairn and J. Shang. “A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring”. In: *Multitemporal Remote Sensing: Methods and Applications*. Ed. by Y. Ban. Remote Sensing and Digital Image Processing. Cham: Springer International Publishing, 2016, pp. 317–340. ISBN: 978-3-319-47037-5. DOI: [10.1007/978-3-319-47037-5_15](https://doi.org/10.1007/978-3-319-47037-5_15).
- [41] C. Pathe, W. Wagner, D. Sabel, M. Doubkova, and J. B. Basara. “Using ENVISAT ASAR Global Mode Data for Surface Soil Moisture Retrieval Over Oklahoma, USA”. In: *IEEE Transactions on Geoscience and Remote Sensing* 47.2 (Feb. 2009), pp. 468–480. ISSN: 1558-0644. DOI: [10.1109/TGRS.2008.2004711](https://doi.org/10.1109/TGRS.2008.2004711).
- [42] B. Bauer-Marschallinger, V. Freeman, S. Cao, C. Paulik, S. Schaufler, T. Stachl, S. Modanesi, C. Massari, L. Ciabatta, L. Brocca, and W. Wagner. “Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.1 (Jan. 2019), pp. 520–539. ISSN: 1558-0644. DOI: [10.1109/TGRS.2018.2858004](https://doi.org/10.1109/TGRS.2018.2858004).
- [43] M. El Hajj, N. Baghdadi, M. Zribi, and H. Bazzi. “Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas”. In: *Remote Sensing* 9.12 (Dec. 2017), p. 1292. ISSN: 2072-4292. DOI: [10.3390/rs9121292](https://doi.org/10.3390/rs9121292).
- [44] B. A. M. Bouman, D. W. G. van Kraalingen, W. Stol, and H. J. C. van Leeuwen. “An Agroecological Modeling Approach to Explain ERS SAR Radar Backscatter of Agricultural Crops”. In: *Remote Sensing of Environment* 67.2 (Feb. 1999), pp. 137–146. ISSN: 0034-4257. DOI: [10.1016/S0034-4257\(98\)00079-0](https://doi.org/10.1016/S0034-4257(98)00079-0).
- [45] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. M. De Lannoy. “Optimisation of AquaCrop Backscatter Simulations Using Sentinel-1 Observations”. In: *Remote Sensing of Environment* 294 (Aug. 2023), p. 113621. ISSN: 0034-4257. DOI: [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621).

- [46] G. Hoogenboom. “Contribution of Agrometeorology to the Simulation of Crop Production and Its Applications”. In: *Agricultural and Forest Meteorology* 103.1 (June 2000), pp. 137–157. ISSN: 0168-1923. DOI: [10.1016/S0168-1923\(00\)00108-8](https://doi.org/10.1016/S0168-1923(00)00108-8).
- [47] J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. “The DSSAT Cropping System Model”. In: *European Journal of Agronomy. Modelling Cropping Systems: Science, Software and Applications* 18.3 (Jan. 2003), pp. 235–265. ISSN: 1161-0301. DOI: [10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7).
- [48] G. Hoogenboom, C. H. Porter, K. J. Boote, V. Shelia, P. W. Wilkens, U. Singh, J. W. White, S. Asseng, J. I. Lizaso, and L. P. Moreno. “The DSSAT Crop Modeling Ecosystem”. In: *Advances in Crop Modelling for a Sustainable Agriculture*. Burleigh Dodds Science Publishing, 2019. ISBN: 978-0-429-26659-1.
- [49] A. V. M. Ines, N. N. Das, J. W. Hansen, and E. G. Njoku. “Assimilation of Remotely Sensed Soil Moisture and Vegetation with a Crop Simulation Model for Maize Yield Prediction”. In: *Remote Sensing of Environment* 138 (Nov. 2013), pp. 149–164. ISSN: 0034-4257. DOI: [10.1016/j.rse.2013.07.018](https://doi.org/10.1016/j.rse.2013.07.018).
- [50] Y. Xie, P. Wang, X. Bai, J. Khan, S. Zhang, L. Li, and L. Wang. “Assimilation of the Leaf Area Index and Vegetation Temperature Condition Index for Winter Wheat Yield Estimation Using Landsat Imagery and the CERES-Wheat Model”. In: *Agricultural and Forest Meteorology* 246 (Nov. 2017), pp. 194–206. ISSN: 0168-1923. DOI: [10.1016/j.agrformet.2017.06.015](https://doi.org/10.1016/j.agrformet.2017.06.015).
- [51] W. Zhuo, J. Huang, L. Li, X. Zhang, H. Ma, X. Gao, H. Huang, B. Xu, and X. Xiao. “Assimilating Soil Moisture Retrieved from Sentinel-1 and Sentinel-2 Data into WOFOST Model to Improve Winter Wheat Yield Estimation”. In: *Remote Sensing* 11.13 (Jan. 2019), p. 1618. ISSN: 2072-4292. DOI: [10.3390/rs11131618](https://doi.org/10.3390/rs11131618).
- [52] X. Jin, L. Kumar, Z. Li, H. Feng, X. Xu, G. Yang, and J. Wang. “A Review of Data Assimilation of Remote Sensing and Crop Models”. In: *European Journal of Agronomy* 92 (Jan. 2018), pp. 141–152. ISSN: 1161-0301. DOI: [10.1016/j.eja.2017.11.002](https://doi.org/10.1016/j.eja.2017.11.002).
- [53] J. Wang, Y. Wang, and Z. Qi. “Remote Sensing Data Assimilation in Crop Growth Modeling from an Agricultural Perspective: New Insights on Challenges and Prospects”. In: *Agronomy* 14.9 (Sept. 2024), p. 1920. ISSN: 2073-4395. DOI: [10.3390/agronomy14091920](https://doi.org/10.3390/agronomy14091920).
- [54] H. Lievens, B. Martens, N. E. C. Verhoest, S. Hahn, R. H. Reichle, and D. G. Miralles. “Assimilation of Global Radar Backscatter and Radiometer Brightness Temperature Observations to Improve Soil Moisture and Land Evaporation Estimates”. In: *Remote Sensing of Environment* 189 (Feb. 2017), pp. 194–210. ISSN: 0034-4257. DOI: [10.1016/j.rse.2016.11.022](https://doi.org/10.1016/j.rse.2016.11.022).

- [55] D. Rains, H. Lievens, G. J. M. De Lannoy, M. F. McCabe, R. A. M. de Jeu, and D. G. Miralles. "Sentinel-1 Backscatter Assimilation Using Support Vector Regression or the Water Cloud Model at European Soil Moisture Sites". In: *IEEE Geoscience and Remote Sensing Letters* (2021), pp. 1–5. ISSN: 1558-0571. DOI: [10.1109/LGRS.2021.3073484](https://doi.org/10.1109/LGRS.2021.3073484).
- [56] G. J. M. D. Lannoy and R. H. Reichle. "Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation". In: *Journal of Hydrometeorology* 17.2 (Feb. 2016), pp. 669–691. ISSN: 1525-7541, 1525-755X. DOI: [10.1175/JHM-D-15-0037.1](https://doi.org/10.1175/JHM-D-15-0037.1).
- [57] E. P. W. Attema and F. T. Ulaby. "Vegetation Modeled as a Water Cloud". In: *Radio Science* 13.2 (Mar. 1978), pp. 357–364. ISSN: 1944-799X. DOI: [10.1029/RS013i002p00357](https://doi.org/10.1029/RS013i002p00357).
- [58] M. Bracaglia, P. Ferrazzoli, and L. Guerriero. "A Fully Polarimetric Multiple Scattering Model for Crops". In: *Remote Sensing of Environment* 54.3 (Dec. 1995), pp. 170–179. ISSN: 0034-4257. DOI: [10.1016/0034-4257\(95\)00151-4](https://doi.org/10.1016/0034-4257(95)00151-4).
- [59] B. A. Forman, R. H. Reichle, and C. Derksen. "Estimating Passive Microwave Brightness Temperature Over Snow-Covered Land in North America Using a Land Surface Model and an Artificial Neural Network". In: *IEEE Transactions on Geoscience and Remote Sensing* 52.1 (Jan. 2014), pp. 235–248. ISSN: 1558-0644. DOI: [10.1109/TGRS.2013.2237913](https://doi.org/10.1109/TGRS.2013.2237913).
- [60] X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. "Towards Constraining Soil and Vegetation Dynamics in Land Surface Models: Modeling ASCAT Backscatter Incidence-Angle Dependence with a Deep Neural Network". In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. ISSN: 0034-4257. DOI: [10.1016/j.rse.2022.113116](https://doi.org/10.1016/j.rse.2022.113116).
- [61] S. M. Ghosh and M. D. Behera. "Aboveground Biomass Estimates of Tropical Mangrove Forest Using Sentinel-1 SAR Coherence Data - The Superiority of Deep Learning over a Semi-Empirical Model". In: *Computers & Geosciences* 150 (May 2021), p. 104737. ISSN: 0098-3004. DOI: [10.1016/j.cageo.2021.104737](https://doi.org/10.1016/j.cageo.2021.104737).
- [62] B. Thapa, S. Lovell, and J. Wilson. "Remote Sensing and Machine Learning Applications for Aboveground Biomass Estimation in Agroforestry Systems: A Review". In: *Agroforestry Systems* 97.6 (Aug. 2023), pp. 1097–1111. ISSN: 1572-9680. DOI: [10.1007/s10457-023-00850-2](https://doi.org/10.1007/s10457-023-00850-2).
- [63] S. S. Virnodkar, V. K. Pachghare, V. C. Patil, and S. K. Jha. "Remote Sensing and Machine Learning for Crop Water Stress Determination in Various Crops: A Critical Review". In: *Precision Agriculture* 21.5 (Oct. 2020), pp. 1121–1155. ISSN: 1573-1618. DOI: [10.1007/s11119-020-09711-9](https://doi.org/10.1007/s11119-020-09711-9).
- [64] E. Panwar, A. N. J. Kukunuri, D. Singh, A. K. Sharma, and H. Kumar. "An Efficient Machine Learning Enabled Non-Destructive Technique for Remote Monitoring of Sugarcane Crop Health". In: *IEEE Access* 10 (2022), pp. 75956–75970. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2022.3190716](https://doi.org/10.1109/ACCESS.2022.3190716).

- [65] M. Alami Machichi, I. E. Mansouri, Y. Imani, O. Bourja, O. Lahlou, Y. Zennayi, F. Bourzeix, I. Hanadé Houmma, and R. Hadria. “Crop Mapping Using Supervised Machine Learning and Deep Learning: A Systematic Literature Review”. In: *International Journal of Remote Sensing* 44.8 (Apr. 2023), pp. 2717–2753. ISSN: 0143-1161. DOI: [10.1080/01431161.2023.2205984](https://doi.org/10.1080/01431161.2023.2205984).
- [66] T. van Klompenburg, A. Kassahun, and C. Catal. “Crop Yield Prediction Using Machine Learning: A Systematic Literature Review”. In: *Computers and Electronics in Agriculture* 177 (Oct. 2020), p. 105709. ISSN: 0168-1699. DOI: [10.1016/j.compag.2020.105709](https://doi.org/10.1016/j.compag.2020.105709).
- [67] A. E. Maxwell, T. A. Warner, and F. Fang. “Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review”. In: *International Journal of Remote Sensing* 39.9 (May 2018), pp. 2784–2817. ISSN: 0143-1161. DOI: [10.1080/01431161.2018.1433343](https://doi.org/10.1080/01431161.2018.1433343).
- [68] A. Orynbaikyzy, U. Gessner, and C. Conrad. “Crop Type Classification Using a Combination of Optical and Radar Remote Sensing Data: A Review”. In: *International Journal of Remote Sensing* 40.17 (Sept. 2019), pp. 6553–6595. ISSN: 0143-1161. DOI: [10.1080/01431161.2019.1569791](https://doi.org/10.1080/01431161.2019.1569791).
- [69] F. Reuß, I. Greimeister-Pfeil, M. Vreugdenhil, and W. Wagner. “Comparison of Long Short-Term Memory Networks and Random Forest for Sentinel-1 Time Series Based Large Scale Crop Classification”. In: *Remote Sensing* 13.24 (Jan. 2021), p. 5000. ISSN: 2072-4292. DOI: [10.3390/rs13245000](https://doi.org/10.3390/rs13245000).
- [70] G. J. M. De Lannoy, M. Bechtold, C. Albergel, L. Brocca, J.-C. Calvet, A. Carrassi, W. T. Crow, P. de Rosnay, M. Durand, B. Forman, G. Geppert, M. Grotto, H.-J. Hendricks Franssen, T. Jonas, S. Kumar, H. Lievens, Y. Lu, C. Massari, V. R. N. Pauwels, R. H. Reichle, and S. Steele-Dunne. “Perspective on Satellite-Based Land Data Assimilation to Estimate Water Cycle Components in an Era of Advanced Data Availability and Model Sophistication”. In: *Frontiers in Water* 4 (2022). ISSN: 2624-9375.
- [71] A. Karpatne, I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar. “Machine Learning for the Geosciences: Challenges and Opportunities”. In: *IEEE Transactions on Knowledge and Data Engineering* 31.8 (Aug. 2019), pp. 1544–1554. ISSN: 1558-2191. DOI: [10.1109/TKDE.2018.2861006](https://doi.org/10.1109/TKDE.2018.2861006).
- [72] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. “Physics-Informed Machine Learning”. In: *Nature Reviews Physics* 3.6 (June 2021), pp. 422–440. ISSN: 2522-5820. DOI: [10.1038/s42254-021-00314-5](https://doi.org/10.1038/s42254-021-00314-5).
- [73] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar. “Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data”. In: *IEEE Transactions on Knowledge and Data Engineering* 29.10 (Oct. 2017), pp. 2318–2331. ISSN: 1558-2191. DOI: [10.1109/TKDE.2017.2720168](https://doi.org/10.1109/TKDE.2017.2720168).

- [74] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. “Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems”. In: *ACM Comput. Surv.* 55.4 (Nov. 2022), 66:1–66:37. ISSN: 0360-0300. DOI: [10.1145/3514228](https://doi.org/10.1145/3514228).
- [75] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar. *Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling*. Sept. 2021. DOI: [10.48550/arXiv.1710.11431](https://doi.org/10.48550/arXiv.1710.11431). arXiv: [1710.11431](https://arxiv.org/abs/1710.11431) [physics, stat].
- [76] Z. Zhang, Q. Pan, Z. Yang, and X. Yang. “Physics-Informed Deep Learning Method for Predicting Tunnelling-Induced Ground Deformations”. In: *Acta Geotechnica* 18.9 (Sept. 2023), pp. 4957–4972. ISSN: 1861-1133. DOI: [10.1007/s11440-023-01874-9](https://doi.org/10.1007/s11440-023-01874-9).
- [77] K. Kashinath, M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmailzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and n. Prabhat. “Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194 (Feb. 2021), p. 20200093. DOI: [10.1098/rsta.2020.0093](https://doi.org/10.1098/rsta.2020.0093).

2

ON THE VALUE OF SENTINEL-1 INSAR COHERENCE TIME-SERIES FOR VEGETATION CLASSIFICATION

SAR acquisitions are mainly deemed suitable for mapping dynamic land-cover and land-use scenarios due to their timeliness and reliability. This particularly applies to Sentinel-1 imagery. Nevertheless, the accurate mapping of regions characterized by a mixture of crops and grasses can still represent a challenge. Radar time-series have to date mainly been exploited through backscatter intensities, whereas only fewer contributions have focused on analyzing the potential of interferometric information, intuitively enhanced by the short revisit. In this study, we evaluate, as primary objective, the added value of short-temporal baseline coherences over a complex agricultural area in the São Paulo state, cultivated with heterogeneously (asynchronously) managed annual crops, grasses for pasture and sugarcane plantations. We also investigated the sensitivity of the radar information to the classification methods as well as to the data preparation and sampling practices. Two supervised machine learning methods—namely support vector machine (SVM) and random forest (RF)—were applied to the Sentinel-1 time-series at the pixel and field levels. The results highlight that an improvement of 10 percentage points (p.p.) in the classification accuracy can be achieved by using the coherence in addition to the backscatter intensity and by combining co-polarized (VV) and cross-polarized (VH) information. It is shown that the largest contribution in class discrimination is brought during winter, when dry vegetation and bare soils can be expected. One of the added values of coherence was indeed identified in the enhanced sensitivity to harvest events in a small but significant number of cases.

Parts of this chapter have been published in: Nikaein, T., Iannini, L., Molijn, R.A. and Lopez-Dekker, P., 2021. On the Value of Sentinel-1 InSAR Coherence Time-Series for Vegetation Classification. *Remote Sensing*, 13(16), p.3300.

2.1. INTRODUCTION

Space-borne radar sensors are deemed to play an important role in agriculture and land cover monitoring, mainly due to their potential to provide images independently of the weather and solar illumination conditions, but also for their complementary sensitivity to physical retrievables with respect to optical sensors. The use of SAR data to discriminate different land cover types was already demonstrated using ERS-1/2 data [1, 2]. The Sentinel-1 mission [3] provided, for the first time, dense systematic time-series of radar scattering and interferometric coherences in C-Band and dual-polarization (VV and VH) with a repeat-sampling interval of 6 or 12 days. Its interferometric wide (IW) swath mode provides data with swath widths of approximately 250 km at 5 m by 20 m single-look spatial resolution.

Land use/land cover (LULC) mapping using SAR data is commonly implemented using data-driven methods [4–7] which do not require the statistical modeling of the land cover signatures and of their patterns in time, often characterized by significant complexities. Data-driven methods, such as random forest (RF), support vector machine (SVM) and neural network classifiers, can account for underlying relationships between features in dense data series in a cost- and performance-effective way. Dense time series are known to be the key for reliable mapping as they enable the exploitation of the dissimilarities in the signature of different LULC classes during specific days of the year, particularly useful for vegetated classes with dynamic phenology such as crops [8].

Most of the land cover mapping studies exploit SAR intensity. However, the complex-valued correlation coefficients between SAR images, i.e., the interferometric coherence, also provide information about the land cover classes' characteristics [9]. The time-series of coherence images can provide information about events, such as mowing events [10, 11], which can serve as the *smoking gun* that distinguishes one LULC type from another.

Considering InSAR information as an input feature for land cover classification is not a new concept. Previous studies have already confirmed the potential of InSAR coherence for LULC classification, e.g., using time-series of ERS data with a one day revisit time [1, 12] and a stack of 12 days of Sentinel-1 images [13]. Single-pass interferometric coherence acquired by TanDEM-X and repeat-pass coherence from the TerraSAR-X mission was employed for crop-type mapping in [14] and [15], respectively. Furthermore, in [9], the authors showed that the temporal dynamics and spatial context of the multi-temporal InSAR coherence can enhance the performance of land cover classification.

In addition, the study by [16] utilized backscattering with estimated parameters from a temporal decorrelation model as input features for large scale land cover mapping over short-time-series of Sentinel-1 images. The study by [6] presented the LULC classification map by applying SVM and RF on different combinations of Sentinel-1 attributes including backscattering, interferometric coherence and polarimetric $H-\alpha$ decomposition attributes of two single look complex (SLC) images. It showed that the best performance was achieved by providing all the available features to an RF classifier. Moreover, mean backscatter, backscatter difference and the coherence information of two SLC Sentinel-1 images were employed to separate the water, barren, vegetation and

built-up classes through maximum likelihood classification [17]. However, despite the noteworthy efforts, the benefits of using the coherence time-series for LULC mapping have not been fully understood and exploited yet. The identification of the physical events, highlighted by the coherence, which help in discriminating between different land cover types, still represents an open investigation area.

The objective of this chapter is then to shed further light on the added value of coherence, when combined with the backscattered intensities, for mapping naturally vegetated and cultivated areas with dual-polarized (VV and VH) Sentinel-1 data time-series. The use of coherence information heavily increases the numbers of features in the classification problem. Whereas only one value per acquisition shall be considered for amplitudes, and therefore a total of N values for N acquisitions, the number of image pairs that can be potentially addressed for coherence features is $N(N-1)/2$. However, since previous efforts [13] showed that most of the sensitive information is carried in short-term coherences, only consecutive acquisitions will be used to estimate the cross-correlation between SAR images. The objective of this study is to evaluate the added value of coherence in SAR-based land cover mapping. Our aim is then to identify which kind of exploitable information, complementary to the VH and VV amplitude, the coherence can provide. Coherences are typically low, which leads to significant uncertainties in their estimates. Therefore, we also explore and compare two different approaches to estimate the coherences: one based on a standard fixed-resolution multi-looking approach, and one taking advantage of contextual information by averaging per field.

2.2. STUDY AREA AND DATA

The study area is situated near Campinas in the São Paulo state, Brazil. São Paulo has a tropical and subtropical climate with long and hot summers. The mean temperature reaches 30 °C in the warmest months with heavy rainfall. Conversely, the winter months are mostly dry. The vegetation shows lower biomass and also lower greenness due to these seasonality changes. Crop and pasture fields are commonly rain-fed. Irrigation can be occasionally applied to annual crops at the beginning of the growth cycle [18]. The reference dataset consists of the LULC information collected both from ground surveying activities in 2015 and from the visual inspection and interpretation of 2016–2017 high resolution optical imagery such as Google Earth imagery, Landsat and MODIS time series, manually digitized into polygons. The fields are grouped into five LULC classes/categories: crop, forest, pasture, sugarcane and urban. The crop class includes mostly soybean, wheat and corn. The forest class contains native and production forests. Pasture fields, used for cattle grazing, and sugarcane, with a crop cycle of 12–18 months, expand over grasslands [19]. As conveyed by Table 2.1, the sugarcane cycle typically starts between September and November, whereas annual crops are characterized by two different emerging periods throughout the year, in autumn (April–June) and in spring (November–December), and can be subject to double cropping practices. The spatial map of the reference LULC and some of the polygons' characteristics, arranged per class, are shown in Figure 2.1.

A total of 28 available Sentinel-1 (S1) Interferometric Wide (IW) acquisitions were

Table 2.1.: Distribution of the start of growth dates throughout the year (percentages are reported) for annual crops and sugarcane fields.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sugarcane	3	0	0	1	1	0	1	1	10	35	36	12
Crop	2	3	3	8	17	18	5	2	6	4	14	18

employed, covering one year between November 2016 and October 2017. Since only one satellite (Sentinel-1B) is active in IW mode over this region, the revisit time is 12 days. The study area is illuminated with an average incidence angle of 35° . The sensor has a resolution of 20 m (in the azimuth direction) \times 4.5 m (in the ground range direction). As a result, approximately 4.5 looks are available in a 20×20 m cell and 110 over a 100×100 m cell. Although the classification is only performed by using radar features, the normalized difference vegetation index (NDVI) from Landsat-8 data is also employed in this study for the visual interpretation of the Sentinel-1 behavior. The NDVI index expresses the greenness of canopies and can hence be readily related to crop cycles and plant status throughout the seasons. According to [20, 21], the NDVI can be affected by topography but in our study, for visual inspection purposes, the impact of this variable can be considered negligible. The Landsat-8 surface reflectance products, provided by USGS, have been retrieved using the Google Earth Engine (GEE). Only Landsat products with less than 40% of the total tile area covered by clouds were used in order to ensure that the collected NDVI series is only negligibly affected by atmospheric effects. Although it will not include all the cloud-free acquisitions, such an arbitrary choice is deemed a convenient compromise for our visual interpretation purposes.

2.3. METHODS

2.3.1. PRE-PROCESSING

The S1 IW acquisitions are downloaded in their SLC product format from the ESA hub. The data processing is performed using the Radar Interferometric Parallel Processing Lab (RIPPL), a TU Delft's in-house Sentinel-1 InSAR processing tool. Figure 2.2 provides an overview of the processing steps performed to obtain the interferometric coherences. Only the interferograms between the consecutive (12-days interval) image pairs are formed and subsequently geo-referenced. The backscatter intensity computation includes the radiometric calibration and terrain correction steps. SRTM-3 [22] is used as an external digital elevation model (DEM) for processing. For both backscatter and coherences, two types of outputs are produced by the final spatial averaging step: one by using the conventional boxcar filter for spatial multi-looking (for the pixel-based classification); and the other by averaging within the reference polygons (for object/polygon-based classification) addressed in Section 2.3.3. For the pixel-based approach, three filter sizes, i.e., 40, 100 and 200 m, were tested in order to evaluate the most convenient trade-off between the radiometric quality (as a reference, 100 independent looks would lead to a precision of approximately 0.5 dB) and the mixed

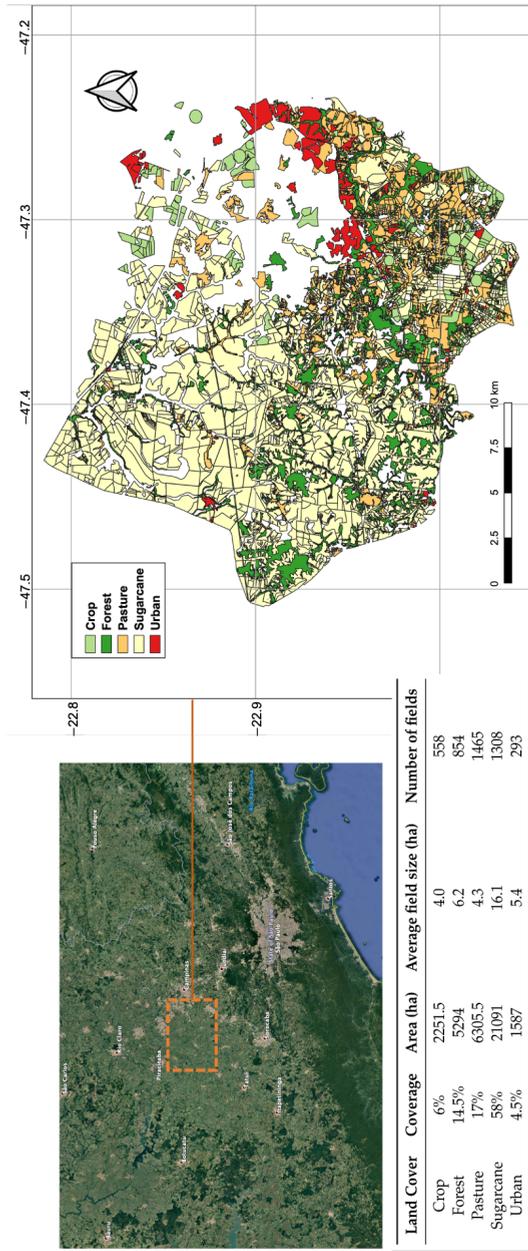


Figure 2.1.: Location of the study area (**top left**); spatial map of the reference LULC (**right**); and characteristics of the LULC based on the collected ground truth data (**bottom left**).

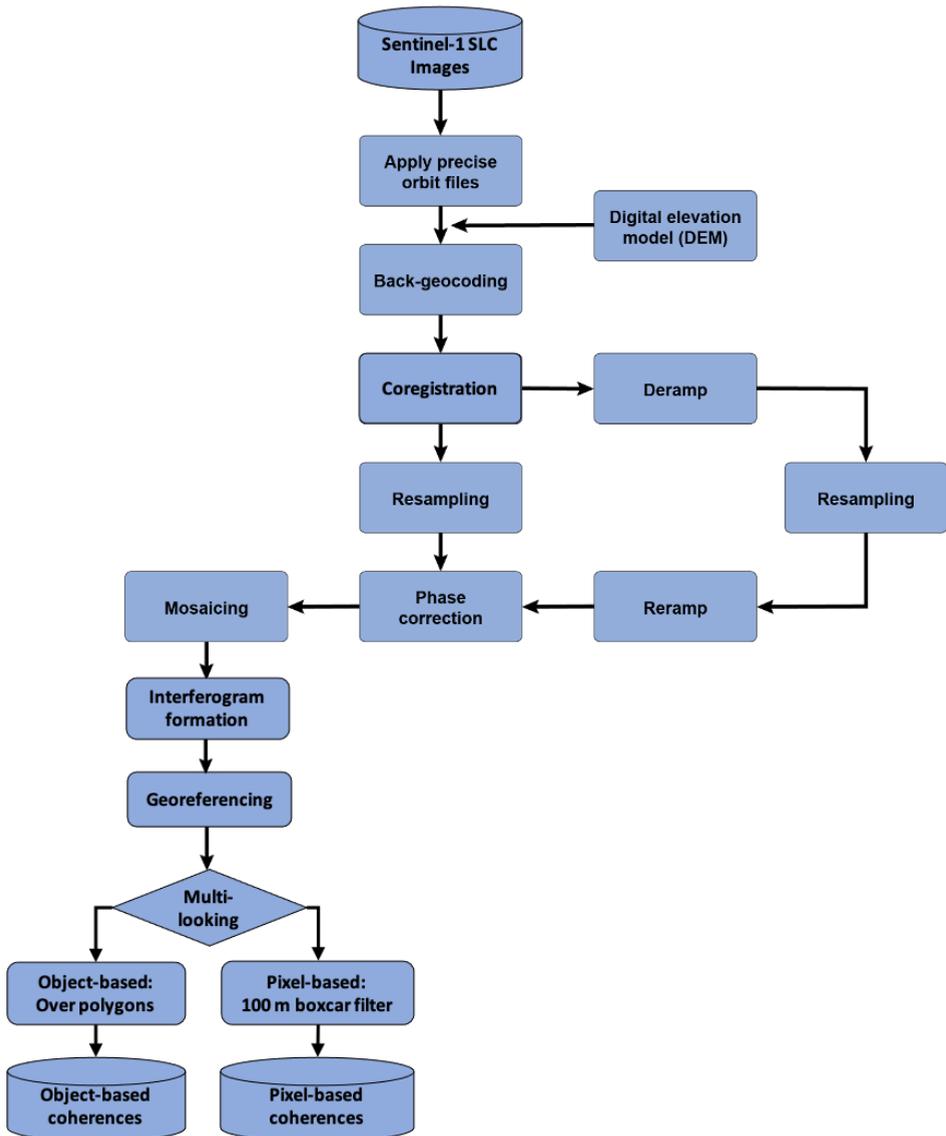


Figure 2.2.: Processing chain for coherence feature extraction from the SLC Sentinel-1 images.

field effects (most of the fields are larger than 2–3 ha, as can be seen in Figure 2.1).

Figure 2.3 shows the temporal signatures of the coherence and backscatter intensity for the vegetation classes extracted from the object-based features. The statistics strongly convey the need for multi-temporal classification as the distance between the classes, evaluated on single features per epoch, is insufficient. This is particularly true

for the coherence that presents inter-class dynamics only at the end of the dry season (June–September).

2.3.2. INTERFEROMETRIC COHERENCE

The interferometric coherence, which is commonly used as an indicator of the quality of the interferometric phase, is defined as the normalized cross-correlation between two coregistered SAR images. The absolute value of the coherence varies between 0 and 1. Here, for each new image, the coherence is computed with respect to the previous image. The coherence measures the relative stability of the scattering mechanisms within a spatial neighborhood between a pair of images. If all the backscattering elements maintain their relative position and scattering strength during the 12-day time-interval, the coherence will be high. This is typically the case for bare soil and urban areas. In contrast, if the elements move or alter their microwave signature, a low value of coherence will be the outcome. This typically occurs for vegetated surfaces with high fractional canopy cover and for water. The interferometric coherence is defined as

$$\gamma = \frac{E\{S_1 S_2^*\}}{\sqrt{E\{|S_1|^2\} E\{|S_2|^2\}}}, \quad (2.1)$$

where S_1 and S_2 represent two coregistered complex images, $E\{\cdot\}$ represents the mathematical expectation, and $*$ denotes the complex conjugate operator. Following common practice in the InSAR literature (e.g., [23]), the coherence is estimated by replacing the expectation operator by a sampled average over a given spatial window. This assumes that the signal is ergodic and locally homogeneous:

$$\hat{\gamma} = \frac{\langle S_1 S_2^* \rangle}{\sqrt{\langle |S_1|^2 \rangle \langle |S_2|^2 \rangle}}. \quad (2.2)$$

The average number of samples and the coherence map resolution have a significant impact on the coherence magnitude estimate accuracy [24]. The estimated coherence is typically noisy with a large estimation uncertainty, particularly for the small averaging windows and low coherence values.

It is worth noting that computing space-averaged coherence magnitude over entire fields improves the estimation of the coherence as the number of averaged samples increases and the quality of the estimator now depends on the field extent. Specifically, under the assumption of homogeneity, larger fields have higher coherence estimation precision and lower bias, while the smaller fields provide more bias [25]. The bias of the coherence for a homogeneous area is given by [26]

$$E(|\hat{\gamma}|) = \frac{\Gamma(L)\Gamma(1+1/2)}{\Gamma(L+1/2)} \times {}_3F_2(3/2, L, L; L+1/2, 1; |\gamma|^2) \times (1-|\gamma|^2)^L, \quad (2.3)$$

where ${}_3F_2$ is the hypergeometric function and L is the number of independent samples. Figure 2.4 shows the coherence magnitude estimate as a function of coherence magnitude ($|\gamma|$) for the number of pixels for each sugarcane field used in this study.

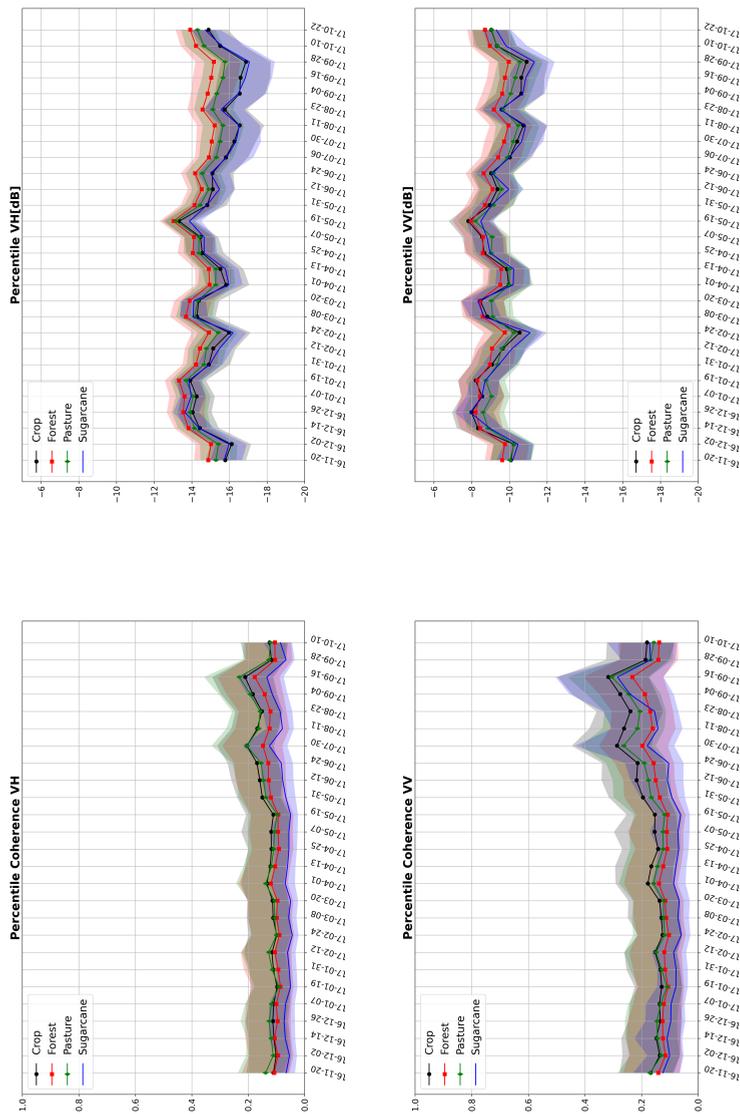


Figure 2.3.: Temporal statistics of the backscatter intensity (**top**) and coherence (**bottom**) for the vegetation classes. The 50th percentiles (solid lines) and the 20th–80th percentiles (transparent buffer) are based on the features derived from the field sampling approach.

We can see that the coherence magnitude estimate is positively biased, especially for low coherences. Bias and/or coherence estimation uncertainty reduces for fields with more independent samples.

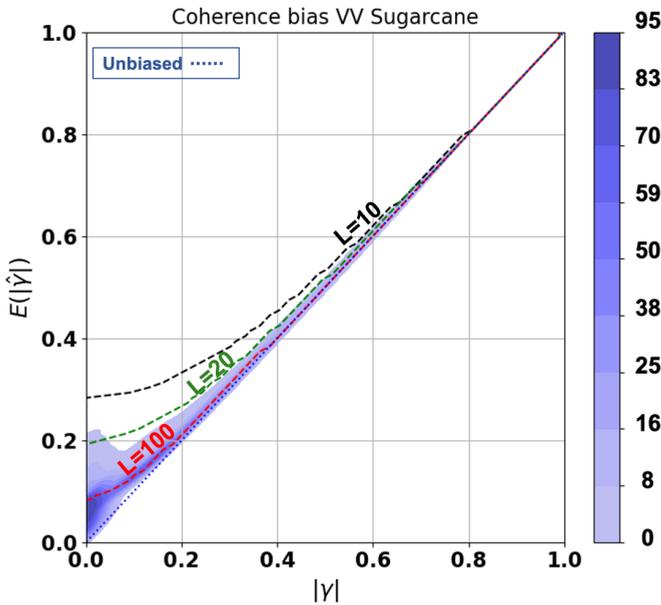


Figure 2.4.: Coherence magnitude bias of sugarcane fields with different number of samples L . The color refers to the number of fields.

2.3.3. LAND COVER CLASSIFICATION

In this study, the supervised classifications are performed at a pixel- and object-level. The pixel-based approach has a clear advantage in terms of implementation, as no further processing steps in addition to the spatial multi-looking in the earlier data preparation stages is required. In the most common case, a moving average filter with the desired spatial support is used for both amplitude and coherence. The major drawback of pixel-based approaches is that the spatial context of the scene is not fully exploited [27, 28]. This aspect would be particularly relevant over distributed scatterers, due to their intrinsic noisy nature. Identifying homogeneous segments, i.e., groups of pixels that share similar land cover, in order to average as many looks as possible, naturally leads to the so-called object-based classification.

We consider three schemes to generate training and validation sub-sets, the first two associated to the pixel-based approach and the third to the object-based approach. A visual representation of these schemes is provided in Figure 2.5. For all three schemes, a common cross-validation procedure is used to avoid positively biased results. Algorithms are therefore trained with 70% of the dataset, while the remaining

30% is used for testing. More specifically, the three addressed strategies are:

1. **Random-Pixel Sampling:** The pixel samples are randomly assigned to the training and test sets without any spatial context constraint. The outcome is that any arbitrary field is allowed to have part of its pixels in the training set and part in the test set. This is expected to lead to a positive bias in accuracy due to eventual overfitting, which occurs when the intra-polygon variability is lower than the variability between polygons of the same class. The risk in this random sampling approach is therefore that the algorithm learns the behavior of the individual fields rather than modeling their common statistical traits.
2. **Field-Pixel Sampling:** In this approach, the pixels from the same polygon are entirely assigned either to the training or the test set. For each class, the training set is built by iterative growth, i.e., by adding a field at a time to the set until 70% of the total pixels are allocated. The pixels from the excluded fields are assigned to the test set.
3. **Field Sampling:** This refers to object-based classification, as the samples correspond to the polygons themselves. The coherence magnitude and the backscatter intensity features are computed through multi-looking over the entire field. The differences with the field-pixels sampling are found in the impact of intra-field heterogeneities and in the different sensitivity to speckle noise. We are using the digitalized polygons (or objects or segments) from our ground surveying activities.

For each of these schemes, SVM [29] and RF [30] supervised classification are individually tested on the intensity and on the coherence stacks and then applied to the combination of the intensity and coherence in the two polarizations. The classification methods are implemented in Python using the scikit-learn package [31]. With concern to the SVM, the radius basis function (RBF) kernel [32] has been used. Two important parameters for the RBF kernel must be considered: the trade-off between margin and misclassification (C) and the kernel width (γ) that controls the influence of the feature data point on the decision boundary. In our study, SVM was run with $C = 1$ and $\gamma = 1$. As regards the RF, the algorithm was applied by adopting a number of trees (N_{tree}) equal to 100 and default values for the other parameters.

The accuracy assessment was carried out by analyzing the overall accuracy, the kappa index, and the producer's and user's accuracies [33]. The producer's accuracy is related to the omission error and is defined as the number of correctly classified samples divided by the total number of reference samples in the given class. The user's accuracy instead represents the commission error and is defined as the number of correctly classified samples divided by the total number of classified samples in the given class. Such well-known metrics are computed after normalizing the confusion matrix by the number of samples for each class, therefore forcing an equivalent true area for all the classes. The rationale is to prevent the accuracy being dominated by the classes with larger coverage.

To fairly compare the results of the field sampling with the two pixel sampling strategies, we consider the statistics based on the number of pixels, i.e., considering the

classified area. As a result, larger objects have more weight than smaller ones in the validation but not in the training part. This represents a reasonable evaluation practice, provided that the objective of the mapping application is to minimize the misclassified area rather than the number of objects. The steps of the methodology so far discussed are summarized in Figure 2.5.

2.3.4. FEATURE RELEVANCE

In order to assess the set of available features, we carried out a feature relevance analysis. This can help in identifying the physical processes that make a given feature useful, and lead to more robust or more optimal classification strategies. Feature relevance can be evaluated using different metrics such as the correlation between the feature and the target variable (class membership), mutual information or information gain. These metrics are independent of the classification algorithm used [34]. The most common feature selection methods are based on mutual information. However, often they do not address the correlation between features that causes feature redundancy. In this study, we adopted the minimum-redundancy–maximum-relevance (mRMR) algorithm. The algorithm searches for the subset of features (S), containing n features ($x_i, i = 1, 2, \dots, n$), that maximizes the dependency (D) of the feature set on the target class through the mean value of the mutual information:

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \quad (2.4)$$

where c is the target class and $|S|$ is the size (number of elements) of S . I is the mutual information, defined by

$$I(x; c) = \int \int p(x, c) \log \frac{p(x, c)}{p(x)p(c)} dx dc \quad (2.5)$$

with $p(x, c)$ being the joint probability density function (PDF) of the two variables and $p(x)$ and $p(c)$ standing for the corresponding marginal PDFs.

Merely maximizing D is likely to result in sets of features that are highly correlated. This redundancy can be quantified using the mean mutual information between the features within the set:

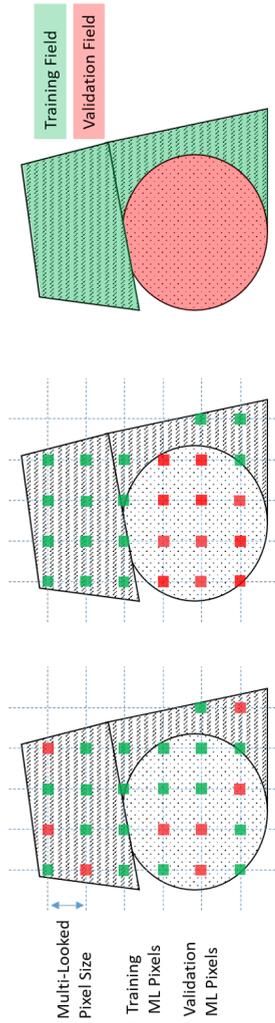
$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad (2.6)$$

which should be minimized. The mRMR algorithm combines the two mentioned constraints [35, 36] by maximizing $D - R$.

2.4. RESULTS AND DISCUSSION

2.4.1. QUANTITATIVE ACCURACY

The impact on the SAR classification performance of the three factors of interest, i.e., the sampling scheme (random pixel, field pixel, field), the classification method (SVM and RF) and the radar feature set (amplitude, coherences and their combination)



	Random Pixel Sampling	Field Pixel Sampling	Field Sampling
S1 Pre-processing	Geocoded intensities and 12-day coherences		
Feature generation	Average intensities and compute coherences on 40, 100, 200 m square windows		Average intensities and compute coherences on the pixels within the polygon perimeter
Data partitioning (training and validation sets)	Training: 70% samples, Validation: 30% samples No constraint.	Constraint: disjoint polygons for training and validation	No constraint
Classification	SVM and RF • Amplitudes • Coherences • Amplitudes & Coherences		
Accuracy assessment	Normalized accuracy on pixels		<ul style="list-style-type: none"> • Conversion to pixels • Normalized accuracy on pixels

Figure 2.5.: **(Top)** Sketched illustration of the pixel-based and field-based sampling schemes for three polygons belonging to the same class. The different patterns in the polygon background represent intra-class diversity; **(Bottom)** For each sampling scheme, the methodology steps from image pre-processing to accuracy evaluation are briefly summarized.

is herein analyzed. The overall accuracies (OA) and kappa indices for the most effective configurations in terms of performance are presented in Table 2.2. It can be observed that the SVM classification approach has better performance than the RF method. In the SVM case with random-pixel sampling, when the amplitude of only one polarization is given as an input to the classifier, the overall accuracy (OA) and kappa coefficient are approximately 60% and 0.50, respectively, for both VV and VH. When using both polarization intensities, the algorithm has an improvement of roughly 8 p.p., reaching 68% OA. The VV and VH coherences, γ_{vv}^0 , γ_{vh}^0 , add a further 2 p.p. enhancement to the overall accuracy, and a 0.03 increase in the kappa coefficient. A similar behavior is also observed in the RF case with random-pixel sampling, although with lower accuracies.

In the field-pixel sampling configurations, the accuracies are lower compared to the random-pixel sampling for both classifiers. This is indeed in line with our expectations. Such an approach is nevertheless deemed more reliable, since the chances of model overfitting are lower. In the field sampling approach, i.e., using averages based on polygons instead of averages based on pixels, the accuracy of SVM experiences a 7 p.p. increase when compared to the field-pixel approach. Still with reference to Table 2.2, we observe that the incorporation of the coherences leads to a statistically significant improvement in accuracy in all configurations. The added value of the coherence in crop mapping was also registered by TanDEM-X data covering a shorter time interval (three months), in [14], and by Sentinel-1 time-series covering a different agricultural environment in [37].

As mentioned in Section 2.3.1, the impact of the multi-looking window size for the pixel-based approaches was evaluated. The overall accuracy achieved by the three window sizes (40×40 m, 100×100 m and 200×200 m) with the SVM classifier is

Table 2.2.: The calculated overall accuracy and kappa coefficient for different features with SVM and RF

Sampling	Features	SVM		RF	
		OA	Kappa	OA	Kappa
Random pixel	σ_{VV}	0.61	0.51	0.54	0.43
	σ_{VH}	0.60	0.50	0.54	0.42
	$\sigma_{VV\&VH}$	0.68	0.60	0.59	0.49
	$\sigma_{VV\&VH} \& \gamma_{VV\&VH}$	0.70	0.63	0.62	0.52
Field pixel	σ_{VV}	0.58	0.47	0.53	0.42
	σ_{VH}	0.57	0.46	0.53	0.41
	$\sigma_{VV\&VH}$	0.64	0.55	0.58	0.47
	$\sigma_{VV\&VH} \& \gamma_{VV\&VH}$	0.68	0.60	0.60	0.50
Field	σ_{VV}	0.65	0.56	0.63	0.53
	σ_{VH}	0.57	0.46	0.55	0.44
	$\sigma_{VV\&VH}$	0.65	0.56	0.63	0.53
	$\sigma_{VV\&VH} \& \gamma_{VV\&VH}$	0.75	0.69	0.68	0.60

compared in Figure 2.6. The figure shows that when 40×40 m windows are used, the accuracies are low due to speckle noise. The accuracies for the 200×200 m windows are then lower compared to the 100×100 m configuration due to an increased amount of mixed pixel, i.e., pixels that cover two or more neighboring fields.

Although the overall accuracy gives a general understanding of a classifier's performance, it does not reveal any information about the error partition among the classes, e.g., whether some land covers are identified more correctly than others. Confusion matrices, producer and user accuracy are then used to provide more insight. As SVM performed better than RF, in the following, only the results of SVM are presented.

From the producer and user accuracies, as shown in Figure 2.7, it is straightforward to notice that the intensities perform significantly better than the coherences. By looking at their combination, it can be seen, however, that the coherence has added value for each land cover class. The largest benefits are registered by the polygon-based classification, with a fundamental 6–10 p.p. user accuracy increment for the crop and pasture classes, that are the lowest scoring classes in absolute terms.

The confusion matrices, as shown in Figure 2.8, provide a more detailed picture, also including the urban/built-in class. In each cell, the upper value (light green) corresponds to the results obtained using only the backscattered intensities, while the lower value (dark green) corresponds to the joint use of intensities and coherences. As already specified in Section 2.3.3, the columns were normalized by the number of samples of the corresponding class. With such a setting, the diagonal cells contain the producer accuracy. As expected, the dark green cells show higher values than light green ones. For the object-based classification, this difference is more apparent. The matrices also confirm that the most significant accuracy issues regard the crop omission and pasture commission errors. A relevant percentage of crop fields (>40%) is indeed classified either as pasture or as sugarcane. This is mainly due to the similarities between the seasonality behavior of the pasture and crop growth cycles. Pasture is then also receiving misclassified samples from forest and urban areas. The largest part of such errors can be justified by the broad range of vegetation typologies included in the pasture class. On the one hand, it can include shrubs and tall grasses that can be easily confused with forest or even with sugarcane when mature. On the other hand, it can include short grasses or degraded land that leads to the omission errors for urban/built-in areas. However, it cannot be excluded that a minor part could be due to errors in labeling the reference data (i.e., in the ground truth). Notice that the omission error for urban areas is lower in the field-pixel sampling than in field/polygon sampling approach. This is probably due to the fact that the urban polygons are highly heterogeneous and could contain vegetation patches. These latter would be more correctly filtered out by the pixel-based classification.

2.4.2. SPATIAL ANALYSIS

We analyze in more explicit spatial detail the output from the two most relevant feature configurations: (1) $\sigma_{vv}^0, \sigma_{vh}^0$ —only backscattering coefficients are used; and (2) $\sigma_{vv}^0, \sigma_{vh}^0, \gamma_{vv}^0, \gamma_{vh}^0$ —a combination of all SAR features is used. The classified maps are shown in Figure 2.9 for field pixel and field sampling schemes and the results related

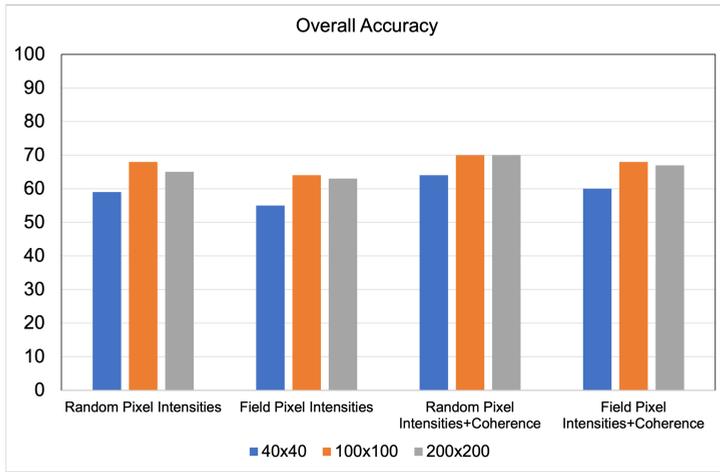


Figure 2.6.: The comparison between the impact of the different multi-looking window sizes on overall accuracy achieved by SVM in pixel sampling.

to the pixel sampling are not presented due to the fact that they include overfitting. As field sampling performs better according to the OA and confusion matrices, it is considered a reference for the remaining analysis in this chapter. The blue circles in Figure 2.9b highlight two examples of fields presenting mixed pasture and forest pixels in the pixel-based classification maps, possibly denoting spatial heterogeneities not properly accounted for within the reference polygons. It can, however, be observed that the integration of the coherence in the field-based approach allows one to correctly identify the land cover majority, i.e., pasture cover for the upper field and forest for the lower one. From a qualitative standpoint, the classified maps convey that the two approaches are characterized by a substantial agreement which manifests in the difficulty to spot a total classification mismatch on large polygons. The differences between the pixel-based and the field-based polygons shall be rather found in single pixel errors (similarly to a salt and pepper noise). It is for instant evident in the presence of yellow (sugarcane) pixels in areas where no sugarcane is expected. This issue could be partly mitigated by the application of a majority filter as a post-processing step [38]. The use of such a filter on areas with small and medium parcels (compared to the sensor resolution), such as the one shown in Figure 2.9, however, could be detrimental and its impact should be more carefully investigated in future works.

The impact of the coherence on the field-based classification is further highlighted by the differential map in Figure 2.10. The figure shows the whole area with four colors; the orange stands for those fields correctly classified only with the combined use of coherence and amplitude, where the purple represents correctly classified fields only by the exclusive use of the amplitude, and the yellow and the blue indicate the areas correctly and incorrectly classified in both configurations, respectively. The map conveys that the integration of the coherence does not only yield positive changes. Several fields are indeed correctly classified only when the single amplitude is used.

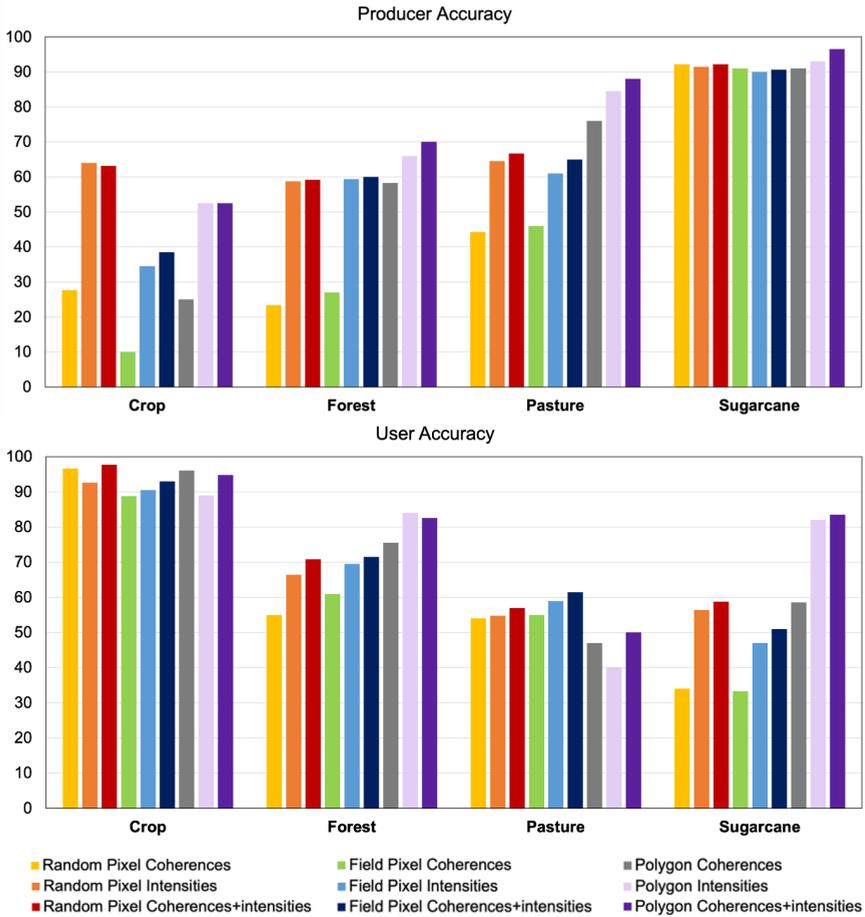


Figure 2.7.: Producer (**top**) and user (**bottom**) accuracy achieved by the SVM classification for different sampling schemes and feature configurations.

In accordance with the performance in Table 2.2, it can be therefore inferred that the coherence introduces a small but significant noise in the classification output but that its effect is overall positive.

2.4.3. FEATURE RELEVANCE ANALYSIS

To illustrate the importance of the use of InSAR coherence information in land cover classification, the time-series of coherence and backscattered intensity in both channels was analyzed for those correctly classified fields only through the use of coherence in Figure 2.11a. The radar cross-polarized backscatter (VH) is known to be sensitive to the canopy volume, whereas it is less sensitive to the soil surface. The soil backscatter is stronger on the co-polarized signal (VV), which hence becomes a better proxy for soil

		True classes				
		Crop	Forest	Pasture	Sugarcane	Urban
Predicted classes	Crop	34.5	0	1	2	0.5
		38.5	0	1	1.5	0
	Forest	2	59.5	14	4	5.5
		2.5	59.5	14.5	4	2
	Pasture	13	15.5	61	4	9
		14	14.5	65	3	9
	Sugarcane	48.5	24	21.5	90	7.5
		43	25	17	91	2.5
	Urban	1.5	0.5	2	0	77
		1.5	0.5	2	0	86.5

(a)

Figure 2.8.: *Cont.*

		True classes				
		Crop	Forest	Pasture	Sugarcane	Urban
Predicted classes	Crop	52.5	0.5	1	1	4
		52.5	0.5	0	0.5	1.5
	Forest	1	66	9.5	1	0.5
		5	70	8	1	0
	Pasture	34	30	84.5	4.5	57
		29	26.5	88	2	30
	Sugarcane	12	3	4.5	93	1
		13	2.5	3	96.5	0
	Urban	0	0	0	0	37
		0	0.5	0.5	0	68

(b)

Figure 2.8.: Confusion matrix of (a) field-pixel and (b) field sampling; with SVM classifier: each cell included two values, where upper ones are only based on using intensity while the lower values are related to using intensity and coherence together.

moisture. However, both polarizations are sensitive to the water content of the medium, either in the soil or in the canopy, showing fluctuations after watering events (rain in our studied area) that can be used to infer information on the land cover conditions, as proven in [18]. Such fluctuations are clearly visible in Figure 2.11. From the panels, it

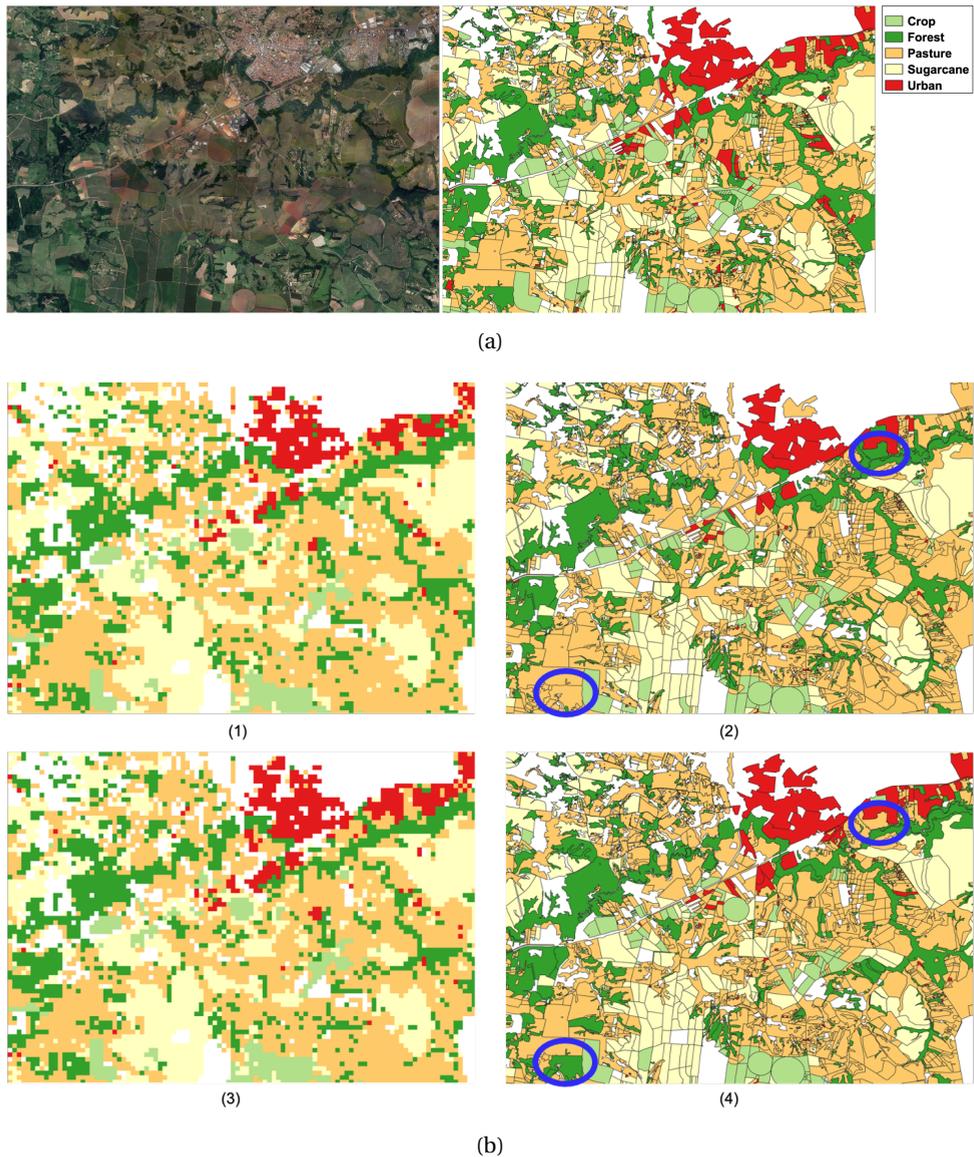


Figure 2.9.: Comparison of the SVM classifier maps over a 10×7 km close-up of the study area: (a) reference data and (b) results from: (1) amplitude and field pixel sampling; (2) amplitude and field sampling; (3) amplitude + coherence and field pixel sampling; (4) amplitude + coherence and field sampling.

is evident that the radar has a clear advantage over optical sensors in terms of temporal coverage, as the crop cycle is missed for some months by Landsat (the filled NDVI in the plot is the data that are estimated by interpolation). It can also be noticed that the

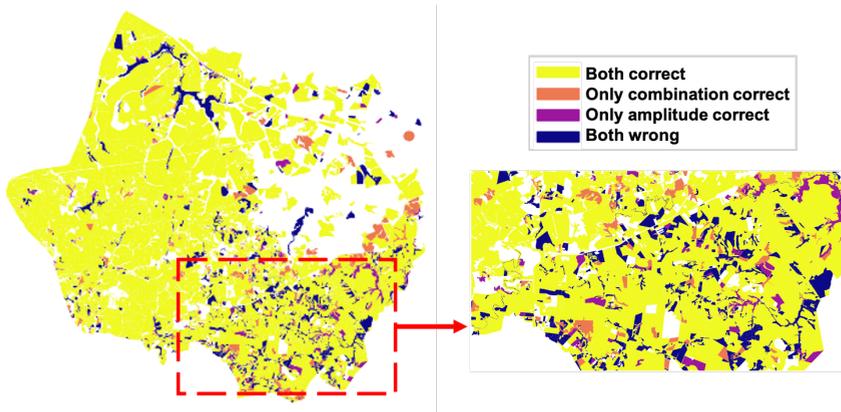


Figure 2.10.: Representation of match and mismatch, with respect to the ground truth, for the SVM-based maps obtained from intensity only and from the combination of intensities and coherence.

radar signal is noisier, although, as already specified, part of the fluctuations has to be considered as water-related signal.

This first panel in Figure 2.11 shows that the coherence is sensitive to the harvest event in a crop field at the end of 2017, which appears as a sharp increase from a near zero value to approximately 0.4. Such clear change (a large backscatter drop is expected) cannot be found in the amplitude, which probably remains high either due to the straws in the field, or to possible enhanced Bragg scattering effects (although unlikely, since only VV would be expected to suffer from it) or to high moisture in the soil. It is likely that such sensitivity in the coherence is the key factor for enabling correct labeling from the algorithm.

Figure 2.11b corresponds to a pasture field that is classified correctly only with amplitudes. It is indeed confused by the algorithm with a crop field when the coherence is integrated. In light of the previous example, the confusion is introduced by the strong spike in coherence in September 2017, which is more characteristic for crop fields in bare soil state than for pasture fields. From the NDVI value closest to the spike, amounting to approximately 0.4, it can be inferred that the vegetation has sparse, dry or underdeveloped canopy, but it is not in bare soil state. The last panel in Figure 2.11c illustrates the time-series of a crop field which classified correctly both with and without the use of the coherence. In this case, both the coherence and the amplitude behavior clearly reveal the crop cycles, with two harvest events, the first in summer (at the end of March) and the second in winter (September). Notice a similar sensitivity in summer would also be expected from the first time-series (Figure 2.11a), as two distinct cycles can be identified from the NDVI. However, due to the fact that the harvest is performed on a portion of the field at a time, the bare soil condition is never reached on the whole reference polygon.

The coherences and the intensities in the two winter months of August and September are illustrated in Figure 2.12 with the aim of conveying, qualitatively, their partial

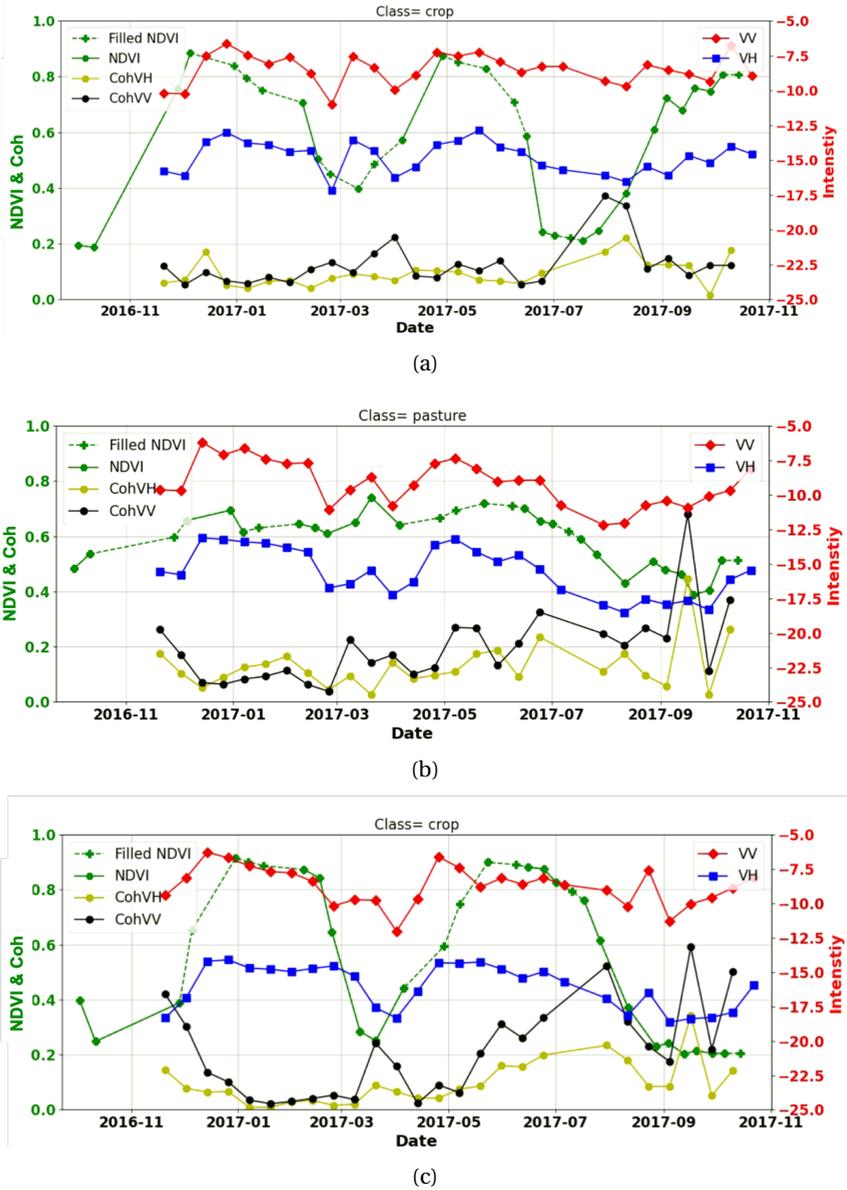


Figure 2.11.: SAR and optical time-series over different land covers: (a) for a field that is only correctly classified with the combined use of coherence and amplitude; (b) for a field that is only correctly classified by the exclusive use of the amplitude; and (c) for a field that is correctly classified with both configurations.

complementary. At the beginning of August, the VH intensities and the coherences appear in a large part inversely correlated. However, it is possible to spot a few field locations with simultaneous moderate coherence and intensity values. This conveys that a minor, although significant, number of fields in senescence and post-harvest conditions can be exclusively identified through the coherence. The different sensitivity of the two features is further confirmed by the last acquisition of August (second column in the figure), where high backscatter values are registered over the whole image (possibly due to a rain event), including bare soil areas. It is interesting to notice that the coherence is instead only marginally affected, revealing, in this particular circumstance, an improved robustness.

In order to have a deeper understanding of the added value of the feature integration, the feature selection is applied through mRMR based on the mutual information, as it was explained in Section 2.3.4. Table 2.3 reports the first four features selected by mRMR. For coherence, the date of the first SAR acquisition in the pair is reported. Notice that three out of the first four features are associated to winter acquisitions. The winter season is in fact the time of the year where the classes are most different.

Figure 2.13 illustrates the distributions of the high ranking features in Table 2.3 for the vegetation classes. The amplitude histograms in the first and in the fourth panel clearly convey that annual crops and sugarcane fields respond with lower backscatter, on average, than pasture fields. This offset is mainly due to the harvesting and ploughing operations that are often carried out on temporary and permanent crops. The two amplitude histograms similarly show that distinguishing between sugarcane and crop from a single amplitude image in winter is not possible, whereas the pasture and the forest have more distinct profiles.

The two coherence panels are associated with a dry period in summer (12–24 February) and a rain event in winter (11–23 August). The histogram of coherence in August is related to the period during which most of the annual crops and sugarcane are either harvested or in senescent conditions. As conveyed by Figure 2.3, the second acquisition in the pair (23 August) experiences an increase in amplitude that can be interpreted as the effect of rain during previous days. The historic rain data (both from weather stations and satellite) indeed confirm the occurrence of precipitations on the 19th and the 20th of August. The effect on VH is a generalized drop in the sugarcane coherence, whereas the impacts are not so evident for the other classes. For the sugarcane fields that are already harvested, this is due to the residues left on the ground, whereas for the fields in dry senescent state, the drop is caused by an increase in the canopy returns. Such a generalized response of sugarcane is less visible in the VV channel, where a significant portion of the harvested field manages to retain some coherence (see Figure 2.3). A similar behavior in the histograms can be observed for the February coherence, although with less separability between forest, pasture and crops. The reasons for the discrepancy between sugarcane and the other classes are, however, opposite in this feature, as the scene is illuminated at the end of a dry period in summer. The coherence values in summer are in fact extremely low for all classes and they incrementally rise during temporary droughts or after harvest events (for annual crop fields with double cropping management). Sugarcane fields are, however, less affected by such events since most of the fields are either in the vegetative or

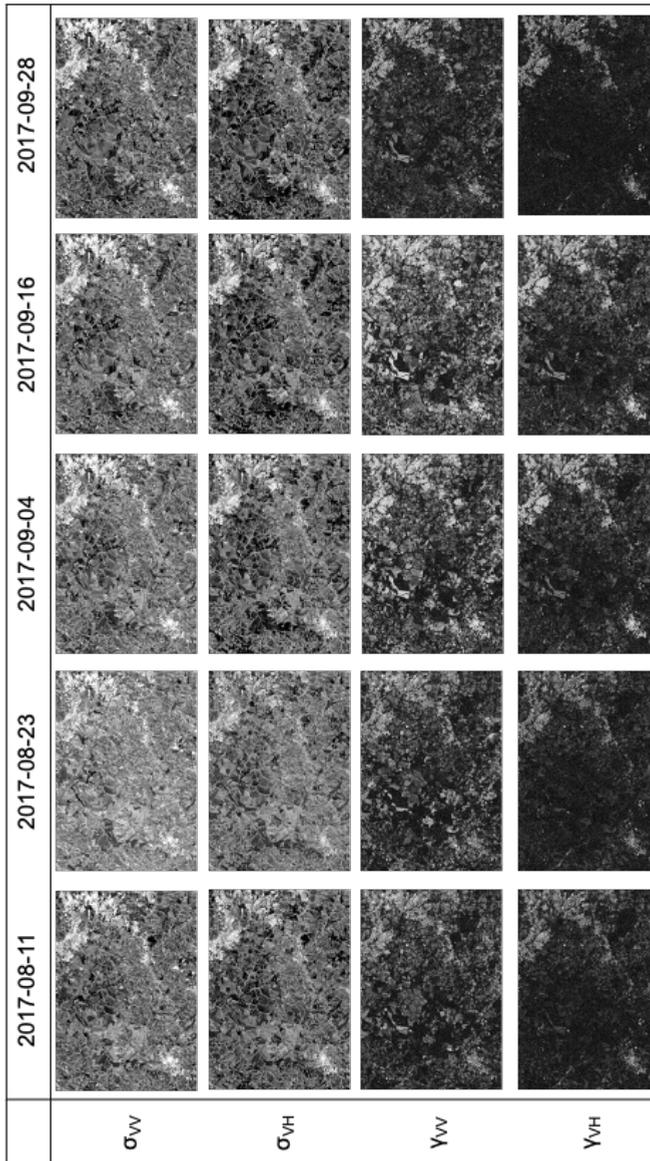


Figure 2.12.: Images of coherence and intensities in the two winter months of August and September for both polarization. The coherence colors range from 0 (black) to 1 (white). The intensity values are shown in decibels. For the VV channel, the color range is set between -15 and -5 and for the VH channel it is set between -20 and -10 . In all the images, lighter colors indicate higher values.

grand growth stages, characterized by high biomass and canopy water content [18]. The values therefore remain extremely low or null.

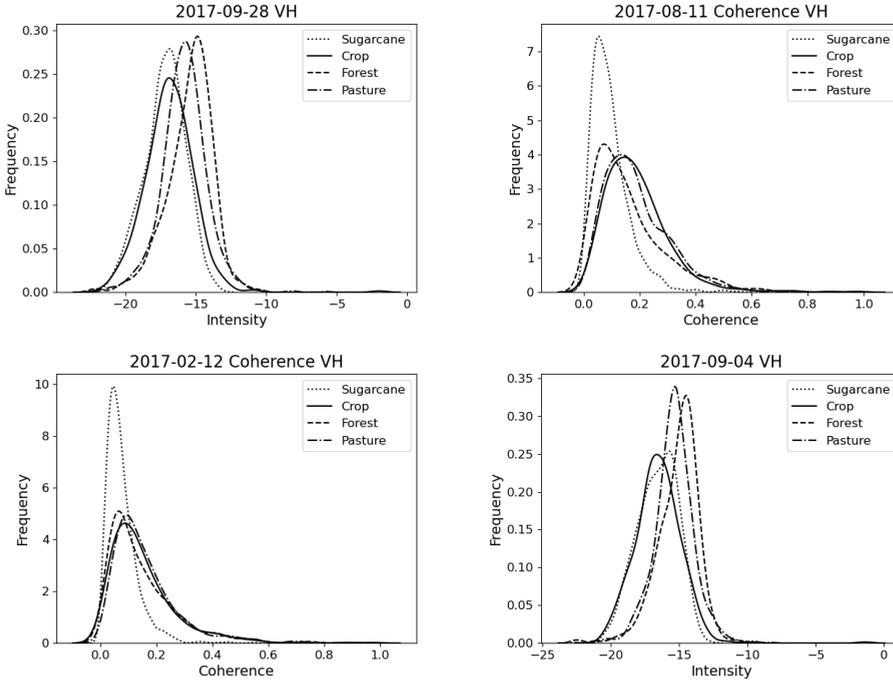


Figure 2.13.: Histograms of the vegetated classes for the four most relevant features, as can be seen in Table 2.3, as selected by the mRMR method.

Table 2.3.: First four selected features through the mutual information based mRMR.

Number	Date	Feature	Channel
1	28 September 2017	Amplitude	VH
2	11 August 2017	Coherence	VH
3	12 February 2017	Coherence	VH
4	4 September 2017	Amplitude	VH

Two further aspects are worth noting. The first is that the analysis elaborated thus far led to general principles that hold for different areas and different years, as well as to quantitative outcomes that are strongly dataset dependent. For instance, the high relevance of the winter acquisitions is easily applicable to different case studies and dataset in Brazil. The double cropping consideration for the summer/spring acquisitions also fall in this category, when considered in a broad sense. However, the score and the exact date of these features shall be intended as area-, year- and

dataset-specific and cannot be generalized. The second consideration concerns the low-class separability on single images, expressed by Figure 2.13. It is for instance not possible to distinguish between sugarcane and forest in winter, since not all the sugarcane fields are harvested at the same time. Such an issue conveys the need to use multi-temporal datasets and properly exploit the non-linear information of land cover events. Among these, the harvest event is deemed the key for the classification performance of both amplitudes and coherences.

2.5. CONCLUSIONS

In this study, the added value of short-term coherence information in discriminating vegetation land covers was evaluated for Sentinel-1 dual-polarized SAR. The work was conducted on a site characterized by native vegetation and rain-fed pasture and crops, with a critical overlap of the class signatures in every amplitude and coherence feature.

Consistently with previous work, we found that the use of InSAR coherence leads to a significant improvement of the classification performances, for example, with improvements in the user accuracy for most classes considered in the order of 5 p.p. However, contrary to the results reported in [13], for our case study, we observed that the radar intensities guarantee higher separability than coherences by themselves.

The most sensitive information brought by coherence is found in the winter months, when crops are harvested, and during the short droughts in summer. In winter, the coherence increase experienced by annual crops help in discriminating them from forest and high grass pastures. During dry summer periods, the crops and the low grass pastures are then more likely to stand out from the near-null coherences of sugarcane.

The analysis further revealed that the capability of classifiers to exploit such a marginal amount of informative interferometric pairs (for sugarcane and pasture time-series, a single non-null coherence feature can be for instance observed) can significantly vary. In our case study, it was found that SVM classifiers are more effective than RF algorithms, although the improvements are only incremental. On such a note, it shall be specified that the impact of year-specific variables, i.e., the season-dependent weather and field management practices on the results is still poorly addressed, as only one season has been processed. The events leading to the complementary coherence information has been in fact effectively outlined, however, the extent of the associated performance improvement shall be object of further assessments.

A point of attention in the use of coherence time-series is that the coherences remain generally low throughout the time-series, making its estimation unreliable if the number of samples averaged to estimate it is low. The estimation of coherence on fields brings substantial performance improvement over conventional fixed-filter multi-looking.

A final recommendation from the study concerns the generation of training and validation data sets for pixel-based classification. Our position is that it is not sufficient to select a disjoint set of pixels for training and validation, but that the pixels used for training and validation should correspond to different fields. The analysis indeed conveys that failing to do so leads to substantial positive biases.

REFERENCES

- [1] T. Strozzi, P. Dammert, U. Wegmuller, J.-M. Martinez, J. Askne, A. Beaudoin, and N. Hallikainen. “Landuse Mapping with ERS SAR Interferometry”. In: *IEEE Transactions on Geoscience and Remote Sensing* 38.2 (Mar. 2000), pp. 766–775. ISSN: 1558-0644. DOI: [10.1109/36.842005](https://doi.org/10.1109/36.842005).
- [2] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann. “An Advanced System for the Automatic Classification of Multitemporal SAR Images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 42.6 (June 2004), pp. 1321–1334. ISSN: 1558-0644. DOI: [10.1109/TGRS.2004.826821](https://doi.org/10.1109/TGRS.2004.826821).
- [3] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L’Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan. “GMES Sentinel-1 Mission”. In: *Remote Sensing of Environment. The Sentinel Missions - New Opportunities for Science* 120 (May 2012), pp. 9–24. ISSN: 0034-4257. DOI: [10.1016/j.rse.2011.05.028](https://doi.org/10.1016/j.rse.2011.05.028).
- [4] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. “An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 67 (Jan. 2012), pp. 93–104. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2011.11.002](https://doi.org/10.1016/j.isprsjprs.2011.11.002).
- [5] Y. Shao and R. S. Lunetta. “Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 70 (June 2012), pp. 78–87. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2012.04.001](https://doi.org/10.1016/j.isprsjprs.2012.04.001).
- [6] J. M. F. d. S. Diniz, F. F. Gama, and M. Adami. “Evaluation of Polarimetry and Interferometry of Sentinel-1A SAR Data for Land Use and Land Cover of the Brazilian Amazon Region”. In: *Geocarto International* 0.0 (May 2020), pp. 1–19. ISSN: 1010-6049. DOI: [10.1080/10106049.2020.1773544](https://doi.org/10.1080/10106049.2020.1773544).
- [7] M. Gašparović and D. Dobričić. “Comparative Assessment of Machine Learning Methods for Urban Vegetation Mapping Using Multitemporal Sentinel-1 Imagery”. In: *Remote Sensing* 12.12 (Jan. 2020), p. 1952. DOI: [10.3390/rs12121952](https://doi.org/10.3390/rs12121952).
- [8] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. Liu, and K. Papathanassiou. “Radar Remote Sensing of Agricultural Canopies: A Review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2249–2273. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2016.2639043](https://doi.org/10.1109/JSTARS.2016.2639043).

- [9] S. Ge, O. Antropov, W. Su, H. Gu, and J. Praks. “Deep Recurrent Neural Networks for Land-Cover Classification Using Sentinel-1 INSAR Time Series”. In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. July 2019, pp. 473–476. DOI: [10.1109/IGARSS.2019.8900088](https://doi.org/10.1109/IGARSS.2019.8900088).
- [10] K. Zalite, O. Antropov, J. Praks, K. Voormansik, and M. Noorma. “Monitoring of Agricultural Grasslands With Time Series of X-Band Repeat-Pass Interferometric SAR”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.8 (Aug. 2016), pp. 3687–3697. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2015.2478120](https://doi.org/10.1109/JSTARS.2015.2478120).
- [11] T. Tamm, K. Zalite, K. Voormansik, and L. Talgre. “Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands”. In: *Remote Sensing* 8.10 (Oct. 2016), p. 802. DOI: [10.3390/rs8100802](https://doi.org/10.3390/rs8100802).
- [12] M. Engdahl and J. Hyypä. “Land-Cover Classification Using Multitemporal ERS-1/2 InSAR Data”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.7 (July 2003), pp. 1620–1628. ISSN: 1558-0644. DOI: [10.1109/TGRS.2003.813271](https://doi.org/10.1109/TGRS.2003.813271).
- [13] A. W. Jacob, F. Vicente-Guijalba, C. Lopez-Martinez, J. M. Lopez-Sanchez, M. Litzinger, H. Kristen, A. Mestre-Quereda, D. Ziólkowski, M. Lavallo, C. Notarnicola, G. Suresh, O. Antropov, S. Ge, J. Praks, Y. Ban, E. Pottier, J. J. Mallorquí Franquet, J. Duro, and M. E. Engdahl. “Sentinel-1 InSAR Coherence for Land Cover Mapping: A Comparison of Multiple Feature-Based Classifiers”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 535–552. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2019.2958847](https://doi.org/10.1109/JSTARS.2019.2958847).
- [14] M. Busquier, J. M. Lopez-Sanchez, A. Mestre-Quereda, E. Navarro, M. P. González-Dugo, and L. Mateos. “Exploring TanDEM-X Interferometric Products for Crop-Type Mapping”. In: *Remote Sensing* 12.11 (Jan. 2020), p. 1774. DOI: [10.3390/rs12111774](https://doi.org/10.3390/rs12111774).
- [15] R. Sonobe, H. Tani, X. Wang, N. Kobayashi, and H. Shimamura. “Discrimination of Crop Types with TerraSAR-X-Derived Information”. In: *Physics and Chemistry of the Earth, Parts A/B/C. Emerging Science and Applications with Microwave Remote Sensing Data* 83–84 (Jan. 2015), pp. 2–13. ISSN: 1474-7065. DOI: [10.1016/j.pce.2014.11.001](https://doi.org/10.1016/j.pce.2014.11.001).
- [16] F. Sica, A. Pulella, M. Nannini, M. Pinheiro, and P. Rizzoli. “Repeat-Pass SAR Interferometry for Land Cover Classification: A Methodology Using Sentinel-1 Short-Time-Series”. In: *Remote Sensing of Environment* 232 (Oct. 2019), p. 111277. ISSN: 0034-4257. DOI: [10.1016/j.rse.2019.111277](https://doi.org/10.1016/j.rse.2019.111277).
- [17] R. Z. Khalil and Saad-ul-Haque. “InSAR Coherence-Based Land Cover Classification of Okara, Pakistan”. In: *The Egyptian Journal of Remote Sensing and Space Science. EJRS Special Issue: Microwave Remote Sensing in Honor of Professor Adel Yehia* 21 (July 2018), S23–S28. ISSN: 1110-9823. DOI: [10.1016/j.ejrs.2017.08.005](https://doi.org/10.1016/j.ejrs.2017.08.005).

- [18] R. A. Molijn, L. Iannini, P. López Dekker, P. S. G. Magalhães, and R. F. Hanssen. “Vegetation Characterization through the Use of Precipitation-Affected SAR Signals”. In: *Remote Sensing* 10.10 (Oct. 2018), p. 1647. DOI: [10.3390/rs10101647](https://doi.org/10.3390/rs10101647).
- [19] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen. “Author Correction: Ground Reference Data for Sugarcane Biomass Estimation in São Paulo State, Brazil”. In: *Scientific Data* 6.1 (Dec. 2019), p. 306. ISSN: 2052-4463. DOI: [10.1038/s41597-019-0309-x](https://doi.org/10.1038/s41597-019-0309-x).
- [20] B. Matsushita, W. Yang, J. Chen, Y. Onda, and G. Qiu. “Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest”. In: *Sensors* 7.11 (Nov. 2007), pp. 2636–2651. DOI: [10.3390/s7112636](https://doi.org/10.3390/s7112636).
- [21] D. Hao, J. Wen, Q. Xiao, S. Wu, X. Lin, D. You, and Y. Tang. “Modeling Anisotropic Reflectance Over Composite Sloping Terrain”. In: *IEEE Transactions on Geoscience and Remote Sensing* 56.7 (July 2018), pp. 3903–3923. ISSN: 1558-0644. DOI: [10.1109/TGRS.2018.2816015](https://doi.org/10.1109/TGRS.2018.2816015).
- [22] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf. “The Shuttle Radar Topography Mission”. In: *Reviews of Geophysics* 45.2 (2007). ISSN: 1944-9208. DOI: [10.1029/2005RG000183](https://doi.org/10.1029/2005RG000183).
- [23] M. Seymour and I. Cumming. “Maximum Likelihood Estimation for SAR Interferometry”. In: *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 4. Aug. 1994, 2272–2275 vol.4. DOI: [10.1109/IGARSS.1994.399711](https://doi.org/10.1109/IGARSS.1994.399711).
- [24] R. Touzi, A. Lopes, J. Bruniquel, and P. Vachon. “Coherence Estimation for SAR Imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* 37.1 (Jan. 1999), pp. 135–149. ISSN: 1558-0644. DOI: [10.1109/36.739146](https://doi.org/10.1109/36.739146).
- [25] C. López-Martínez and E. Pottier. “Coherence Estimation in Synthetic Aperture Radar Data Based on Speckle Noise Modeling”. In: *Applied Optics* 46.4 (Feb. 2007), pp. 544–558. ISSN: 2155-3165. DOI: [10.1364/AO.46.000544](https://doi.org/10.1364/AO.46.000544).
- [26] R. Touzi and A. Lopes. “Statistics of the Stokes Parameters and of the Complex Coherence Parameters in One-Look and Multilook Speckle Fields”. In: *IEEE Transactions on Geoscience and Remote Sensing* 34.2 (Mar. 1996), pp. 519–531. ISSN: 1558-0644. DOI: [10.1109/36.485128](https://doi.org/10.1109/36.485128).
- [27] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley. “Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 80 (June 2013), pp. 91–106. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2013.03.006](https://doi.org/10.1016/j.isprsjprs.2013.03.006).

- [28] E. A. Addink, F. M. B. Van Coillie, and S. M. De Jong. “Introduction to the GEOBIA 2010 Special Issue: From Pixels to Geographic Objects in Remote Sensing Image Analysis”. In: *International Journal of Applied Earth Observation and Geoinformation*. Special Issue on Geographic Object-Based Image Analysis: GEOBIA 15 (Apr. 2012), pp. 1–6. ISSN: 0303-2434. DOI: [10.1016/j.jag.2011.12.001](https://doi.org/10.1016/j.jag.2011.12.001).
- [29] C. Cortes and V. Vapnik. “Support-Vector Networks”. In: *Machine Learning* 20.3 (Sept. 1995), pp. 273–297. ISSN: 1573-0565. DOI: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018).
- [30] L. Breiman. “Random Forests”. In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. ISSN: 1573-0565. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau. “Scikit-Learn: Machine Learning in Python”. In: *MACHINE LEARNING IN PYTHON* (), p. 6.
- [32] T. Kavzoglu and I. Colkesen. “A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification”. In: *International Journal of Applied Earth Observation and Geoinformation* 11.5 (Oct. 2009), pp. 352–359. ISSN: 0303-2434. DOI: [10.1016/j.jag.2009.06.002](https://doi.org/10.1016/j.jag.2009.06.002).
- [33] S. V. Stehman. “Selecting and Interpreting Measures of Thematic Classification Accuracy”. In: *Remote Sensing of Environment* 62.1 (Oct. 1997), pp. 77–89. ISSN: 0034-4257. DOI: [10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7).
- [34] Hanchuan Peng, Fuhui Long, and C. Ding. “Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (Aug. 2005), pp. 1226–1238. ISSN: 1939-3539. DOI: [10.1109/TPAMI.2005.159](https://doi.org/10.1109/TPAMI.2005.159).
- [35] C. Ding and H. Peng. “Minimum Redundancy Feature Selection from Microarray Gene Expression Data”. In: *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*. Aug. 2003, pp. 523–528. DOI: [10.1109/CSB.2003.1227396](https://doi.org/10.1109/CSB.2003.1227396).
- [36] C. Ding and H. Peng. “Minimum Redundancy Feature Selection from Microarray Gene Expression Data”. In: *Journal of Bioinformatics and Computational Biology* 03.02 (Apr. 2005), pp. 185–205. ISSN: 0219-7200. DOI: [10.1142/S0219720005001004](https://doi.org/10.1142/S0219720005001004).
- [37] A. Mestre-Quereda, J. M. Lopez-Sanchez, F. Vicente-Guijalba, A. W. Jacob, and M. E. Engdahl. “Time Series of Sentinel-1 Interferometric Coherence and Backscatter for Crop-Type Mapping”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2020), pp. 1–1. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2020.3008096](https://doi.org/10.1109/JSTARS.2020.3008096).
- [38] L. E. Cué La Rosa, R. Queiroz Feitosa, P. Nigri Happ, I. Del’Arco Sanches, and G. A. Ostwald Pedro da Costa. “Combining Deep Learning and Prior Knowledge for Crop Mapping in Tropical Regions from Multitemporal SAR Image Sequences”. In: *Remote Sensing* 11.17 (Jan. 2019), p. 2029. DOI: [10.3390/rs11172029](https://doi.org/10.3390/rs11172029).

3

MODELING SAR OBSERVABLES BY COMBINING A CROP-GROWTH MODEL WITH MACHINE LEARNING

Our aim is to estimate Synthetic Aperture Radar (SAR) observables, such as backscatter in VV and VH polarizations, as well as the VH/VV ratio, cross-ratio (CR), and interferometric coherence in VV, from agricultural fields. In this study, we use the Decision Support System for Agrotechnology Transfer (DSSAT) crop growth simulation model to simulate parcel-level phenological and growth parameters for over 1500 parcels of silage maize in the Netherlands. The crop model was calibrated using field data, including silage maize phenological phases, leaf area index (LAI), and above-ground dry biomass (AGB). The simulations incorporate fine-resolution gridded precipitation data and soil parameters to model the interaction between soil-plant-atmosphere and genotype in DSSAT. The crop variables produced by DSSAT are then used as inputs to a Support Vector Regression (SVR) model. This model is trained to simulate SAR observables in 2017, 2018, and 2019, and its performance is evaluated using independent fields in each of these years. The results show a close fit between modeled and observed SAR C-band observables. The importance of vegetation variables in the estimation of SAR observables is assessed. The AGB showed significant importance in the estimation of backscatter. This study demonstrates the potential value of combining crop growth simulation models and machine learning to simulate SAR observables. For example, the SVR model developed here could be used as an observation operator in an assimilation context to constrain vegetation and soil water dynamics in a crop growth model.

Parts of this chapter have been published in: Nikaein, T., Lopez-Dekker, P., Steele-Dunne, S., Kumar, V. and Huber, M., 2023. Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

3.1. INTRODUCTION

The launch of the Sentinel-1 Synthetic Aperture Radar (SAR) constellation [1] has provided unprecedented opportunities for agricultural monitoring. This is due to its dense time series of radar scattering and interferometric coherence in C-band, which offers a short revisit time of 6 days in Europe and 12 days globally. Additionally, the dataset is freely accessible. Satellite observations from active and passive microwave sensors are sensitive to vegetation structure [2] and water content [3].

Space-borne SAR observations have been used for many years for agricultural applications and vegetation monitoring because of microwave signal sensitivity to changes during the crop growth period [4, 5]. Several studies highlighted the potential of Sentinel-1 SAR data for crop growth monitoring [6–8], crop water content estimation and soil moisture mapping [9, 10] and parameter retrieval [11]. Vreugdenhil et al. [12] demonstrated using in situ observations that C-band backscatter from Sentinel-1 is sensitive to vegetation parameters such as vegetation water content (VWC), biomass, height, and leaf area index (LAI).

In addition to monitoring dynamics directly, Sentinel-1 data could be assimilated into a crop growth model to constrain the estimates of bio-geophysical variables. The assimilation of Earth Observation (EO) data in agricultural models has been demonstrated in several studies. For example, the assimilation of LAI and soil moisture derived from Sentinel-1 and Sentinel-2 data into the World Food Studies (WOFOST) model to estimate crop yield was studied in [13]. Assimilating LAI and dry biomass from optical and SAR data into a model to estimate soybean yield was studied in [14]. In another study, [15] they assimilated LAI from SAR product to Decision Support System for Agrotechnology Transfer (DSSAT) for rice yield estimation. These studies collectively demonstrate that the assimilation of EO data or products can be used to improve yield estimates. In addition, assimilation could provide improved estimates of the growth and development of the crop to support agricultural management and decision-making.

To integrate SAR observations with modeled soil moisture and vegetation, two approaches can be employed. The first approach involves converting the SAR signal into retrievals of geophysical variables using change detection algorithms [16] or other methods [17]. However, this approach is limited by the availability of high-resolution retrievals from the C-band with global coverage. The second approach focuses on estimating and combining the backscatter components of soil moisture and vegetation to simulate the expected signal at the sensor level using a backscatter model as a forward model. This study specifically adopts the second approach, utilizing a forward operator to convert model simulations into a backscatter signal. Our aim is to assimilate Sentinel-1 observables directly rather than retrieved parameters. This approach allows us to utilize all the information related to the incidence angle dependence while avoiding any potential for cross-correlated errors between retrievals and model simulations [18]. The objective is to prepare for a data assimilation system, where SAR observations will be utilized to update crop model simulations of soil moisture and biomass [19].

To exploit Sentinel-1 backscatter and coherence for agricultural monitoring, we need to be able to model the observables given a description of the soil and vegetation.

Furthermore, it is essential to be able to understand and quantify the effect of crop bio-geophysical variables on SAR observables. The most common way to model radar observables is using a model, the complexity and requirements of which can vary considerably. The most widely used model is the Water Cloud Model (WCM). The WCM is a semi-empirical parameterized model to relate the Normalized Radar Cross Section (NRCS) to the characteristics of the vegetation and the surface. The model is relatively simple, as it models the vegetation as a collection of identical water droplets randomly distributed within the canopy [20, 21]. As a recent example, [22] used the WCM as a measurement operator to assimilate Sentinel-1 data into the Global Land Evaporation Amsterdam Model (GLEAM). More sophisticated models, such as Michigan microwave canopy scattering (MIMICS) [23], can also be used to simulate microwave observables [24]. However, they require descriptions of vegetation geometry, architecture, and dielectric properties that are seldom available. An alternative to using the WCM as a forward operator is employing machine learning techniques. Machine learning offers advantages due to its flexibility, adaptability, and ability to handle nonlinear relationships. It can extract meaningful insights directly from data and generalize well to new instances. It is important to note that they also have limitations. They may require large amounts of high-quality labeled data for training.

This study aims to provide a model capable of simulating Sentinel-1 observables that require readily-available crop descriptors. We propose circumventing the limited availability of in-situ data by using a crop growth model instead. This is a significant step towards developing a robust and reliable system for assimilating high-resolution Sentinel-1 observables over vegetation areas. Davitt et al. [24] demonstrated that DSSAT could be used to provide a description of the growing crop to be used as input to MIMICS. Here, instead, we will use this description as input to a machine-learning model to map the crop descriptors to the SAR observables. The biophysical parameters generated by the crop-growth model reflect the state of the crop at a given time, which reflects all the past inputs to the model. The radar observables depend only on the state of the crop and the soil at the time the data are acquired. Our approach, of using modeled land surface states as input to a machine-learning model has been used to map snow and land parameters to brightness temperature [25], to map land surface variables to Advanced Scatterometer (ASCAT) backscatter, slope, and curvature [26] and recently, to map daily 1- km AquaCrop model biomass and surface soil moisture to backscatter [27].

The objectives of this study are to:

- Demonstrate that a crop-growth model can be used together with a machine-learning model to simulate SAR observables to optimize the future assimilation of SAR observables into the regional DSSAT model.
- Show that the relationships between crop bio-geophysical variables and modeled SAR observables are physically plausible.
- Discuss/identify the potential limitations of applying this technique in agricultural applications.

The analysis will be conducted over 1500 maize fields in the Netherlands. The DSSAT model will be used to simulate a description of the growing vegetation in terms of LAI,

AGB, surface soil moisture (SM_S), and root zone soil moisture (SM_R). To establish the link between the observables and the biophysical parameters we use a Support Vector Regression (SVR) model. Feature analysis, the Minimum Redundancy Maximum Relevance (MRMR), will be used to examine the sensitivity of the SAR observables to the DSSAT variables.

3.2. STUDY AREA

The study is performed in the Noord-Brabant province (5081 km²) in the South of The Netherlands. Province boundaries and location of agricultural fields with their associated crop types were retrieved from the Basisregistratie Gewaspercelen (BRP) [28], an open national database of crop parcel boundaries and crop type in The Netherlands. The average temperature in this area varies between 2°C to 24°C and the average annual precipitation is around 646 mm. Mean monthly sunshine hours range from 5 to 14 hours [29].

The study area location and the spatial distribution of the maize fields are shown in Fig. 3.1. Almost all maize grown by farmers in The Netherlands is silage maize. It is planted between mid-April and the beginning of May, and the emergence is in mid-May. It is left to ripen in the field, and is harvested in September [30]. Silage maize is grown in approximately 20% (about 20000#) of parcels in Noord-Brabant. 10% of these fields were used for this study, ensuring that they were randomly spread all over the province (see Fig. 3.1).

For calibration, data were available at an experimental site in Reusel, 51.319°N, 5.173°E, in 2019 [31]. The field is on sandy soil and can be irrigated with a gun sprinkler irrigation system. In 2019, the maize was watered twice during the summer, relying on the rain during the rest of the growth period.

3.3. DATA AND METHODOLOGY

Fig. 3.2 provides an overview of the workflow followed in this study.

3.3.1. DSSAT

Crop growth models, such as the Decision Support System for Agrotechnology Transfer (DSSAT) model, simulate the interaction between plants and their environment, in daily steps, to estimate the growth, development, and yield of different crops [32, 33]. Basically, they are scientific tools with a set of equations that take environmental information as inputs and determine the phenological development, and growth stage as outputs and these require precise parameterization. DSSAT has been used for years for different crops and regions across the world [34]. The CERES-Maize module (Crop Environment Resource Synthesis) [35] is part of DSSAT v4.7, and is one of the most broadly used maize models. The CERES-Maize module has been used and evaluated in several studies in different parts of the world. For instance, in [36] DSSAT performance was validated and then used to determine the best management practices of nitrogen fertilization and irrigation. In another study, the impact of climate variability on maize in the semi-arid area by CERES-Maize module was assessed [37].

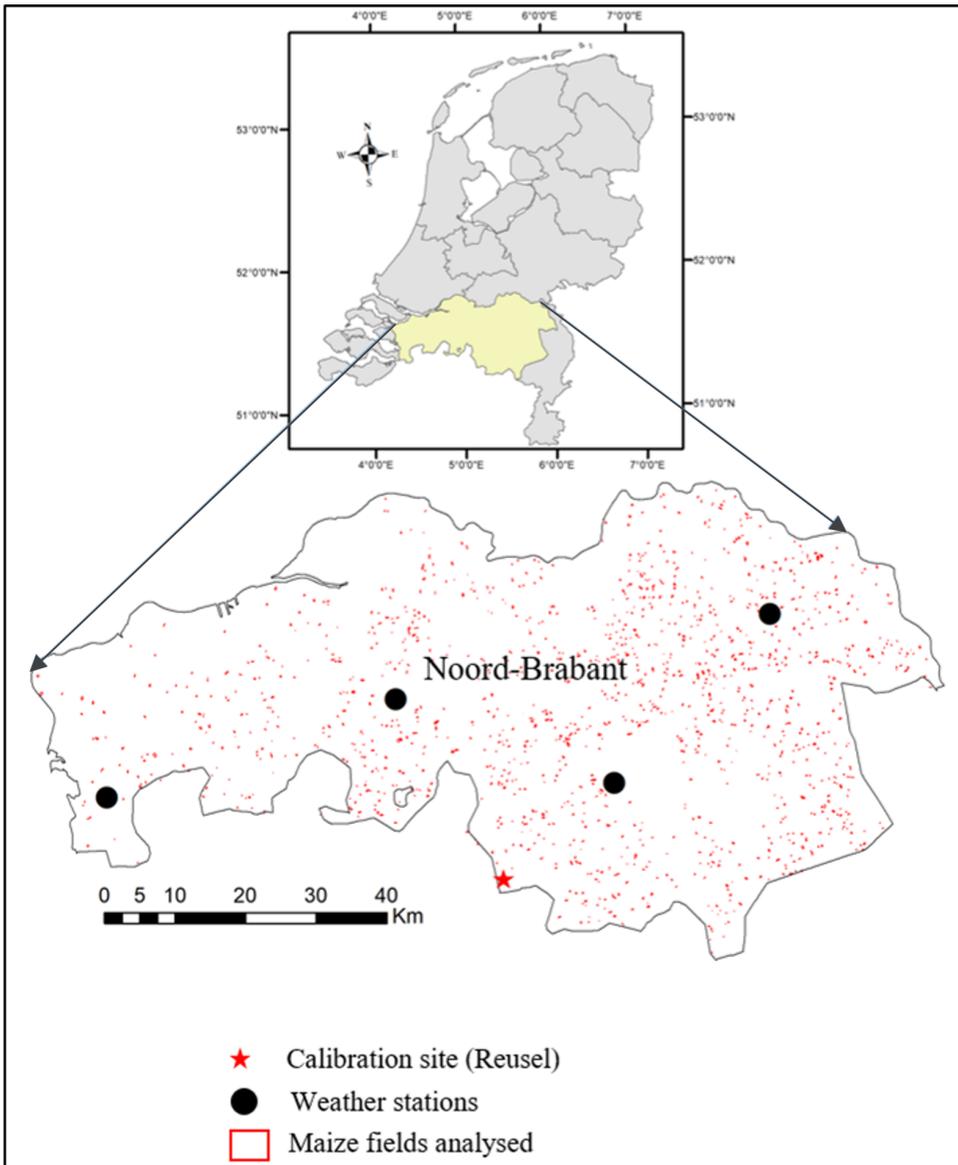


Figure 3.1.: Location of the study area. The maize fields analysed in this study are shown in red. Black circles display the location of four automatic meteorological stations over the province. The red star indicates the location of the site used for calibration.

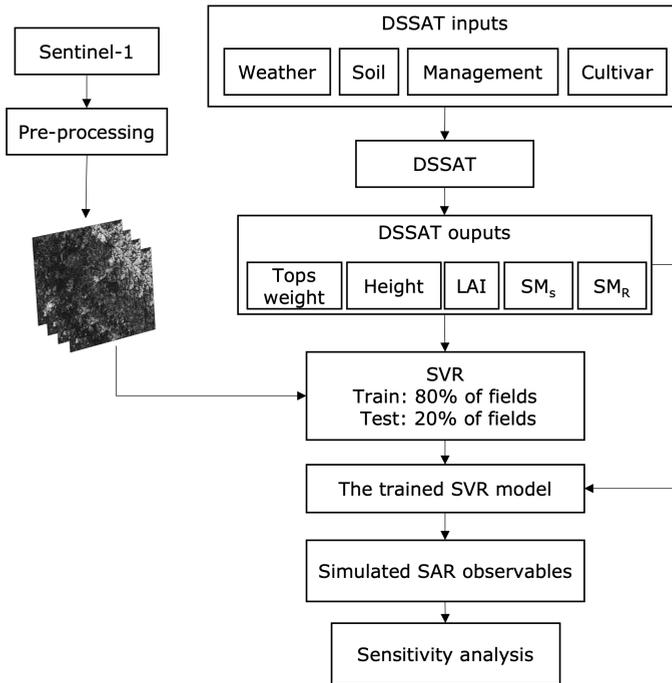


Figure 3.2.: DSSAT-SVR SAR observables simulation workflow.

3.3.2. INPUT DATA FOR DSSAT

DSSAT requires a minimum set of input data to simulate crop growth: daily weather information during the growth period; soil characteristics of the area; crop management data (sowing and harvest date, row spacing, irrigation, and fertilization information); and cultivar coefficients.

WEATHER DATA

The required weather data including daily solar radiation (SRAD), precipitation, and maximum and minimum air temperature (TMAX, TMIN), were obtained from Royal Dutch Meteorological Institute (KNMI) weather station data [29]. There are over 40 weather stations in The Netherlands, of which four are within the Noord-Brabant province. The point-wise meteorological data was interpolated to raster format using Inverse Distance Weighting (IDW) spatial interpolation.

Precipitation is a key parameter. So, as an alternative to interpolated gauge data at a limited number of locations, the daily precipitation sum data in gridded format measured on approximately 300 locations of a voluntary network over The Netherlands has been used to obtain precipitation data with higher spatial resolution [38]. The gridded data is not available for all the other features. Daily weather files in the required DSSAT file format were generated for each crop parcel, and each year.

SOIL DATA

Bulk density, soil organic carbon, sand, silt, clay fractions, coarse fragments, soil pH, and total nitrogen were obtained from the ISRIC's global Soil Information System (SoilGrids) [39]. Soil data at six different depth layers 0-5, 5-10, 15-30, 30-60, 60-100, and 100-200 cm were obtained with a 250 m grid spacing from SoilGrids250m [40].

Soil properties required by DSSAT that are not available in SoilGrids250m, are obtained from HarvestChoice HC27 [41]. HC27 is a soil database containing 27 soil profiles that are provided by considering only three criteria: soil texture, organic carbon content (proxy for soil fertility), and rooting depth (water availability proxy). In this study, values for soil color, albedo, evaporation limit, drainage rate, runoff curve number, mineralization factor, photosynthesis factor, pH in buffer determination method, extractable phosphorus determination code, and potassium determination method were obtained from the corresponding HC27 soil profiles.

We calculated the values for Saturated hydraulic conductivity (SSKS), Saturation (SSAT), Drained Upper limit (SDUL), and Lower limit (SLLL) by pedo-transfer functions following [42–44].

Table 3.1 lists the most relevant soil parameters and the range of values assumed in this study.

Table 3.1.: Soil parameters for DSSAT.

Parameters	Range
Depth (<i>cm</i>)	5-200
Porosity (cm^3/cm^3)	0.376-0.417
Field capacity (cm^3/cm^3)	0.155-0.187
Wilting point (cm^3/cm^3)	0.069-0.097
Bulk density (g/cm^3)	1.11-1.51
Organic carbon (%)	1.33-6.55
Total nitrogen (%)	0.16-0.48

MANAGEMENT AND GENETICS

Table 3.2 provides details on planting, emergence, harvest dates, and planting density assumed in this study. In Noord-Brabant, they prepare the field in April, the growing season of maize starts in May and lasts until the end of September. During the simulation, water stress simulation is enabled but all the other nutrient stresses are turned off as all the crops growing over the area, are well fertilized. The planting density was measured by [31], and these are typical values for maize planted in this area.

The DSSAT model requires some calibration. Specifically, we need to define a set of genetic coefficients controlling the growth, development, and yield of the crop such as ecotype and cultivar coefficients [35]. Ecotype coefficients are a set of coefficients for a group of cultivars (cultivar means a type of cultivated crop) that show similar responses to environmental conditions. Genetic coefficients in DSSAT have been calibrated for maize grown in tropical and semi-arid areas [45]. However, the day length, radiation use efficiency, and temperature mean that different kinds of maize are

Table 3.2.: Maize input management parameters into DSSAT model from Noord-Brabant province.

Parameters	Value
Density ($\#/m^2$)	8
Row Spacing (cm)	75
Depth (cm)	12
Planting date	Beginning of May
Flowering date	End of July
Physiological maturity date	End of August
Harvest date	End of September

grown in Northern Europe. Furthermore, the maize grown in the Netherlands is grown for silage. Therefore, some ecotype coefficients such as radiation use efficiency (RUE) [46] and canopy light extinction coefficient for daily PAR (photosynthetically active radiation) (KCAN) [47, 48], have to be tuned. These are important drivers controlling the maximum value of the growth parameters [49].

The *medium season* cultivar, was selected as an initial maize cultivar in the area and the calibration started based on those primary values. The cultivar coefficients are calibrated by comparison between observed and simulated variables. First, we calibrate the phenology coefficients (P1, P2, and P5 in Table 3.3) by finding the values that provide the best agreement between simulated and observed flowering and physiological maturity dates. Then, we calibrate the growth coefficients (G2 and G3) in order to minimize the root mean square difference between observed and simulated LAI and dry biomass for the calibration field. Calibration data were available at an experimental site in Reusel [31]. Table 3.3 shows the calibrated cultivar coefficients for CERES-Maize.

Table 3.3.: DSSAT Maize cultivar coefficients definition and calibrated values.

Cultivar parameters	Value
P1: Thermal time from emergence to end of juvenile phase.	140
P2: Photoperiod sensitivity coefficient.	0.856
P5: Thermal time from silking to physiological maturity.	890.7
G2: Potential kernel number.	983.3
G3: Potential kernel growth rate (mg/day).	7.09
PHINT: Phylochron interval; the interval in thermal time (degree days) between successive leaf tip appearances.	38.9

The key biophysical parameters of maize, like LAI, canopy height, AGB, SM_S , and SM_R provided by DSSAT, are entered as inputs to the SVR forward model to simulate the SAR observables.

3.3.3. SAR AND OPTICAL DATA

Sentinel-1 C-band Interferometric Wide (IW) swath data with a 6-day repeat cycle (relative orbit 37) have been used. For this relative orbit, the incidence angle range varies between 36 to 40 degrees over the study area. We used the spatially averaged parcel-level NRCS in VV and VH polarizations (σ_{VV}^0 and σ_{VH}^0), the cross-ratio ($CR = \sigma_{VH}^0 / \sigma_{VV}^0$) and the interferometric coherence in VV polarization. All these radar observables were extracted from the Agricultural SandboxNL database [50, 51]. This database was generated by utilizing the openly available annual BRP vector layers and Sentinel-1 SAR ground range detected (GRD) data over The Netherlands in the Google Earth Engine (GEE) using data collected between 2017 and 2019. The GEE-provided Sentinel-1 GRD images are pre-processed with orbit file update, radiometric calibration, border, thermal noise correction, and terrain correction [52]. The Agricultural SandboxNL database contains parcel averaged Sentinel-1 backscatter (σ_{VV}^0 , σ_{VH}^0 and CR) and associated attributes (local incidence angle, azimuth look angle, and pixel count) for six different relative orbits (37, 110, 139, 15, 88, and 161).

The interferometric coherence is defined as the normalized cross-correlation between two coregistered SAR images. The coherence is defined as [53]:

$$\gamma_{est} = \frac{\langle S_1 S_2^* \rangle}{\sqrt{\langle |S_1|^2 \rangle \langle |S_2|^2 \rangle}}, \quad (3.1)$$

where S_1 and S_2 represent two co-registered complex images, and $*$ denotes the complex conjugate operator, and $\langle \cdot \rangle$ represents the spatial averaging operator. We have estimated the coherence values implementing standard InSAR pre-processing steps[54] in ESA SNAP software[55] for VV polarization with a 6-day repeat cycle. The parcel-level spatially averaged coherence product is generated similarly to the Sentinel-1 SAR backscatter in the Agricultural SandboxNL database.

The coherence values range from 0 to 1, where low values correspond to high decorrelation between the two acquisitions. High coherence is achieved when the physical properties and position of scatterers remain the same between two acquisitions. This happens during the bare soil or after harvest, which makes the coherence a valuable indicator to detect agricultural events.

The SandboxNL was extended to include parcel-level averaged Sentinel-2 optical data. In this study, we used these optical data to obtain estimates of LAI over our study area. The parcel-level LAI values for maize were estimated from the Normalized Difference Vegetation Index (NDVI) using the relationship for maize adopted from Kang et. al. 2016 [56, 57].

3.3.4. SUPPORT VECTOR REGRESSION (SVR)

To model the relation between the biophysical parameters and SAR observables (NRCS and coherence), we used a SVR [58] algorithm. SVR has been employed in previous studies for different applications. For example, in [22], the performance of the WCM and SVR as a forward model was compared, in order to assimilate Sentinel-1 data into the Global Land Evaporation Amsterdam Model (GLEAM). Their results show the capability of machine learning as an alternative to semi-empirical models to predict backscatter.

Additionally, soil moisture was retrieved by [59] through the inversion of both the theoretical integral equation model (IEM) and the semi-empirical model (Oh), and the results were compared with SVR. They showed that the data-driven machine-learning approach outperforms the other mentioned models. As mentioned in [60] SVR has limited complexity in the training phase and produces high accuracy with the less computational load. The purpose of this study is not to compare machine-learning algorithms nor to determine the best algorithm. In this study, we chose to use a proven machine learning algorithm as a data-driven forward modeling technique to predict SAR observables based on its positive track record in similar applications [22, 25].

The SVR model is fed with LAI, dry biomass, SM_S and SM_R from the DSSAT model to predict NRCS in VV and VH polarizations, and the cross-ratio. In the case of coherence, canopy height from DSSAT is also added to the inputs, and SVR is fed with the values corresponding to the dates of the first acquisition in the interferometric pair and with the differences between the values of the first and second acquisition.

We trained and tested the SVR model in three ways:

1. Training with data of individual years and testing on the same year (Same-year).
2. Training on individual years and testing on multiple years (Cross-year).
3. Training with multiple-year data and testing on multiple years (Multi-year).

In all cases, we follow standard practices by first training the algorithm to find the best model fit and then testing the model on the separate test data sets. In cases 1 and 3, the algorithm is trained using 80% of the fields, and the remaining 20% fields are used for test of the model, with the fields being randomly assigned to the training and testing sets.

Each field at each Sentinel-1 acquisition time counts as an individual sample. During the growth period of each year (planting to harvest), 25 SAR acquisitions are available. We applied 10-fold cross-validation to avoid overfitting. It means that the model is trained using 9 of the folds and validated on the remaining part. Still, the 20% test data are kept and used exclusively for the final assessment. Grid-search is used in order to obtain the hyper-parameters with the highest cross-validation accuracy. The cost parameter C and the hyperparameter γ have to be tuned. The best (C, γ) values are used to train and generate the model. The values of all input variables are scaled so that they are between 0 and 1. The optimum kernel was the radial basis function (RBF) kernel which is defined as

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \quad (3.2)$$

where x_i and x_j are the a pair of samples.

3.3.5. SURFACE ROUGHNESS

Rough surface scattering contributes to the observed radar signal. This contribution is dominant at the beginning of the growth period, when fields are mostly bare or covered with a small amount of vegetation [21]. Rough surface scattering is controlled by the dielectric constant, which in turn depends on the soil composition and the soil moisture content, and on the roughness spectrum. The latter is not represented in any

of the DSSAT variables, which implies that field-to-field variations of the rough surface scattering term cannot be modeled with the available biophysical parameters. This contributes to inter-field differences in the observed NRCS values during the planting and emergence period.

Conceptually, we can assume that the rough surface contribution to the radar observables can be directly measured in the period between planting and emergence. Therefore, we assume that the properties controlling this rough surface component, with the exception of the soil moisture, remain constant during the growth season (in particular during the initial period).

We consider a reference backscatter value for each parcel as a proxy for the effect of roughness/geometry on the variability between parcels and included that as a label for each parcel (along with other parameters for the parcel). This reference backscatter value is calculated as the mean NRCS value for the parcel in the three acquisitions following the planting date.

3.3.6. EVALUATION OF MODEL PERFORMANCE

Five error metrics are used to evaluate the performance of the regression model. The simulated and observed backscatter and coherence values are compared using standard statistical metrics:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3.3)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (3.4)$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.5)$$

where y_i represents the i^{th} observation, \hat{y}_i is the i^{th} predicted value, n is the number of observations. The Mean Absolute Error (MAE) gives the average absolute difference between the predicted and the actual values. The Mean Square Error (MSE) is used to measure the error of the model in simulating SAR observables [61] that is magnifying large errors. The coefficient of determination, R^2 , is used to show the capability of the model in explaining the variation of the actual data. In an ideal case, R^2 is equal to 1. In addition, we use Pearson and Spearman's correlation, which shows the correlation between predicted and observed backscatter.

3.3.7. FEATURE ANALYSIS

In order to understand and quantify the importance of different features in defining the regression model, we used a feature analysis algorithm. The Minimum Redundancy Maximum Relevance (MRMR) algorithm is applied to maximize the relevance of a feature set with the dependent variable and minimize the redundancy in a feature set [62]. The MRMR algorithm searches to find an optimal subset of features (S)

that maximize V , the relevance of a feature set with response variable (y), where this relevance is quantified through the mean value of the mutual information (I),

$$V = \frac{1}{|S|} \sum_{x \in S} I(x, y) \quad (3.6)$$

where $|S|$ is the number of features in S . The redundancy, W , is quantified by the mean of the mutual information between the features within the set

$$W = \frac{1}{|S|^2} \sum_{x, z \in S} I(x, z) \quad (3.7)$$

The MRMR algorithm ranks features by using the mutual information quotient (MIQ) value:

$$\text{MIQ} = \frac{V}{W} \quad (3.8)$$

3.4. RESULTS AND DISCUSSION

3.4.1. DSSAT CALIBRATION AND PERFORMANCE

Fig. 3.3 shows three years of meteorological data, including TMIN, TMAX, SRAD, and cumulative precipitation in daily steps. DSSAT needs to be calibrated for the specific crop variety of interest. The model was calibrated using data collected during field experiment in Reusel, Noord-Brabant (See Figure 3.1 and Section 3.2.)

As illustrated in Fig. 3.4, after calibration, DSSAT simulated the LAI and biomass with 95% and 98% accuracy, respectively. Fig. 3.4a compares the time series of predicted and observed LAI and AGB. LAI estimates were derived from Sentinel-2 NDVI observation and from in-situ measurements. NDVI-derived and simulated LAI have a downward trend after the end of August while the LAI value from field measurement increases. As reported in [63], the field measurements of LAI were obtained by multiplying the averaged leaf area by the plant density. Consequently, the loss of LAI (related to primary productivity) due to leaf degradation after the crop reaches maturity is not reflected in the field data. The quality of NDVI-derived LAI estimates is limited by the spatial resolution of the Sentinel-2 data and by the density of the crop, as the visible and near-infrared reflectances depend also on bare soil exposure [64]. Fig. 3.4b displays the simulated LAI and biomass versus the predicted values. Generally, there is a good agreement between the simulated and observed values for LAI and biomass, showing that the model is well-calibrated.

Fig. 3.5 provides an inter-annual comparison between DSSAT-simulated crop bio-geophysical parameter outputs of the maize fields for normal (2017) and drought-affected (2018 and 2019) years. In 2018, Europe experienced a very hot and dry summer. Drought in 2018 and 2019 affected agricultural production in The Netherlands. The 2018 drought influenced the groundwater levels, and hence crop production [65]. The influence of the drought on Sentinel-1 SAR observables was discussed by [30].

Starting in July 2018, reduced soil moisture levels lead to a significant drop in the simulated LAI and the rate of growth of the simulated dry biomass. The maximum daily

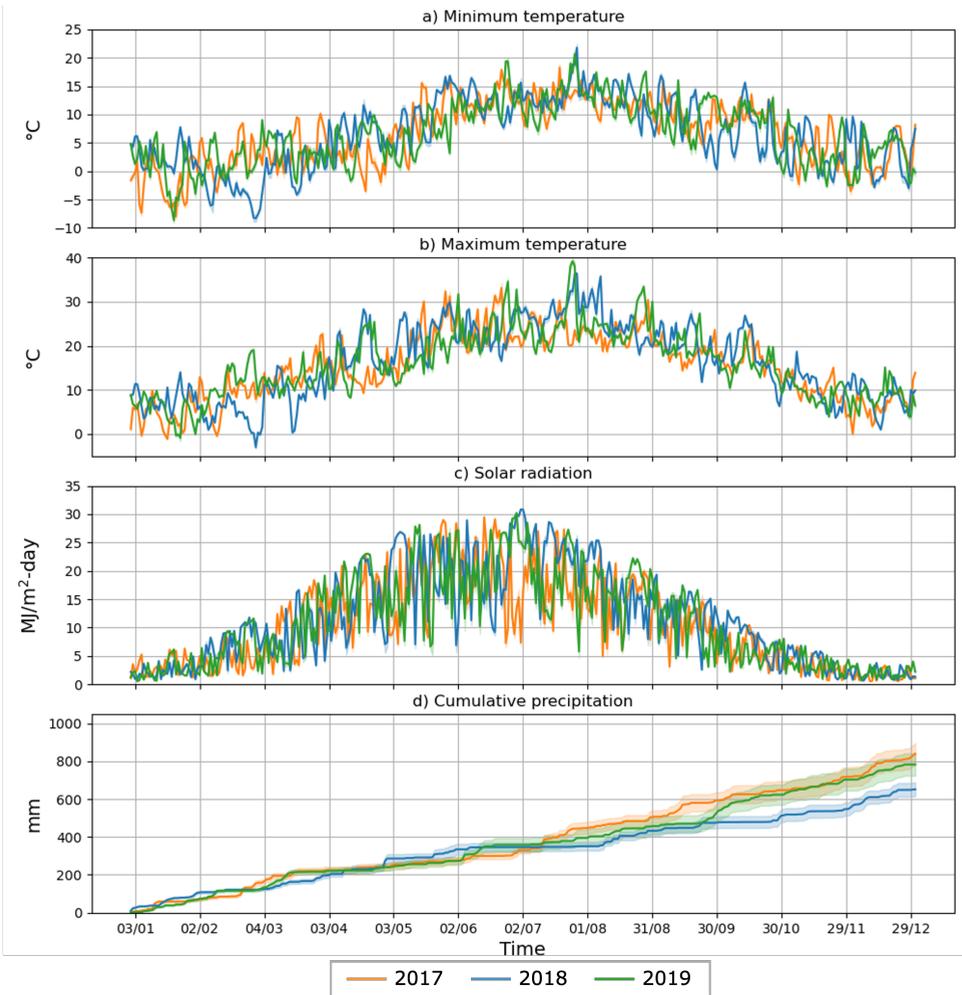


Figure 3.3.: Daily time-series of meteorological forcing inputs to the DSSAT model. (a) Minimum temperature. (b) Maximum temperature. (c) Solar radiation. (d) Cumulative precipitation. Solid lines show the mean value of the variable over studied maize fields, and the bounded area represents one standard deviation of the variables.

average of LAI reaches approximately 3.5, compared to around 4.5 in 2017. Similarly, the biomass accumulation was substantially lower. Fig. 3.5 also illustrates that, in our case, the simulated LAI and AGB have higher variances during drought periods than in normal conditions. For example, this is clearly visible in 2018 for LAI, after it reaches its highest value in July. This illustrates that the anomaly in root zone soil moisture was sufficient to constrain crop growth.

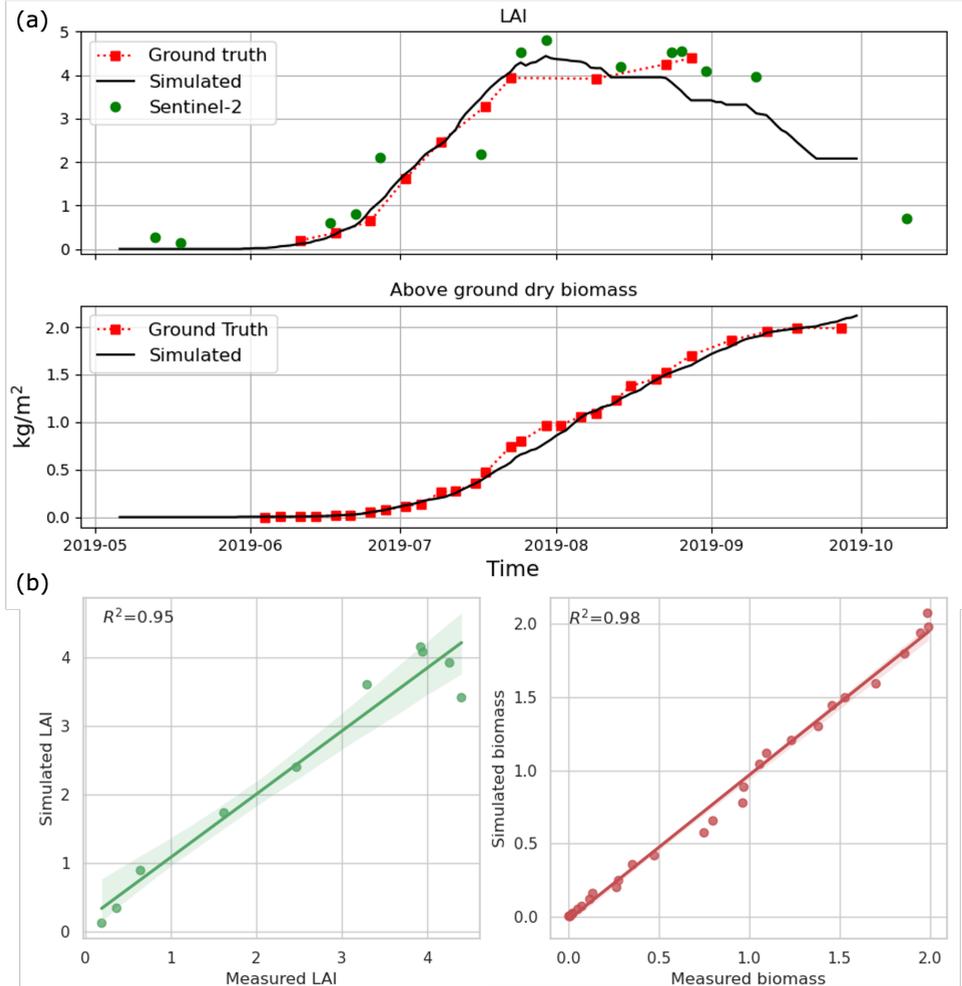


Figure 3.4.: DSSAT model calibration. (a) Time-series fit between (**top**): estimated LAI from Sentinel-2 data, simulated LAI from DSSAT model and in-situ measurements, (**down**) simulated biomass from DSSAT and ground measurements. (b) Comparison between simulated and measured LAI and biomass.

3.4.2. MODELED VS MEASURED RADAR OBSERVABLES

Fig. 3.6 provides the comparison of NRCS in VV and VH, cross-ratio (CR, VH/VV), and interferometric coherence VV between DSSAT-SVR estimated (in red) and Sentinel-1 observations (in blue). The results are calculated for the independent test data sets. In this figure, training and test data are from the same year in this case.

Before crop emergence, radar backscatter is controlled by the surface roughness and moisture content of the exposed soil. During this period, sudden variations of the NRCS

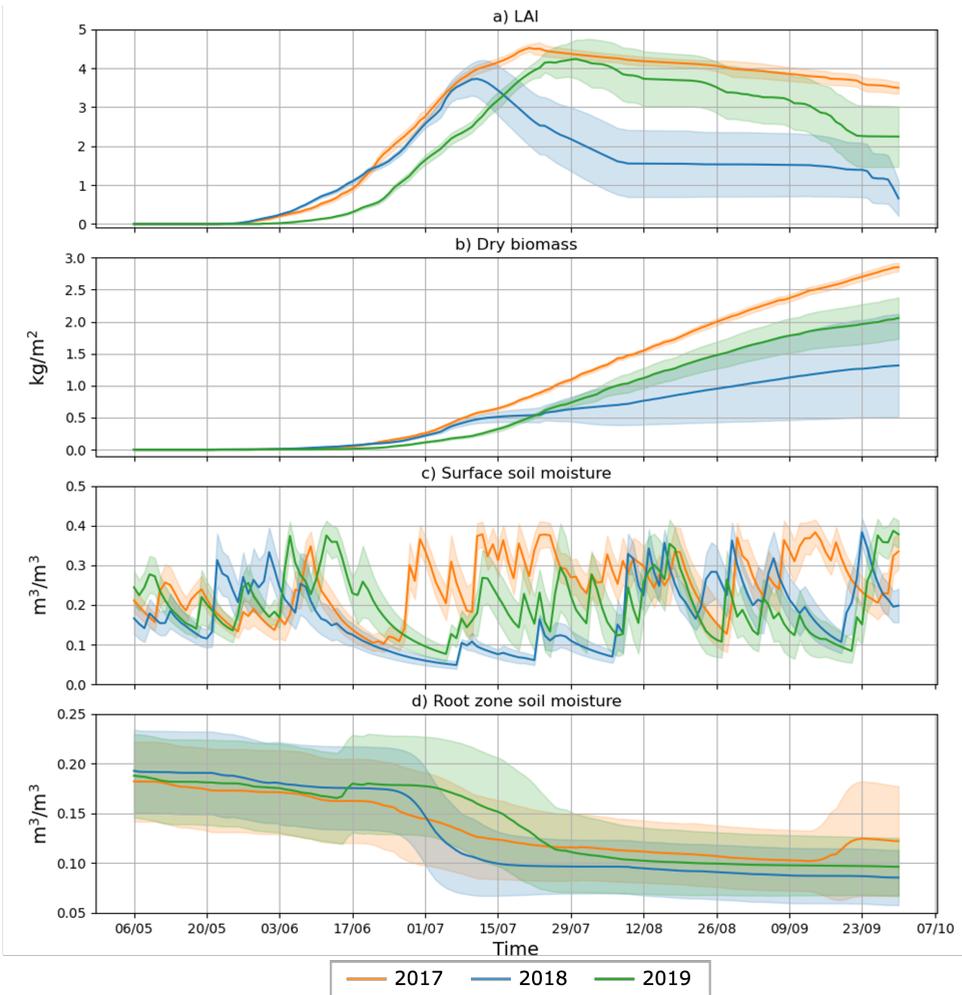


Figure 3.5.: Vegetation and soil parameters during the growing season in daily steps from DSSAT model that are considered relevant inputs to simulate SAR observables. (a) LAI, (b) Above ground dry biomass, (c) Surface soil moisture, (d) Root zone soil moisture. Solid lines show the mean value of the variable over studied maize fields and the bounded area represents one standard deviation of the variable.

are caused by precipitation events. Starting from late May, when maize enters the leaf development stage, radar backscatter increases as the plant grows. During the stem elongation stage to tassel initiation in July, LAI rapidly reaches its maximum, leading to an increase in both co- and cross-polarized NRCS and in the cross-ratio. Once the crop reaches maximum LAI, the sensitivity of backscatter to growth decreases. The fluctuations in backscatter are probably due to the rain events on June 6th 2017, June

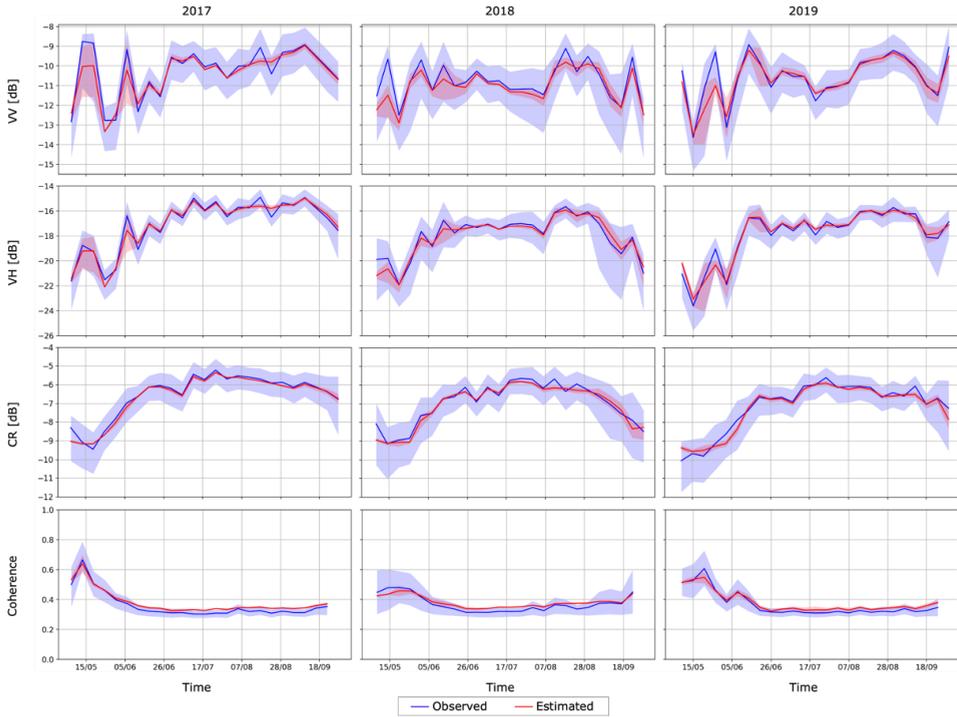


Figure 3.6.: Comparison of NRCS in VV, VH, cross-ratio (CR, VH/VV), and interferometric coherence VV between DSSAT-SVR estimated and Sentinel-1 observations. Each column is associated with a different year. Solid lines indicate the mean value of the feature over maize parcels in the test set, and the bounded area shows the 20th-80th percentiles. Training and test data are from the same year in this case.

1st and 7th 2018, and May 28th and June 5th, 2019. When maize reaches physiological maturity, in the last week of August, radar backscatter begins to decrease until harvest. This happens due to the decrease in vegetation water content. The inter-field variability ranges from 1 to 4 dB. Other studies also established similar temporal behavior of the radar observables for maize fields [24, 66]. In 2018, dry conditions resulted in an earlier ripening and harvest, leading to a shorter growing season compared to 2017 [30]. This is evident in the observed NRCS values, particularly in the cross-pol channel and the CR. The jump in NRCS co-pol in August 2018 was caused by the rain event on 7th August. Fig. 3.5(C) shows that in July and the beginning of August, surface soil moisture was lower than in 2017 and a rain event on 7th August 2018 led to higher soil moisture values.

As expected, coherence values are higher before the maize emergence. After crop emergence, coherence drops quickly due to temporal decorrelation associated with crop growth [67]. The coherence value remains low (<0.3) through the vegetative period

of maize. The lower panels in Fig. 3.6 illustrate the reasonable agreement between estimated and observed SAR observables.

In Section 3.3.5, it is argued that the average backscatter of three acquisitions after planting can serve as an indicator of the rough surface contribution to the backscatter for each parcel. This variable, referred to as offset, is included along with LAI, biomass, SM_S , and SM_R in the analysis. To assess the impact of the offset parameter, the SVR model was trained and tested with and without taking it into account. Fig. 3.7 shows the time-series of the difference between the observed and the estimated VV backscatter with and without using the offset, when the model is trained and tested on normal 2017 data. This figure illustrates that including the offset reduces the difference between the estimated and observed backscatter at the beginning of the vegetative period when the total backscatter is still sensitive to surface scattering. Similar results were obtained in other years and channels, as shown in supplementary Fig. A.1. We calculate statistics for the entire growing period and also for the period where the surface scattering contribution is expected to be significant. This period covers the start of the vegetative stage, from emergence, typically the last week of May, until the beginning of tassel emergence, around the first week of July each year. Including the offset reduces the error, improves the accuracy, and increases the correlation coefficients. For example, the R^2 score increases from 0.45, 0.15, and 0.27 to 0.55, 0.30, and 0.44 at the beginning of the vegetative period for 2017, 2018, and 2019, respectively. Supplementary Fig. A.2 shows standard regression performance metrics for backscatter using train and test sets from the same year, both with and without the offset at the beginning of the vegetative period. Because including the offset to account for e.g. roughness in the surface scattering, all results discussed hereafter have been obtained including the offset parameter.

Tables 3.4 and 3.5 display the MAE and Pearson correlation for different years and channels over the beginning of the vegetative period and the whole growing period (the values within the parentheses). The first column of the tables indicates the year in which the train data was selected. Other statistical metrics are presented in supplementary Table A.1 to A.5.

Given the significantly different conditions during the three years considered, we expect low model accuracy in the case that the training data set does not include data from the year to which the SVR model is applied. This is confirmed by the results presented. For instance, using 2018 VV and VH backscatter data to train the model and applying it to 2017 results in a high MAE and very low correlation. Likewise, using VV and VH backscatter in 2017 as training data to predict backscatter in 2018, produces similar results. This happens because the trained model is applied to sets of inputs for which it has not been trained. As an illustrative example, Fig. 3.8 shows the bivariate histogram of LAI and AGB for each year. The figure shows how in this 2-D subspace of our 4-D parameter space, the crop state follows similar trajectories during the beginning of the vegetative periods, where both LAI and AGB grow rapidly, after which we can observe distinct trajectories. The consequence is that some regions of the parameter space are only visited in particular years.

Training the model with backscatter in 2018 to predict it in 2019, performs reasonably during the start of the vegetative stage of 2019 as these two periods suffer from

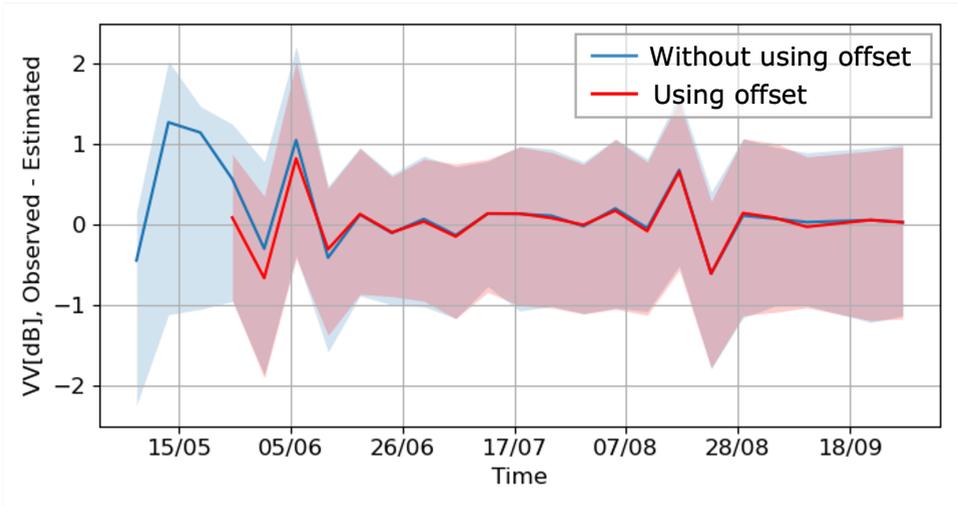


Figure 3.7.: Time series difference between observed and estimated backscatter VV in 2017 with and without using the offset.

drought. The CR produces better results as it suppresses the influence of soil moisture. The CR generally follows the overall temporal behavior of LAI. Notably, year-to-year transferability is improved in the CR results. When the model is trained on multiple years (2017, 2018, and 2019), it performs almost as well as when it is trained on the same year.

The CR is not an independent observable, while we already have σ_{VH}^0 and σ_{VV}^0 . However, this does not necessarily mean that the best possible forward models of σ_{VH}^0 and σ_{VV}^0 would give the best possible model of the CR. In particular, the CR is often used because it partially suppresses some contributions, for example, soil moisture that affects σ_{VH}^0 and σ_{VV}^0 in similar ways. The quality of the forward model for the backscattering coefficients may be limited by its ability to correctly represent the effect of soil moisture effects or by errors in the soil moisture input data, which would result in errors in the ratio of these forward-modeled coefficients. In contrast, the CR forward model will be less affected by these errors.

We see a similar general behavior for the coherence, with a high correlation between the predicted and observed coherences in the Same-year and Multi-year cases but relatively low in the cross-year cases. In particular, the results for 2018 if the model is trained with 2017 or 2019 data are very poor. As discussed in Section 3.4.3, the trained coherence model is sensitive to the dry biomass difference. Indeed, one can expect the coherence to be higher if the biomass is not changing. Fig. 3.5 suggests that in 2018, for a significant number of fields, the dry biomass predicted by DSSAT stagnates around August, which can cause the model to predict high coherences during that period. When we train including 2018 data the model *learns* to put much less emphasis on the biomass difference. In a physical conceptual model, we would take the biomass difference as a good indicator of coherence during the beginning of the

growth period, but not anymore at later stages, where the wind-induced motion of the crop is sufficient to explain a low coherence. Other coherence statistics are presented in supplementary Table A.6.

Table 3.4.: Mean absolute error between estimated and observed σ_{VV}^0 , σ_{VH}^0 , CR and γ_{VV} , over the beginning of the vegetative period and the whole growing period (the values within the parenthesis).

2017 MAE	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	1.01 (0.98)	1.05 (0.93)	0.78 (0.77)	0.05
2018	2.89 (2.64)	2.53 (2.03)	0.91 (0.98)	0.07
2019	1.47 (1.37)	1.40 (1.34)	0.91 (0.92)	0.06
2018 + 2019	1.64 (1.42)	1.53 (1.31)	0.90 (0.89)	0.06
2017+2018+2019	1.17 (1.05)	1.45 (1.32)	0.80 (0.78)	0.05
<hr/>				
2018 MAE				
2017	1.88 (1.91)	1.99 (2.44)	1.09 (1.92)	0.15
2018	1.04 (1.18)	1.18 (1.33)	0.95 (1.02)	0.07
2019	1.38 (1.70)	1.65 (1.70)	1.07 (1.18)	0.10
2017 + 2019	1.42 (1.55)	1.64 (1.60)	1.05 (1.18)	0.10
2017+2018+2019	1.14 (1.23)	1.28 (1.37)	0.95 (1.02)	0.07
<hr/>				
2019 MAE				
2017	1.72 (1.64)	1.93 (1.66)	1.05 (1.05)	0.07
2018	1.46 (1.76)	1.53 (1.66)	0.98 (0.97)	0.06
2019	1.12 (1.00)	1.18 (1.00)	0.82 (0.83)	0.05
2017 + 2018	1.46 (1.30)	1.43 (1.26)	0.98 (0.96)	0.06
2017+2018+2019	1.20 (1.06)	1.25 (1.05)	0.84 (0.85)	0.05

Fig. 3.9 shows the time series of observed backscatter in different years and channels (in blue) and estimated backscatter from the combination of three years (Multi-year) in green and estimated backscatter from a year that behaves differently with the estimating year (Cross-year) 2017 in red and 2018 in black. The transparent buffer shows 20th-80th percentiles. This figure demonstrates the impact of having partially disjoint sets of SVR input parameters (or DSSAT outputs) for different years. In early vegetative stages, inter-annual variation in the SM_S will cause errors in the predicted backscatter. For example, when we apply the 2018-trained model to 2017 inputs, we see large errors in the NRCS at times where the SM_S in 2018 was consistently lower than in 2017 (see Fig. 3.5). This problem does not affect the CR because it is much less sensitive to SM_S variations [68]. At later vegetative stages, the NRCS is much more controlled by LAI and dry biomass. This explains, for example, the large error in the prediction of all the 2018 observables, including CR, for a 2017-trained model: the low LAI values (Fig. 3.8) of 2018 where never encountered in the 2017 training data set. Similar results were obtained using all other combinations of training and testing years (See Supplementary Fig. A.3).

Table 3.5.: Pearson correlation between estimated and observed σ_{VV}^0 , σ_{VH}^0 , CR and γ_{VV} , over the beginning of the vegetative period and the whole growing period (the values within the parenthesis).

2017 Pearson	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	0.75 (0.69)	0.80 (0.80)	0.68 (0.63)	0.77
2018	-0.2 (-0.09)	0.09 (0.40)	0.61 (0.54)	0.50
2019	0.45 (0.33)	0.69 (0.57)	0.56 (0.50)	0.57
2018 + 2019	0.27 (0.30)	0.63 (0.68)	0.57 (0.53)	0.57
2017+2018+2019	0.67 (0.63)	0.67 (0.70)	0.67 (0.63)	0.75

2018 Pearson	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	0.17 (-0.02)	0.16 (0.08)	0.49 (0.15)	0.10
2018	0.55 (0.53)	0.61 (0.60)	0.58 (0.55)	0.50
2019	0.33 (0.21)	0.42 (0.33)	0.50 (0.35)	0.04
2017 + 2019	0.21 (0.21)	0.33 (0.32)	0.50 (0.36)	0.17
2017+2018+2019	0.44 (0.48)	0.53 (0.58)	0.60 (0.56)	0.50

2019 Pearson	σ_{VV}^0	σ_{VH}^0	CR	γ_{VV}
2017	0.44 (0.22)	0.45 (0.39)	0.64 (0.47)	0.54
2018	0.45 (0.19)	0.59 (0.40)	0.65 (0.57)	0.52
2019	0.67 (0.63)	0.74 (0.71)	0.75 (0.67)	0.74
2017 + 2018	0.50 (0.40)	0.62 (0.59)	0.66 (0.58)	0.58
2017+2018+2019	0.62 (0.58)	0.71 (0.69)	0.74 (0.66)	0.70

Fig. 3.10 displays the same time series for the VV coherence. In general, the highest correlation and lowest error are observed when the training and testing data are from the same year or when data from all three years is used. As expected, the model wrongly shows high coherences for part of 2018 growth when we trained the model with 2017. This can be explained by considering the quite constant slope of 2017 biomass time series as compared to the wide range of biomass differences for 2018. Similar results were obtained for coherence from all training and testing year combinations (See Supplementary Fig. A.5).

3.4.3. FEATURE ANALYSIS

The feature analysis aims to understand the drivers of SAR observables and to ensure that they are physically plausible. The MRMR algorithm is applied to assess the importance of the different variables in the regression model.

Fig. 3.11 displays the feature importance scores of the different parameters fed to the SVR model for the estimation of the VV and VH backscatter and the CR for different years, in the Same-year and Multi-year cases. In general, dry biomass and the offset parameter have the highest scores. Significant correlations between dry biomass and C-band radar backscatter were also observed by [66]. In CR, because the sensitivity to soil moisture is minimized, we expect the higher importance to AGB always would

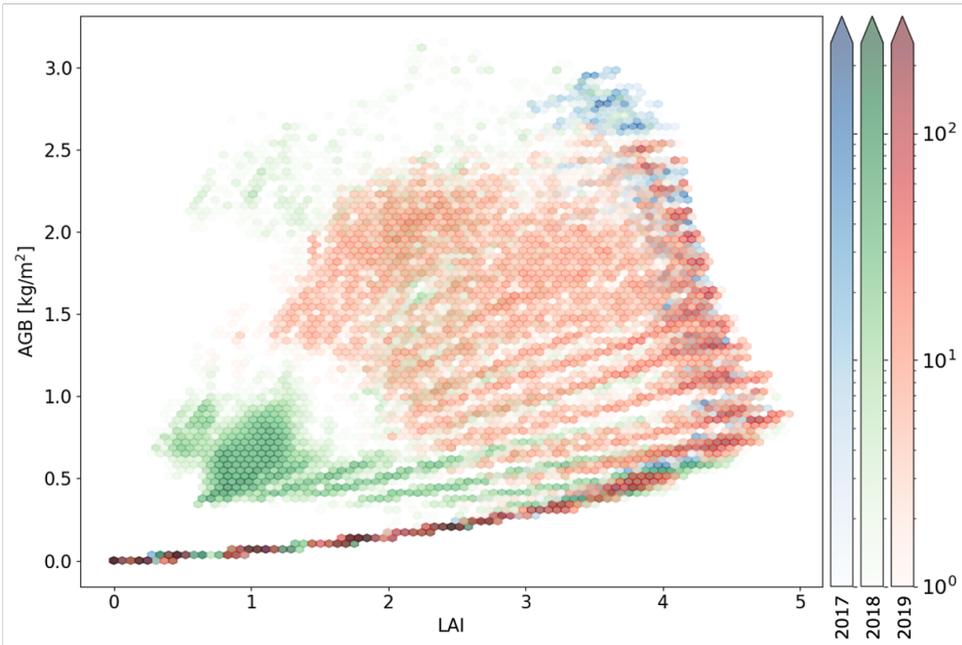


Figure 3.8.: Bivariate histogram of LAI and dry biomass for 2017 (blue), 2018 (green), and 2019 (red). The color bars show the normalized density of samples on a logarithmic scale for each year.

be the case unless LAI and AGB are inter-changeable so that one is picked up in the Same-year and the other in the Multi-year case. The impact of LAI in VV and VH was found to be minimal for all three years, possibly due to the strong correlation between LAI and dry biomass. Similarly, the influence of soil moisture was relatively low, likely because the analysis covers the entire growing season. The influence of soil moisture varies with the growth stage, as backscatter sensitivity to soil moisture decreases with increasing biomass. According to Fig. 3.12, the biomass difference is the most important factor for coherence VV across different years, as this parameter represents the overall growth throughout the entire period, making it a significant indicator. A non-zero difference in biomass implies a low coherence, which happens from emergence till harvest. While surface soil moisture difference has the least impact. (The feature importance scores of other training years are presented in supplementary Fig. A.4 and A.6).

3.5. CONCLUSIONS

This study aimed at demonstrating the feasibility of using machine learning techniques to create a forward model linking bio-geophysical crop parameters to C-band radar observables. This forward model serves as a bridge between the observed SAR data and

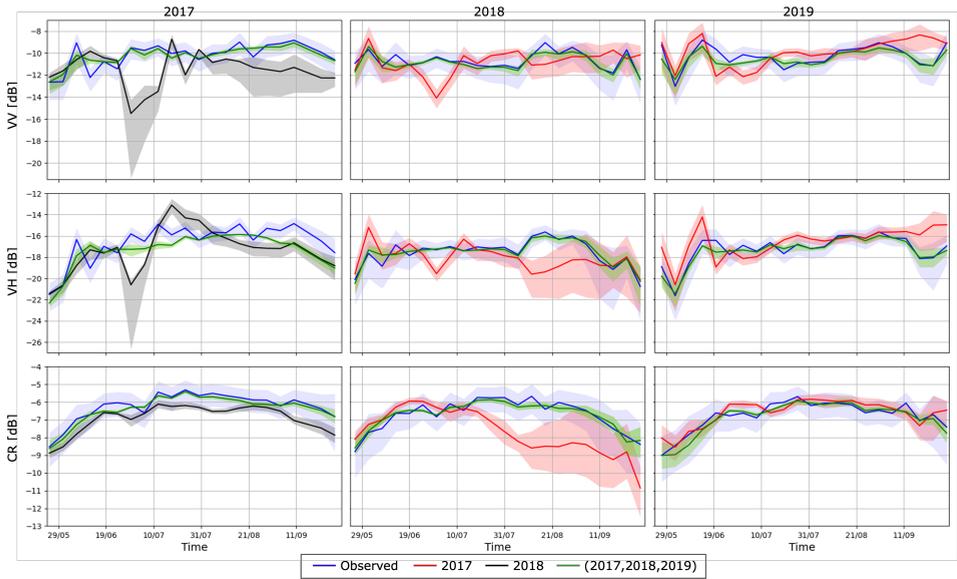


Figure 3.9.: Time-series of observed and estimated backscatter in 2017 (1st column), 2018 (2nd column) and 2019 (3rd column) for different channels. The years that training data are selected from are shown in the legend.

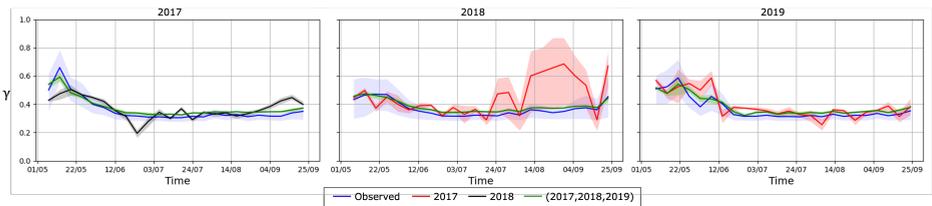


Figure 3.10.: Time-series of observed and estimated coherence VV in different years. The legend shows the year in which the training data belongs. Different year training and test data were used for the estimation of observables.

the crop model simulations, enabling the integration of the two. We used a crop growth model as an alternative to in-situ data to provide crop descriptors over 1500 maize fields in the Netherlands. The Minimum Redundancy Maximum Relevance (MRMR) was used to quantify the sensitivity of the SAR observables to the DSSAT variables. We demonstrate that the connections between crop bio-geophysical variables, such as LAI, AGB, SM_S and SM_R and the modeled SAR observables, such as NRCS and coherence, are plausible and consistent with known physical principles of microwave remote sensing of vegetated surfaces.

In the early season, surface scattering plays an important role in the interaction with the soil surface, so the mean value of backscatter in three acquisitions after the

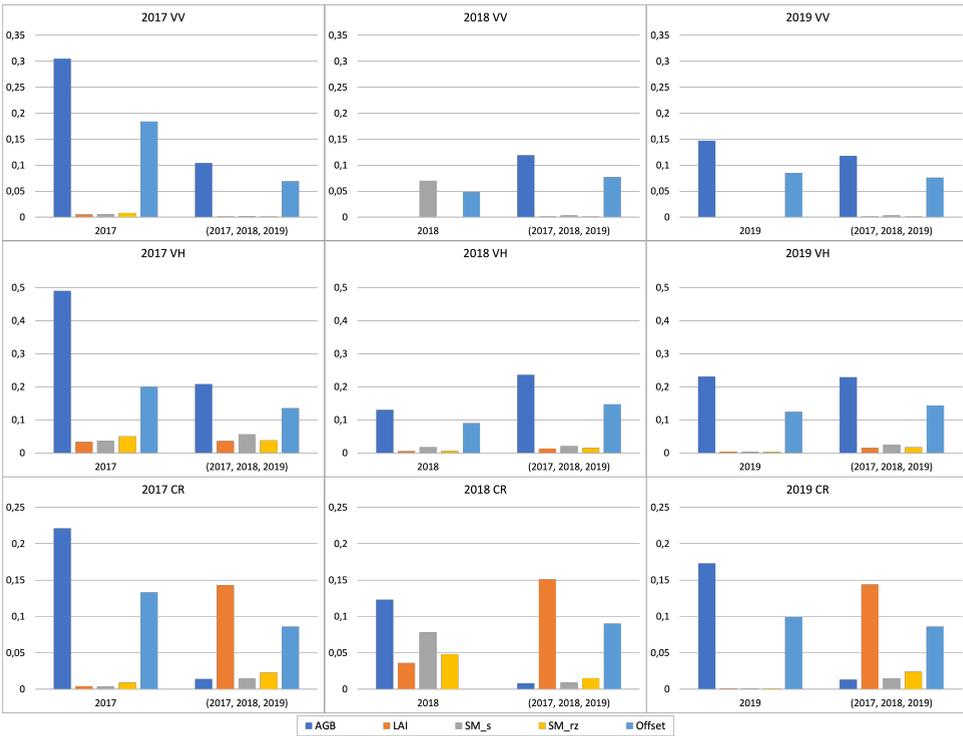


Figure 3.11.: Rank features for SVR with backscatter using the MRMR algorithm. The bars in each plot are positioned on the left side for Same-year and on the right side for Multi-year.

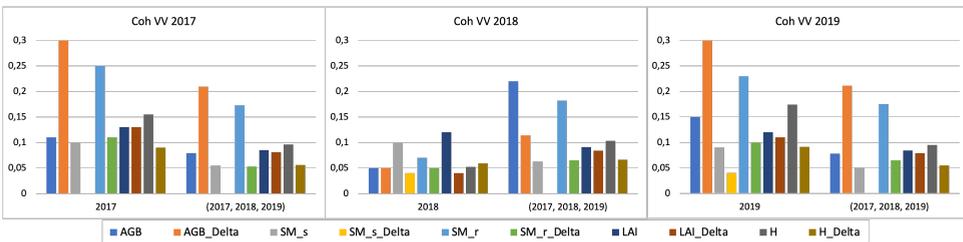


Figure 3.12.: Rank features for SVR with coherence VV using MRMR algorithm. The bars in each plot are positioned on the left side for Same-year and on the right side for Multi-year. The variables with the "Delta" extension represent the difference between the dates of the pair of consecutive acquisitions used to create the interferograms.

planting date in each year is used as a proxy for information about the roughness, row geometry and other static parameters that influence surface scattering. Adding this proxy improves the estimations in the early season. We demonstrate that the difference

in environmental conditions (drought and non-drought situations) affects the model accuracy when the training data set does not include data from the year to which the SVR model is applied. The reason is that there is not enough spatial intra-variability within the area of study due to the similarity in soil texture and rainfall patterns across the province. However, DSSAT-SVR can estimate SAR observables with reasonable accuracy including the effect of surface roughness and using three years of training data. The CR shows better transferability from year to year as it minimizes the influence of soil moisture. Applying the method over a larger area with more heterogeneity or a longer time frame of the observations should result in an improved performance as the model would train with a much wider set of data.

Like all models, crop growth model performance itself depends on the quality of the input data and the calibration of the model parameters. Therefore, there is always some degree of uncertainty associated with them. This uncertainty can be reduced, to some degree, by the availability of accurate meteorological forcing data at suitably fine spatial resolutions as well as in-situ observations of biomass and LAI for model calibration.

Combining crop growth models with machine learning methods has the potential to estimate remote sensing observations without solely relying on ground measurements. Ongoing research will consider the suitability of this approach for anomaly detection in agricultural applications, and its use in a data assimilation context where Sentinel-1 data are used to constrain the state and parameters of the crop growth model.

REFERENCES

- [1] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan. "GMES Sentinel-1 Mission". In: *Remote Sensing of Environment*. The Sentinel Missions - New Opportunities for Science 120 (May 2012), pp. 9–24. ISSN: 0034-4257. DOI: [10.1016/j.rse.2011.05.028](https://doi.org/10.1016/j.rse.2011.05.028).
- [2] G. Macelloni, S. Paloscia, P. Pampaloni, F. Marliani, and M. Gai. "The Relationship between the Backscattering Coefficient and the Biomass of Narrow and Broad Leaf Crops". In: *IEEE Transactions on Geoscience and Remote Sensing* 39.4 (Apr. 2001), pp. 873–884. ISSN: 1558-0644. DOI: [10.1109/36.917914](https://doi.org/10.1109/36.917914).
- [3] A. G. Konings, K. Rao, and S. C. Steele-Dunne. "Macro to Micro: Microwave Remote Sensing of Plant Water Content for Physiology and Ecology". In: *New Phytologist* 223.3 (2019), pp. 1166–1172. ISSN: 1469-8137. DOI: [10.1111/nph.15808](https://doi.org/10.1111/nph.15808).
- [4] H. McNairn and J. Shang. "A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring". In: *Multitemporal Remote Sensing: Methods and Applications*. Ed. by Y. Ban. Remote Sensing and Digital Image Processing. Cham: Springer International Publishing, 2016, pp. 317–340. ISBN: 978-3-319-47037-5. DOI: [10.1007/978-3-319-47037-5_15](https://doi.org/10.1007/978-3-319-47037-5_15).
- [5] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. Liu, and K. Papathanassiou. "Radar Remote Sensing of Agricultural Canopies: A Review". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2249–2273. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2016.2639043](https://doi.org/10.1109/JSTARS.2016.2639043).
- [6] A. Veloso, S. Mermoz, A. Bouvet, T. Le Toan, M. Planells, J.-F. Dejoux, and E. Ceschia. "Understanding the Temporal Behavior of Crops Using Sentinel-1 and Sentinel-2-like Data for Agricultural Applications". In: *Remote Sensing of Environment* 199 (Sept. 2017), pp. 415–426. ISSN: 0034-4257. DOI: [10.1016/j.rse.2017.07.015](https://doi.org/10.1016/j.rse.2017.07.015).
- [7] S. Khabbazan, P. Vermunt, S. Steele-Dunne, L. Ratering Arntz, C. Marinetti, D. van der Valk, L. Iannini, R. Molijn, K. Westerdijk, and C. van der Sande. "Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands". In: *Remote Sensing* 11.16 (Jan. 2019), p. 1887. ISSN: 2072-4292. DOI: [10.3390/rs11161887](https://doi.org/10.3390/rs11161887).
- [8] K. Harfenmeister, D. Spengler, and C. Weltzien. "Analyzing Temporal and Spatial Characteristics of Crop Parameters Using Sentinel-1 Backscatter Data". In: *Remote Sensing* 11.13 (Jan. 2019), p. 1569. ISSN: 2072-4292. DOI: [10.3390/rs11131569](https://doi.org/10.3390/rs11131569).

- [9] D. Han, S. Liu, Y. Du, X. Xie, L. Fan, L. Lei, Z. Li, H. Yang, and G. Yang. “Crop Water Content of Winter Wheat Revealed with Sentinel-1 and Sentinel-2 Imagery”. In: *Sensors* 19.18 (Jan. 2019), p. 4013. ISSN: 1424-8220. DOI: [10.3390/s19184013](https://doi.org/10.3390/s19184013).
- [10] L. Pulvirenti, G. Squicciarino, L. Cenci, G. Boni, N. Pierdicca, M. Chini, C. Versace, and P. Campanella. “A Surface Soil Moisture Mapping Service at National (Italian) Scale Based on Sentinel-1 Data”. In: *Environmental Modelling & Software* 102 (Apr. 2018), pp. 13–28. ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2017.12.022](https://doi.org/10.1016/j.envsoft.2017.12.022).
- [11] D. Mandal, M. Hosseini, H. McNairn, V. Kumar, A. Bhattacharya, Y. S. Rao, S. Mitchell, L. D. Robertson, A. Davidson, and K. Dabrowska-Zielinska. “An Investigation of Inversion Methodologies to Retrieve the Leaf Area Index of Corn from C-band SAR Data”. In: *International Journal of Applied Earth Observation and Geoinformation* 82 (Oct. 2019), p. 101893. ISSN: 1569-8432. DOI: [10.1016/j.jag.2019.06.003](https://doi.org/10.1016/j.jag.2019.06.003).
- [12] M. Vreugdenhil, W. Wagner, B. Bauer-Marschallinger, I. Pfeil, I. Teubner, C. Rüdiger, and P. Strauss. “Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study”. In: *Remote Sensing* 10.9 (Sept. 2018), p. 1396. ISSN: 2072-4292. DOI: [10.3390/rs10091396](https://doi.org/10.3390/rs10091396).
- [13] H. Pan, Z. Chen, A. de Wit, and J. Ren. “Joint Assimilation of Leaf Area Index and Soil Moisture from Sentinel-1 and Sentinel-2 Data into the WOFOST Model for Winter Wheat Yield Estimation”. In: *Sensors (Basel, Switzerland)* 19.14 (July 2019), p. 3161. ISSN: 1424-8220. DOI: [10.3390/s19143161](https://doi.org/10.3390/s19143161).
- [14] J. Betbeder, R. Fieuzal, and F. Baup. “Assimilation of LAI and Dry Biomass Data From Optical and SAR Images Into an Agro-Meteorological Model to Estimate Soybean Yield”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.6 (June 2016), pp. 2540–2553. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2016.2541169](https://doi.org/10.1109/JSTARS.2016.2541169).
- [15] S. Pazhanivelan, V. Geethalakshmi, R. Tamilmounika, N. S. Sudarmanian, R. Kaliaperumal, K. Ramalingam, A. P. Sivamurugan, K. Mrunalini, M. K. Yadav, and E. D. Quicho. “Spatial Rice Yield Estimation Using Multiple Linear Regression Analysis, Semi-Physical Approach and Assimilating SAR Satellite Derived Products with DSSAT Crop Simulation Model”. In: *Agronomy* 12.9 (Sept. 2022), p. 2008. ISSN: 2073-4395. DOI: [10.3390/agronomy12092008](https://doi.org/10.3390/agronomy12092008).
- [16] B. Bauer-Marschallinger, V. Freeman, S. Cao, C. Paulik, S. Schaufler, T. Stachl, S. Modanesi, C. Massari, L. Ciabatta, L. Brocca, and W. Wagner. “Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.1 (Jan. 2019), pp. 520–539. ISSN: 1558-0644. DOI: [10.1109/TGRS.2018.2858004](https://doi.org/10.1109/TGRS.2018.2858004).
- [17] N. Efremova, M. E. A. Seddik, and E. Erten. “Soil Moisture Estimation Using Sentinel-1/-2 Imagery Coupled With CycleGAN for Time-Series Gap Filing”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–11. ISSN: 1558-0644. DOI: [10.1109/TGRS.2021.3134127](https://doi.org/10.1109/TGRS.2021.3134127).

- [18] G. J. M. D. Lannoy and R. H. Reichle. “Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation”. In: *Journal of Hydrometeorology* 17.2 (Feb. 2016), pp. 669–691. ISSN: 1525-7541, 1525-755X. DOI: [10.1175/JHM-D-15-0037.1](https://doi.org/10.1175/JHM-D-15-0037.1).
- [19] G. J. M. De Lannoy, M. Bechtold, C. Albergel, L. Brocca, J.-C. Calvet, A. Carrassi, W. T. Crow, P. de Rosnay, M. Durand, B. Forman, G. Geppert, M. Grotto, H.-J. Hendricks Franssen, T. Jonas, S. Kumar, H. Lievens, Y. Lu, C. Massari, V. R. N. Pauwels, R. H. Reichle, and S. Steele-Dunne. “Perspective on Satellite-Based Land Data Assimilation to Estimate Water Cycle Components in an Era of Advanced Data Availability and Model Sophistication”. In: *Frontiers in Water* 4 (2022). ISSN: 2624-9375.
- [20] E. P. W. Attema and F. T. Ulaby. “Vegetation Modeled as a Water Cloud”. In: *Radio Science* 13.2 (Mar. 1978), pp. 357–364. ISSN: 1944-799X. DOI: [10.1029/RS013i002p00357](https://doi.org/10.1029/RS013i002p00357).
- [21] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote Sensing: Active and Passive. Volume 3 - From Theory to Applications*. Jan. 1986.
- [22] D. Rains, H. Lievens, G. J. M. De Lannoy, M. F. McCabe, R. A. M. de Jeu, and D. G. Miralles. “Sentinel-1 Backscatter Assimilation Using Support Vector Regression or the Water Cloud Model at European Soil Moisture Sites”. In: *IEEE Geoscience and Remote Sensing Letters* (2021), pp. 1–5. ISSN: 1558-0571. DOI: [10.1109/LGRS.2021.3073484](https://doi.org/10.1109/LGRS.2021.3073484).
- [23] F. T. Ulaby, K. Sarabandi, K. McDonald, M. Whitt, and M. C. Dobson. “Michigan Microwave Canopy Scattering Model”. In: *International Journal of Remote Sensing* 11.7 (July 1990), pp. 1223–1253. ISSN: 0143-1161. DOI: [10.1080/01431169008955090](https://doi.org/10.1080/01431169008955090).
- [24] A. Davitt, J. M. Winter, and K. McDonald. “Integrated Crop Growth and Radiometric Modeling to Support Sentinel Synthetic Aperture Radar Observations of Agricultural Fields”. In: *Journal of Applied Remote Sensing* 14.4 (Nov. 2020), p. 044508. ISSN: 1931-3195, 1931-3195. DOI: [10.1117/1.JRS.14.044508](https://doi.org/10.1117/1.JRS.14.044508).
- [25] B. A. Forman and R. H. Reichle. “Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.9 (Sept. 2015), pp. 4431–4441. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2014.2325780](https://doi.org/10.1109/JSTARS.2014.2325780).
- [26] X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. “Towards Constraining Soil and Vegetation Dynamics in Land Surface Models: Modeling ASCAT Backscatter Incidence-Angle Dependence with a Deep Neural Network”. In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. ISSN: 0034-4257. DOI: [10.1016/j.rse.2022.113116](https://doi.org/10.1016/j.rse.2022.113116).

- [27] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. M. De Lannoy. “Optimisation of AquaCrop Backscatter Simulations Using Sentinel-1 Observations”. In: *Remote Sensing of Environment* 294 (Aug. 2023), p. 113621. ISSN: 0034-4257. DOI: [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621).
- [28] <https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp->
- [29] *KNMI*. <https://daggegevens.knmi.nl/>.
- [30] M. Shorachi, V. Kumar, and S. C. Steele-Dunne. “Sentinel-1 SAR Backscatter Response to Agricultural Drought in The Netherlands”. In: *Remote Sensing* 14.10 (Jan. 2022), p. 2435. ISSN: 2072-4292. DOI: [10.3390/rs14102435](https://doi.org/10.3390/rs14102435).
- [31] P. C. Vermunt, S. C. Steele-Dunne, S. Khabbazan, J. Judge, and N. C. van de Giesen. “Extrapolating Continuous Vegetation Water Content to Understand Sub-Daily Backscatter Variations”. In: *Hydrology and Earth System Sciences* 26.5 (Mar. 2022), pp. 1223–1241. ISSN: 1027-5606. DOI: [10.5194/hess-26-1223-2022](https://doi.org/10.5194/hess-26-1223-2022).
- [32] J. W. Jones, G. Y. Tsuji, G. Hoogenboom, L. A. Hunt, P. K. Thornton, P. W. Wilkens, D. T. Imamura, W. T. Bowen, and U. Singh. “Decision Support System for Agrotechnology Transfer: DSSAT V3”. In: *Understanding Options for Agricultural Production*. Ed. by G. Y. Tsuji, G. Hoogenboom, and P. K. Thornton. Systems Approaches for Sustainable Agricultural Development. Dordrecht: Springer Netherlands, 1998, pp. 157–177. ISBN: 978-94-017-3624-4. DOI: [10.1007/978-94-017-3624-4_8](https://doi.org/10.1007/978-94-017-3624-4_8).
- [33] M. K. van Ittersum and M. Donatelli. “Modelling Cropping Systems—Highlights of the Symposium and Preface to the Special Issues”. In: *European Journal of Agronomy*. Modelling Cropping Systems: Science, Software and Applications 18.3 (Jan. 2003), pp. 187–197. ISSN: 1161-0301. DOI: [10.1016/S1161-0301\(02\)00095-3](https://doi.org/10.1016/S1161-0301(02)00095-3).
- [34] Hoogenboom. *DSSAT*. <https://dssat.net/>.
- [35] C. A. Jones and J. R. Kiniry. *CERES-Maize: A Simulation Model of Maize Growth and Development*. Texas A&M University Press, 1986. ISBN: 978-0-89096-269-5.
- [36] W. Malik, R. Isla, and F. Dechmi. “DSSAT-CERES-maize Modelling to Improve Irrigation and Nitrogen Management Practices under Mediterranean Conditions”. In: *Agricultural Water Management* 213 (Mar. 2019), pp. 298–308. ISSN: 0378-3774. DOI: [10.1016/j.agwat.2018.10.022](https://doi.org/10.1016/j.agwat.2018.10.022).
- [37] I. Ahmed, M. H. ur Rahman, S. Ahmed, J. Hussain, A. Ullah, and J. Judge. “Assessing the Impact of Climate Variability on Maize Using Simulation Modeling under Semi-Arid Environment of Punjab, Pakistan”. In: *Environmental Science and Pollution Research* 25.28 (Oct. 2018), pp. 28413–28430. ISSN: 1614-7499. DOI: [10.1007/s11356-018-2884-3](https://doi.org/10.1007/s11356-018-2884-3).
- [38] T. Brandsma. *Comparison of Automatic and Manual Precipitation Networks in the Netherlands*. KNMI, 2014.

- [39] T. Hengl, J. M. de Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. “SoilGrids250m: Global Gridded Soil Information Based on Machine Learning”. In: *PLOS ONE* 12.2 (Feb. 2017), e0169748. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748).
- [40] *Soilgrid*. <https://soilgrids.org/>.
- [41] J. Koo and J. Dimes. *HC27 Generic Soil Profile Database*. 2013. DOI: [10.7910/DVN/90WJ9W](https://doi.org/10.7910/DVN/90WJ9W).
- [42] K. E. Saxton and W. J. Rawls. “Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions”. In: *Soil Science Society of America Journal* 70.5 (2006), pp. 1569–1578. ISSN: 1435-0661. DOI: [10.2136/sssaj2005.0117](https://doi.org/10.2136/sssaj2005.0117).
- [43] E. Han, A. V. M. Ines, and J. Koo. “Development of a 10-Km Resolution Global Soil Profile Dataset for Crop Modeling Applications”. In: *Environmental Modelling & Software* 119 (Sept. 2019), pp. 70–83. ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2019.05.012](https://doi.org/10.1016/j.envsoft.2019.05.012).
- [44] E. Han, A. Ines, and J. Koo. *Global High-Resolution Soil Profile Database for Crop Modeling Applications*. Oct. 2015. DOI: [10.13140/RG.2.1.1303.8804](https://doi.org/10.13140/RG.2.1.1303.8804).
- [45] J. Morel, D. Parsons, M. A. Halling, U. Kumar, A. Peake, G. Bergkvist, H. Brown, and M. Hetta. “Challenges for Simulating Growth and Phenology of Silage Maize in a Nordic Climate with APSIM”. In: *Agronomy* 10.5 (May 2020), p. 645. ISSN: 2073-4395. DOI: [10.3390/agronomy10050645](https://doi.org/10.3390/agronomy10050645).
- [46] R. S. Loomis and J. S. Amthor. “Yield Potential, Plant Assimilatory Capacity, and Metabolic Efficiencies”. In: *Crop Science* 39.6 (1999), pp. 1584–1596. ISSN: 1435-0653. DOI: [10.2135/cropsci1999.3961584x](https://doi.org/10.2135/cropsci1999.3961584x).
- [47] F.X. López-Cedrón, K. J. Boote, J. Piñeiro, and F. Sau. “Improving the CERES-Maize Model Ability to Simulate Water Deficit Impact on Maize Production and Yield Components”. In: *Agronomy Journal* 100.2 (2008), pp. 296–307. ISSN: 1435-0645. DOI: [10.2134/agronj2007.0088](https://doi.org/10.2134/agronj2007.0088).
- [48] J. Lacasa, T. J. Hefley, M. E. Otegui, and I. A. Ciampitti. “A Practical Guide to Estimating the Light Extinction Coefficient with Nonlinear Models—a Case Study on Maize”. In: *Plant Methods* 17.1 (June 2021), p. 60. ISSN: 1746-4811. DOI: [10.1186/s13007-021-00753-2](https://doi.org/10.1186/s13007-021-00753-2).
- [49] F. Sau, K. J. Boote, W. McNair Bostick, J. W. Jones, and M. Inés Mínguez. “Testing and Improving Evapotranspiration and Soil Water Balance of the DSSAT Crop Models”. In: *Agronomy Journal* 96.5 (2004), pp. 1243–1257. ISSN: 1435-0645. DOI: [10.2134/agronj2004.1243](https://doi.org/10.2134/agronj2004.1243).
- [50] V. Kumar, M. Huber, B. Rommen, and Steele-Dunne, Susan C. *Agricultural SandboxNL Database V1.0*. July 2021. DOI: [10.4121/14438750.V1](https://doi.org/10.4121/14438750.V1).

- [51] V. Kumar, M. Huber, B. Rommen, and S. C. Steele-Dunne. “Agricultural SandboxNL: A National-Scale Database of Parcel-Level Processed Sentinel-1 SAR Data”. In: *Scientific Data* 9.1 (July 2022), p. 402. ISSN: 2052-4463. DOI: [10.1038/s41597-022-01474-4](https://doi.org/10.1038/s41597-022-01474-4).
- [52] https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD.
- [53] R. Touzi, A. Lopes, J. Bruniquel, and P. Vachon. “Coherence Estimation for SAR Imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* 37.1 (Jan. 1999), pp. 135–149. ISSN: 1558-0644. DOI: [10.1109/36.739146](https://doi.org/10.1109/36.739146).
- [54] A. Braun and L. Veci. “TOPS Interferometry Tutorial”. In: *ESA: Paris, France* (2021).
- [55] ESA. *SNAP Toolbox*. <https://step.esa.int/main/download/snap-download/>.
- [56] Y. Kang, M. Özdoğan, S. C. Zipper, M. O. Román, J. Walker, S. Y. Hong, M. Marshall, V. Magliulo, J. Moreno, L. Alonso, A. Miyata, B. Kimball, and S. P. Loheide. “How Universal Is the Relationship between Remotely Sensed Vegetation Indices and Crop Leaf Area Index? A Global Assessment”. In: *Remote Sensing* 8.7 (July 2016), p. 597. ISSN: 2072-4292. DOI: [10.3390/rs8070597](https://doi.org/10.3390/rs8070597).
- [57] S. I. Jiménez-Jiménez, M. d. J. Marcial-Pablo, W. Ojeda-Bustamante, E. Sifuentes-Ibarra, M. A. Inzunza-Ibarra, and I. Sánchez-Cohen. “VICAL: Global Calculator to Estimate Vegetation Indices for Agricultural Areas with Landsat and Sentinel-2 Data”. In: *Agronomy* 12.7 (July 2022), p. 1518. ISSN: 2073-4395. DOI: [10.3390/agronomy12071518](https://doi.org/10.3390/agronomy12071518).
- [58] M. C. Mozer, M. I. Jordan, and T. Petsche. *Advances in Neural Information Processing Systems 9: Proceedings of the 1996 Conference*. MIT Press, 1997. ISBN: 978-0-262-10065-6.
- [59] J. Ezzahar, N. Ouaadi, M. Zribi, J. Elfarkh, G. Aouade, S. Khabba, S. Er-Raki, A. Chehbouni, and L. Jarlan. “Evaluation of Backscattering Models and Support Vector Machine for the Retrieval of Bare Soil Moisture from Sentinel-1 Data”. In: *Remote Sensing* 12.1 (Jan. 2020), p. 72. DOI: [10.3390/rs12010072](https://doi.org/10.3390/rs12010072).
- [60] D. Rains, R. Sabia, D. Fernández-Prieto, M. Marconcini, and T. Katagis. “Extended Analysis of SMOS Salinity Retrieval by Using Support Vector Regression (SVR)”. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. July 2014, pp. 2265–2268. DOI: [10.1109/IGARSS.2014.6946921](https://doi.org/10.1109/IGARSS.2014.6946921).
- [61] C. J. Willmott. “Some Comments on the Evaluation of Model Performance.” In: *Bulletin of the American Meteorological Society* 63 (Nov. 1982), pp. 1309–1369. DOI: [10.1175/1520-0477\(1982\)063<1309:SCOTED>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTED>2.0.CO;2).
- [62] C. Ding and H. Peng. “Minimum Redundancy Feature Selection from Microarray Gene Expression Data”. In: *Journal of Bioinformatics and Computational Biology* 3.2 (Apr. 2005), pp. 185–205. ISSN: 0219-7200. DOI: [10.1142/s0219720005001004](https://doi.org/10.1142/s0219720005001004).
- [63] P. C. Vermunt, S. Khabbazan, S. C. Steele-Dunne, J. Judge, A. Monsivais-Huertero, L. Guerriero, and P.-W. Liu. “Response of Subdaily L-Band Backscatter to Internal and Surface Canopy Water Dynamics”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9 (Sept. 2021), pp. 7322–7337. ISSN: 1558-0644. DOI: [10.1109/TGRS.2020.3035881](https://doi.org/10.1109/TGRS.2020.3035881).

- [64] S.-Y. Hong, K.-A. Sudduth, N.-R. Kitchen, C.-W. Fraisse, H.-L. Palm, and W.-J. Wiebold. “Comparison of Remote Sensing and Crop Growth Models for Estimating Within-Field LAI Variability”. In: *Korean Journal of Remote Sensing* 20.3 (2004), pp. 175–188. ISSN: 1225-6161. DOI: [10.7780/kjrs.2004.20.3.175](https://doi.org/10.7780/kjrs.2004.20.3.175).
- [65] M. H. J. van Huijgevoort, J. A. de Wit, and R. P. Bartholomeus. “Evaluation of 2018 Drought and Effectiveness of Adaptation Measures in the Netherlands”. In: (May 2020), p. 9166. DOI: [10.5194/egusphere-egu2020-9166](https://doi.org/10.5194/egusphere-egu2020-9166).
- [66] G. Wiseman, H. McNairn, S. Homayouni, and J. Shang. “RADARSAT-2 Polarimetric SAR Response to Crop Biomass for Agricultural Production Monitoring”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.11 (Nov. 2014), pp. 4461–4471. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2014.2322311](https://doi.org/10.1109/JSTARS.2014.2322311).
- [67] Y. Morishita and R. F. Hanssen. “Temporal Decorrelation in L-, C-, and X-band Satellite Radar Interferometry for Pasture on Drained Peat Soils”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.2 (Feb. 2015), pp. 1096–1104. ISSN: 1558-0644. DOI: [10.1109/TGRS.2014.2333814](https://doi.org/10.1109/TGRS.2014.2333814).
- [68] M. Vreugdenhil, C. Navacchi, B. Bauer-Marschallinger, S. Hahn, S. Steele-Dunne, I. Pfeil, W. Dorigo, and W. Wagner. “Sentinel-1 Cross Ratio and Vegetation Optical Depth: A Comparison over Europe”. In: *Remote Sensing* 12.20 (Jan. 2020), p. 3404. ISSN: 2072-4292. DOI: [10.3390/rs12203404](https://doi.org/10.3390/rs12203404).

4

PHYSICS-GUIDED MACHINE LEARNING BASED FORWARD-MODELING OF RADAR OBSERVABLES: A CASE STUDY ON SENTINEL-1 OBSERVATIONS OF CORN-FIELDS

Artificial neural networks have the potential to model the interaction of radar signals with vegetation but often do not follow the physical rules. This paper aims to develop a new physics-guided machine learning approach that combines neural networks and physics-based models to leverage their complementary strengths and improve the modeling of physical processes. We propose a data-driven framework to model Synthetic Aperture Radar (SAR) observables by incorporating physical knowledge in two ways: through the network architecture and the loss function. A key aspect of our approach is its ability to integrate knowledge encoded in physics-based models. The results show that by using scientific knowledge to guide the construction and learning of the neural network, we can provide a framework with better generalizability and stability.

Parts of this chapter have been published in: Nikaein, T., Lopez-Dekker, P., 2025. Physics-Guided Machine Learning Based Forward-Modeling of Radar Observables: a Case Study on Sentinel-1 Observations of Corn-Fields. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

4.1. INTRODUCTION

Forward models, or observation operators, are important for the analysis and interpretation of remote sensing data, for the conceptualization and development of observational concepts, and for the assimilation of measurements in numerical models. As in many other fields of study, traditional approaches use either simple empirical models, simplified physical models, or a mixture of both (e.g. physical models with empirically tuned parameters). While robust and easy to work with, these types of models often fail to account for many of the phenomena present in the full physical system.

To address this shortcoming, remote sensing scientists are increasingly adopting machine learning (ML) algorithms. These algorithms can learn complex relations and patterns that are not well captured by theoretical models due to the complexity of the underlying physics [1, 2]. On the other hand, supervised machine learning algorithms, particularly deep learning networks, often require vast amounts of training data and can sometimes yield results that, while statistically accurate, may not always align with physical laws [3]. Another issue with purely data-driven models is that they may work well for the region of input vector-values covered by the training data, but often fail to generalize for input values outside this region [4].

An emerging trend in physical sciences is to use the robust theoretical foundations of physics to guide and constrain machine learning models, leading to predictions that are not only more accurate but also physically plausible. In this work, we follow this approach to model satellite-based radar observables over crop fields. While we apply the methodology to the particular case of Sentinel-1[5] Normalized Radar Cross Section (NRCS) over corn fields, the approach should apply to other crops and other observables.

Recently, there has been an increasing interest in the integration of physics with machine learning, as discussed in detail in [6, 7]. Previous studies demonstrated that physics-informed machine learning can improve the accuracy and generalizability of the model in different applications. For example, in [8] an improvement in the prediction of the chemical reflectance signature was studied using a physics-guided neural network (PGNN). In another study to predict lake temperature [9], PGNN was used in two different approaches; 1) the simulated output of the physics-based model was fed into the neural network as additional inputs and 2) including physics knowledge into the loss function. Their results showed better accuracy and lower physical inconsistency. Jia et al. [10] pre-trained a model using simulated data from a generic physics-based model to improve prediction accuracy with limited observed data. An effective method for guiding the initialization process to aid in model training and avoiding local minima is to employ transfer learning, an ML approach. With transfer learning, a model can be first pre-trained using simulated data from a physics-based model and subsequently fine-tuned with a limited amount of training data to adapt to the specific task at hand. Their results show that using physical model data for pre-training, even with imperfect parameters, can reduce the training data requirements. They incorporate the knowledge encoded in the physical model with a recurrent neural network (RNN) model to leverage their complementary strengths to predict lake water temperature. Zhong et al. [11] developed a physics-informed deep

learning model to simulate runoff changes in alpine catchments under climate change, outperforming traditional models. Their model combines deep learning techniques with the physics of hydrological processes, providing more credible projections. Previous studies showed the efficiency of including physical knowledge into the architecture of the model, for example in [12, 13], where prior system knowledge was incorporated in the architecture, for lake temperature modeling and dynamic system modeling, respectively. In all state-of-the-art applications, the integration of ML with physical knowledge has led to significant improvements in adaptability. To the best of our knowledge, the application of these advancements to modeling microwave signals from vegetation remains unexplored. This gap presents a unique opportunity for research, where methodologies developed in other contexts could be adapted to enhance the accuracy and efficiency of remote sensing in vegetative environments.

The advantages of direct assimilation of microwave satellite observation, which circumvents the need for retrievals, were studied in [14, 15]. Forward models are required to map biogeophysical parameters to satellite observations. In the context of this paper, our aim is to model the dependence of NRCS, σ_0 , on crop and soil-related parameters, in order to use this model for direct assimilation. A commonly used radiative transfer model is the so-called water cloud model (WCM) [16], which often serves as a forward operator. There are more complex models to simulate radar backscatter, like the Tor Vergata model [17] and the Michigan microwave canopy scattering model (MIMICS) [18]. While these models offer detailed simulations, their use is often limited due to the difficulty in parameterizing them accurately. These models require a large number of input parameters, many of which are challenging to measure or estimate with high precision in real-world conditions. More recently, the advantage of using machine learning as an observation operator has been studied by [19–22]. The challenge in accurately predicting NRCS lies in the complex interplay of numerous factors influencing the returned signal, including surface roughness, moisture content, vegetation cover, and geometric properties of the observed scene. The study by Nikaein et al. [22], highlights the difficulties in using data-driven models alone to simulate Synthetic Aperture Radar (SAR) observables, such as backscatter, during anomalous conditions for example drought years. This challenge arises when the model encounters scenarios for which it has not been trained, such as vegetation parameters under dry conditions, resulting in predictions that are not representative of the actual conditions. This research builds on these insights and seeks to address these limitations by demonstrating how the integration of physical knowledge into ML models can improve their performance as an observation operator. Specifically, the focus is on the development of physics-guided machine learning frameworks that incorporate domain knowledge into both the architecture and learning process of ML models. By doing so, this approach not only ensures that the predictions remain consistent with fundamental physical principles but also enhances the ability of the model to generalize to previously unseen conditions, such as drought or other environmental anomalies. We used machine learning as an observation operator to map biogeophysical parameters from crop growth models, such as the Decision Support System for Agrotechnology Transfer (DSSAT), to SAR observables. This approach provides a more robust framework for simulating SAR observables.

The main aim of the paper is to investigate how incorporating physical constraints into ML models can enhance their robustness, generalizability, and interpretability. Rather than focusing on developing the best-performing ML model, our objective is to understand how physical principles can be embedded into the modeling process. To achieve this, we explore two complementary approaches. The first approach incorporates physical knowledge directly into the learning process by adding a custom constraint to the loss function. A positive gradient constraint enforces consistency with the expected relationship between radar backscatter and soil moisture. The second approach mirrors the structure of the WCM, decomposing radar backscatter into physically meaningful components (soil and vegetation) and constraining their contributions to the total signal. This implementation enables explicit modeling of attenuation effects, providing deeper insights into the physical processes governing SAR observables.

The contributions of this work are multifold:

1. We demonstrate that incorporating physical constraints improves model robustness and transferability across different environmental conditions, such as year-to-year variability.
2. We highlight the potential of gradient-based constraints, as an effective way to embed physical principles into ML models for modeling radar observables.
3. We show that mirroring the WCM structure in the neural network architecture allows for intermediate outputs that are physically interpretable, enabling the analysis of specific contributions from soil and vegetation to backscatter.
4. By using synthetic data derived from the WCM, we validate the accuracy and behavior of the proposed models under controlled conditions, providing a benchmark for real-world applications.

This paper provides a practical framework for integrating physical principles into ML models, with the dual goal of improving performance and gaining deeper insights into the underlying processes. The findings contribute to advancing the state of physics-informed machine learning and its applications in remote sensing.

4.2. STUDY AREA AND DATA

Building on the work presented in [22], we select maize fields in the province of Noord-Brabant, The Netherlands for our study. The CERES (Crop Environment Resource Synthesis)-Maize model, which is among the various crop models included in the DSSAT [23, 24], were used to simulate crop growth for each field. This model uses input data on soil characteristics, climatic conditions, crop genetics, and management practices to simulate daily growth stages, biomass development, and crop yield. Our research focuses on key biophysical parameters of maize, such as the Leaf Area Index (LAI), Above-Ground Biomass (AGB), surface soil moisture (SM_S), and root zone soil moisture (SM_R), to simulate SAR observables. More details, including details about the study area and the crop growth modeling steps, can be found in [22].

Following [22], we utilize Sentinel-1 C-band data acquired in the Interferometric Wide Swath (IW) mode with a 6-day repeat cycle (relative orbit 37). NRCS values in both VV and VH polarizations were retrieved from the Agricultural SandboxNL database[25]. In order to validate the rationale of our proposed approach, we generate synthetic data based on the principles of WCM. This synthetic data allows us to test and validate the accuracy, robustness, and ability of the model to generalize across different scenarios in an idealized case.

4.3. METHODOLOGY

In this section, we describe the two main approaches to incorporate physical knowledge into neural networks: 1) physics-based loss function and 2) physics-guided network topology.

4.3.1. PHYSICS-BASED LOSS FUNCTION

One way to incorporate physical knowledge into ML algorithms is by enforcing constraints on the outputs of these algorithms. This can be accomplished through the formulation of a custom loss function, as shown in [9]. Consider a learning system characterized by a function f , which operates on a set of input parameters \mathbf{X} that possess a physical relationship with the target variable Y . In this context, we can express the relationship as $\hat{Y} = f_{\text{NN}}(\mathbf{X})$, where f_{NN} denotes the neural network function approximating the mapping from \mathbf{X} to \hat{Y} across our training samples. In the conventional training paradigm, the goal is to minimize the discrepancy between the predicted values (\hat{Y}) and the observed values (Y). However, while this standard approach is effective in reducing predictive error, it may not ensure that predictions are in accordance with the underlying physical principles. To address this gap, the custom loss function comes into play, integrating physical constraints directly into the learning process. By doing so, the loss function not only penalizes deviations from observed data points but also incorporates penalties for violations of known physical laws. This dual-purpose loss function ensures that the learning process is not merely data-driven but is also guided by the underlying physical principles. The modified learning objective, incorporating this physical constraint, is defined as

$$\arg \min_f (L_{\text{data}}(Y, \hat{Y}) + \lambda L_{\text{phys}}(\hat{Y}))$$

with

$$L_{\text{data}}(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.1)$$

where f is the model, L_{data} is the data term of the loss function, L_{phys} is the physical constraint term and λ is a hyperparameter that balances the contributions of L_{data} and L_{phys} to the overall loss function. For the data term of the loss function, we use the mean square error (MSE) function. To operationalize this physical constraint within our machine learning model, we developed a custom loss function with an additional term specifically designed to ensure the partial derivative of backscatter with respect to surface soil moisture remains positive [26], as

$$L_{\text{phys}}(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n \text{ReLU}\left(-\frac{\partial \hat{y}_i}{\partial \text{SM}_s}\right), \quad (4.2)$$

where index i iterates over the training samples, and $\text{ReLU}(\cdot)$ is a Rectified Linear Unit function applied within the L_{phys} term to enforce the positive partial derivative constraint by penalizing negative values of the predicted partial derivative.

This adjustment is critical for maintaining the physical integrity of the predictions of the model, ensuring they are consistent with the known behavior of microwave radar signals interacting with varying levels of soil moisture.

4.3.2. PHYSICS GUIDED NETWORK TOPOLOGY

In this case, we are constraining the internal architecture of ML models with physical insights to enhance their interpretability. This approach involves integrating physical principles directly into the structure of neural network, as shown in Fig. 4.1. By doing so, we give an implicit physical meaning to some intermediate outputs, which also means that we can apply physical constraints to them. In this framework, we try to limit the existing freedom of a standard neural network to simulate NRCS. For this approach, we tried two steps: 1) bound the freedom of the model in the architecture and 2) incorporate the physical knowledge through the loss function. The network topology tested follows the architecture of the WCM [16], a widely used model for backscatter. Here, the total NRCS during the growth period (from planting to harvest) is decomposed in a vegetation component, and underlying soil term, and a term representing their interaction. The general form of the WCM equations is represented in equations (4.3) to (4.6), where the WCM neglects the interactions between the ground and vegetation, implicitly assuming that it is small compared to the other terms.

$$\sigma_{\text{total}}^0 = \sigma_{\text{veg}}^0 + \sigma_{\text{soil}}^0 T^2, \quad (4.3)$$

$$\sigma_{\text{veg}}^0 = AV_1 \cos\theta(1 - T^2) \quad (4.4)$$

$$T^2 = \exp\left(\frac{-2BV_2}{\cos\theta}\right) \quad (4.5)$$

$$\sigma_{\text{soil}}^0 = C + D \cdot \text{SM}_s \quad (4.6)$$

Here, θ is the incidence angle, σ_{total}^0 is the total backscattering coefficient, σ_{veg}^0 is the backscatter contribution from the vegetation, σ_{soil}^0 is the backscatter contribution from the soil and T^2 is the two-way transmissivity of the vegetation layer. There are more sophisticated ways to estimate σ_{soil}^0 (e.g. [27]) but generally, it is a reasonable assumption that there is a linear relationship between backscattering coefficient and soil moisture over bare soil. σ_{soil}^0 influenced by soil moisture, surface roughness, and the incidence angle of the radar signal. The attenuation term depends on the density and water content of the vegetation. Scattering from the vegetation depends on vegetation water content, structure, and orientation of leaves and stems [28]. The WCM contains four coefficients, A, B, C and D related to vegetation scattering, vegetation

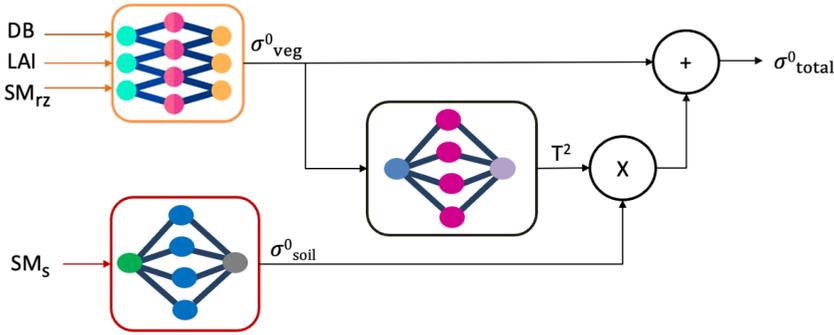


Figure 4.1.: A depiction of the WCM-inspired NN, where prior knowledge from physics is embedded in a structured format.

attenuation, surface roughness, and soil moisture respectively. All of these coefficients are polarization- and frequency-dependent. A and B depend on vegetation type, while C and D are related to soil texture. Several quantities can be used to describe the vegetation by setting V_1 and/or V_2 to quantities such as vegetation water content (VWC), vegetation optical depth (VOD), AGB or LAI. As LAI is readily available from the DSSAT model outputs, we follow [29, 30] by assuming $V_1 = V_2 = \text{LAI}$.

We defined a multiple input network that separates the inputs for the vegetation component from the inputs for the soil component. The vegetation components inputs are LAI, AGB, and SM_R (as an indication of the availability of root zone moisture to replace transpired water) and SM_S is the input for the soil component. As illustrated in Fig. 4.1, the network produces internal outputs denoted as σ_{veg}^0 and σ_{soil}^0 . We anticipate that any increase in σ_{veg}^0 will result in a diminished direct impact from σ_{soil}^0 . This modulation effect is represented by the term T^2 within the framework and multiplied by NRCS from the soil term. Note that in the attenuation term of the constrained model architecture, we prescribe a behavior inspired by the reverse exponential nature of attenuation and vegetation effects as it is shown in equation (4.7). This term has a trainable coefficient α , which is defined by the training procedure.

$$T^2 = \exp(-\alpha \sigma_{\text{veg}}^0) \quad (4.7)$$

The proposed framework incorporates physics-based prior knowledge into the structure of the neural network and imposes constraints on the internal states of the model and the output values. In this architecture, the loss function constraint is applied to σ_{soil}^0 rather than σ_{total}^0 . This choice is based on the observation that the partial derivative with respect to SM_S for σ_{soil}^0 is consistently positive. The sensitivity of σ_{total}^0 to SM_S can approach zero when the LAI is high, so in the presence of noise, this can lead to excessive penalties for slightly negative values.

The remainder of the paper is devoted to analyzing the outcomes, testing, and comparing the four combinations discussed: Standard NN and WCM-inspired NN, each with either a regular or modified loss function. As mentioned in Section 4.2, synthetic data were generated using the WCM formula, which allows the extraction of

intermediate layer output. This data enables us to validate the performance of the WCM-inspired NN model. In this study, we formulate three key hypotheses that we aim to test and validate through our results:

1. When training and testing on consistent data sets, adding any constraints would lead to a higher MSE, given that this is our default (unconstrained) loss function.
2. Both our positive partial derivative and topology constraint should enhance the robustness of the model and, consequently, its transferability.
3. We anticipate that our WCM-inspired neural network will learn to produce meaningful intermediate results.

4

4.3.3. NN IMPLEMENTATION

We implemented the neural network models using the Keras library [31]. The dataset was partitioned field-wise into training and testing subsets, following a 70:30 ratio to ensure independence between the two datasets. We used the Adaptive Moment Estimation (Adam) optimization algorithm to minimize the loss function. To mitigate the risk of overfitting, we incorporated an early stopping mechanism, adjusting the patience parameter to 50. The input features were rescaled between 0 and 1 to prevent saturation at the tails of the activation functions; the same transformation was applied to the test data. Our fully connected network architecture consists of 3 hidden layers and neuron numbers of 32, 16, 4 with ReLU activation functions in each hidden layer. In the constrained model architecture, ReLU was used in the hidden layers, while the sigmoid activation function was used in each output layer to produce outputs between 0 and 1. The weights of the neural network were randomly initialized and each experiment was run 50 times where the standard deviation of accuracy was around 2%.

We present results and compare outcomes in the following scenarios: standard NN, standard NN with constraints, WCM-inspired NN, and WCM-inspired with constraints.

4.4. EXPERIMENT WORKFLOW

In this section, we present the results from each approach for two scenarios: 1) when the model is tested on data from the same year, incorporating the environmental conditions present in the training phase, and 2) when the model is tested on data from a different year, which includes different environmental conditions not represented in the training data. The details of these environmental differences are detailed in [22]. Fig. 4.2 illustrates the general workflow of our proposed framework, demonstrating the simulation of NRCS using machine learning for corn parcels. The dashed arrows in this figure highlight the added value of this method to the data-driven study, enhancing it with physical principles not addressed in our previous work [22].

As mentioned in the previous section, these models aim to improve transferability. Specifically, the proposed method is tested in two distinct situations: training and testing on the same year and testing on a different year with varying meteorological conditions. The lack of separate signals for soil and vegetation in satellite data limits the ability to validate the model effectively. To address this limitation, synthetic data can be

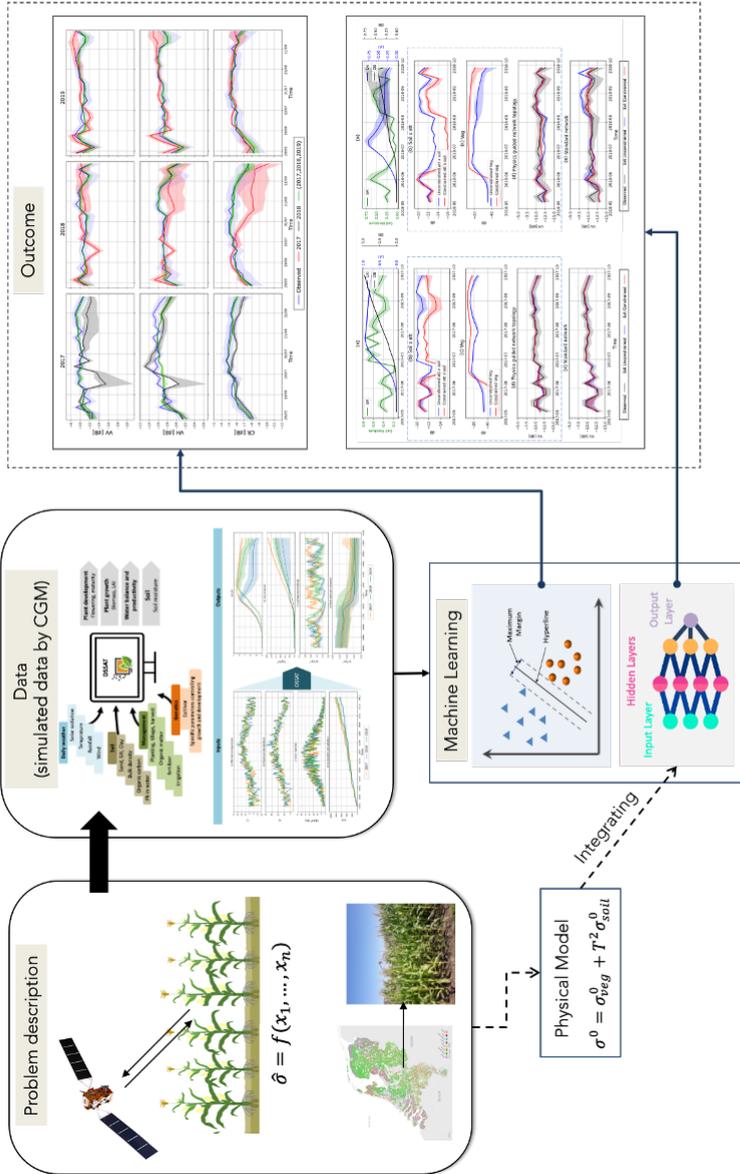


Figure 4.2.: A general workflow of our proposed framework. The simulated data from the crop growth model was fed to the model. The details of data simulation and modeling the NRCS using machine learning are discussed in [22]. (The map of The Netherlands is sourced from [25]).

used to understand better whether the model behaves as expected. We expect that the model performs optimally under these controlled conditions. We generate synthetic data by using soil and vegetation states modeled by DSSAT for both years as inputs to the WCM. The values for WCM parameters were optimized by the range provided in [21].

4.5. RESULTS

4.5.1. SYNTHETIC DATA

Fig. 4.3 shows the results obtained by training the physics-guided model on synthetic data for 2017, testing on data for the same year Fig. 4.3(i) and on data for 2018 Fig. 4.3(ii). This figure illustrates how constraints on the network topology by physical properties such as soil moisture and vegetation characteristics can improve remote sensing models' accuracy. The figure provides insight into the contributions to total backscatter. Fig. 4.3(ii)a illustrates a rapid decrease in LAI during July, along with lower maximum AGB and greater variability across the parcels (indicated by the shaded area). In Fig. 4.3(ii)b, $T^2\sigma_{\text{soil}}^0$ is overestimated in late-June/early-July; however, this has limited impact on σ_{total}^0 since σ_{veg}^0 dominates during this period. Similarly, while σ_{veg}^0 is overestimated in June, it has a negligible effect on σ_{total}^0 because σ_{veg}^0 is much smaller than σ_{soil}^0 at that time. The graph highlights that the largest deviation between the estimated and synthetic truth σ_{total}^0 occurs from mid-July onwards, coinciding with an anomaly in the LAI, which leads to a poor estimate of σ_{veg}^0 . However, this difference remains minimal, as shown in Fig. 4.3(ii)(d). The inclusion of a constraint in the loss function helps to reduce this deviation, bringing the estimate slightly closer to the synthetic truth, and improving the accuracy of the model despite the anomaly. In subplot (c), while the difference between synthetic (WCM) and estimated backscatter looks large prior to June 2017, note that in linear units, these initial values are all close to zero, so the absolute difference is very small.

The physics-guided network topology (Fig. 4.3d) with $R^2 \approx 0.99$ and an MSE close to zero, effectively captures the interactions between the soil and attenuation (Fig. 4.3b) and the vegetation (Fig. 4.3c). This integration is consistent with the principles of the WCM. The variability and trends observed in Fig. 4.3d are direct results of the dynamic changes in soil moisture and vegetation properties captured in Fig. 4.3b and c.

Maybe the most salient observation is that our WCM-inspired model produces accurate values of σ_{soil}^0 and σ_{veg}^0 without being trained with corresponding data. This happens because the WCM-inspired topology cannot *find* another way to minimize the MSE in a situation where the NN topology is a perfect match to the actual model. It would be risky to conclude that the same behavior will automatically happen with real data given the simplifying assumptions of the WCM.

4.5.2. SENTINEL-1 DATA

After assessing the performance of the proposed method on synthetic data, we now proceed to evaluate its performance using satellite data. Fig. 4.4 provides a comparison of the NRCS in VV polarization between the two approaches that incorporate physical

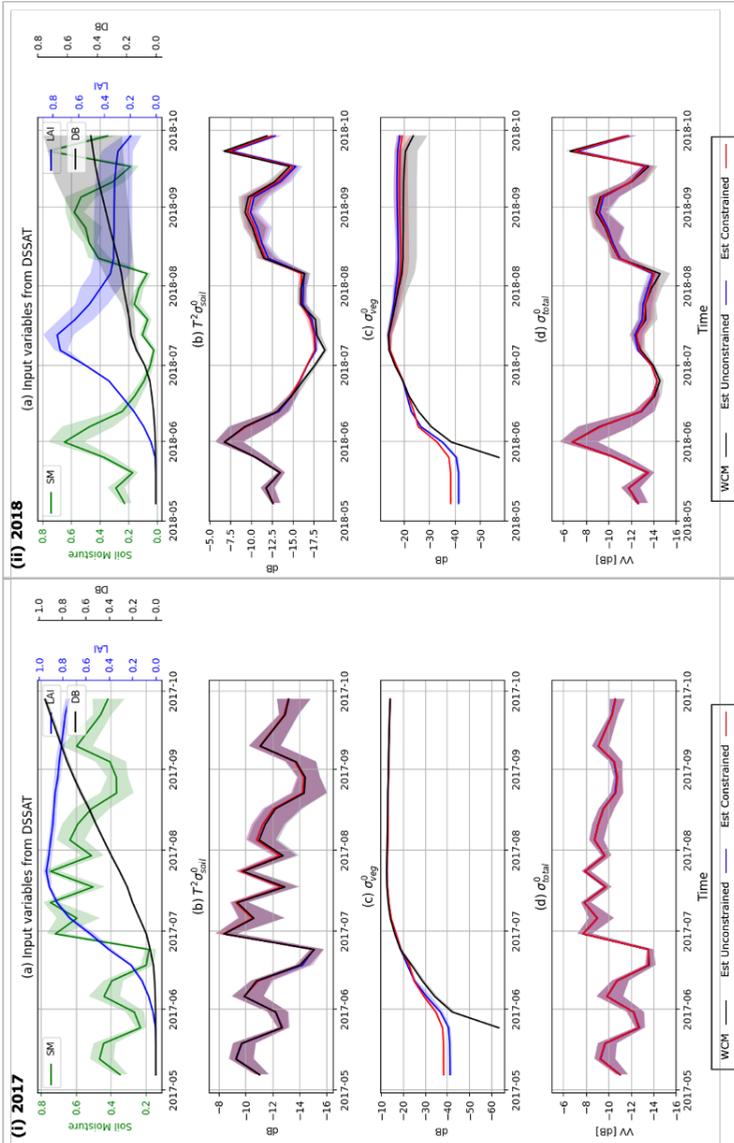


Figure 4.3.: Time series of synthetic and estimated NRCS values are presented, based on training using synthetic data generated from 2017 vegetation and soil parameters. The models were tested on (i) synthetic data from the same year (2017) and (ii) synthetic data generated from 2018 vegetation and soil parameters. (b) to (d) illustrate the following: blue lines represent unconstrained estimations, red ones indicate constrained estimations and the black lines represent the estimated backscatter from WCM.

knowledge into neural networks, with each model trained and tested on data from 2017. Fig. 4.4a illustrates the simulated biogeophysical parameters of maize fields in 2017, which are important to understand the subfigures (b), (c), and (d). Fig. 4.4 (b-d) show the $T^2\sigma_{\text{soil}}$, σ_{veg} and σ_{tot} generated by our trained WCM-inspired neural network. Fig. 4.4e shows the total NRCS for the *regular* fully-connected dense network.

Fig. 4.4b shows the contributions of NRCS from the interaction between soil and attenuation. The blue line corresponds to the model trained without physical constraints, while the red line corresponds to the model trained with the modified loss function. Both models follow a similar trend, but the constrained model shows more variability, particularly from early June onward. This suggests that the constrained model is more responsive to fluctuations in soil moisture. Fig. 4.4c shows the NRCS contributions from vegetation. Both models show an increasing trend from June, are aligned with the growing season, and follow the same behavior as LAI. Fig. 4.4d shows the overall NRCS simulated by the physics-guided network topology. The variability seen in subplot (b) due to soil and attenuation interactions directly impacts the overall NRCS in subplot (d). Peaks and troughs in soil moisture (subplots (a) and (b)) correspond to similar variations in the total NRCS, highlighting the sensitivity of the model to soil moisture dynamics. The increasing trend observed in subplot (d) from mid- to late-2017 aligns with the progression of the growth season, as indicated by increasing LAI and AGB in subplot (a). This demonstrates that the physics-guided network effectively integrates the seasonal growth patterns of vegetation and changes in soil conditions. Fig. 4.4e illustrates the NRCS values obtained using a standard neural network. It compares the observed NRCS (black line) with estimates from the standard neural network without constraints (blue line) and with constraints (red line). The constrained model shows a closer alignment with the observed NRCS data. The constraint enforces a positive gradient with respect to soil moisture, improving the responsiveness of the model to changes in soil conditions. Interestingly, the total NRCS from the WCM-inspired NN does not improve much with the additional constraint. However, the primary effect of incorporating the constraint into the loss function is to reduce the $T^2\sigma_{\text{soil}}^0$ term and increase σ_{veg}^0 as soon as LAI begins to rise. Notably, throughout most of the growing season, the estimates of σ_{total}^0 from both the WCM-inspired NN and the standard NN are quite similar. The notable exception occurs during the bare soil period, where the standard NN appears to perform slightly better.

Now we turn our attention to how the different models, trained on 2017 data, behave and perform on test data corresponding to 2018. As discussed before, the 2018 period includes combinations of input values not seen in 2017. This cross-year analysis helps to evaluate the transferability and robustness of the model under different environmental conditions.

Fig. 4.5 shows the model inputs and outputs in this case. Fig. 4.5a presents the DSSAT generated SM_s , LAI, and AGB over time in 2018. A comparison with Fig. 4.4a immediately reveals the severe drought during the summer months, as noted in [32], and its impact on the LAI and AGB. Panel (b) shows that the constrained WCM-inspired model predicts significantly lower values for the $T^2\sigma_{\text{soil}}^0$ term during the drought period. Subplot (c) shows that, as in the 2017 case, the constrained model produces higher contributions of the vegetation to the total NRCS and, more importantly, a sharper

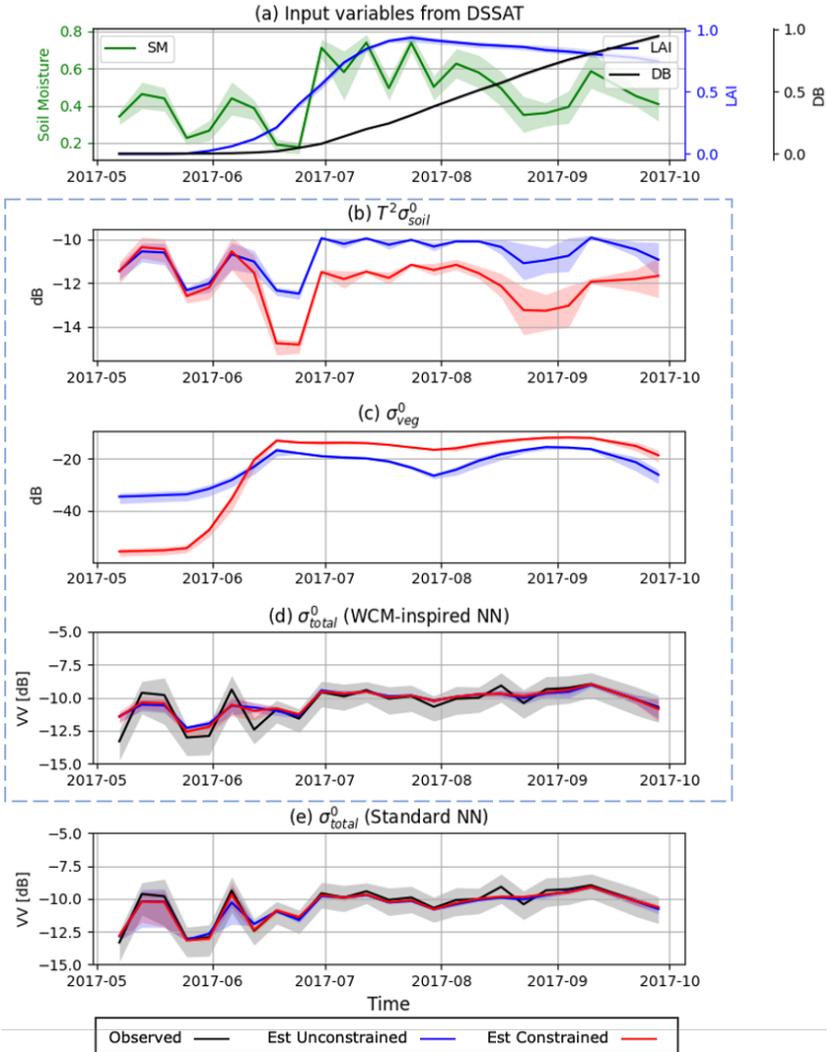


Figure 4.4.: Time-series of observed and estimated NRCS for training and test data on the same year data (2017). Solid lines indicate the mean value of the feature over maize parcels in the test set, and the bounded area shows the 20th-80th percentiles. (a) Vegetation and soil parameters during the growing season in daily steps in 2017. (b) to (e) illustrate the following: blue lines represent unconstrained estimations, red ones indicate constrained estimations, in (d) and (e), the black lines represent the observed backscatter. The dashed rectangle around subplots (b), (c), and (d) highlights that these are related to the physics-guided network topology method.

contrast between the bare-soil period and the growth period. The estimated NRCS without constraint in Fig. 4.5d follows the observed data but shows some deviations, especially between mid-June and August. The constrained model aligns well with the observed data, demonstrating improved accuracy in capturing NRCS variations due to changes in soil moisture and vegetation. The NRCS values obtained using the standard neural network are shown in Fig. 4.5e. In the standard NN, the estimate is poor during the LAI anomaly, and the constraint brings the estimate closer to the observed σ_{total}^0 . When both NNs include a constraint in the loss function, the overall performance is better for the WCM-inspired NN.

Table 4.1 provides the performance metrics for different models for both years. As it was shown and discussed in [22], if the trained model is applied to sets of inputs corresponding to conditions (e.g. severe drought) for which it has not been trained we expect decreased model performance. The fully connected dense neural network demonstrates higher accuracy in predicting NRCS values when trained and tested on data from the same year. Although constraints improve the performance of the standard network, the improvement is less pronounced than in the physics-guided network case. When trained on 2017 data and tested on 2018 data, the WCM-inspired network predicts NRCS values with a higher accuracy. The correlation coefficients for 2018 are relatively low. However, it is crucial to emphasize that these metrics are calculated for the entire period under consideration. Notably, a significant improvement in these metrics is observed during the anomaly in late summer 2018. Constraints further improve both models, but the physics-guided network remains superior in terms of accuracy and robustness.

The results indicate that our physics-constrained method improves the forward modeling of SAR observables, demonstrating the potential to combine machine learning with scientific knowledge for advanced remote sensing applications.

4.6. CONCLUSION

In this paper, we introduced a physics-guided neural network to model SAR observables over vegetation. Unlike traditional black-box neural networks, our approach integrates physical principles directly into the network architecture and the loss function, resulting in a model that is not only data-driven but also physically consistent. Specifically, we incorporated physics-guided constraints into the neural network by (1) adding a physics-based term to the loss function and (2) modifying the network architecture to reflect the underlying physical processes.

Our proposed network topology follows that of the WCM, which is widely used to model NRCS over crop fields. Through data-driven training, the model learns behaviors that are not reflected in a standard analytical WCM formulation. At the same time, the model inherits some simplifications embedded in the WCM. For example, the WCM does not represent double bounce (e.g. stem-ground) scattering. It is therefore possible that a network topology inspired by more sophisticated physical models would produce better results. However, a more complex topology may gravitate towards a fully connected dense NN, which as our results showed, is harder to train and generalizes worse.

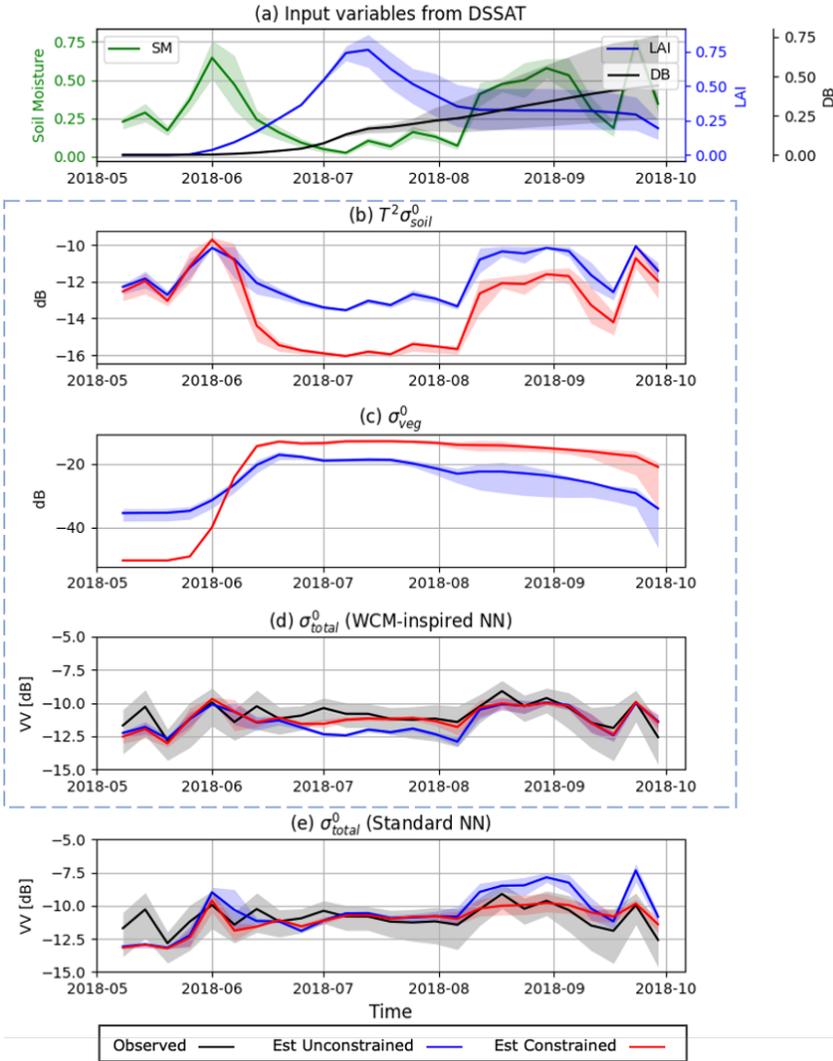


Figure 4.5.: Time-series of observed and estimated NRCS for training on 2017 data and testing on 2018 data. (a) Vegetation and soil parameters during the growing season in daily steps in 2018. (b) to (e) illustrate the following: blue lines represent unconstrained estimations, red ones indicate constrained estimations, in (d) and (e), the black lines represent the observed backscatter. The dashed rectangle around subplots (b), (c), and (d) highlights that these are related to the physics-guided network topology method.

Table 4.1.: Comparison of model's performance with and without constraints for testing on data from 2017 and 2018.

Year	Model	Constrained	MAE	MSE	R2	Pearson	Spearman
2017	Standard	No	1.06	1.83	0.45	0.67	0.62
		Yes	1.03	1.76	0.47	0.69	0.63
	Topology	No	1.18	2.28	0.32	0.60	0.54
		Yes	1.14	2.16	0.35	0.62	0.58
2018	Standard	No	1.64	4.43	-0.35	0.34	0.33
		Yes	1.39	3.32	-0.01	0.30	0.31
	Topology	No	1.41	3.18	0.03	0.34	0.33
		Yes	1.33	2.94	0.10	0.38	0.36

For our physics-guided loss function, we added a single constraint, requiring the partial derivative of the NRCS with respect to the soil moisture to be positive. This constraint makes sense from a physical modeling point of view and is consistent with the WCM topology. It is also a very simple constraint, and therefore easy to implement, which we assume to be valid at all times. However, in reality, there can be a correlation between the SM_S and the VWC which can affect the NRCS in complex ways. More importantly, there is additional physical knowledge that could be incorporated into the loss function.

Back to our key hypotheses, according to our results, we can see that:

1. In line with expectations, the WCM-inspired model does perform worse than the regular dense network when trained and tested on data from the same year. The WCM-inspired limits what the model can learn, for example, hidden correlations between SM_S and VWC. However, contrary to our expectations, but in line with the literature, the results produced by the constrained models have a lower MSE for both training and test datasets. This suggests that these constraints not only prevent overfitting but also aid the training process. Our interpretation is that constraints can effectively reshape the loss landscape, potentially smoothing out poor local minima and leading the optimizer toward more generalizable and accurate solutions. Thus, when constraints align well with the underlying physical or statistical realities of the data, they can be a powerful tool to enhance the learning process and overall model performance.
2. As expected, both proposed approaches produce models that generalize better to completely unseen situations.
3. With synthetic data, the results are encouraging; the simplicity of the WCM model suggests that the neural network should learn it perfectly, yet achieving this was nontrivial and required considerable effort. This success highlights the capabilities of the network. With actual data, the intermediate outputs appear reasonable and generally consistent with typical WCM behavior. In particular, the outputs in the constrained case are quantitatively better, indicating a potential advantage of incorporating constraints.

Looking ahead, one promising direction for further research is the application of transfer learning techniques, which could leverage pre-trained models to improve performance on new datasets or under different environmental conditions. In addition, addressing the challenge of local minima in neural network training remains a critical area of focus. Various strategies, including fine-tuning the network architecture and optimizing hyperparameters, may mitigate these issues and lead to more optimal solutions.

We also recommend that future studies research expand this work by extending the area of interest to cover a broader range of agricultural landscapes and by including multiple seasons to better capture seasonal dynamics. Furthermore, future efforts could focus on integrating the developed method in data assimilation or anomaly detection to enhance the practical applications of the model in agricultural monitoring. If possible, simulate the model under controlled conditions in which the expected behavior of the soil and vegetation components is known. This can help identify whether the model systematically biases towards one component or another.

REFERENCES

- [1] B. A. Forman, R. H. Reichle, and C. Derksen. “Estimating Passive Microwave Brightness Temperature Over Snow-Covered Land in North America Using a Land Surface Model and an Artificial Neural Network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.1 (Jan. 2014), pp. 235–248. ISSN: 1558-0644. DOI: [10.1109/TGRS.2013.2237913](https://doi.org/10.1109/TGRS.2013.2237913).
- [2] B. A. Forman and R. H. Reichle. “Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.9 (Sept. 2015), pp. 4431–4441. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2014.2325780](https://doi.org/10.1109/JSTARS.2014.2325780).
- [3] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat. “Deep Learning and Process Understanding for Data-Driven Earth System Science”. In: *Nature* 566.7743 (Feb. 2019), pp. 195–204. ISSN: 1476-4687. DOI: [10.1038/s41586-019-0912-1](https://doi.org/10.1038/s41586-019-0912-1).
- [4] S. Zhong, K. Zhang, M. Bagheri, J. G. Burken, A. Gu, B. Li, X. Ma, B. L. Marrone, Z. J. Ren, J. Schrier, W. Shi, H. Tan, T. Wang, X. Wang, B. M. Wong, X. Xiao, X. Yu, J.-J. Zhu, and H. Zhang. “Machine Learning: New Ideas and Tools in Environmental Science and Engineering”. In: *Environmental Science & Technology* 55.19 (Oct. 2021), pp. 12741–12754. ISSN: 0013-936X. DOI: [10.1021/acs.est.1c01339](https://doi.org/10.1021/acs.est.1c01339).
- [5] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L’Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan. “GMES Sentinel-1 Mission”. In: *Remote Sensing of Environment. The Sentinel Missions - New Opportunities for Science* 120 (May 2012), pp. 9–24. ISSN: 0034-4257. DOI: [10.1016/j.rse.2011.05.028](https://doi.org/10.1016/j.rse.2011.05.028).
- [6] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. “Integrating Physics-Based Modeling with Machine Learning: A Survey”. In: *ArXiv* (Mar. 2020).
- [7] M. J. Zideh, P. Chatterjee, and A. K. Srivastava. “Physics-Informed Machine Learning for Data Anomaly Detection, Classification, Localization, and Mitigation: A Review, Challenges, and Path Forward”. In: *IEEE Access* 12 (2024), pp. 4597–4617. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2023.3347989](https://doi.org/10.1109/ACCESS.2023.3347989).
- [8] C. P. Murphy and J. P. Kerekes. “Physics-Guided Neural Network for Predicting Chemical Signatures”. In: *Applied Optics* 60.11 (Apr. 2021), pp. 3176–3181. ISSN: 1539-4522. DOI: [10.1364/AO.420688](https://doi.org/10.1364/AO.420688).

- [9] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar. *Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling*. Sept. 2021. DOI: [10.48550/arXiv.1710.11431](https://doi.org/10.48550/arXiv.1710.11431). arXiv: [1710.11431](https://arxiv.org/abs/1710.11431) [physics, stat].
- [10] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar. “Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles”. In: *ACM/IMS Transactions on Data Science* 2.3 (May 2021), 20:1–20:26. ISSN: 2691-1922. DOI: [10.1145/3447814](https://doi.org/10.1145/3447814).
- [11] L. Zhong, H. Lei, and B. Gao. “Developing a Physics-Informed Deep Learning Model to Simulate Runoff Response to Climate Change in Alpine Catchments”. In: *Water Resources Research* 59.6 (2023), e2022WR034118. ISSN: 1944-7973. DOI: [10.1029/2022WR034118](https://doi.org/10.1029/2022WR034118).
- [12] A. Daw, R. Q. Thomas, C. C. Carey, J. S. Read, A. P. Appling, and A. Karpatne. “Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling”. In: *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*. Proceedings. Society for Industrial and Applied Mathematics, Jan. 2020, pp. 532–540. DOI: [10.1137/1.9781611976236.60](https://doi.org/10.1137/1.9781611976236.60).
- [13] F. Djeumou, C. Neary, E. Goubault, S. Putot, and U. Topcu. “Neural Networks with Physics-Informed Architectures and Constraints for Dynamical Systems Modeling”. In: *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*. PMLR, May 2022, pp. 263–277.
- [14] H. Lievens, A. A. Bitar, N. E. C. Verhoest, F. Cabot, G. J. M. D. Lannoy, M. Drusch, G. Dumedah, H.-J. H. Franssen, Y. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, M. J. van den Berg, H. Vereecken, J. P. Walker, E. F. Wood, and V. R. N. Pauwels. “Optimization of a Radiative Transfer Forward Operator for Simulating SMOS Brightness Temperatures over the Upper Mississippi Basin”. In: *Journal of Hydrometeorology* 16.3 (June 2015), pp. 1109–1134. ISSN: 1525-7541, 1525-755X. DOI: [10.1175/JHM-D-14-0052.1](https://doi.org/10.1175/JHM-D-14-0052.1).
- [15] H. Lievens, B. Martens, N. E. C. Verhoest, S. Hahn, R. H. Reichle, and D. G. Miralles. “Assimilation of Global Radar Backscatter and Radiometer Brightness Temperature Observations to Improve Soil Moisture and Land Evaporation Estimates”. In: *Remote Sensing of Environment* 189 (Feb. 2017), pp. 194–210. ISSN: 0034-4257. DOI: [10.1016/j.rse.2016.11.022](https://doi.org/10.1016/j.rse.2016.11.022).
- [16] E. P. W. Attema and F. T. Ulaby. “Vegetation Modeled as a Water Cloud”. In: *Radio Science* 13.2 (Mar. 1978), pp. 357–364. ISSN: 1944-799X. DOI: [10.1029/RS013i002p00357](https://doi.org/10.1029/RS013i002p00357).
- [17] M. Bracaglia, P. Ferrazzoli, and L. Guerriero. “A Fully Polarimetric Multiple Scattering Model for Crops”. In: *Remote Sensing of Environment* 54.3 (Dec. 1995), pp. 170–179. ISSN: 0034-4257. DOI: [10.1016/0034-4257\(95\)00151-4](https://doi.org/10.1016/0034-4257(95)00151-4).
- [18] F. T. ULABY, S. SARABANDI, S. McDONALD, S. WHITT, and M. C. DOBSON. “Michigan Microwave Canopy Scattering Model”. In: *International Journal of Remote Sensing* 11.7 (July 1990), pp. 1223–1253. ISSN: 0143-1161. DOI: [10.1080/01431169008955090](https://doi.org/10.1080/01431169008955090).

- [19] D. Rains, H. Lievens, G. J. M. De Lannoy, M. F. McCabe, R. A. M. de Jeu, and D. G. Miralles. “Sentinel-1 Backscatter Assimilation Using Support Vector Regression or the Water Cloud Model at European Soil Moisture Sites”. In: *IEEE Geoscience and Remote Sensing Letters* (2021), pp. 1–5. ISSN: 1558-0571. DOI: [10.1109/LGRS.2021.3073484](https://doi.org/10.1109/LGRS.2021.3073484).
- [20] X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. “Towards Constraining Soil and Vegetation Dynamics in Land Surface Models: Modeling ASCAT Backscatter Incidence-Angle Dependence with a Deep Neural Network”. In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. ISSN: 0034-4257. DOI: [10.1016/j.rse.2022.113116](https://doi.org/10.1016/j.rse.2022.113116).
- [21] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. M. De Lannoy. “Optimisation of AquaCrop Backscatter Simulations Using Sentinel-1 Observations”. In: *Remote Sensing of Environment* 294 (Aug. 2023), p. 113621. ISSN: 0034-4257. DOI: [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621).
- [22] T. Nikaevin, P. Lopez-Dekker, S. C. Steele-Dunne, V. Kumar, and M. Huber. “Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), pp. 7763–7776. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2023.3301124](https://doi.org/10.1109/JSTARS.2023.3301124).
- [23] C. A. Jones and J. R. Kiniry. *CERES-Maize: A Simulation Model of Maize Growth and Development*. Texas A&M University Press, 1986. ISBN: 978-0-89096-269-5.
- [24] J. W. Jones, G. Y. Tsuji, G. Hoogenboom, L. A. Hunt, P. K. Thornton, P. W. Wilkens, D. T. Imamura, W. T. Bowen, and U. Singh. “Decision Support System for Agrotechnology Transfer: DSSAT V3”. In: *Understanding Options for Agricultural Production*. Ed. by G. Y. Tsuji, G. Hoogenboom, and P. K. Thornton. Systems Approaches for Sustainable Agricultural Development. Dordrecht: Springer Netherlands, 1998, pp. 157–177. ISBN: 978-94-017-3624-4. DOI: [10.1007/978-94-017-3624-4_8](https://doi.org/10.1007/978-94-017-3624-4_8).
- [25] V. Kumar, M. Huber, B. Rommen, and S. C. Steele-Dunne. “Agricultural SandboxNL: A National-Scale Database of Parcel-Level Processed Sentinel-1 SAR Data”. In: *Scientific Data* 9.1 (July 2022), p. 402. ISSN: 2052-4463. DOI: [10.1038/s41597-022-01474-4](https://doi.org/10.1038/s41597-022-01474-4).
- [26] M. C. Dobson and F. T. Ulaby. “Active Microwave Soil Moisture Research”. In: *IEEE Transactions on Geoscience and Remote Sensing* GE-24.1 (Jan. 1986), pp. 23–36. ISSN: 1558-0644. DOI: [10.1109/TGRS.1986.289585](https://doi.org/10.1109/TGRS.1986.289585).
- [27] K. Chen, T.-D. Wu, L. Tsang, Q. Li, J. Shi, and A. Fung. “Emission of Rough Surfaces Calculated by the Integral Equation Method with Comparison to Three-Dimensional Moment Method Simulations”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.1 (Jan. 2003), pp. 90–101. ISSN: 1558-0644. DOI: [10.1109/TGRS.2002.807587](https://doi.org/10.1109/TGRS.2002.807587).

- [28] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. Liu, and K. Papathanassiou. “Radar Remote Sensing of Agricultural Canopies: A Review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2249–2273. ISSN: 1939-1404. DOI: [10.1109/JSTARS.2016.2639043](https://doi.org/10.1109/JSTARS.2016.2639043).
- [29] I. Champion and G. Guyot. “Generalized Formulation for Semi-Empirical Radar Models Representing Crop Backscattering”. In: (1991).
- [30] L. Prévot, I. Champion, and G. Guyot. “Estimating Surface Soil Moisture and Leaf Area Index of a Wheat Canopy Using a Dual-Frequency (C and X Bands) Scatterometer”. In: *Remote Sensing of Environment* 46.3 (Dec. 1993), pp. 331–339. ISSN: 0034-4257. DOI: [10.1016/0034-4257\(93\)90053-Z](https://doi.org/10.1016/0034-4257(93)90053-Z).
- [31] F. Chollet. *Keras*. <https://github.com/fchollet/keras>. 2015.
- [32] M. Shorachi, V. Kumar, and S. C. Steele-Dunne. “Sentinel-1 SAR Backscatter Response to Agricultural Drought in The Netherlands”. In: *Remote Sensing* 14.10 (Jan. 2022), p. 2435. ISSN: 2072-4292. DOI: [10.3390/rs14102435](https://doi.org/10.3390/rs14102435).

5

CONCLUSIONS AND RECOMMENDATIONS

In this final chapter, we examine how the findings of this research address the initial research questions presented in Chapter 1. Additionally, recommendations for future research are proposed.

5.1. CONCLUSIONS AND MAIN CONTRIBUTIONS

The main goal of this research is to explore the use of SAR data in agricultural monitoring by exploiting the potential of machine learning, crop growth models, and the analysis of InSAR coherence time series.

The main findings of this research can be summarized as:

- Including interferometric coherence time series alongside backscatter intensity in both polarizations improves land cover classification. This is mainly due to the sensitivity of coherence to seasonal transitions, especially during winter months (Chapter 2).
- A crop-growth model can be used to estimate bio-geophysical parameters as a valid alternative to in-situ data, capturing the dynamic growth patterns of crops as well as the short and long-term impact of weather (Chapter 3).
- Machine learning techniques can be employed to develop a forward model that links bio-geophysical crop parameters to radar observables (Chapter 3).
- Physical knowledge can be incorporated into ANNs by mirroring the structure of physical models in the network architecture, and by penalizing non-physical model behaviors through the loss function used for training. This improves the performance of the machine learning-based observation operator, particularly in terms of generalizability (Chapter 4).

The findings of this research address the main research questions:

1. *To what extent does the inclusion of short-term interferometric coherence time series improve the accuracy of vegetation mapping?*

In Chapter 2, we evaluated the added value of Sentinel-1 InSAR coherence time series for vegetation classification in São Paulo, Brazil. It demonstrated that integrating coherence and backscatter intensities from SAR data improves the accuracy of land cover mapping in complex agricultural regions by 10 percentage points (p.p). The largest contribution to class discrimination occurred in winter, a period characterized by dry vegetation and bare soils.

As expected, InSAR coherence showed high sensitivity to physical changes in the landscape resulting from activities like harvesting. Despite low coherence in vegetated areas, its temporal evolution contributed to improving classification accuracy. We examined the impact of different classification methods and data preparation on classification results. Multi-looking and spatial averaging at the field level improved the quality of the observables, which in turn improved the accuracy of the classification. This approach maximizes the use of spatial context by preserving the inherent relationships within each field, thereby enabling the model to learn more generalized patterns and reducing the likelihood of memorizing noise associated with individual pixel variations. This approach was an important aspect that was applied consistently across other chapters. We further investigated the effects of various training and testing dataset selection strategies—random-pixel, field-pixel, and field-based sampling—on our analysis. Among these strategies, field-pixel and field-based sampling ensure that data from the same field are exclusively allocated to training or validation sets, preventing the model from having prior exposure to test data. This setup is crucial for maintaining truly independent test datasets. The size of the multilooking window was also crucial, with a 100×100 m window providing an optimal balance between radiometric quality and the mixed-pixel effect. While field-level sampling sometimes resulted in slightly lower accuracy, it proved to be more reliable and unbiased. This suggests that incorporating spatial context, such as averaging over fields, is key to reducing misclassification errors. Furthermore, the findings of our research indicated that coherences enhance the classification matrix by highlighting temporal and spatial changes in vegetation, unseen by backscatter intensities alone. The study also highlighted the temporal dynamics and spatial heterogeneities in the agricultural region, stressing the importance of seasonal variability in classification tasks. The integration of coherence information helps to correctly identify the predominant land cover within reference polygons. Even though the overall improvement resulting from the addition of coherence is limited, we observed that it helps produce the right classification in mixed land cover polygons, where classification based only on backscatter intensity is, logically, more prone to errors.

2. *How can we address the unavoidable scarcity of field measurement data to train a data-driven forward model that relates crop bio-geophysical parameters to radar observables?*

Machine learning approaches rely on extensive datasets for training, testing, and validation. Considering that, as discussed in Chapter 2, SAR observables start to

become reliable when we average them down to field-level scales. Training with in-situ observations of the relevant bio-geophysical parameters, and accounting of their spatio-temporal variability, would require an unrealistic amount of samples, in space and in time. To address this, Chapter 3 uses a crop-growth model as an alternative to relying solely on in-situ measurements. This approach generates the necessary crop bio-geophysical parameters, which reflect the dynamic states of crops and soil conditions over time, thereby compensating for limited observational data.

We initially set up a crop growth model in part of The Netherlands to simulate corn growth and development, considering factors like soil characteristics, weather conditions, crop genetics, and specific farming practices for each field. The model generated outputs such as LAI, AGB, surface soil moisture, and root zone moisture (which can be used as a proxy for VWC). This crop-growth model was fine-tuned using detailed field measurement data, including LAI and AGB measurements. Subsequently, a forward model employing machine learning was developed to link SAR observations with the simulated crop growth variables.

The combination of crop-growth model outputs with machine learning also offers the advantage of reduced dependency on field measurements, though it introduces limitations, including the need for well-calibrated model parameters and high-quality meteorological inputs, and potential biases or inaccuracies from the crop model itself.

We developed a forward model within a controlled environment featuring a monocultural agricultural system to understand how various conditions and changes affect the SAR signal. While accurate outputs from the crop-growth model are essential for training a reliable forward model, they are less critical when the primary goal is operational tasks like anomaly detection or data assimilation. In these applications, even if the outputs of the crop-growth model have calibration issues, the combined model can still effectively identify deviations from the expected crop behavior. This capability makes the integration of radar observables with crop-growth model outputs valuable for detecting anomalies and continuously refining model predictions through data assimilation.

Chapter 3 demonstrates the integration of a crop-growth model with a machine learning algorithm to model SAR observables, including NRCS, CR, and coherence. This approach effectively bridges the gap between agricultural modeling and remote sensing observations by simulating bio-geophysical parameters of maize fields and linking them with SAR data from Sentinel-1. As established in Chapter 2, coherence provides valuable supplementary information in vegetation monitoring alongside intensities, hence its inclusion in the analysis.

The combined DSSAT-SVR model predicts SAR observables with a close fit to observed values, especially when trained and tested using data acquired during the same growth season. However, model accuracy declines when applied to data spanning across different growing seasons, particularly in extreme conditions, such as during the drought of 2018. Expanding the training dataset to include data from multiple growing seasons enhances the adaptability of the model, enabling

it to better account for varied environmental conditions and improve year-to-year transferability.

We demonstrated that the CR observable, which minimizes sensitivity to soil moisture, shows better year-to-year transferability compared to individual polarization backscatter channels. Coherence was found to be highly sensitive to changes in biomass, highlighting its potential for monitoring crop development. Additionally, including a proxy for surface roughness, calculated by averaging the early-season backscatter values, enhances model accuracy during early vegetative stages, suggesting that surface roughness is important for accurate radar signal interpretation.

The methodology provides a foundational step toward integrating SAR data directly into crop models, enabling better estimation of crop growth and soil moisture dynamics.

3. *How can we guide neural network architectures with physical principles to improve the prediction of SAR observables for vegetation?*

5

In Chapter 4 we investigate two complementary approaches to physically guide an artificial neural network trained to predict SAR observables, to improve the accuracy and generalization capabilities of the resulting PGNN while also improving the physical interpretability of its behavior. This approach addresses some of the limitations highlighted in Chapter 3. Incorporating physical principles into neural networks was approached in two ways: 1) using a physics-based loss function and 2) through a physics-guided network topology. This approach was demonstrated through a case study of corn fields in The Netherlands, showcasing how the PGNN can improve modeling SAR observables.

For the first approach, a straightforward constraint was introduced by adding a term to ensure that the partial derivative of the NRCS with respect to surface soil moisture remains positive. In the second approach, we designed the network topology and chose the activation functions aiming to mimic the structure of the widely used WCM. This integration of data-driven methods with physical principles allows the model to learn behaviors from data that standard physical models do not explicitly capture, while remaining physically grounded. Integrating physical principles into the network topology constrains the hidden relationships between SM_5 and VWC that the model might otherwise learn purely from data. This approach, however, enhances the training process by reducing overfitting and smoothing out local minima of the loss function with respect to the model coefficients, leading to a more robust learning model.

Performance evaluations on synthetic datasets demonstrate that the PGNN not only reproduces the reference values with high accuracy but also predicts correct intermediate products. When trained with real observations, the model also produces intermediate outputs that are consistent with our expectations, even though they cannot be verified independently. In our case study, physics-constrained methods improve the prediction of SAR observables, which should lead to better performances when used for data assimilation, anomaly detection, or in cases

where understanding of the underlying physical processes is crucial for the interpretation of the observed data. However, the study also acknowledges certain limitations:

- a) Simplifications in network topology: while the PGNN benefits from the inclusion of physical principles, its current topology may be overly simplified. Future work could explore more complex network structures that can capture a broader range of interactions and dependencies within the data.
- b) Training complexities: the training of PGNNs involves balancing the traditional data-driven learning objectives with the need to adhere to physical constraints. This can introduce complexities in network optimization and require more sophisticated training algorithms to ensure convergence and stability.

5.2. FURTHER RESEARCH

While this dissertation demonstrates the potential of machine learning for modeling SAR observables, the focus has been primarily on developing and validating the methodology. However, for practical applications, the model would benefit from a more extensive dataset that captures the range of environmental conditions encountered in real-world agricultural settings. To achieve this, future research should prioritize expanding the temporal and spatial scope of the data used to train these models. Extending the temporal scope to capture multi-year variability in vegetation growth and land management practices would allow the model to account for inter-annual fluctuations and seasonal patterns. Additionally, expanding the geographic range to encompass a broader range of agricultural landscapes and crop types would further validate the adaptability and robustness of the model. A major effort associated with this extension relates to the calibration and validation of the crop growth models used for this purpose. It is important to note that the in-situ measurements are critical for the development of remote sensing products to support agricultural applications.

Following the work presented in Chapter 3 and 4, the next logical step is to integrate the resulting forward models into a data assimilation framework, where real-time SAR data could be used to update and constrain the state variables of the crop-growth model, improving the accuracy of predictions for crop yield and water use. This method was explored by [1] who assimilated normalized backscatter and slope data from ASCAT into the ISBA LSM to better constrain soil water and vegetation dynamics. A question to be answered is if the introduction of physical constraints can lead to better Jacobians, and therefore to improved assimilation.

Chapter 4 demonstrates the effectiveness of adding physics-guided constraints in neural networks. Future research should extend the method to accommodate additional constraints, possibly valid only during certain periods. Additionally, integrating uncertainty quantification techniques within the PGNN framework could enhance its robustness [2]. Exploring a PGNN framework inspired by more sophisticated physical models could further improve modeling, allowing for a more detailed representation of the complex interactions. However, increasing complexity too much can lead to fully connected network architectures, hereby losing the benefits of incorporating physical constraints.

Explore the modeling of different radar wavelengths, particularly transitioning from C-band to L-band SAR data as crops mature. While C-band is more responsive to smaller vegetation structures and upper canopy dynamics early in the growth cycle, integrating L-band data as crops mature could provide improved sensitivity to the full canopy structure and underlying soil moisture. This transition could improve the overall predictive capability of models. This approach aligns well with existing L-band satellites and upcoming missions like ROSE-L, which aim to provide extensive L-band radar coverage for monitoring land and vegetation dynamics. This can become particularly interesting if we consider a unified network to predict all SAR observables simultaneously or across different wavelengths, incorporating constraints based on known interdependencies between observables.

Although backscatter is highly sensitive to water content, current crop model outputs lack information on VWC, which is important for accurately modeling SAR observables. Addressing this gap by incorporating VWC and/or wet biomass into SAR observable modeling can improve predictions related to drought stress [3].

Given the promising results demonstrated by the PGNN framework, future research should focus on expanding its application to encapsulate more complex physical relationships across various scientific domains beyond just vegetation studies. For instance, applying the PGNN framework to model SAR observables over the ocean could represent a compelling research direction. This approach leverages the framework's strength in integrating domain-specific knowledge with data-driven learning, ensuring accurate modeling across diverse contexts.

REFERENCES

- [1] X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. “Assimilating ASCAT Normalized Backscatter and Slope into the Land Surface Model ISBA-A-gs Using a Deep Neural Network as the Observation Operator: Case Studies at ISMN Stations in Western Europe”. In: *Remote Sensing of Environment* 308 (July 2024), p. 114167. ISSN: 0034-4257. DOI: [10.1016/j.rse.2024.114167](https://doi.org/10.1016/j.rse.2024.114167).
- [2] A. Daw, R. Q. Thomas, C. C. Carey, J. S. Read, A. P. Appling, and A. Karpatne. “Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling”. In: *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*. Proceedings. Society for Industrial and Applied Mathematics, Jan. 2020, pp. 532–540. DOI: [10.1137/1.9781611976236.60](https://doi.org/10.1137/1.9781611976236.60).
- [3] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. M. De Lannoy. “Optimisation of AquaCrop Backscatter Simulations Using Sentinel-1 Observations”. In: *Remote Sensing of Environment* 294 (Aug. 2023), p. 113621. ISSN: 0034-4257. DOI: [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621).

A

APPENDIX-A: SUPPLEMENTARY MATERIAL

A.1. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

2017 MAE	VV	VH	CR
2017 without offset	1.10 (1.09)	1.09 (1.02)	0.84 (0.86)
2017	1.01 (0.98)	1.05 (0.93)	0.78 (0.77)
2018	2.89 (2.64)	2.53 (2.03)	0.91 (0.98)
2019	1.47 (1.37)	1.40 (1.34)	0.91 (0.92)
2018 + 2019	1.64 (1.42)	1.53 (1.31)	0.90 (0.89)
2017+2018+2019	1.17 (1.05)	1.45 (1.32)	0.80 (0.78)
2018 MAE			
2018 without offset	1.11 (1.26)	1.21 (1.41)	0.95 (1.09)
2017	1.88 (1.91)	1.99 (2.44)	1.09 (1.92)
2018	1.04 (1.18)	1.18 (1.33)	0.95 (1.02)
2019	1.38 (1.70)	1.65 (1.70)	1.07 (1.18)
2017 + 2019	1.42 (1.55)	1.64 (1.60)	1.05 (1.18)
2017+2018+2019	1.14 (1.23)	1.28 (1.37)	0.95 (1.02)
2019 MAE			
2019 without offset	1.20 (1.11)	1.23 (1.10)	0.96 (0.95)
2017	1.72 (1.64)	1.93 (1.66)	1.05 (1.05)
2018	1.46 (1.76)	1.53 (1.66)	0.98 (0.97)
2019	1.12 (1.00)	1.18 (1.00)	0.82 (0.83)
2017 + 2018	1.46 (1.30)	1.43 (1.26)	0.98 (0.96)
2017+2018+2019	1.20 (1.06)	1.25 (1.05)	0.84 (0.85)

2017 MAE	VV	VH	CR
----------	----	----	----

Table A.1.: Mean absolute error for backscatter. The presented stats belong to the vegetative period, the stats within the parenthesis belong to the whole growth period.

2017 MSE	VV	VH	CR
2017 without offset	2.13 (2.16)	2.31 (2.05)	1.24 (1.50)
2017	1.73 (1.60)	2.23 (1.76)	1.07 (1.07)
2018	18.60 (12.98)	13.5 (7.87)	1.42 (1.59)
2019	3.48 (3.09)	3.50 (3.52)	1.41 (1.46)
2018 + 2019	4.23 (3.23)	3.77 (2.83)	1.39 (1.36)
2017+2018+2019	2.27 (1.84)	3.49 (2.84)	1.09 (1.09)
2018 MSE			
2018 without offset	2.22 (2.85)	2.69 (3.89)	1.80 (2.32)
2017	5.70 (6.00)	6.41 (9.82)	2.10 (6.31)
2018	1.81 (2.34)	2.57 (3.47)	1.73 (1.95)
2019	3.28 (5.06)	4.88 (5.46)	2.03 (2.47)
2017 + 2019	3.35 (4.08)	4.82 (5.18)	1.97 (2.42)
2017+2018+2019	2.17 (2.52)	3.05 (3.72)	1.72 (1.95)
2019 MSE			
2019 without offset	2.70 (2.30)	2.65 (2.35)	1.78 (1.73)
2017	4.54 (4.61)	6.28 (5.17)	1.85 (1.97)
2018	3.42 (5.53)	3.74 (4.97)	1.66 (1.63)
2019	2.12 (1.76)	2.47 (1.98)	1.20 (1.22)
2017 + 2018	3.36 (2.75)	3.38 (2.80)	1.64 (1.58)
2017+2018+2019	2.34 (1.91)	2.68 (2.10)	1.23 (1.26)

Table A.2.: Mean squared error of backscatter.

2017 R2	VV	VH	CR
2017 without offset	0.45 (0.40)	0.62 (0.68)	0.37 (0.50)
2017	0.55 (0.47)	0.63 (0.64)	0.46 (0.40)
2018	-3.87 (-3.38)	-1.24 (-0.61)	0.30 (0.12)
2019	0.08 (-0.04)	0.41 (0.27)	0.30 (0.19)
2018 + 2019	-0.10 (-0.09)	0.37 (0.41)	0.31 (0.25)
2017+2018+2019	0.41 (0.40)	0.43 (0.42)	0.44 (0.40)
2018 R2			

2017 R2	VV	VH	CR
2018 without offset	0.15 (0.22)	0.35 (0.40)	0.31 (0.35)
2017	-1.13 (-0.85)	-0.50 (-0.83)	0.14 (-1.31)
2018	0.30 (0.28)	0.38 (0.36)	0.33 (0.30)
2019	-0.22 (-0.56)	-0.14 (-0.02)	0.17 (0.09)
2017 + 2019	-0.24 (-0.25)	-0.13 (0.03)	0.20 (0.11)
2017+2018+2019	0.17 (0.22)	0.26 (0.31)	0.33 (0.30)
2019 R2			
2019 without offset	0.27 (0.36)	0.51 (0.65)	0.34 (0.50)
2017	-0.16 (-0.55)	-0.15 (-0.30)	0.34 (0.14)
2018	0.12 (-0.87)	0.31 (-0.25)	0.41 (0.29)
2019	0.44 (0.39)	0.54 (0.50)	0.56 (0.45)
2017 + 2018	0.13 (0.06)	0.38 (0.30)	0.41 (0.31)
2017+2018+2019	0.38 (0.34)	0.50 (0.47)	0.54 (0.44)

Table A.3.: R2 score of backscatter.

2017 Pearson	VV	VH	CR
2017 without offset	0.67 (0.63)	0.79 (0.82)	0.61 (0.71)
2017	0.75 (0.69)	0.80 (0.80)	0.68 (0.63)
2018	-0.20 (-0.09)	0.09 (0.40)	0.61 (0.54)
2019	0.45 (0.33)	0.69 (0.57)	0.56 (0.50)
2018 + 2019	0.27 (0.30)	0.63 (0.68)	0.57 (0.53)
2017+2018+2019	0.67 (0.63)	0.67 (0.70)	0.67 (0.63)
2018 Pearson			
2018 without offset	0.38 (0.47)	0.60 (0.64)	0.56 (0.59)
2017	0.17 (-0.02)	0.16 (0.08)	0.49 (0.15)
2018	0.55 (0.53)	0.61 (0.60)	0.58 (0.55)
2019	0.33 (0.21)	0.42 (0.33)	0.50 (0.35)
2017 + 2019	0.21 (0.21)	0.33 (0.32)	0.50 (0.36)
2017+2018+2019	0.44 (0.48)	0.53 (0.58)	0.60 (0.56)
2019 Pearson			
2019 without offset	0.53 (0.60)	0.72 (0.81)	0.59 (0.71)
2017	0.44 (0.22)	0.45 (0.39)	0.64 (0.47)
2018	0.45 (0.19)	0.59 (0.40)	0.65 (0.57)
2019	0.67 (0.63)	0.74 (0.71)	0.75 (0.67)
2017 + 2018	0.50 (0.40)	0.62 (0.59)	0.66 (0.58)
2017+2018+2019	0.62 (0.58)	0.71 (0.69)	0.74 (0.66)

Table A.4.: Pearson correlation of backscatter.

A

2017 Spearman	VV	VH	CR
2017 without offset	0.69 (0.61)	0.78 (0.78)	0.58 (0.64)
2017	0.75 (0.63)	0.79 (0.72)	0.61 (0.54)
2018	-0.07 (-0.12)	0.10 (0.39)	0.52 (0.41)
2019	0.34 (0.19)	0.54 (0.36)	0.44 (0.38)
2018 + 2019	0.18 (0.20)	0.41 (0.49)	0.47 (0.40)
2017+2018+2019	0.66 (0.58)	0.46 (0.51)	0.54 (0.52)
2018 Spearman			
2018 without offset	0.40 (0.48)	0.56 (0.63)	0.55 (0.59)
2017	0.09 (-0.03)	0.08 (0.01)	0.42 (0.10)
2018	0.49 (0.50)	0.56 (0.56)	0.53 (0.52)
2019	0.30 (0.20)	0.31 (0.25)	0.40 (0.29)
2017 + 2019	0.12 (0.19)	0.16 (0.22)	0.40 (0.30)
2017+2018+2019	0.32 (0.44)	0.40 (0.52)	0.53 (0.52)
2019 Spearman			
2019 without offset	0.50 (0.58)	0.70 (0.75)	0.60 (0.65)
2017	0.37 (0.17)	0.32 (0.27)	0.56 (0.32)
2018	0.37 (0.20)	0.51 (0.37)	0.57 (0.42)
2019	0.60 (0.60)	0.72 (0.67)	0.66 (0.54)
2017 + 2018	0.40 (0.31)	0.50 (0.46)	0.57 (0.41)
2017+2018+2019	0.53 (0.53)	0.65 (0.62)	0.64 (0.52)

Table A.5.: Spearman correlation of backscatter

2017	MAE	MSE	R2	Pearson	Spearman
2017	0.05	0.005	0.56	0.77	0.54
2018	0.07	0.009	0.22	0.50	0.45
2019	0.06	0.008	0.30	0.57	0.29
2018 + 2019	0.06	0.008	0.29	0.57	0.44
2017+2018+2019	0.05	0.005	0.54	0.75	0.53
2018					
	MAE	MSE	R2	Pearson	Spearman
2017	0.15	0.04	-3.12	0.10	0.15
2018	0.07	0.008	0.22	0.50	0.44
2019	0.10	0.01	-0.84	0.04	0.04
2018 + 2019	0.10	0.01	-0.84	0.17	0.17
2017+2018+2019	0.07	0.008	0.22	0.50	0.43
2019					
	MAE	MSE	R2	Pearson	Spearman
2017	0.07	0.01	0.08	0.54	0.40
2018	0.066	0.008	0.26	0.52	0.47

2017	MAE	MSE	R2	Pearson	Spearman
2019	0.05	0.006	0.52	0.74	0.57
2018 + 2019	0.06	0.008	0.24	0.58	0.48
2017+2018+2019	0.05	0.006	0.47	0.70	0.55

Table A.6.: Stats of coherence VV.

A

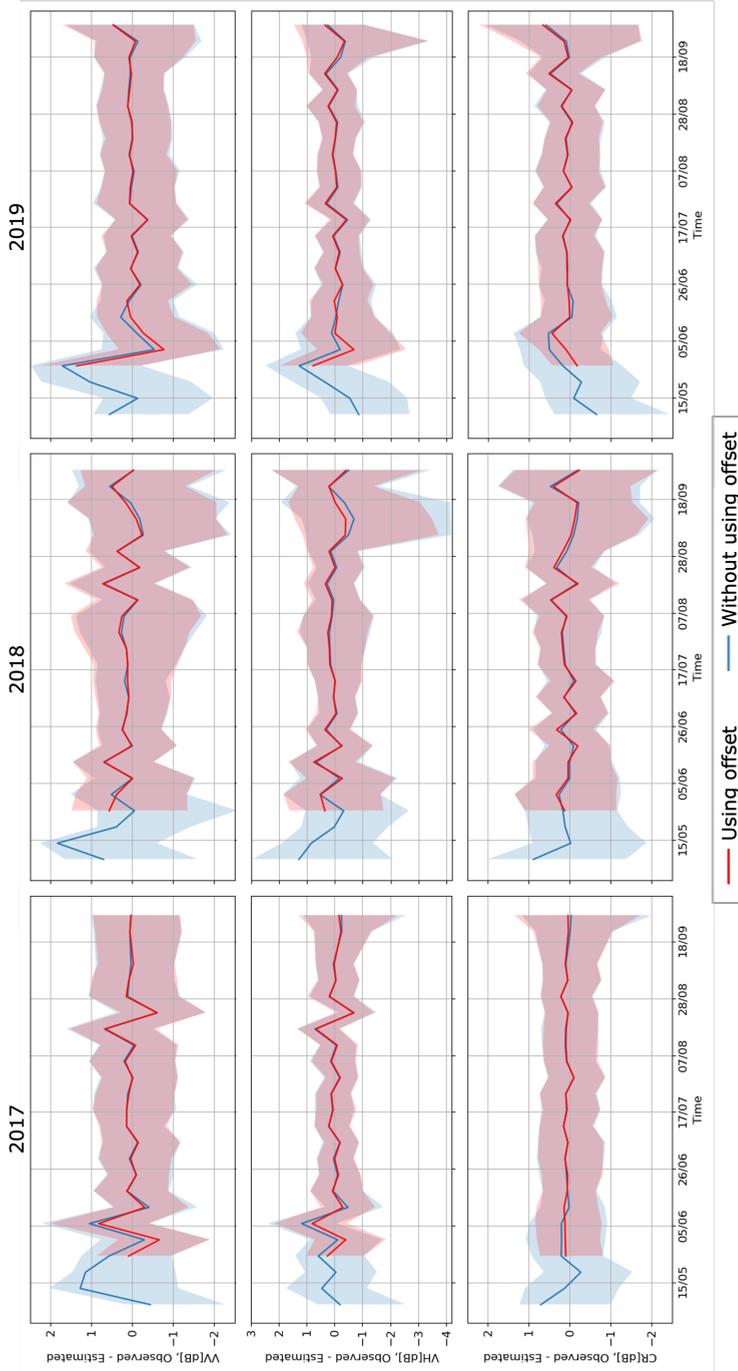


Figure A.1.: Time series difference between observed and estimated backscatter with and without using the offset in different years and channels.

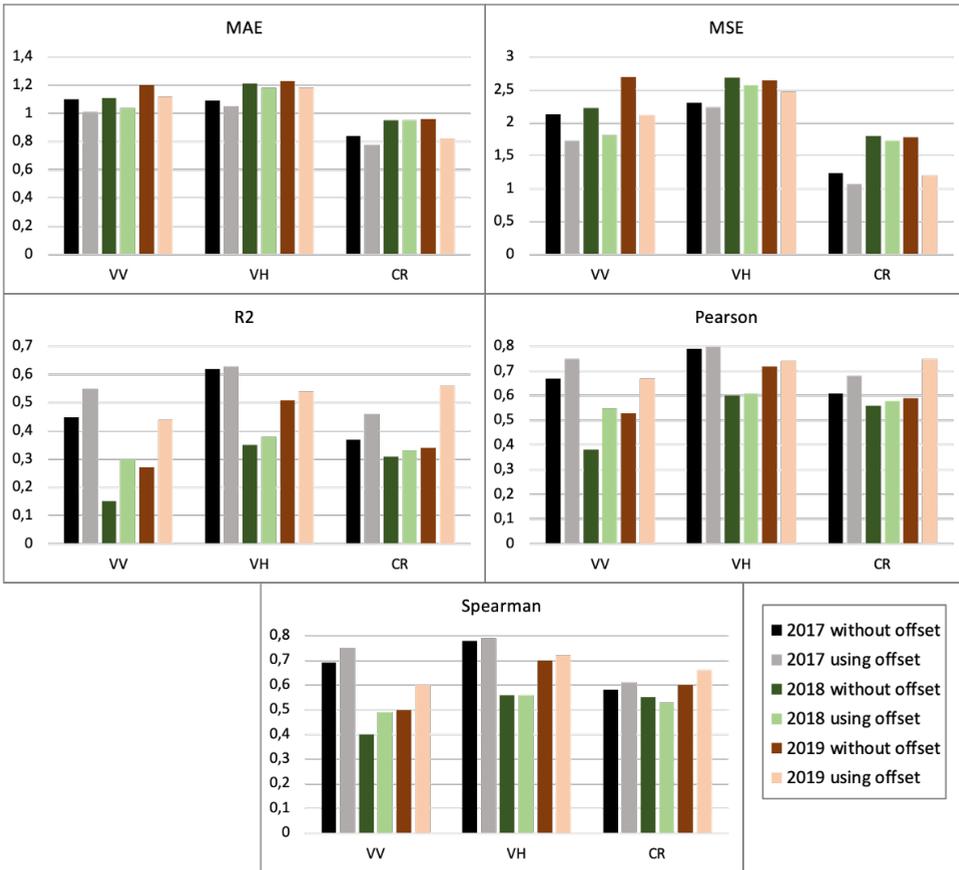


Figure A.2.: Standard metrics to evaluate the regression performance for backscatter when training and test data set belong to the same year with and without offset in the beginning of the vegetative period.

A

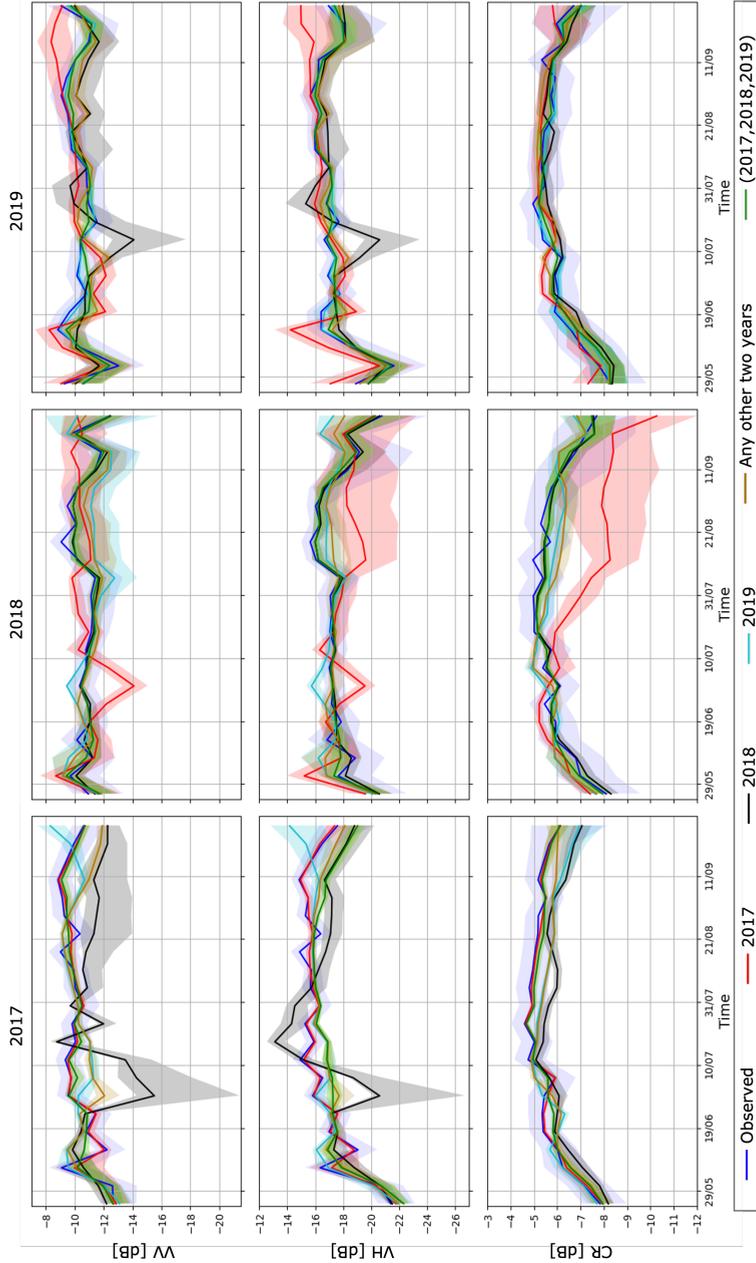


Figure A.3.: Time-series of observed and estimated backscatter in 2017 (1st column), 2018 (2nd column) and 2019 (3rd column) for different channels. The years that training data are selected from are shown in the legend.

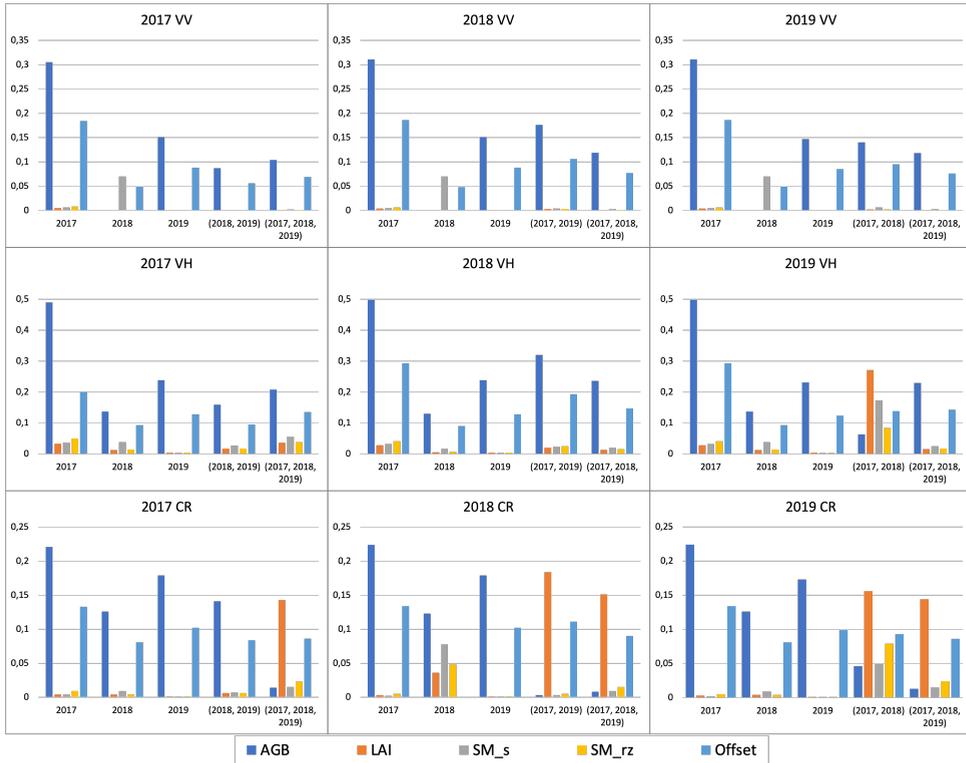


Figure A.4.: Rank features for SVR with backscatter using the MRMR algorithm.

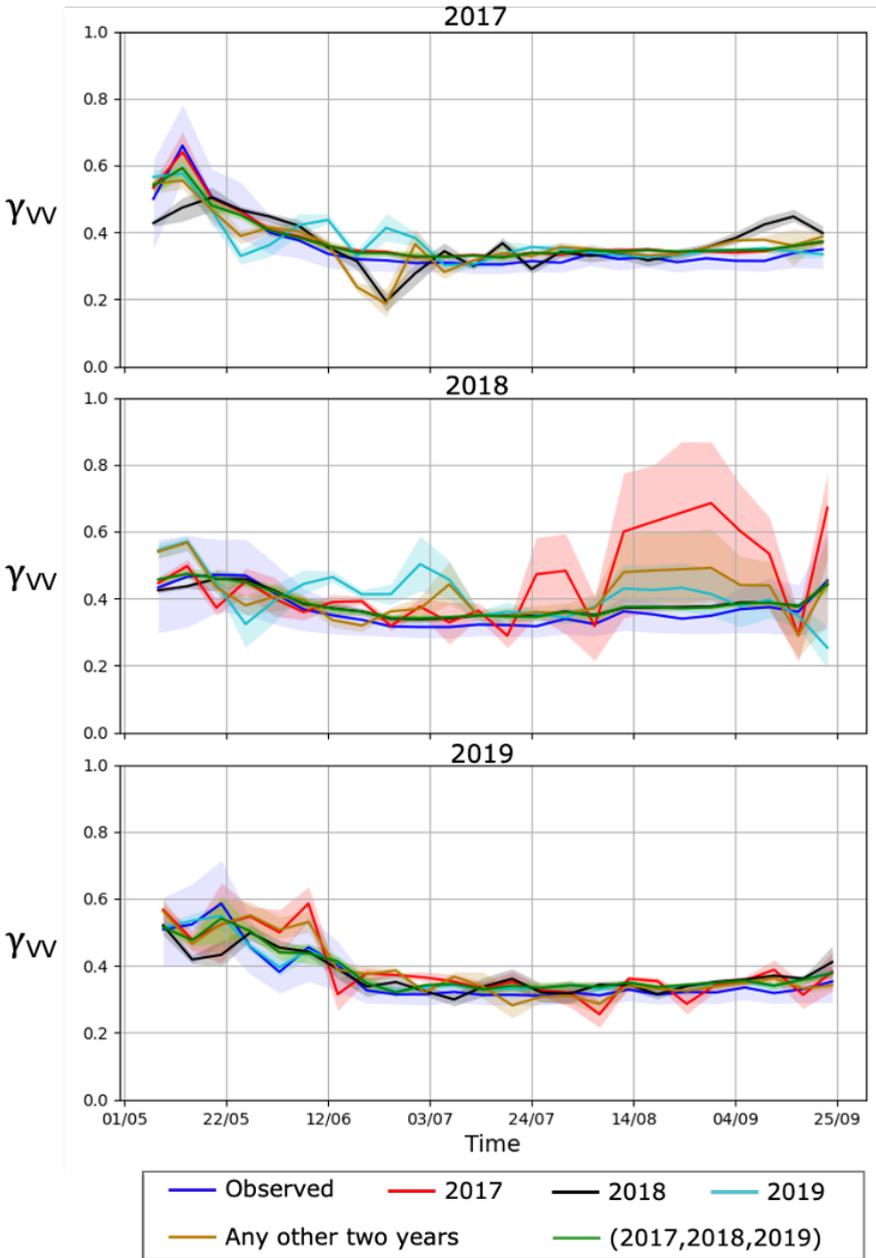


Figure A.5.: Time-series of observed and predicted coherence V_w . The legend shows the year in which the training data belongs.

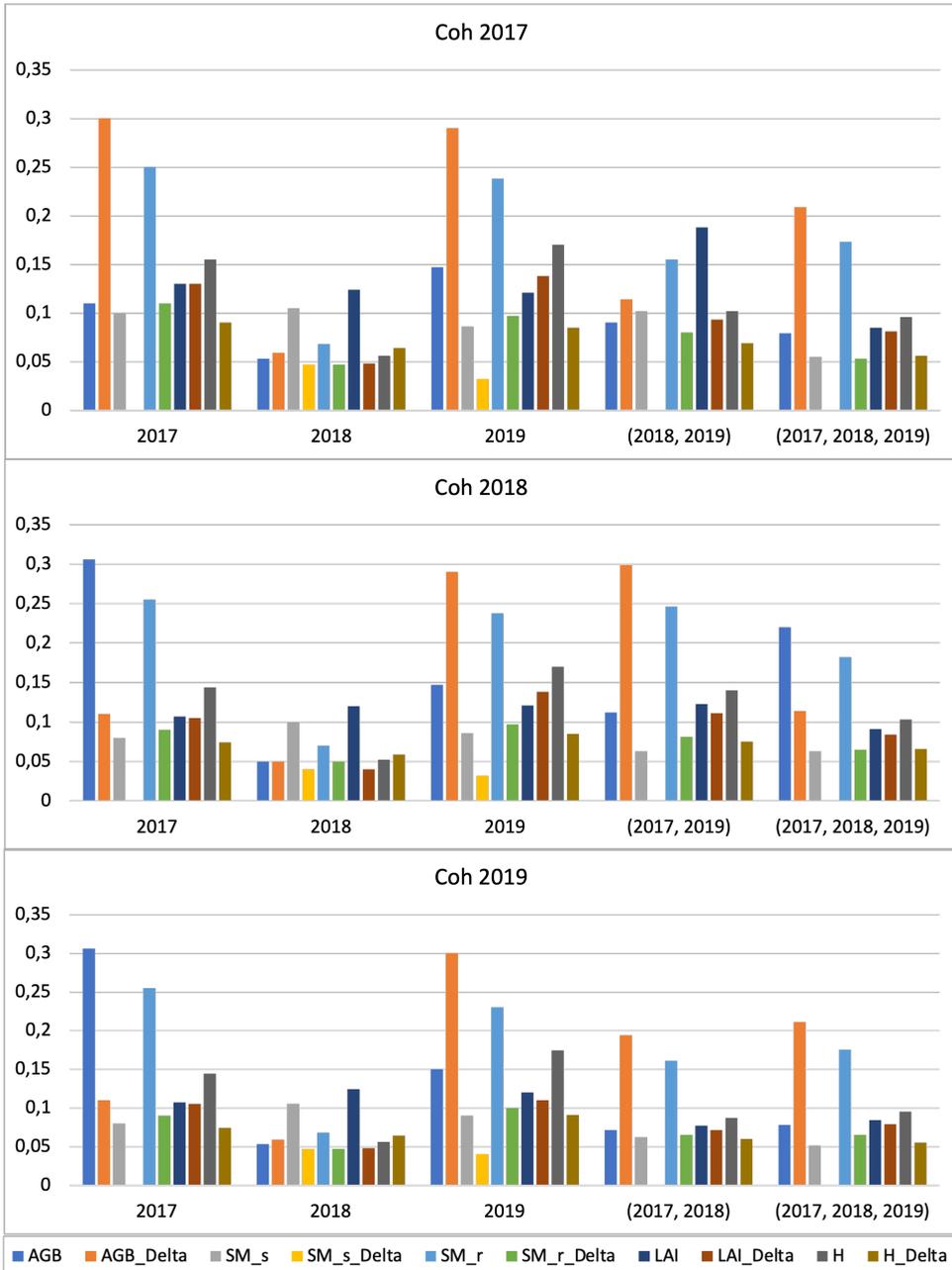


Figure A.6.: Rank features for SVR with coherence VV using MRMR algorithm. The variables with the "Delta" extension represent the difference between the dates.

ACKNOWLEDGEMENTS

Those who know me well are aware that my PhD journey has been anything but easy, and not just because of the demands of the PhD itself, life had its trials waiting in the wings. These experiences have only deepened my appreciation for the people who stood by me throughout.

In reverse chronological order, I would like to express my heartfelt gratitude to everyone who supported me, both scientifically and personally. First and foremost, I want to thank Paco. Working with you has been one of the most formative and meaningful aspects of this journey. I am deeply grateful for your support through all my ups and downs. Pursuing a PhD is as much about personal growth as it is about scientific development, and I grew alongside you. You taught me not only how to be a better researcher but also how to be a better person. Even if I sometimes fall short, I know you have set a standard I deeply admire. I will always carry your words with me: “We learn how to enjoy life even when we are unhappy.” Thank you for your patience and for giving me the space I needed when times were hard. You never made things harder than they already were, and I am profoundly grateful for that. I extend my sincere appreciation to Susan. Thank you for being a truly unforgettable part of this journey. Your warmth, support, and kindness have meant so much to me. Thank you for always being a safe and welcoming presence and opening your door when I needed it most. I am also grateful to Ramon for welcoming me into the SkyGeo office during the challenging Covid period, and thanks for our scientific discussions.

My sincere thanks also go to my committee members, Prof. Juanma López-Sánchez, Prof. Jasmeet Judge, Dr. Armando Marino, and Dr. Riccardo Taormina, for their time, feedback, and willingness to be part of this important milestone.

I want to express my heartfelt thanks to Lorenzo for his support at the beginning of my PhD journey and for introducing me to the fascinating topic of agriculture. Your guidance during those early days is a memory I will always cherish. I am also deeply grateful to Ramses, who took the time to contribute to this work despite his busy schedule, and to Dr. Ali Mousivand for his valuable advice during the initial phase of my research.

A special thanks to the M-WAVE group, particularly Vineet and Manuel, for your efforts on Sandbox NL—it was a pleasure collaborating with you.

I would like to thank the current and former GRS staff for all their support, especially Debbie, Natasha, Suzanne, and Lidwien. Debbie, I appreciate the time you spent with me and our walks together.

I am also glad for sharing the office with Gert and Ou Ku, and for the nice moments we had together. A big thank you to everyone in the radar group. Your innovative ideas and collaborative spirit have made this journey not only richer but also more inspiring.

A special thanks goes to my friends, whose support throughout this journey has been invaluable. You made this experience joyful and fun, and I truly appreciate your presence

along the way.

To my colleagues and friends at S4G, I am deeply grateful for your understanding of the challenges I faced in balancing work and my PhD. Your constant motivation and encouragement have been a source of strength.

I want to thank my cheerful partner's family, who made the Netherlands feel like home for me. Tineke, you stepped into the role of a mom here and taught me the importance of self-care. Thank you for always looking out for me and providing such warmth and kindness.

To Maman, Baba, Shaghayegh, and Amirali, we made it through all the unexpected storms along the way. I love you more than words can express. I am deeply grateful for the motivational calls and the unwavering belief you always had in me. Maman and Baba, I know that if not for your sacrifices, I wouldn't be here today.

I owe these years to the support of someone who took care of my heart and embraced all the fears, failures, and happiness along the way. Nard, you are the greatest gift that coming to the Netherlands has given me. Without you, this journey would have been impossible. Thank you for your patience and for accepting the challenges of life with someone pursuing a PhD and a full-time job. Your confidence in me gave me the strength to keep going.

*Tina Nikaein
Delft, 2024*

CURRICULUM VITÆ

Tina NIKAEIN

22-06-1993 Born in Iran.

EDUCATION

2011–2015 Surveying and Geomatics
University of Tehran, Iran

2015–2018 M.Sc. in Remote Sensing
University of Tehran, Iran

Thesis: Improvement of Permanent Scatterers Detection Using Potential of Polarimetric Data

Supervisors: Prof. dr. ir. H. Arefi

2018–2023 PhD. Geoscience and Remote Sensing
Delft University of Technology

Thesis: From InSAR Time-Series to Crop Growth: Machine Learning and Physics-Guided Models for Radar-Based Vegetation Analysis

Promotor: Prof. dr. ir. R.E Hanssen
Prof. dr. ir. P López-Dekker

EXPERIENCE

2023-present Radar specialist at Space4Good, Den Haag, The Netherlands

2012-2013 Executive Engineer in Tehran Rasta surveying consulting Engineers (N.G.O Representative), Tehran, Iran

LIST OF PUBLICATIONS

9. T. Nikaein and P. Lopez-Dekker. “Physics-Inspired Neural Network for Forward Modeling of Sentinel-1 Radar Observables”. In: *Living Planet Symposium*. 2025
8. T. Nikaein and P. Lopez-Dekker. “Physics-Guided Machine Learning Based Forward-Modeling of Radar Observables: A Case Study on Sentinel-1 Observations of Corn-Fields”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2025), pp. 1–11. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2025.3543238](https://doi.org/10.1109/JSTARS.2025.3543238).
7. T. Nikaein, P. Lopez-Dekker, S. C. Steele-Dunne, V. Kumar, and M. Huber. “Modeling SAR Observables by Combining a Crop-Growth Model With Machine Learning”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), pp. 7763–7776. ISSN: 2151-1535. DOI: [10.1109/JSTARS.2023.3301124](https://doi.org/10.1109/JSTARS.2023.3301124).
6. T. Nikaein, V. Kumar, S. Steele-Dunne, and P. Lopez-Dekker. “Crop-Growth Driven Forward-Modeling of Sentinel-1 Observables Using Machine-Learning”. In: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. July 2022, pp. 5961–5964. DOI: [10.1109/IGARSS46834.2022.9883154](https://doi.org/10.1109/IGARSS46834.2022.9883154).
5. T. Nikaein, L. Iannini, R. A. Molijn, and P. Lopez-Dekker. “On the Value of Sentinel-1 InSAR Coherence Time-Series for Vegetation Classification”. In: *Remote Sensing* 13.16 (Jan. 2021), p. 3300. DOI: [10.3390/rs13163300](https://doi.org/10.3390/rs13163300).
4. T. Nikaein, L. Iannini, and P. Lopez Dekker. “Agricultural Land Cover Mapping Based on Sentinel-1 Coherence Time-Series”. In: *EUSAR 2021; 13th European Conference on Synthetic Aperture Radar*. Mar. 2021, pp. 1–4
3. L. Iannini, R. Molijn, T. Nikaein, R. Hanssen, P. Lopez-Dekker, J. V. Rocha, R. Lamparelli, and P. S. G. Magalhães. “C-Band Radar and Optical Remote Sensing for a Sustainable Agriculture in Brazil”. In: *Living Planet Symposium*. 2019
2. T. Nikaein, H. Arefi, and V. Akbari. “Improvement of Persistent Scatterers Detection Using Polarimetric Capability of UAVSAR Sensor”. In: *Journal of Geomatics Science and Technology* 8.3 (Feb. 2019), pp. 163–178.
1. T. Nikaein, H. Arefi, and V. Akbari. “Added Value of Multitemporal Polarimetric UAVSAR Data for Permanent Scatterers Detection”. In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. July 2018, pp. 653–656. DOI: [10.1109/IGARSS.2018.8519543](https://doi.org/10.1109/IGARSS.2018.8519543).

