Deep Learning and Earth Observation for the Study of West African Rainfall
Observing rainfall processes through the lens of AI

Estebanez Camarena, M.

DOI
10.4233/uuid:6bbe6005-d020-4cde-a737-bc75b084e1c4

Publication date
2025

Document Version
Final published version

Citation (APA)
Estebanez Camarena, M. (2025). *Deep Learning and Earth Observation for the Study of West African Rainfall: Observing rainfall processes through the lens of AI*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:6bbe6005-d020-4cde-a737-bc75b084e1c4

Deep Learning and Earth Observation for the Study of West African Rainfall

Observing rainfall processes through the lens of AI

Mónica Estébanez Camarena

Deep Learning and Earth Observation for the Study of West African Rainfall

Observing rainfall processes through the lens of AI

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
Friday 25, April 2025 at 10:00 o'clock

by

Mónica ESTÉBANEZ CAMARENA

MSc in Space Navigation, University of Cranfield, United Kingdom
MPhil in Space Studies, University of Cape Town, South Africa
born in Madrid, Spain

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Dr. ir. J.A.E. ten Veldhuis	Delft University of Technology, *promotor*
Prof. dr. ir. N. C. van de Giesen	Delft University of Technology, *promotor*
Independent members:	
Prof. dr. B.L. Lamptey	University of Leeds, United Kingdom
Prof. dr. M.E. Mc Clain	IHE Delft, Delft
Dr. M.L. Moel	IHE Delft, Delft
Prof. dr. ir. H.W.J. Russchenberg	Delft University of Technology
Dr. R. Taormina	Delft University of Technology

The work leading to these results has received funding from the European Commission's Horizon 2020 Programme (2018-2022) under grant agreement No. 776691 (TWIGA) and the Horizon Europe Research & Innovation Programme 2021-2027 (grant agreement no. 101086209 – TEMBO Africa). The opinions expressed in the document are of the authors only and in no way reflect the European Commission's opinions. The European Union is not liable for any use that may be made of the information. This work has also received funds from the European Space Agency through the Citizen Science Earth Observation Lab (CSEOL) program, and from the TU Delft | Global Initiative, a program of the Delft University of Technology to boost Science and Technology for Global Development.

TWIGA

TU Delft | Global Initiative

Science for the benefit of people. All people. Worldwide.

To my mother

CONTENTS

FOREWORD

The book that you have in your hands is my PhD dissertation. I have called it "Deep Learning and Earth Observation for the Study of West African Rainfall" and it contains part of the research I have conducted and the learnings I have gained at the TU Delft during the last six years. I say part–and not all–because in no way does it cover everything that my PhD has taught me, or all the results it has yielded. In fact, like someone told me once, *the most important output of my PhD is not my dissertation, but who I have become.* During these last years I have had the chance to travel often to Ghana, to work with schools, research institutions and smallholder farmers. I have experienced what tropical rainfall feels like, I have seen seen the impact of climate change on people's livelihoods with my own eyes, I have reflected on intercultural collaboration, and I have learnt so much. I have been able to submerge myself in the Ghanaian culture - during a fieldwork month in Tamale I even took daily evening classes of "Bamaya", a traditional harvest dance from the Dagomba people of Northern Ghana, performed to celebrate the rains - and learnt some indigenous knowledge. I have learnt to really care about the outcome of my research, more than for the scientific value of it, for the societal impact.

On an academic side, publishing and defending my dissertation marks the completion of not only my last six years, but also my (almost 15-years) university education. I never thought I would do a PhD related to weather. I never even liked rainfall! But step by step, I got myself here. Looking back I can only see how much sense the whole puzzle made. I started off with my love for space, studying Aerospace and then Space Engineering. I was first passionate about space exploration, and then I decided to use space to look back at Earth. Utilizing space for the benefit of society and the planet became then my main interest. That is how I ended up, after leaving Spain, and passing by South Africa and England, in the Water Management Department of the TU Delft, the Netherlands. It has been a long way. And no words can express the happiness of getting here.

During my first PhD years, I used to say assertively that I would not submit until I was able to tell Ghanaian farmers when to plant, to help them cope with climate change. I am grateful that although my research took another direction, I ended up getting there with my later job. And not only in Ghana, but in many other (African) countries.

At the end, following your intuition–and putting your heart in it–is always the best way forward.

Mónica Estébanez Camarena
March 2025, Delft

SUMMARY

West African food and economic safety are heavily reliant on agriculture, most of which is rainfed. Changing rainfall patterns induced by global warming jeopardize yields by unpredictable water availability. At the same time, a rapidly growing population leads to rising demands for food production. Accurate rainfall information is essential for farmers to adjust their crop management practices and avoid yield, thereby improving the overall resilience of the region. However, this information is largely lacking due to a sparse rain gauge distribution, limited resources and data transmission challenges. Added to this, existing satellite rainfall products show a particularly poor correlation with ground observations in West Africa.

This dissertation aims at improving rainfall information for farmers in the Sudanian Savanna bioclimatic region of West Africa. The Sudanian Savanna stretches across Africa as a broad belt, covering roughly from southern Mali in the north to northern Ghana in the south. Its West African area expands from Senegal in the west to Chad in the easte. Improving rainfall information will support climate resilience and food and economic safety in the region. This research leverages on the unique potential of Earth Observation satellites to provide rainfall information everywhere, due to their global coverage and ability to track atmospheric processes. To tackle the general poor performance of existing rainfall information products, this work proposes an alternative avenue. Particularly, it investigates the potential of Deep Learning (DL) methods to extract relationships between meteorological variables and raw satellite data that might be overlooked by traditional satellite rainfall retrieval methods.

However, DL models are data intensive, that is, they require large amounts of data to be developed. Therefore, doing so over a data-scarce context such as the Sudanian Savanna presents challenges of its own. To overcome this, this dissertation starts by proposing a methodology to develop DL models for data-scarce areas.

As a first demonstration of the potential of DL to model rainfall in West Africa for a scarce set of target data, two models – called Rain-Runner and Rainrunner+ – are developed for binary classification (rain / no-rain) based on standard DL architectures: Convolutional Neural Networks (CNN) and Convolutional Long-Short Term Memory (ConvLSTM). The input data for both models are thermal infrared (TIR) observations from the Meteosat Second Generation satellites. The choice of these DL architectures is based on their ability to capture spatial and temporal relations in the data. As target data, we use rainfall observations from the Trans-African Hydro-Meteorological Observatory (TAHMO). Both models are trained over the north of Ghana using data from only 8 stations over 2.5 years, with 20.4% of the data missing. The Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System (PERSIANN-CCS) and the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG) products are used as benchmarks. The first results are promising as they compare well against these state-of-the-art products. Our models consistently outperform PERSIANN-CCS and during the second half of the rainy season they even outperform the much more complex IMERG.

Of the two models, the CNN-based model performed best and this is developed further by adding two information layers to the model, water vapor (WV) and temporal information related to the time of the observation (hour and month). WV data is included because of its importance for rainfall processes in West Africa, where most rainfall is convective and therefore highly depends on available atmospheric moisture. Moreover, literature suggests that one reason for the poor regional performance of satellite rainfall products is a drier atmosphere than in other regions of the world. Temporal information is included to account for the diurnal and seasonal patterns of rainfall. The results show that addition of WV observations highlights areas of strong convection and discards non-precipitating low-level features that introduce false alarms for methods based only on TIR data. This is especially beneficial in areas like the tropics where most rainfall is convective. Furthermore, it enables detection of dry air masses advected from the Sahara Desert, that produce discontinuities in rainfall events. Overall, the resulting model generates fewer false alarms and lower rainfall overdetection (FBias < 2.0) compared to the benchmark, IMERG Final Run.

In a third step, the models trained over the north of Ghana were applied to stations across the wider Sudanian Savanna region. This yields valuable information about the generalization capabilities of the models and the contribution of each information type (i.e., temperature, water vapor and time of the observation) to satellite rainfall retrieval in the region. From this, implications for the poor performance and possible improvements of existing satellite rainfall products in the area can be inferred. Results showed that the models have good generalization capabilities, achieving similar performance across the Sudanian Savanna compared to the north of Ghana, where they are trained. The observed effect of adding WV and temporal information is also similar. Furthermore, analyzing this effect in a larger area and throughout the year highlights that WV information is especially relevant during the first half of the rainy season (March to June). We suggest that this is related to a larger atmo-

spheric process that dictates West African climate and rainfall dynamics during the year and that is often referred to as the movement of the Intertropical Convergence Zone (ITCZ). This is the discontinuity area between the northeasterly dry and dusty Sahara air and the southwesterly humid oceanic air. In the period from March to June, the Sahara air reaches further south and has a higher probability of causing rainfall inhibition in our area of interest. This phenomenon triggers false alarms in TIR methods, but is corrected for by WV observations.

Lastly, this work proposes other future research avenues for advancing the study of rainfall and satellite rainfall retrieval over West Africa, based on insights gained during the research. Some of these avenues are employing additional satellite observations, more advanced Deep Learning architectures, and Citizen Science.

One of these avenues that received special attention during the development of this research is Citizen Science. A Citizen Science project called Schools and Satellites (SaS) was deployed in the north of Ghana between 2019 and 2021. SaS created a Citizen Observatory formed by high schools and farmers that became the densest ground observation network in the area during the time of the project. Although it is only briefly referred to in the main body of this dissertation, additional information can be found in the Appendix.

Overall, this work advances satellite rainfall retrieval in the Sudanian Savanna region of West Africa in that: (1) It demonstrates the great potential of DL methods for satellite rainfall retrieval over data-scarce areas (2) it explains TIR rainfall over-detection and debunks traditional methods based on TIR-only rainfall retrieval and (3) it provides evidence in support of regional models over global models. Particularly, we demonstrate that locally training a DL model achieves comparable performances to much more complex global models, even when developed with a small training dataset and based on simple DL structures.

Furthermore, this work opposes the “black-box” narrative against DL models for meteorology and showcases how meteorology knowledge can be used to analyze the results of a DL model, and explain them from a physical perspective.

Finally, I hope that the insights and recommendations extracted from this dissertation can assist future researchers to further improve the needed reliable rainfall information that can benefit the population and agriculture and ecosystems management in Sub-Saharan Africa. Only from there will we be able to progress towards a climate-resilient and food- and economically-secure Africa.

SAMENVATTING

De West-Afrikaanse voedselvoorziening en economische veiligheid zijn sterk afhankelijk van de landbouw, waarvan de watervoorziening grotendeels gebaseerd op neerslag, zonder toegevoegde beregening. Veranderende neerslagpatronen als gevolg van de opwarming van de aarde brengen de opbrengst in gevaar door onvoorspelbaarheid van beschikbaar water. Tegelijkertijd neemt de vraag naar landbouwproducten toe door de snel groeiende bevolking. Nauwkeurige neerslaginformatie is essentieel voor boeren om hun gewasbeheer te kunnen aanpassen en verliezen te voorkomen, waarmee de algehele veerkracht van de regio verbetert. Deze informatie ontbreekt echter grotendeels door de lage dichtheid van het netwerk van regenmeters, beperkte middelen en problemen met data-communicatie. Daar komt bij dat de bestaande satellietproducten voor neerslag juist slecht correleren met de waarnemingen op de grond in West-Afrika.

Dit proefschrift is gericht op het verbeteren van neerslaginformatie voor boeren in de Soedanese Savanne-regio in West-Afrika, een regio die zich uitstrekt van Senegal in het westen tot .. in het oosten. ter ondersteuning van klimaatbestendigheid en voedsel- en economische veiligheid in de regio. Het maakt gebruik van het unieke potentieel van aardobservatiesatellieten, vanwege hun wereldwijde dekking en vermogen om atmosferische processen te meten. In dit werk wordt een alternatieve aanpak voorgesteld ten opzichte van de bestaande producten. Deze maakt gebruik van het potentieel van Deep Learning (DL) methoden om relaties tussen meteorologische variabelen en ruwe satellietgegevens te naar boven te halen die mogelijk over het hoofd worden gezien door traditionele satellietmethoden voor neerslaginformatie.

DL-modellen zijn echter gegevensintensief, dat wil zeggen dat er grote hoeveelheden gegevens voor nodig zijn om ze te ontwikkelen. Daarom vormt het ontwikkelen van DL-modellen in een context van data-schaarste, zoals de Soedanese savanne, op zichzelf al een uitdaging. Om dit te ondervangen begint dit proefschrift met het ontwikkelen van een methodologie om DL modellen te ontwikkelen voor gebieden met weinig gegevens.

Als eerste demonstratie van het potentieel van DL voor het modelleren van regenval in West-Afrika zijn twee modellen ontwikkeld - RainRunner en RainRunner+ genaamd - voor binaire classificatie van regenval (regen / geen regen) op basis van standaard DL-architecturen: Convolutionele neurale netwerken (CNN) en Convolutioneel langetermijngeheugen (ConvLSTM). De invoergegevens voor beide modellen zijn thermisch infrarood (TIR) waarnemingen van tweede generatie Meteosat satellieten. De keuze van deze DL-architecturen is gebaseerd op hun vermogen om ruimtelijke en temporele relaties in de gegevens vast te leggen. We gebruiken de neerslagwaarnemingen van het Trans-African Hydro-Meteorological Observatory (TAHMO) als referentie voor de modellen. Beide modellen zijn getraind over het noorden van Ghana met behulp van gegevens van slechts 8 stations gedurende 2,5 jaar, waarbij 20,4% van de gegevens ontbraken. De bestaande neerslag-producten Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System (PERSIANN-CCS) en Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement mission (IMERG) worden gebruikt als benchmarks. De eerste resultaten zijn veelbelovend omdat de modellen vergelijkbaar presteren met de state-of-the-art producten. Onze modellen presteren consistent beter dan PERSIANN-CCS en voor de tweede helft van het regenseizoen zijn ze zelfs beter dan het veel complexere IMERG.

Van de twee modellen presteerde het op CNN gebaseerde model het best en dit is verder ontwikkeld. Hiervoor zijn er twee informatielagen aan het model toegevoegd, waterdamp (WV) en temporele informatie met betrekking tot het tijdstip van de waarneming (uur en maand). WV-gegevens worden toegevoegd vanwege het belang ervan in neerslagprocessen in West-Afrika, waar de meeste neerslag convectief is en daardoor sterk afhankelijk is van de beschikbare atmosferische vochtigheid. Bovendien suggereert de literatuur dat een van de redenen voor de slechte regionale prestaties van satellietneerslagproducten een drogere atmosfeer is dan in andere delen van de wereld. Temporele informatie houdt rekening met de dag- en seizoenspatronen van neerslag. De toevoeging van WV-data legt meer nadruk op gebieden met sterke convectie en verwijdert neerslag-achtige kenmerken op laag niveau die niet uitregenen. Deze veroorzaken valse neerslag-detectiesbij methoden die alleen op TIR-gegevens zijn gebaseerd. Deze toevoeging is vooral gunstig in gebieden zoals de tropen waar de meeste neerslag convectief is. Bovendien helpt het om droge luchtmassa's te detecteren, aangevoerd vanuit de Sahara-woestijn, die discontinuïteiten kunnen veroorzaken in regenbuien. Over het geheel genomen geeft het nieuwe model minder valse alarmen en minder overdetectie van neerslag (FBias < 2,0) vergeleken met de IMERG Final Run.

Als derde stap is het nieuwe model dat is getraind voor het noorden van Ghana toegepast op stations in de bredere regio van de Soedanese Savanne. Dit levert waardevolle informatie op over de generalisatiemogelijkheden van de modellen en de bijdrage van elk informatietype (d.w.z. temperatuur, waterdamp en tijdstip van de waarneming) aan het correct afleiden vanneerslag uit satellietmetingen in deze regio. Ook leiden we hieruit verdere implicaties af voor de slechte prestaties en mogelijke verbeteringen van bestaande satellietproducten voor neerslag in het gebied. De modellen blijken een goed generalisatievermogen te hebben en presteren even goed in de Soedanese savanne als in het noorden van Ghana, waar ze getraind waren. De toegevoegde waarde WV en temporele informatie is ook vergelijkbaar. Bovendien laat de regionale analyse zien dat WV-informatie relevanter is tijdens de eerste helft van het regenseizoen (maart tot juni). We vermoeden dat dit verband houdt met een groter atmosferisch proces dat het West-Afrikaanse klimaat en de neerslagdynamiek gedurende het jaar dicteert en vaak wordt aangeduid als: de beweging van de Intertropische Convergentiezone (ITCZ). Dit is het overgangsgebied tussen de noordoostelijke droge en stoffige Sahara-lucht en de zuidwestelijke vochtige oceaanlucht. In de periode van maart tot juni reikt de Sahara-lucht verder naar het zuiden en is er een grotere kans op regenvalonderdrukking in ons onderzoeksgebied. Dit fenomeen zou valse alarmen genereren bij TIR-methoden, maar wordt gecorrigeerd door de WV-waarnemingen.

Tot slot stelt dit werk toekomstige onderzoekspaden voor om het begrip van neerslagpatronen en afleiden van neerslag uit satellietmetingen boven West-Afrika te verbeteren, gebaseerd op de inzichten die tijdens het onderzoek zijn opgedaan. Enkele van deze mogelijkheden zijn het gebruik van nog meer typen satellietwaarnemingen zoals van aerosolen, toepassen van geavanceerdere Deep Learning-architecturen en de inzet van Citizen Science.

Eén van deze onderwerpen kreeg speciale aandacht tijdens de ontwikkeling van dit onderzoek: Citizen Science. Tussen 2019 en 2021 is in het noorden van Ghana het Citizen Science-project Schools and Satellites (SaS) uitgevoerd. SaS heeft een Citizen Observatory opgezet met middelbare scholen en boeren dat het dichtste grondobservatienetwerk in het gebied werd gedurende de looptijd van het project. Hoewel er slechts kort naar wordt verwezen in de hoofdtekst van dit proefschrift, is aanvullende informatie te vinden in de Appendix.

Samenvattend draagt dit werk bij aan het verbeteren van neerslaginformatie uit satellietwaarnemingen voor de Soedanese Savanne regio van West-Afrika omdat: (1) het toont het grote potentieel aan van DL methoden voor neerslaginformatie uit satellieten voor gebieden met data schaarste (2) het verklaart de over-detectie van TIR-gebaseerde methoden en ontkracht daarmee traditionele methoden die hierop gebaseerd zijn en (3) het toont de meerwaarde aan van regionale modellen ten opzichte van globale modellen. Specifiek tonen we aan dat het lokaal trainen van een DL-model vergelijkbare prestaties oplevert als veel complexere mondiale modellen, zelfs wanneer dit wordt gedaan met een kleine trainingsdataset en op basis van eenvoudige DL-structuren.

Verder verzet dit werk zich tegen de stelling dat DL slechts “black-box” modellen oplevert door te laten zien hoe meteorologische kennis gebruikt kan worden om de resultaten van een DL-model te analyseren en deze vanuit een natuurkundig perspectief te verklaren.

Tot slot hoop ik dat de inzichten en aanbevelingen uit dit proefschrift toekomstige onderzoekers zullen helpen bij het verkrijgen van de benodigde betrouwbare neerslaginformatie die ten goede kan komen aan de bevolking en het beheer van landbouw en ecosystemen in Sub-Sahara Afrika. Alleen op die manier kunnen we vooruitgang boeken in de richting van een klimaatbestendig en voedsel- en economisch veilig Afrika.

1

INTRODUCTION

Climate change is more than statistics,
it’s more than data points.
It’s more than net-zero targets.
It’s about the people,
it’s about the people that are affected right now.

Vanessa Nakate, Climate Activist, Uganda

1.1. GEOGRAPHICAL AND RESEARCH CONTEXT

1.1.1. RESEARCH AREA: NORTHERN GHANA, SUDANIAN SAVANNA

Ghana, on the southern coast of West Africa, is a vast country with diverse cultures and climates. Only one day drive from the South to the North of the country takes you through different bioclimatic regions, from forest to savanna Figure 1.1. This dissertation begins in northern Ghana and extends to the larger bioclimatic region in which it is located: the Sudanian Savanna.

Figure 1.1: Bioclimatic regions of West Africa [1].

From the Atlantic Ocean to the Sahara Desert, West Africa has one of the most extreme climatic gradients in the world, with the most significant climatic variable for ecosystems and the population being precipitation. Annual precipitation rates change from 200 mm in the Sahel to over 2000 mm on the coast [2]. Two air masses - humid and cold air from the ocean and dry and hot air from the desert – meet each other in what is known as the Intertropical Convergence Zone (ITCZ).The ITCZ and the West African Monsoon – a cool air blowing inland from the Atlantic Ocean - are the main drivers of rainfall dynamics in West Africa, since they control the timing, intensity and distribution of precipitation during the rainy season [3]. However, rainfall mechanisms here are complex and there are still many unknowns [2].

Like in the rest of West Africa, food and economic safety in the region are heavily dependent on agriculture, most of which is rainfed (i.e., only reliant on rainfall) and is threatened by climate change and population growth. Accurate rainfall information is essential to face these challenges.

1.1.2. THE CHALLENGE OF RAINFALL INFORMATION IN WEST AFRICA

Changing rainfall patterns, more frequent extreme weather events like droughts and floods, along with inaccurate and inaccessible rainfall information, make farmers - most of whom are smallholders - vulnerable to reduced yields and significant losses. In fact, many smallholder farmers engaged during this research claimed facing difficulties in adapting to climate change, and suffering losses in yield, seeds and investments made in their farms, which had important repercussions in their personal life Figure 1.2. Access to accurate rainfall information is, therefore, essential to ensure the preparedness and resilience of the farmers.

Operational weather and seasonal forecasts help smallholder farmers make informed decisions and strengthen their resilience to climate change. However, these models struggle to perform well in West Africa. Improving and reliably assessing weather models requires accurate and dense ground-based rainfall data, which remains insufficient due to sparse rain gauge networks and data transmission challenges. Satellites, with their global coverage, offer a potential solution, but existing satellite products also show poor performance in the region.

Figure 1.2: A smallholder farmer stands in the middle of her tomato field, dried shortly after planting because of an unexpected drought. Photographed by the author in Wa, Ghana, February 2020. Her face has been covered for privacy.

1.1.3. DEVELOPMENTS IN SATELLITE RAINFALL RETRIEVAL METHODS

Satellite rainfall retrieval methods employ mainly two types of Earth Observation (EO) data: thermal infrared (TIR) and passive microwave (PMW) data, from Geostationary (GEO) and Low-Earth orbit (LEO) satellites, respectively. Rainfall is inferred from TIR imagery by studying the temperature of the clouds. The most established TIR-based method is the Cold Cloud Duration (CCD) method, that establishes a relationship between how long a certain pixel remains under a temperature threshold, and rainfall rates on the ground. This method is based on the assumption that rain-bearing clouds have high tops, below a certain atmospheric temperature, which makes it especially relevant for areas with deep convection. It is used by two important rainfall products in Africa, the Climate Hazards Infrared Precipitation with Stations (CHIRPS) [4] - that also uses rain gauge data for bias correction - and the Tropical Applications of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT) [5]. Although the use of TAMSAT is currently less extensive than before, CHIRPS remains the preferred product for many African meteorological agencies. PMW data retrieves rainfall by measuring the natural microwave radiation emitted by hydrometeors in the atmosphere. Although TIR methods have the disadvantage of being a less direct measure of rainfall than PMW ones, they have the large advantage of providing a constant coverage of the Earth. This is because GEO satellites are over the same point of the Earth surface continuously, resulting in constant observations able to track atmospheric movements. On the contrary, PMW observations are made by LEO satellites and therefore only provide intermittent information.

In general, satellite rainfall products perform poorly over West Africa [6–10]. Literature suggests that the reason for this poor performance are a sparse distribution of rain gauges, a higher aerosol concentration and a drier atmosphere compared to other regions in the World. The higher aerosols concentration would be due to dust from the Sahara desert, biomass burning in sub-Saharan Africa and decomposing vegetation in equatorial forests of Africa [11]. Therefore, to improve satellite rainfall retrieval over West Africa, these characteristics should be taken into account.

Machine Learning (ML) is a field that has gained popularity across disciplines in the last decades, because of its ability to simulate physical processes accurately. Its first application for satellite rainfall retrieval was in the late 1900s, with the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIAN) [12], based on GEO TIR data. Since then, ML-based rainfall retrieval methods have rapidly evolved in parallel with advancements in the field of machine learning.

In 2012 Deep Learning (DL) started to become popular, finding its first application in medical image diagnosis, autonomous vehicles and speech recognition. The main advantage of DL models is that they are "universal approximators", they can approximate almost any function, being able to model complex and non-linear processes, such as rainfall [13]. From a data point of view, they can handle large amounts of data, deal with noise, and incorporate new data easily [13]. Furthermore, they are able to automatically extract features from raw input data, as opposed to shallow networks for which features have to be hand-engineered and passed as inputs to the network. By automatically extracting features, the network can capture spatial and temporal structures hidden in the data, potentially leading to the discovery of new concepts and relations, correcting human errors, and being free from preconceived notions about the processes [14]. One common criticism of DL models is that they are black boxes and bring little new understanding about the physical processes. However, by carefully evaluating the model results in relation with the inputs, it is possible to extract valuable knowledge about the physical relationships between them [12–14].

A technical challenge of DL for rainfall retrieval in West Africa and other data-scarce contexts is its data-intensive requirements. Furthermore, existing ML-based satellite rainfall retrieval models are trained on dense ground data, which is missing in this region.

1.1.4. RESEARCH OBJECTIVE

The overall goal of this research is to improve rainfall information over the Sudanian Savanna region of West Africa, as a first step towards improved weather forecasts models to support climate resilience and food and economic security in the region.

To tackle this problem, the global coverage of Earth Observation satellites - especially relevant for data-scarce contexts such as West Africa – is combined with the ability of Deep Learning to learn the relationships between meteorological variables and rainfall directly from raw data.

Sub-objectives are (1) to investigate the potential of DL-based methods for satellite rainfall retrieval over the Sudanian Savanna, (2) to propose a methodology to develop DL models for data-scarce contexts, (3) to assess the contributions of water vapor (WV) and temporal information to improve rainfall information retrieval compared to TIR-only methods and (4) to use DL as a diagnostic tool to understand the reasons behind the poor performance of TIR-based rainfall retrieval methods over the Sudanian Savanna, as well as how it could be improved.

1.1.5. RESEARCH QUESTIONS

The main research question of this thesis is: **How can Deep Learning exploit thermal infrared, water vapor and temporal EO data for satellite rainfall retrieval in the Sudanian Savanna region of West Africa?**

And it can be divided in the following sub-questions:

RQ1. How can DL be exploited to improve satellite rainfall retrieval in data-scarce contexts? (Chapter 2)

RQ2. What role can water vapor observations and temporal information, added to thermal infrared information, play in satellite rainfall retrieval? (Chapter 3)

RQ3. Can a DL satellite rainfall retrieval model developed for the north of Ghana be extrapolated to the wider Sudanian Savanna bioclimatic region? (Chapter 4)

An overarching question that runs through this thesis is to what extent can DL inform us about the underlying physics of rainfall processes.

1.1.6. RESEARCH METHODOLOGY

The basic concept of this research methodology is to develop a data-driven model to learn the relationship between EO data and rainfall on the ground. In the absence of gridded ground data on which to train the model, we use point data Figure 1.3. This results in the output of the model not being a precipitation map (gridded), but precipitation for the center of the input image.

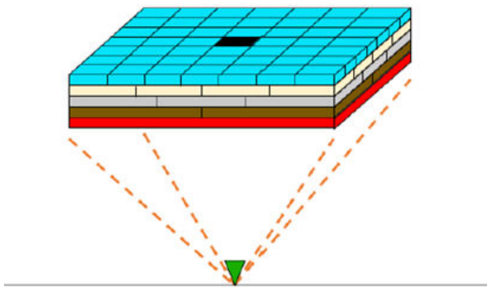


Figure 1.3: Concept of the link between multiple satellite information layers and one point rainfall measurement.

In essence, such a methodology has three components:

1. **EO data:** Based on existing satellite rainfall retrieval methods, the first EO data used in this thesis is Meteosat TIR data. Next, water vapor (WV) data from the same Meteosat observation, collected simultaneously and covering the same pixels as the TIR data but at a different wavelength, is added as a second layer to the model. This is because of the crucial role that WV plays in rainfall processes in West Africa, as well as because, as mentioned previously, a drier atmosphere might be one of the reasons for a poor performance of satellite rainfall products over West Africa. Lastly, temporal information related to the observation (time of the day and month) is included to account for the diurnal and seasonal patterns of rainfall.

2. **Ground data:** One of the basic requirements for DL models is to have extensive training and validation datasets. However, this is challenging in a data-scarce context such as the Sudanian Savanna. Therefore, we evaluated the possibility of expanding our training and validation dataset, from the Trans-African Hydro-Meteorological Observatory (TAHMO) [15] rain gauge network with Citizen Science (CS) data. To this end, we ran a CS project in the north of Ghana from July 2019 to December 2021, called Schools and Satellites, that collected daily rainfall data. However, at the end, these data are not contained in the main body of this dissertation for not being deemed of sufficient scientific quality.

3. **DL model:** In this dissertation we develop two models – that we call RainRunner - based on standard DL architectures: Convolutional Neural Networks (CNN) and Convolutional Long-Short Term Memory (ConvLSTM). With these two architectures we aim to capture spatial (CNN and ConvLSTM) and temporal (ConvLSTM) patterns in the data related to rainfall. We start by comparing the performance of both models, and continue the development of the thesis only with the preferred one.

At all stages of model development, performance is evaluated by comparing the results with state-of-the-art satellite rainfall products: the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG) [16] and Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System (PERSIANN-CCS) [17].

By adding information layers in a consecutive manner and comparing the performances of the different models, we can assess the contribution of each one of them. Further performance evaluation includes comparing it across factors such as rain intensity and season (rainy / dry).

Once the final model has been developed using data from the north of Ghana, its generalization capabilities are analyzed by testing it on ground stations across the wider Sudanian Savanna.

Meteorological knowledge is used in both the north of Ghana and the Sudanian Savanna to study the reasons behind the difference in performance of the models developed using each type of information and behind the poor performance of TIR-based satellite rainfall products.

1.1.7. A NOTE ON SCHOOLS AND SATELLITES

Schools and Satellites (SaS) was one of the Citizen Science Earth Observation Lab (CSEOL) pilot projects. CSEOL was an initiative funded by the European Space Agency to foster ideas that combined space big data and CS. Schools and Satellites was a collaboration between TU Delft, PULSAQUA, TAHMO Ghana, Smartphones4Water and the Ghana Meteorological Agency. I was the project lead, after proposing the project and successfully securing the funding through a competitive process in 2019.

The project had the goal to develop a rainfall retrieval model through the combination of DL, EO and CS. This last component aimed at creating an extensive training and validation dataset to support the development of a DL model. In fact, the first version of the DL model presented in this thesis was developed within SaS.

SaS worked with farmers and high schools in Northern Ghana to create a Citizen Observatory to measure daily rainfall. Our Citizen Observatory was formed by 51 citizen scientists across the five northern regions of Ghana, and it became the densest rainfall observation network in the region. In schools, the rainfall measurement was embedded into an optional educational module on climate change and the water cycle. The project had a high social and educational value, as communicated by the participants and their communities. However, the COVID-19 pandemic caused important disruptions in the program, including the impossibility of delivering the planned intensive training to the citizen scientists before the data collection campaign. At the end, the data collected was inconsistent and not deemed to be of a quality good enough to be used to develop the DL model. Therefore, this project has been left out of the main body of the thesis. To learn more about SaS, the reader is referred to the Appendix A, which contains key sections of two of the project deliverables: the report of work package 1 - ground data collection - and work package 2 - algorithm development. More information on the project can be found on this link: `https://new.tahmo.org/schoolandsatellites/`. Learnings from the Citizen Science campaign were shared in the paper “Leveraging Citizen Science for Sustainable Development Education and Water Security in Northern Ghana”, and was published in 2022 in UNESCO’s Youth and Water Security in Africa report.

1.2. THESIS OUTLINE

This thesis is organized as follows:

Chapter 1 introduces the problem tackled in this thesis, its objectives, the research questions to be addressed, the methodology to do so, and the outline of the thesis.

Chapter 2 investigates the potential of DL to model rainfall in West Africa and proposes a methodology to develop DL models in data-scarce areas. It does so by developing two DL models – RainRunner - for rainfall binary classification (rain / no-rain) over northern Ghana from Meteosat TIR data, using CNNs and ConvLSTMs. This chapter uses the north of Ghana as a case study.

Chapter 3 develops further the CNN-based RainRunner model over the north of Ghana by adding two information layers to the TIR-only model. This is, a water vapor (WV) layer and a layer containing temporal information, particularly time of the day and month. In this way, it develops three models - based on (1) only WV data, (2) TIR and WV data and (3) TIR, WV and temporal information, and compares their performances to that of the TIR-only model developed in Chapter 2. The contribution of each information type is then evaluated based on Atmospheric Science of the region.

Chapter 4 extends the four RainRunner models developed using data from the north of Ghana to the wider Sudanian Savanna region. It assesses the generalization capability of the model as well as whether the insights about the contribution of the different information types (i.e. temperature, water vapor and time of the day and of the year) to satellite rainfall retrieval hold true across the wider region. In addition, it investigates what this analysis can teach us about the poor performance of existing satellite rainfall products in the region, and possible improvements.

Chapter 5 suggests future research avenues for advancing the study of West African rainfall, based on insights acquired during this research.

Finally, **Chapter 6** reflects on the main findings and scientific contributions of this dissertation.

1.3. SUPPLEMENTARY MATERIAL

All research data and code supporting the findings described in this thesis are available in 4TU.ResearchData at: DOI: 10.4121/6e101d26-8067-4455-b465-78ce8f6a601d.

2

RAINRUNNER(-TIR). A DEEP LEARNING APPROACH TO SATELLITE RAINFALL RETRIEVAL OVER DATA-SCARCE AREAS

This chapter has been published in Remote Sensing 15(7), 1922. as *The Potential of Deep Learning for Satellite Rainfall Detection over Data-Scarce Regions, the West African Savanna*. (2023). Authors: Estébanez-Camarena, M., Taormina, R., van de Giesen, N., and ten Veldhuis, M. -C., doi: 10.3390/rs15071922 [18].

2.1. ABSTRACT

Food and economic security in West Africa rely heavily on rainfed agriculture and are threatened by climate change and demographic growth. Accurate rainfall information is therefore crucial to tackling these challenges. Particularly, information about the occurrence and length of droughts as well as the onset date of the rainy season is essential for agricultural planning. However, existing rainfall models fail to accurately represent the highly variable and sparsely monitored West African rainfall patterns. In this paper, we show the potential of deep learning (DL) to model rainfall in the region and propose a methodology to develop DL models in data-scarce areas. We built two DL models for satellite rainfall (rain/no-rain) detection over northern Ghana from Meteosat TIR data based on standard DL architectures: Convolutional neural networks (CNNs) and convolutional long short-term memory neural networks (ConvLSTM). The Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG) and Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System (PERSIANN-CCS) products are used as benchmarks. We use rain gauge data from the Trans-African Hydro-Meteorological Observatory (TAHMO) for model development and performance evaluation. We show that our models compare well against existing products despite being considerably simpler, developed with a small training dataset - i.e., 8 stations covering 2.5 years with 20.4% of the data missing - and using TIR data alone. Concretely, our models consistently outperform PERSIANN-CCS for rain/no-rain detection at a sub-daily timescale. While IMERG is the overall best performer, the DL models perform better in the second half of the rainy season despite their simplicity (i.e., up to 120 k parameters). Our results suggest that DL-based regional models are a promising alternative to state-of-the-art global products for providing regional rainfall information, especially in meteorologically complex regions such as the (sub)tropics, which are poorly covered by ground-based rainfall observations.

2.2. INTRODUCTION

Food and economic safety in West Africa depend heavily on rainfed agriculture and, therefore, on rainfall. In this context, accurate rainfall information is essential to ensure food security. Uncertainty in West African rainfall and the associated vulnerability of smallholder farmers have been documented since the last century. In the 1970s, the Sahelian Drought was socially and agriculturally devastating. It was reported to have produced 100,000 deaths by 1973 and was followed by continuous droughts in the next two decades [19, 20]. Currently, climate change and global population growth [21], the two great threats of this century, exacerbate these problems. Sub-Saharan Africa will account for most of this century's population growth and will become the world's most populous area by the late 2060s [22]. Climate change is changing the onset of the rainy season over the Sahel [23] and causing more frequent droughts in most of Africa, which is severely increasing food insecurity [24, 25]. Rainfall detection is essential to monitor these changes, characterize rainfall patterns, and supply the information needed for efficient agricultural planning. However, a sparse, unevenly distributed, and inconsistently reported rain gauge network poses a major challenge to studying rainfall variability in this region and has been a persistent problem since the last century [26].

Satellite rainfall products are of special relevance for areas with sparse rain gauge networks, such as sub-Saharan Africa, because of their global coverage. In fact, satellite rainfall retrieval and its application over Africa have been in constant development since the late 1960s [26–29]. However, existing satellite products show a poor correlation with ground measurements in the region. For example, the Africa Climate Hazards Infrared Precipitation with Stations (CHIRPS) [4] and the Tropical Applications of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT) [5], particularly developed for Africa based on the Cold Cloud Duration method, show daily Kling–Gupta Efficiency values below 0.4 [6, 30]. The most widely used machine learning-based product, Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS) [17], tends to have a high false alarm ratio (FAR) and to overestimate rainfall both globally and in Africa [7, 8]. Lastly, the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) [16], which combines both physical and ML-based methods and has been developed to become the longest and most detailed rainfall data set, show a weaker correlation with ground measurements in West Africa than in other regions of the world [9, 10].

The literature suggests that an important reason for the poor performance of satellite rainfall estimates over West Africa is the sparse rain gauge distribution, leading to underrepresentation in the training or calibration data for the modeling algorithms. Additionally, atmospheric conditions differ from other regions in the world, as there are higher aerosol concentrations, higher land surface temperatures, and a generally drier atmosphere [11]. Furthermore, the generalization performance of existing ML rainfall retrieval models trained on dense gridded rainfall data [17, 31–33] may decrease for areas with less training data and different atmospheric conditions.

Deep learning (DL) is becoming increasingly popular in the field of environmental remote sensing because of its ability to learn complex patterns and features from data [34]. DL methods exploit spatial and sequential inductive biases to improve performance by incorporating the assumption that nearby pixels in an image and nearby elements in a sequence have more relevance to the output, which allows the network to learn more effectively and generalize to new examples.

In this work, we investigate whether locally training a deep learning model can overcome the limitations of global products in capturing the complex rainfall dynamics of this region. We develop two models based on CNN and ConvLSTM for rain/no-rain detection in the data-scarce region of northern Ghana, West Africa. Both models have been trained on a small regional dataset, representative of data availability in the region. The focus of this paper is on rain/no-rain detection, i.e., binary classification, as a first step towards rainfall intensity estimation. In Section 2, we present the data and study area and introduce our methodology; in Section 3, we report our results; in Section 4, we compare our findings with those of other studies; and in Section 5 we draw the main conclusions of this study and propose future work beyond this paper.

2.3. MATERIALS AND METHODS

2.3.1. MODEL DEVELOPMENT DATASETS

The input to the model is level 1.5 data from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) instrument aboard the Meteosat Second Generation (MSG) satellite. Concretely, we use data from the 10.8 μm channel (channel 9 of SEVIRI), a window channel in the thermal infrared (TIR) region that is widely employed for rainfall estimation from cloud top temperature [35]. The spatial resolution over our study area is 3.1 km x 3.1 km [36]. The temporal resolution is 15 min.

2.3.2. TARGET DATA: TAHMO RAIN GAUGE DATA

To develop the models, we used hourly rain gauge data from the Trans-African Hydro-Meteorological Observatory (TAHMO) [15] as target data. TAHMO provides quality-controlled rainfall data, available in near real time. There are eight TAHMO stations in our study area during the research period (July 2018-December 2020). Their locations and characteristics are displayed in Figure 2.1 and Table 2.1, respectively. Table 1 also includes the number of rain events per station per year. Here, a rain event is defined as an uninterrupted time period of over-zero rainfall measurements with a cumulative rainfall of at least 1 mm, and there is a 1 h separation window between rain events.

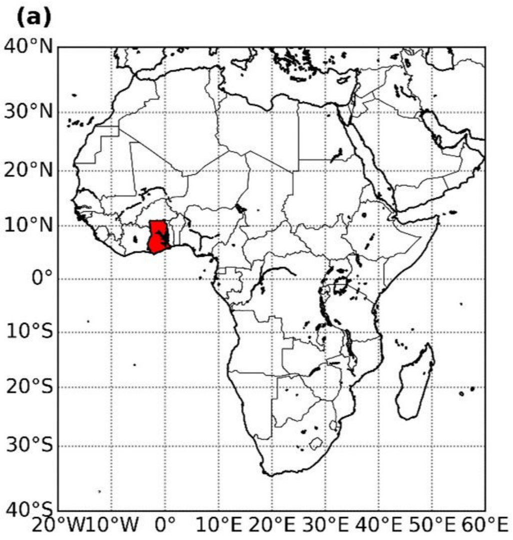


Figure 2.1: (a) Ghana located in West Africa and (b) TAHMO stations considered in this study. UTM coordinates.

Station Code	Station Name	Latitude	Longitude	Elevation (m MSL)	Data Gaps 2018 [%]	Data Gaps 2019 [%]	Data Gaps 2020 [%]	Number of Rain Events 2018	Number of Rain Events 2019	Number of Rain Events 2020
TA00136	Notre Dame Seminary/SHS, Navrongo	10.88°N	1.07°W	187	85.39	20.76	4.22	6	31	83
TA00251	Daffiama SHS, Daffiama	10.42°N	2.55°W	330	0	0	0	71	47	68
TA00253	Han SHS, Han	10.67°N	2.46°W	320	0.43	5.01	37.31	62	25	59
TA00254	Bongo SHS, Bongo	10.91°N	0.81°W	223	0	0	24.19	59	38	52
TA00259	Kpandai SHS, Kpandai	8.48°N	0.03°W	215	92.62	0	0.01	0	50	68
TA00260	Bimbilla SHS, Bimbilla	8.86°N	0.05°W	195	65.13	20.21	0.69	4	64	92
TA00264	Gbewaa College of Education, Pusiga	11.07°N	0.11°W	260	0	0	2.74	50	79	74
TA00616	CSIR-SARI, Nyankpala –Tamale	9.40°N	1.0°W	191	100	31.20	0.02	0	70	31

Table 2.1: Description of the considered TAHMO stations, data gap percentage at hourly timescale and number of rain events during the study period (1 July 2018-31 December 2020).

2.3.3. BENCHMARK PRODUCTS

We used two benchmark products for performance evaluation [Table 2.2]: PERSIANN-CCS [17], as a reference operational ML-based satellite rainfall product, and IMERG [16], as a very high-quality global satellite rainfall product.

Product	Temporal Resolution	Spatial Resolution	Input Data
IMERG	30 min	$0.1^{\circ} \times 0.1^{\circ}$ (10 km × 10 km)	TIR and PMW satellite data, gauge analysis, and additional input data
PERSIANN-CCS	1 h	$0.04^{\circ} \times 0.04^{\circ}$ (4 km × 4 km)	TIR satellite data

Table 2.2: Summary of the characteristics of the benchmark products used in this study.

PERSIANN–CCS builds on its predecessor, PERSIANN [12], and estimates rainfall from GEO IR images. First, the model segments and classifies clouds into cloud patches based on manually selected features such as cloud texture or geometry. Second, it learns the relationship between brightness, temperature, and rainfall rates for each cloud patch [17]. PERSIANN-CCS has a latency of approximately 3 h. A possible limitation of this method lies in the human-assigned features and group definitions of cloud patches, which may be reductionist or faulty in representing physical (rainfall) processes that are not yet fully understood.

IMERG has been developed by NASA and is available as three different products with varying latency times and more data being incorporated in successive runs of the algorithm: Early run, with a 4 h latency time; Late run, with a 12 h latency time; and Final run, with a latency time of 3.5 months. NASA advises using the Final Run as research-ready data. Here, we evaluated all three products. The latest algorithm upgrade of IMERG at the time of writing this paper was version 06.

IMERG relies on multiple data sources and algorithms: It employs GEO IR satellites, "as many as possible" opportunistic LEO satellites, and monthly gauge analyses [16]. The LEO satellites provide PMW rainfall estimates that are propagated forwards and backwards in time using estimated rainfall motion vectors. GEO IR estimates are added using the PERSIANN-CCS algorithm to fill in the gaps between LEO PMW estimates. The Early Run of the algorithm only has forward propagation, whereas the Late Run has both forward and backward propagation, allowing for interpolation. Furthermore, the longer latency time allows for lagging data transmissions that might have been missed in the Early Run to be incorporated in the Late Run. The gauge analyses from the Global Precipitation Climatology Centre (GPCC) are used to regionalize and correct biases in the final stage of the algorithm. Other input data are the GPM Combined Radar-Radiometer (CORRA) rainfall estimates, Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2), and Goddard Earth Observing System model (GEOS) Forward Processing (FP) precipitable water vapor data [16].

2.3.4. STUDY AREA: NORTH OF GHANA

Northern Ghana, defined here as the northern part of Ghana comprising the five northern regions and not the northern region alone, lies between latitudes 8°N and 11°N and longitudes 3° W and 0°30E and is situated in the Savanna climatic zone. It is heavily affected by high variability in climate and hydrological fluxes, with frequent floods and droughts accompanied by high temperatures. This produces frequent crop failures or losses, outbreaks of diseases, and dislocation of human populations, with major economic repercussions [37]. Over 70% of employment in Ghana is in near-subsistence agriculture in rural areas [38].

Ghana's climate is characterized by markedly seasonal rainfall with high interannual variability. Rainfall seasons are determined by the movement of the intertropical convergence zone (ITCZ), which oscillates between the north and south tropics throughout the year [38]. The ITCZ separates a cold, moist air mass moving northward from the Atlantic and a dry, hot, and dusty air mass from the Sahara Desert. As opposed to the south of Ghana, which has two annual rainy seasons, northern Ghana has a unimodal rainfall regime, with a rainy season from March to October, when the ITCZ is in its northernmost position [38, 39]. Figure 2.2 shows the average monthly temperature and rainfall in Bawku, in the upper-east region of Ghana, which is representative of the climatology of the region. Over 75% of rainfall in this area is due to deep convection, most of it organized as large mesoscale convective systems [40]. Intense and short-lived events as a result of deep convection characterize the diurnal rainfall variation in this region [41]. For example, over 80% of rain events present in our development dataset last less than 3 h.

2.3.5. DATA PREPROCESSING

Figure 2.3 shows the flow diagram of the overall methodology presented in this research, with special detail given to the data preprocessing stage.

One-hour TAHMO and thirty minute IMERG data were accumulated in 3 h intervals, while PERSIANN-CCS data were directly obtained with a 3 h resolution. All three products were classified as rain/no-rain using a 1 mm/3 h threshold.

Data scarcity poses an obstacle to DL-based rainfall estimation or prediction, in that the existence of densely gridded data to use as training data during model development is a prerequisite for most

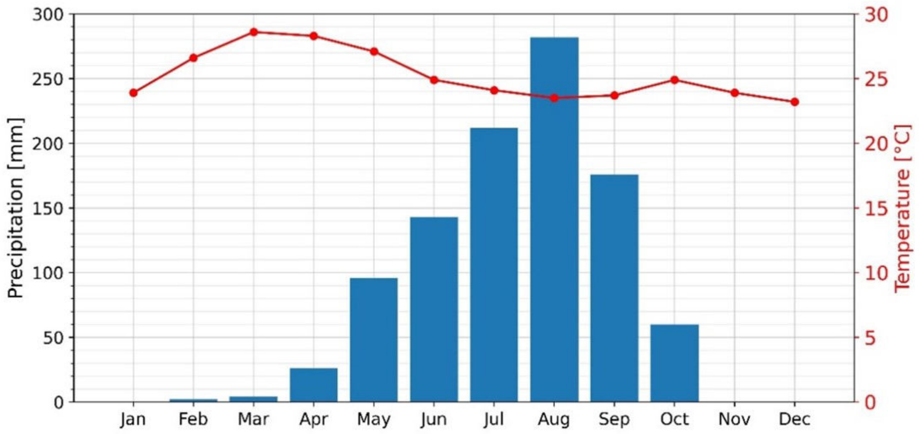


Figure 2.2: Average monthly temperature and rainfall in Bawku, upper-east region, Ghana (1993–2011). Data adapted from [36]. Bawku is representative of the climatology of our study area: the five northern regions of Ghana.

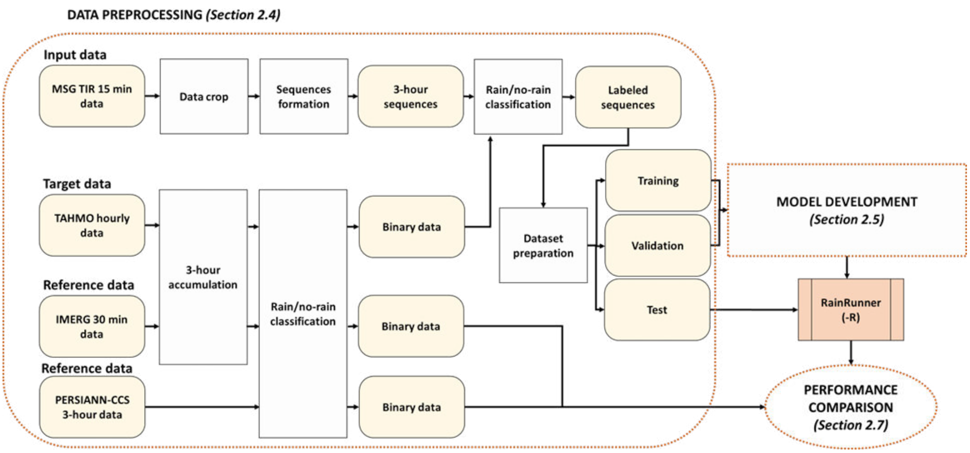


Figure 2.3: Overall flow diagram of the methodology followed in this study, from data preprocessing to performance comparison.

existing approaches [24, 31–33]. Our study area has a sparse rain gauge distribution, with distances between stations too large to allow reliable interpolation, especially considering the highly localized rainfall patterns in West Africa. We employ a methodology to overcome this obstacle by using point-based instead of gridded data as the output of the model. RainRunner utilizes an image-to-point approach: the model is trained only with point-based rainfall data, corresponding to the center of the input image. Some studies [42, 43] have used a similar methodology, cropping satellite data around rain gauge measurements used as target data before being input to a CNN in a DL model to estimate rainfall. However, both approaches use other rainfall measurements present in the cropped scene - and other data sources - as input to the models. Muraux et al. (2021) [42] uses all rain gauges present in the scene, and Wu et al. (2020) [43] uses TRMM 34B2 precipitation data. In our case, MSG TIR images are the only model input, and they were cropped to create 32 pixels × 32 pixels (i.e., approx. 96 km × 96 km area) images centered on each TAHMO station as shown in Figure 2.4. Images were cropped in a way to ensure that the corresponding station fell in a "center square", defined as a square with sides of length equal to the pixel size and with center on the geometrical center of the image. In this way, the model's spatial resolution is the pixel size, i.e., approx. 3 km.

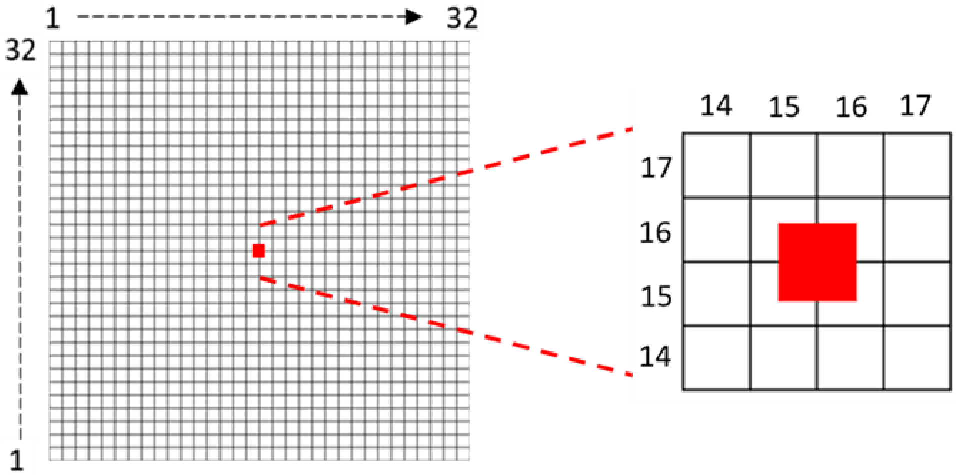


Figure 2.4: Center square in a 32 × 32 pixels image. Here, pixels are numbered from bottom to top and from left to right.

Cropped MSG TIR images were grouped in 3 h sequences (i.e., groups of 12 images). Sequences were then classified as rain/no-rain according to the corresponding TAHMO data. Incomplete sequences due to gaps in TAHMO or MSG data were discarded. We chose a 3 h temporal resolution according to the short-lived rainfall events characteristic of this area. We expect this resolution to be able to capture the daily rainfall dynamics and deem a finer temporal resolution not needed for the end goal of our research, which is to improve the quality of rainfall information for agricultural applications.

To prepare the model development datasets, first we resampled the training dataset to deal with the data imbalance characteristic of rainfall binary classification [44]. We employed a 4:1 dry/rain ratio. Validation and test datasets were created with the same dry/rain ratio as the full 2020 data, i.e., 28.2:1, in order to be representative of reality. The data distribution is presented in Table 2.3.

Dataset	**Total Data Samples**	**Dry Data Samples**	**Rain Data Samples**
Training (2018, 2019, and 2020)	5317	4248	1069
Validation (2020)	7304	7054	250
Test (2020)	7303	7053	250

Table 2.3: Dataset distribution in training, validation, and test.

We assigned all 2018 and 2019 data to the training dataset. Out of the 2020 data, we randomly selected two sets of 250 rain sequences for the validation and test datasets; the rest were assigned to the training dataset.

After model development, its performance on the test dataset was evaluated through comparison to IMERG and PERSIANN-CCS. However, IMERG Final run and PERSIANN-CCS presented data gaps in the validation and test datasets. IMERG Final run had 241 gaps in the validation dataset and 229 in the test dataset, while PERSIANN-CCS only had 110 gaps in the validation dataset. Conveniently, so as not to penalize further the minority class, all corresponded to sequences recorded as dry by TAHMO stations. For a fair comparison, these sequences were removed during result evaluation.

2.3.6. DEEP LEARNING MODEL

We framed the rainfall binary classification as a supervised binary classification problem. We developed two model architectures: RainRunner, based only on convolutional neural networks (CNN), and RainRunner-R, which incorporates a convolutional long short-term memory (ConvLSTM) architecture. Both models have the same input (sequences of 12 TIR images taken every 15 min) and output (rain/no-rain classification).

CNNs are deep neural networks with convolutional layers that exploit symmetries in gridded data by recognizing similar patterns and features, achieving efficient processing and generalization by reducing the number of learned parameters [14]. CNNs treat pixels as connected to their neighborhood instead of independent from each other through convolution and pooling operations [45]. This enables them to account for spatial correlations in rainfall. They are more computationally efficient than multi-layer perceptrons (MLPs) [8]. LSTM architectures are improved recurrent neural networks that incorporate a sequential inductive bias by means of memory cells and gates that selectively maintain and propagate important information across timesteps. This allows them to effectively process sequential data such as time series and natural language texts [46]. ConvLSTM [47] is an extension of LSTM to 2D sequences, i.e., images changing in time, instead of point-based time series. As such, they are suitable techniques to capture the spatio-temporal evolution of satellite gridded data.

We selected these relatively simple methods because of the nature of our problem. State-of-the-art (SOTA) methods for image classification and sequential processing based on transformers require tens of millions to a billion parameters to achieve top performances on benchmark datasets [48]. Training is performed using millions to billions (e.g., pretraining) of images and exceptional computing power [49]. These settings are very different from the case study we are considering, where we are dealing with fewer than ten thousand images. As such, it is beyond the scope of our paper to test very large SOTA models. Instead, we focus on more basic DL models capable of dealing with limited data to explore the overall suitability of the DL approach for this context. We employed ConvLSTM to test whether using a more suitable inductive bias to process our sequences would yield better results.

We differentiate two building blocks: CNN and ConvLSTM blocks. A CNN block comprises multiple convolution and pooling layers. The output of a CNN block is a feature map with dimensions greater than or equal to 8×8. The ConvLSTM block consists of ConvLSTM and batch normalization layers, with the output of the block being a 2D tensor. Besides these building blocks, we also used MLPs with a dropout layer between the hidden and output layers.

RAINRUNNER ARCHITECTURE

Upon receiving an image sequence, RainRunner processes each image in parallel through a CNN block and an MLP to produce one bounded real value (0,1) from each one of them. Then, these outputs are concatenated into a fully connected layer and passed through a second MLP to classify the 3 h input sequence as rain/no-rain. Figure 2.5 shows a schematic block diagram of this architecture.

RAINRUNNER-R ARCHITECTURE

RainRunner-R processes all the images as a sequence through a ConvLSTM block. The output of this block is a 2D tensor that is then passed through a CNN block and an MLP to produce a rain/no-rain prediction. This architecture is shown in Figure 2.6. We investigated the effect of bidirectionality on the ConvLSTM architecture. Bidirectional recurrent neural networks allow training a model using both time directions (i.e., past to future, future to past) of the input when a whole sequence is available. While they cannot be used for forecasting

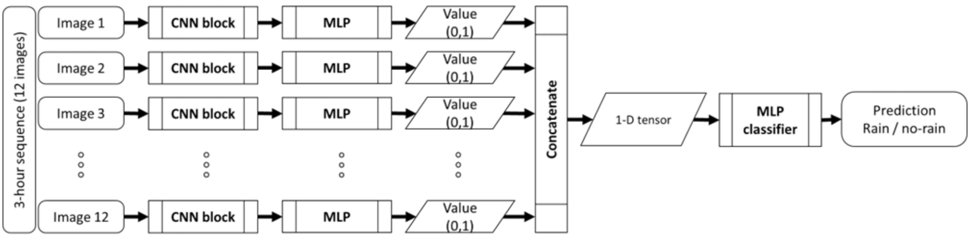


Figure 2.5: Schematic block diagram of the RainRunner model architecture.

purposes, they are particularly suitable for sequence recognition tasks such as ours [50].

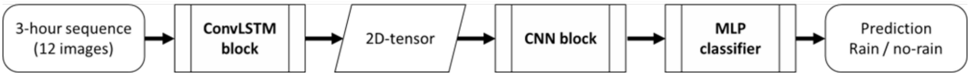


Figure 2.6: Schematic block diagram of the RainRunner-R model architecture.

2.3.7. Training and Hyperparameter Search

To account for data imbalance, we trained the models to minimize a weighted binary cross-entropy loss, where a weight of 0.8 was given to the rain class and 0.2 to the dry class (Equation (2.1)).

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i))] \tag{2.1}$$

where N is the size of the dataset, y is the label / true value (i.e., 0 for no-rain and 1 for rain), and $p(y)$ is the prediction probability (i.e., the estimated probability of each sequence i containing rain).

We trained multiple hyperparameter combinations and chose the best models based on a trade-off between the validation F1-score and the number of trainable parameters. We ran these models ten times and selected the overall best model for both RainRunner and RainRunner-R based on the validation F1-score. Using F1-score as a performance metric helps deal with the rain/dry data imbalance.

2.3.8. Performance Metrics and Misclassification Analysis

We used performance metrics commonly used in the meteorology field as well as the F1-score, a metric commonly used for imbalanced problems in DL, all extracted from the contingency table [Figure 2.7]. Accuracy

(Equation (2.2), where TP denotes True Positives, FP denotes False Positives, FN denotes False Negatives and TN denotes True Negatives) represents the number of correctly classified data samples out of all data samples; probability of detection (POD, Equation (2.3)) measures the ability of the model to correctly detect rain sequences; success rate and false alarm ratio (SR and FAR, Equation (2.4)) are complementary and represent the certainty with which rain sequences are detected; frequency bias (FBias, Equation (2.5)) represents the degree of correspondence between rain predictions and observation; finally, F1-score and critical success index (F1-score and CSI, Equation (2.6) and Equation (2.7)) evaluate at the same time SR and POD.

		Observed	
		Rain	No rain
Predicted	Rain	True positives (Hits)	False positives (False alarms)
	No rain	False negatives (Misses)	True negatives (Correct negatives)

Figure 2.7: Contingency table for the rain/no-rain binary classification problem.

POD, SR, F1score, and CSI can vary from 0 to 1, with 1 being the optimal value. FBias can range from 0 to ∞, with the optimal value being 1. If FBias is below 1, the events are under-forecasted; if it is greater than 1, they are over-forecasted.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \tag{2.2}$$

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.3}$$

$$\text{SR} = 1 - \text{FAR} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.4}$$

$$\text{FBias} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FN}} \tag{2.5}$$

$$\text{F1 score} = \frac{2 \times \text{SR} \times \text{POD}}{\text{SR} + \text{POD}} \tag{2.6}$$

$$\mathrm{CSI} = \frac{1}{\left(\frac{1}{\mathrm{SR}} + \frac{1}{\mathrm{POD}} - 1\right)} \tag{2.7}$$

We present results in three ways: as contingency tables, numerically as the forecast verification metrics, and visually in a Roebber diagram [51] or performance diagram. To assess the generalization ability of the models in the context of the highly localized and seasonal rainfall in northern Ghana, we analyzed their performance depending on factors such as location and time of the year (Table 2.4) in terms of the distribution of misclassified sequences.

Factor	**Possible Values**	**Description**
Station	Bimbilla, Bongo, Daffiama, Kpandai, Han, Navrongo, Pusiga, Tamale	Each one of the 8 TAHMO stations
Month	January to December	Each month of the year
Time of the day	Day	6 AM to 6 PM in the local time (constant throughout the year near the equator)
	Night	6 PM to 6 AM
	Dry	<1 mm/3 h
Rain category	Very light rain	1 mm/3 h to 1 mm/h
	Light rain	<2.5 mm/h
	Moderate rain	2.5 mm/h to 7.6 mm/h

Table 2.4: Factors considered for misclassification analysis. The “rain category” factor follows the definition of rain in the Glossary of Meteorology of the American Meteorological Society, AMS, except the “very light rain” category introduced here.

We compared the performance of RainRunner to that of the benchmark products by computing the forecast verification metrics and performing a misclassification analysis of all products. To assess the difference in performance of the three IMERG products–i.e., Early, Late, and Final Run - we included all of them in the forecast verification metrics computation. For the misclassification analysis, we used IMERG Final Run, as the highest-performing satellite rainfall product. We conducted the performance evaluation based on the results of the test dataset. For reference, we also include the forecast verification metrics of all products on the validation dataset.

2

2.4. RESULTS

2.4.1. SELECTION OF BEST-PERFORMING MODEL ARCHITECTURE

We tested 48 hyperparameter combinations for both RainRunner and RainRunner-R and evaluated learning rates of 0.001, 0.0005, and 0.0001. We used a batch size of 32 and 400 epochs, with an early stopping criterion based on the improvement of the validation loss. We used an Adam optimizer [52].

The F1-score values of the tested model architectures ranged from close to 0 to almost 0.5 (Figure 2.8a), with the best-performing models doing so at the expense of a high number of trainable parameters (Figure 2.8b). Given the limited amount of data available for validation, we selected the simplest, best-performing architectures to reduce the chance of overly optimistic estimates of the model's performance on unseen data [53]. We ran the chosen architectures ten times to select the overall best-performing models (Figure 2.8c). RainRunner achieved an overall higher validation F1-score but a lower median value than RainRunner-R. The model architectures that resulted in the best performances are shown in Figure 2.9. They consist of 120,125 parameters for RainRunner and 21,033 parameters for RainRunner-R, i.e., RainRunner-R has 17.5% of the trainable parameters of RainRunner. The best-performing architectures for both models had two concatenated convolution blocks (i.e., two convolutions + pooling operations). For the RainRunner-R model, a bidirectional ConvLSTM resulted in the best performances, which is in line with the classification task at hand, for which both time directions might contain useful information.

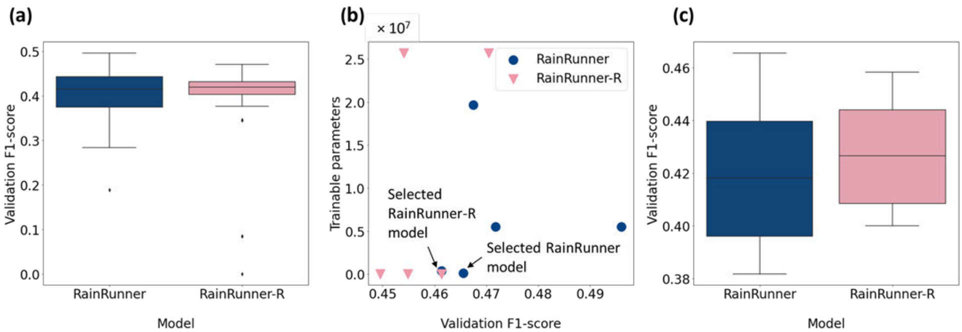


Figure 2.8: Results of the hyperparameter search: (a) performance distribution of all tested model architectures; (b) number of trainable parameters and validation F1-score of the five model architectures with the highest validation F1-score; and (c) performance distribution of the selected models over 10 runs.

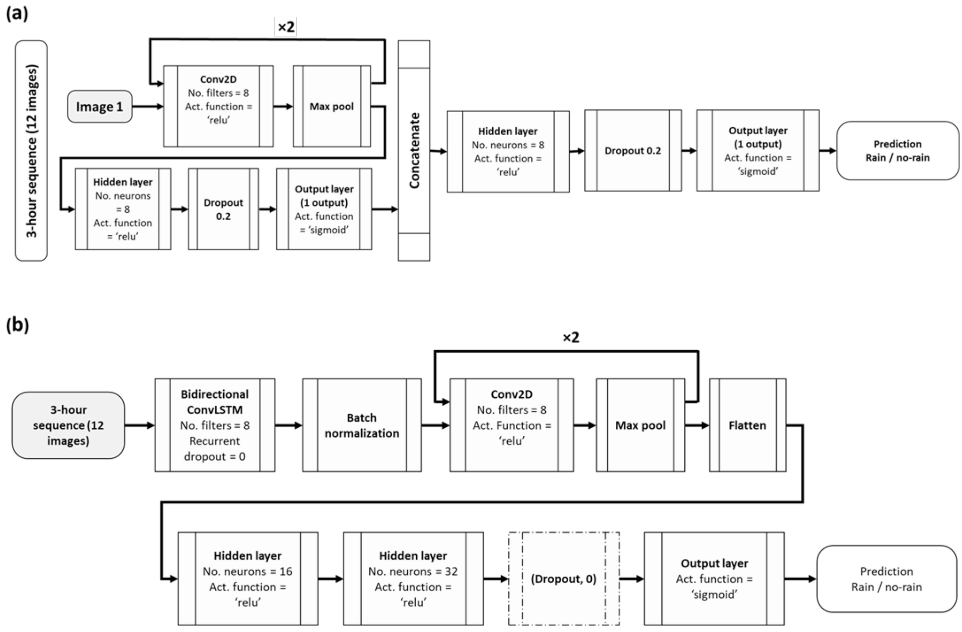


Figure 2.9: Architecture of the best (a) RainRunner model (based on a CNN only) and (b) RainRunner-R model (combining ConvLSTM and CNN). In Figure 9a, only the pipeline of one image is shown, the other 11 images go through a similar CNN block.

2.4.2. MODEL PERFORMANCE EVALUATION

Table 2.5 shows the values of the performance metrics of the selected RainRunner models on our validation dataset, compared to those of PERSIANN-CCS and IMERG Early, Late, and Final run on the same dataset. RainRunner scored higher than PERSIANN-CCS on all metrics except FBias and achieves a POD similar to that of IMERG. The weakest point of RainRunner seems to be the substantially higher FBias.

Figure 2.10 shows the contingency tables of the same models for our independent test dataset. IMERG Final run achieved the overall highest performance among all models, with 95% dry and 82% rain sequences correctly classified. IMERG Late and Early runs followed closely, with 95% (78%) and 95% (76%) dry (rain) sequences correctly classified, respectively. The remaining models achieved a similar performance in dry sequence classification: 94% for RainRunner, RainRunner-R, and PERSIANN-CCS. Lastly, both RainRunner models outperformed PERSIANN-CCS in rain sequence classification, with 74% of rain sequences correctly classified by RainRunner and 73% by RainRunner-R, as compared to 68% by PERSIANN-CCS. Finally, Figure 2.11 summarizes the performance scores in a Roebber diagram. A perfect model - with

Model	Accuracy	F1–score	POD	SR	FBias	CSI
RainRunner	0.94	0.47	0.78	0.33	2.36	0.30
RainRunner-R	0.94	0.46	0.77	0.33	2.34	0.30
PERSIANN-CCS	0.94	0.43	0.63	0.28	2.26	0.24
IMERG Early Run	0.94	0.47	0.73	0.35	2.10	0.31
IMERG Late Run	0.95	0.49	0.78	0.37	2.14	0.33
IMERG Final Run	0.95	0.52	0.82	0.38	2.16	0.35

Table 2.5: Performance metrics on the validation dataset.

POD, CSI, and FBias equal to 1 - would be in the upper-right corner of the diagram. We can see three clusters: the best performance corresponded to the three IMERG products, followed by the RainRunner models, and lastly, PERSIANN-CCS. RainRunner had the largest FBias, which indicates that it over-detected rain more often than the other models. In all the other performance metrics, the two RainRunner models outperformed PERSIANN-CCS on the test dataset.

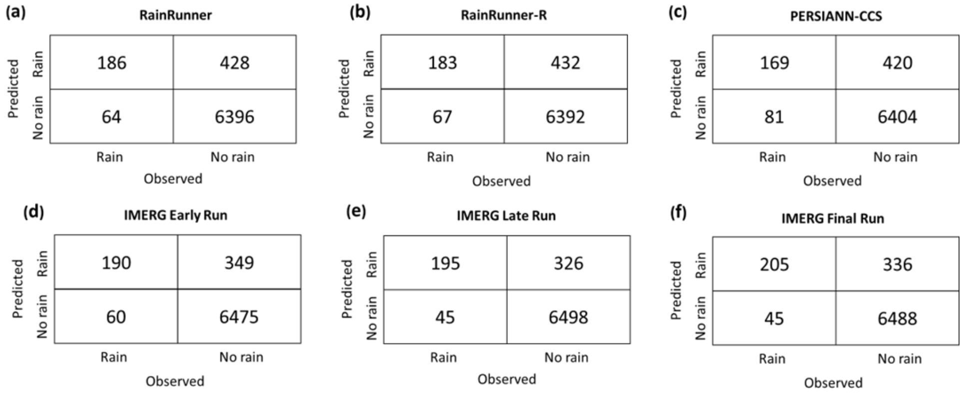


Figure 2.10: Contingency tables of (a) RainRunner, (b) RainRunner-R, (c) PERSIANN-CCS, and (d) IMERG Early, (e) Late, and (f) Final run on the independent test dataset, consisting of 250 rain and 6824 dry sequences.

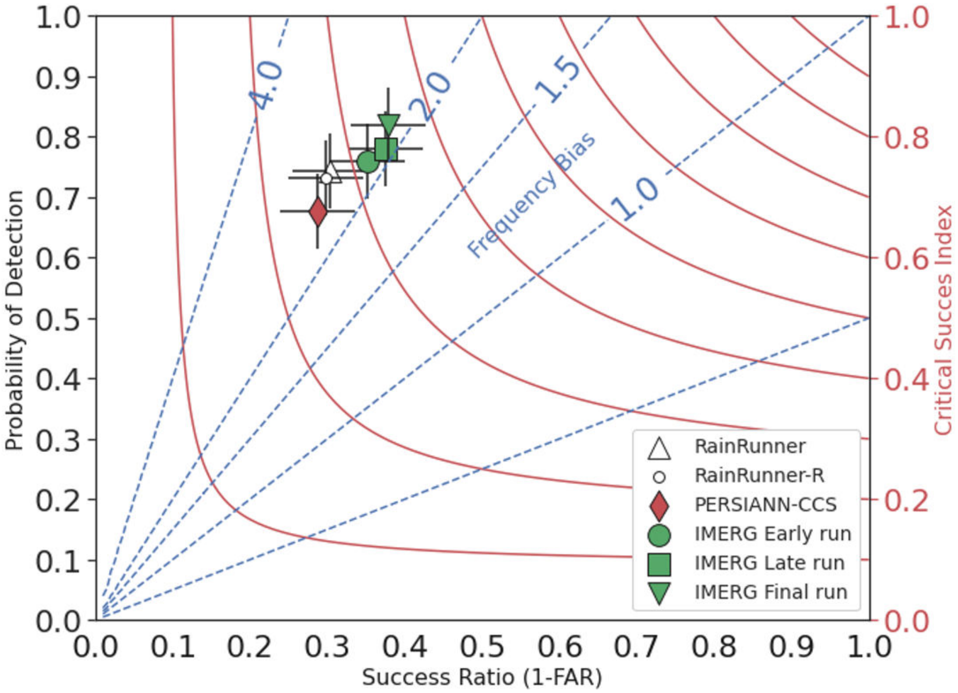


Figure 2.11: Roebber performance diagram on the test dataset.

2.4.3. MISCLASSIFICATION ANALYSIS

Figure 2.12 shows the distribution of misclassified sequences of the test dataset across individual stations, month of the year, and rain categories. About 4% to 10% of the sequences were misclassified across all stations. The southern stations showed a somewhat higher proportion of misclassified sequences than the northern stations. Seasonally, the percentage of misclassifications was much higher in the rainy season than in the dry season, when all models correctly classified nearly all sequences. IMERG performed better in the first half of the rainy season yet misclassified substantially more sequences in the second half of the rainy season, with up to 14% misclassifications for the month of September. For all models, the most challenging events to classify were very light and light rainfall, often misclassified as dry sequences. We also investigated the performance of the models in terms of misclassifications distributed over different times of the day (day/night), but there was no significant difference in the number of misclassifications for day and night.

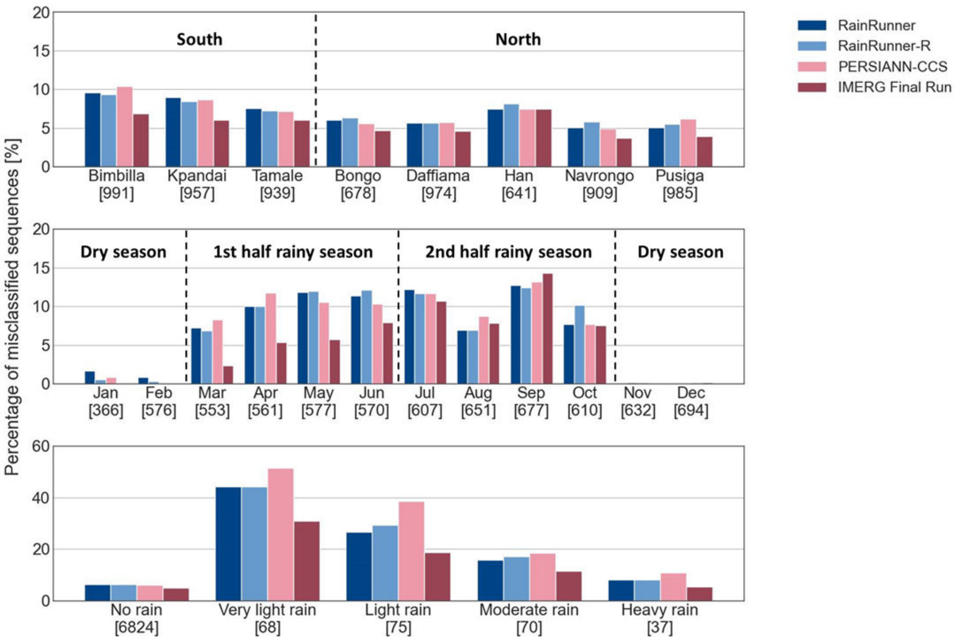


Figure 2.12: Misclassification analysis according to (top) station, (center) month, and (bottom) rainfall intensity. The numbers in brackets below the x-axis represent the number of sequences of each category in the test dataset.

2.5. DISCUSSION

Our findings show that DL models for rainfall binary classification trained with a small local dataset of strictly TIR data compare well to state-of-the-art global products. These results suggest three insights: (1) TIR data are strongly related to rainfall in this region; (2) DL can extract relevant features linking TIR images with rainfall; and (3) locally developing a DL model enables it to capture the characteristics of local processes, in this case, rainfall occurrence, better than some globally trained models.

The strong relationship between brightness temperature (Tb) and rainfall has been extensively studied and used for satellite rainfall retrieval. This relationship is particularly relevant in the Sahel, where around 75% of surface rainfall is due to deep convection that involves cold cloud tops, observable in TIR data [40]. RainRunner surpasses PERSIANN-CCS, which uses machine learning to link TIR data to rainfall through manually extracted features related to cloud properties. This shows that DL methods are able to extract relevant features from data and model natural processes better than expert-based models that rely on manual feature extraction. Especially training the model locally allows it to reproduce regional rainfall patterns more efficiently.

As seen in the Roebber performance diagram Figure 2.10, all models over-predict rainfall with an FBias greater than 2. It is known that TIR-based methods over-predict rainfall because the size of large convection systems is much larger than the surface rainfall area underneath [40]. A further explanation of this over-prediction lies in the characteristic West African rainfall processes. Particularly, the presence of rain-bearing clouds does not necessarily mean rainfall on the ground. Sometimes rainfall does not reach the ground due to the higher concentration of aerosols and associated smaller drops, higher land surface temperature, and drier atmosphere compared to other regions [11]. Therefore, adding other relevant sources of information such as aerosols, land surface temperature, or water vapor data might improve the performance of the models. Furthermore, virga precipitation evaporating before it reaches the ground accounts for 15% of all precipitation profiles in the northern African Savanna (8°–12°N) [41]. Virga has been found to account for 50% of false PMW precipitation results in arid regions [54] and could be a cause for IMERG's rainfall over-prediction. Furthermore, the presence of other MW radiation scatterers, such as dry sand, also results in satellite PMW retrievals over-estimating rainfall [41]. Despite the proven efficiency of DL methods to reproduce physical processes, data scarcity poses a challenge to their employment. To overcome this, we have used an image-to-point methodology that only needs point-based rainfall measurements. Although other studies have applied similar methodologies [42, 43], they required additional rainfall information - additional rain gauges in the study region or a gridded satellite product - as model inputs. Compared to these, our approach has the advantage that it does not require any further rainfall information. Of the two DL architectures we evaluated, results suggest that the temporal inductive bias introduced by the ConvLSTM architectur - processing each image in the 12-image sequence one after the other - does not improve model performance, although it results in a model with fewer trainable parameters (21,033 against 120,125 for RainRunner). The hyperparameter search in model design produced a wide range of performances for both models, which is probably explained by the relatively small training dataset. To investigate the robustness of the models, further research on a range of small to larger datasets would be needed. It is striking that our DL models based on TIR data only, developed with a small dataset and simple model architectures, achieve a performance close to that of IMERG. The high learning efficiency of the DL model, when trained with local data, is promising for the independent application of such models in data-scarce areas such as sub-Saharan Africa. Additionally, it might be interesting to investigate combining the DL model with existing products such as IMERG, where the DL approach can offer complementary insights that help improve performance. For example, substituting the PERSIANN-CCS rainfall estimation scheme from TIR data within IMERG with our better-performing approach might improve IMERG's estimations. With most agriculture in West Africa being rainfed, access to accurate rainfall information is necessary for agricultural productivity. Satellite rainfall products, such as the one developed in this study, that, after training, can be interpolated to areas with no ground observations can play an essential role in overcoming the data scarcity challenge and contributing towards food and economic security.

2.6. CONCLUSION

In this paper, we have developed two DL models based on the CNN and ConvLSTM architectures. The output of our models is a rain/no-rain binary classification of 3 h sequences. We show that our models compare well against existing products despite being considerably simpler, developed with a small training dataset - observations from 8 stations over 2.5 years, with 20.4% data gaps - and using TIR data alone. Specifically, our models consistently outperform PERSIANN-CCS for rain/no-rain detection at a sub-daily timescale. While IMERG is the overall best performer, the DL models perform better than IMERG in the second half of the rainy season despite their simplicity (i.e., up to 120 k parameters). Compared to our models that follow a black-box approach from raw MSG TIR data, IMERG uses data from multiple LEO and GEO satellites, both TIR and PMW, combined with reanalysis and rain gauge data. The high performance that the models are able to reach despite the important challenge of data scarcity shows their high efficiency and, ultimately, the potential of DL to model rainfall in regions with low data availability. We overcome the challenge of data scarcity to develop DL models with an image-to-point methodology that only needs point data instead of densely gridded rainfall information from the ground.

The DL model based on CNN achieved somewhat higher performance than the one including CNN and a ConvLSTM. The temporal structure information brought by the ConvLSTM architecture enables the model to achieve similar performances as when based on CNN, with only 17.5% of the trainable parameters but at the expense of a slower training process.

We suggest that regionally training a DL rainfall model can result in better performances than global models, especially in areas with complex, highly region-specific meteorological characteristics, such as the Savanna region of West Africa.

Further work includes the addition of other EO data as inputs to the model. Particularly, and because of the drier atmosphere characteristic of our study region, the SEVIRI water vapor channel is expected to improve the performance of satellite rainfall estimation. Aerosol data from the Sentinel 5P satellite is also to be added. We expect that the incorporation of these two data products will capture the atmospheric conditions that are the potential causes of rainfall over-detection in West Africa. Furthermore, because the aim of our study was to prove the potential of deep learning methods for providing rainfall information in data-scarce areas, finding the optimal model through a thorough hyperparameter search was out of our scope. However, we believe such a search would improve model performance, and we strongly encourage it. At the same time, we recommend the expansion of the development dataset to cover a longer period and/or a wider region in West Africa, which would allow for the use of more advanced architectures such as ConvNeXt [55] and eventually enable direct rainfall estimation. We expect that the fully data-driven approach can give useful insights into rain processes in the West African savanna.

3

WATER VAPOR AND TEMPORAL DATA TO COMPLEMENT TIR DATA IN THE RAINRUNNER MODEL

This chapter has been published in Atmosphere 14(6), 974. as *The Role of Water Vapor Observations in Satellite Rainfall Detection Highlighted by a Deep Learning Approach*. (2023). Authors: Estébanez-Camarena, M., Curzi, F., Taormina, R., van de Giesen, N., and ten Veldhuis, M. -C., doi: 10.3390/atmos14060974 [56].

3.1. ABSTRACT

West African food systems and rural socio-economics are based on rainfed agriculture, which makes society highly vulnerable to rainfall uncertainty and frequent floods and droughts. Reliable rainfall information is currently missing. There is a sparse and uneven rain gauge distribution and, despite continuous efforts, rainfall satellite products continue to show weak correlations with ground measurements. This paper aims to investigate whether water vapor (WV) observations together with temporal information can complement thermal infrared (TIR) data for satellite rainfall retrieval in a Deep Learning (DL) framework. This is motivated by the fact that water vapor plays a key role in the highly seasonal West African rainfall dynamics. We present a DL model for satellite rainfall detection based on WV and TIR channels of Meteosat Second Generation and temporal information. Results show that the WV inhibition of low-level features enables the depiction of strong convective motions usually related to heavy rainfall. This is especially relevant in areas where convective rainfall is dominant, such as the tropics. Additionally, WV data allow us to detect dry air masses over our study area, that are advected from the Sahara Desert and create discontinuities in precipitation events. The developed DL model shows strong performance in rainfall binary classification, with less false alarms and lower rainfall overdetection (FBias < 2.0) than the state-of-the-art Integrated MultisatellitE Retrievals for GPM (IMERG) Final Run.

3.2. INTRODUCTION

In West Africa, rainfed agriculture is the main pillar of the food system and rural socio-economics. For example, in Ghana, the focus area of this study, agriculture accounts for 54% of the total Gross Domestic Product [57] and is predominantly rainfed small-holder farming. Rainfall in this area is highly uncertain and there are frequent floods and droughts, exacerbated by climate change. Reliable and timely rainfall information is essential to effectively face these challenges and avoid major economic and yield losses. However, a sparse and unevenly distributed rain gauge network-as is typical for tropical areas [58, 59]-and regionally poor-performing satellite rainfall products,hinder the availability of accurate dense rainfall information.

The global coverage of Earth observation satellites can offer a solution for poorly ground-monitored areas. The most widely used methodologies for satellite rainfall retrieval are based on thermal infrared (TIR) from Geostationary (GEO) satellites and passive microwave (PMW) data from Low-Earth Orbit (LEO) satellites. Because of their closer proximity to the Earth's surface, LEO satellites allow for a higher spatial resolution but have the disadvantage of a longer revisit time, which often translates to rainfall events being missed. On the contrary, GEO satellites provide a lower spatial resolution but have the advantage of a constant view of the full Earth disk from their unique position, always above the same point above the Earth's surface. This enables them to have a high temporal resolution and to monitor atmospheric processes like no other satellite platform. This will only become more apparent with Meteosat Third Generation, for which the first satellite has recently been launched [60]. Retrieval methods can be physical- or Machine Learning-based, or a combination of both. Within Machine Learning, Deep Larning (DL) aims to minimize human intervention and facilitate automated feature extraction from large raw datasets [61]. This new data-oriented approach is a promising method to detect and possibly estimate rainfall when theoretical or process-based approaches fail to accurately parameterize such complex atmospheric processes.

Physical-based retrieval methods that use TIR data predominantly employ the Cold Cloud Duration (CCD) method, which correlates the time that a pixel is under a certain temperature threshold with rainfall on the ground. Two examples of this approach are the Tropical Applications of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT) [62] and the Africa Climate Hazards Infrared Precipitation with Stations (CHIRPS) [4] rainfall products, specifically designed for Africa with daily and 6-hourly temporal resolutions, respectively. Results from a calibration of the CCD method in the Sahel region are unreliable due to spatial averaging and temporal aggregation, as well as low gauge density [63]. In West Africa, both TAMSAT and CHIRPS show daily Kling-Gupta Efficiency values below 0.4 [64, 65].

To address the limitations of the CCD method and exploit the benefits of DL, ref. [66] developed a novel DL-based methodology: RainRunner. RainRunner classifies 3 h intervals into rain/no-rain, based only on TIR data. Rainrunner was trained over the North of Ghana using rain gauge data as target, with a very small training dataset-measurements from 8 rain gauges over 2.5 years-with TIR data as the only input and based on standard DL architectures. Nonetheless, this approach showed promising results, reaching near state-of-the-art performances. However, as expected for methods that rely only on TIR data [67], RainRunner heavily overdetected rainfall.

PMW sensors allow for a more direct retrieval of rainfall than TIR sensors because they directly sense hydrometeors in the atmosphere. Using this advantage, the Global Precipitation Measurement (GPM) Integrated Multisatellite Retrievals for GPM (IMERG) rainfall product combines data from TIR and PMW sensors, along with atmospheric reanalysis and rain gauge data. Developed by NASA through the use of physical- and Machine Learning-based algorithms, IMERG aims to become the longest and most detailed rainfall dataset available [68]. Compared with other regions of the world, IMERG shows a weaker correlation with ground measurements in West Africa [69, 70].

The literature suggests that the poor regional performance of satellite rainfall products over West Africa is partly due to sparse rain gauge coverage [63], as has also been observed in other regions of the world [71]. Another reason for this poor performance is the complexity of West African rainfall dynamics. They are governed by the seasonal northward shift of the Intertropical Convergence Zone (ITCZ) and the West African Monsoon (WAM), a low-level south westerly moist flow from the Atlantic Ocean. Wind shear generated by the monsoonal flow creates a strong temperature contrast-especially from June to September-between the dry hot Sahara Desert and the cool moist Guinea coast that favors the formation of the African Easterly Jet. The African Easterly Jet is a unique zonal wind feature located in the midlevel troposphere around 600 hPa (Figure 3.1) and is most intense at the end of August. The jet is caused by a thermal wind balance that promotes the development of the African Easterly waves (AEWs) through baroclinic and barotropic instability [58]. Many studies [72–74] have identified water vapor as a key factor in West African rainfall dynamics. Studies have shown that the main support for the intensification of AEWs is moist convection. At the same time, latent heat release from condensation of atmospheric water vapor and a strong solar irradiation would be the key promoters of unstable atmospheric conditions that lead to sparse but heavy precipitation events in the form of thunderstorms.

In this paper, we build on RainRunner by incorporating water vapor (WV) data as an input to the model. Furthermore, to capture the seasonality and the diurnal cycle of rainfall in this region, we also add the temporal information of the satellite observations as additional input data to the model. The goal of our study is to evaluate the impact of WV observations combined with temporal information on satellite rainfall retrieval in tropical regions and to what extent they can complement TIR data. This paper is organized as follows: First, the data used during our study are introduced together with our study region and research methodology in Section 3.3. Our results are presented in Section 3.4 and subsequently discussed in Section 3.5. Finally, our conclusion and some insights into future work are reported in Section 3.5.

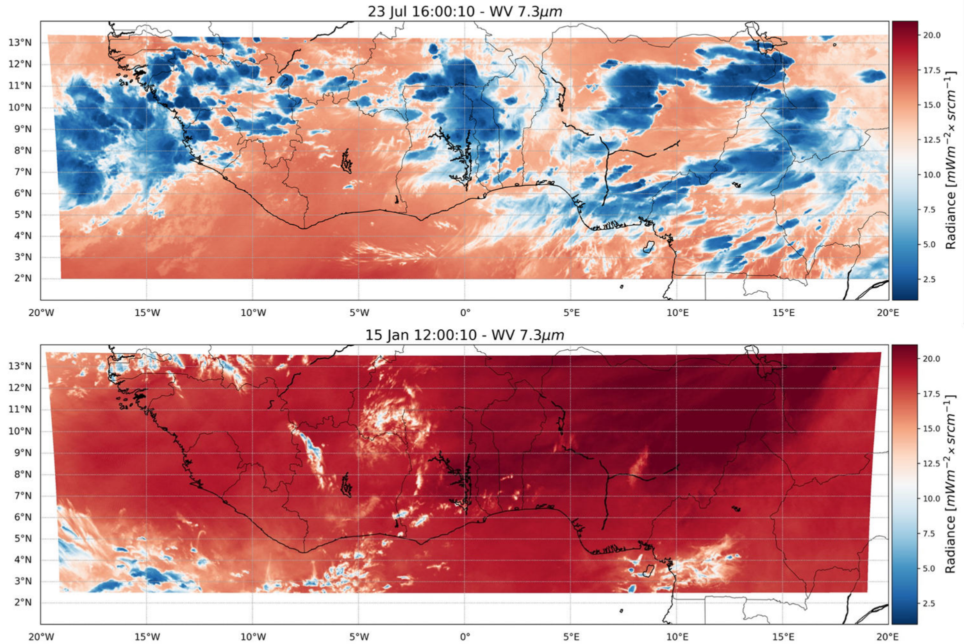


Figure 3.1: Illustration of the difference between rain (**a**) and dry (**b**) season as seen in 7.3 μm imagery: (**a**) Easterly moisture transport during boreal summer under the influence of midlevel jets. (**b**) Dry low-level wind blowing from Sahara desert slightly visible during dry season.

3.3. MATERIALS AND METHODS

3.3.1. DEVELOPMENT DATASET AND BENCHMARK SATELLITE RAINFALL PRODUCTS

The input data to the model are level 1.5 data from two channels of the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) satellite. They have a 15 min temporal resolution and a 3.1 km spatial resolution over our study region [75]. Building on RainRunner [66], we employ 10.8 µm TIR data (channel 9 of SEVIRI). Additionally, we incorporate 7.3 µm WV data (channel 6). Our choice to employ these data instead of 6.2 µm WV data (channel 5), which is the other WV channel of SEVIRI, is based on the fact that channel 6 enables penetration further down into the atmosphere than channel 5, which is situated in the center of the water vapor absorption band. Observing water vapor further down in the atmosphere can be useful to interpret humidity features associated to midlevel jets in a strong convective environment (Figure 3.1). This is very relevant for our study region, where the rainy season is heavily dependent on the African Easterly Jet, which transports moisture horizontally in the middle troposphere. “Further down” is meant in a relative sense. Although the water vapor channel 6 is a thermal band, it does not represent the temperature of the Earth’s surface but the temperature of the so-called effective layer. Only with a very dry troposphere is the WV channel able to reach surface levels (e.g., eastern Sahara desert and Antarctica) [76]. In most circumstances, such as those encountered in the study area, radiation from water in the lower parts of the atmosphere is readily absorbed by water vapor higher up in the atmosphere. Thus, radiation from low liquid water clouds, such as stratocumulus and nimbostratus, does not reach the satellite but is absorbed by water vapor in higher layers. Therefore, channel 6 is not helpful in detecting any rainfall produced by these low clouds. What is observed by the satellite is the temperature of the effective layer, or the layer above which there is insufficient water to absorb radiation from below. The effective layer can include the middle layer in which the all-important African Easterly Jet is situated, typically situated at 3000 masl. A very cold effective layer would indicate the presence of water vapor or ice at high levels in the atmosphere, up to 10,000 masl, which is typically associated with cumulonimbus clouds, which is also relevant for rainfall detection. Finally, the timestamp of MSG data, i.e., date and time of the day of each observation, is also model input. This is to take into consideration the diurnal heating cycle and seasonality patterns closely related to rainfall in this region.

To analyze the added advantage of incorporating WV into the model, we used the same target training data as in our previous work developing RainRunner [66]. That is, hourly data from eight Trans-African

Hydro-Meteorological Observatory (TAHMO) rain gauges in the north of Ghana [77] (Figure 3.2a) over a study period spanning from July 2018 to December 2020, included. Figure 3.2b shows the amount of missing data per station during this time period.

The benchmark satellite rainfall product used in this study is IMERG, developed by NASA, as it is currently the best-performing satellite product over our study region. It combines PMW data from as many Low Earth Orbit (LEO) satellites as possible with TIR data from different Geosynchronous Earth Orbit (GEO) satellites to fill in gaps between PMW measurements and monthly rain gauge data from the Global Precipitation Climatologic Centre (GPCC). TIR estimates are produced using Machine Learning, while PMW estimates through forward and backward propagation using rainfall motion vectors based on atmospheric reanalysis data. IMERG is available in different versions with increasing latency time and model complexity: IMERG Early Run, with a 4 h latency time and only forward propagation; Late Run, with a 12 h latency time and backward propagation; and Final Run, with 3.5-month latency time that is adjusted using gauge data from the Global Precipitation Climatology Centre Full and Monitoring products [78]-hence the longer latency time and higher performance. NASA recommends using the Final Run product for research [79].

3.3.2. STUDY AREA: NORTH OF GHANA

The study area is northern Ghana, between 8° N and 11° N latitude and 3° W and 0°30′ E longitude. The climate in this region corresponds to that of the broader Sudanian savanna agro-ecological zone of West Africa [80]. West Africa has one of the most extreme climatic gradients in the world, where the most significant climatic element is rainfall. The mean annual rainfall steadily increases southward towards the equator, with extremes ranging from near-zero in the arid part of the Sahel up to over 2000 mm/year in the coastal zones [81].

Northern Ghana has a unimodal rainfall regime, with a peak generally occurring during the months of July and August. The dry season in this region starts in November and lasts until late March. During this period of time, there are virtually no significant precipitation events [82]. Rainfall patterns in this area are highly regional and present a strong diurnal cycle. The main characteristics of the rainfall regime in the region of interest are visualized in Figure 3.3. Precipitation displays characteristics of a convective and very heavy rainfall regime: seasonal heavy short-lived thunderstorms (Figure 3.3b,c), short-lived events, with the majority (82%) not lasting more than 3 h, and a close to 20 mm/h median value of the heaviest rainfall events.

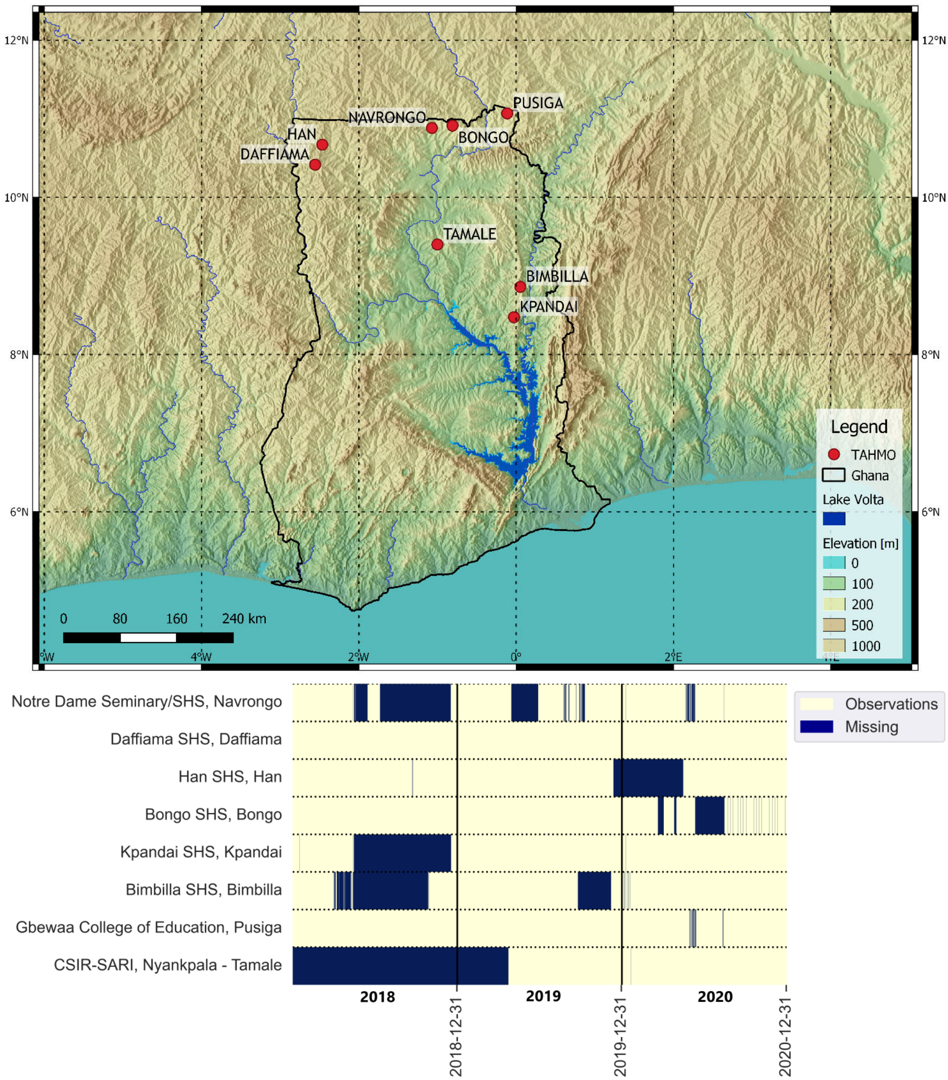


Figure 3.2: (**a**) Digital elevation map of the study area (GRASS QGIS) and locations of the TAHMO stations. Data retrieved from https://www.usgs.gov/ (accessed on 1 June 2022). (**b**) Missing data for each TAHMO station in north Ghana during our study period from July 2018 to December 2020.

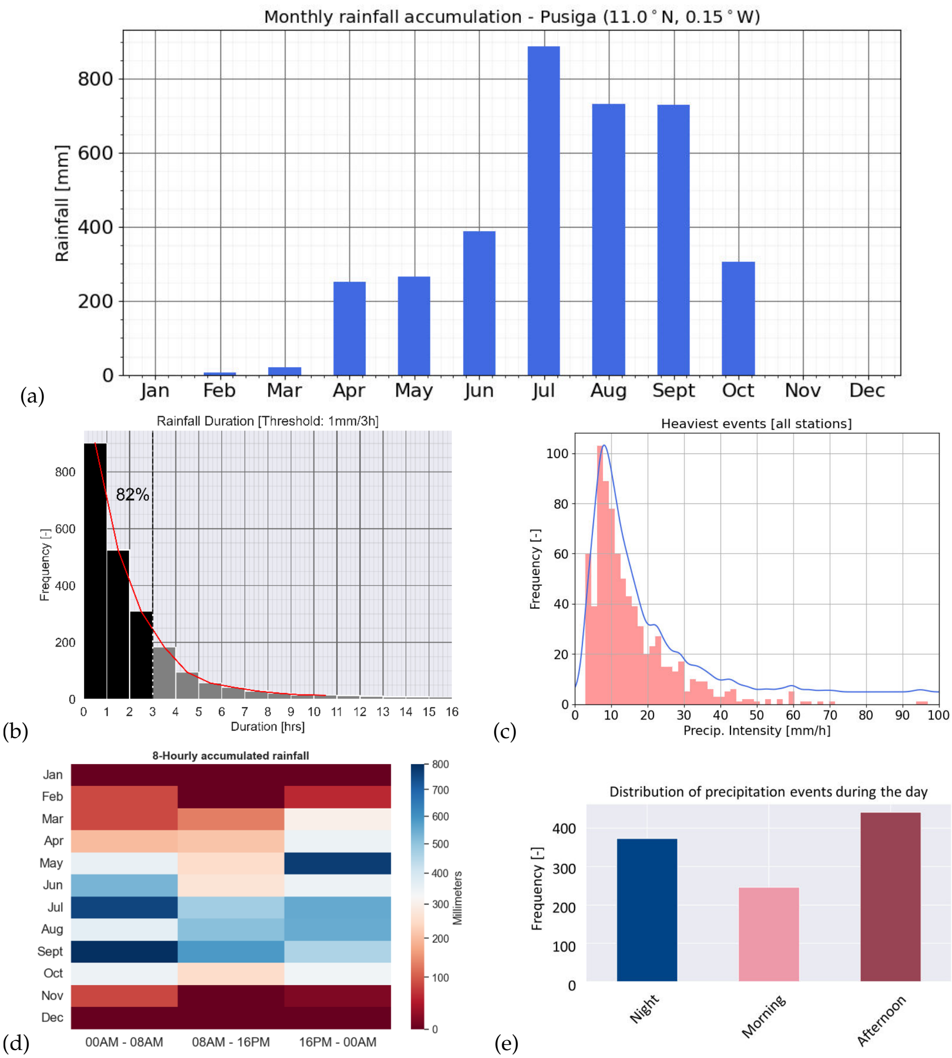


Figure 3.3: Rainfall dynamics in Northern Ghana expressed as (**a**) monthly rainfall patterns, (**b**) frequency distribution of rainfall duration, (**c**) frequency distribution of precipitation intensity of the 100 heaviest rain events at each station, (**d**) seasonal distribution of rainfall accumulation per time of the day, and (**e**) frequency distribution of precipitation events (>1 mm/3h) based on time of the day, depicted as a violin plot. In these graphs, frequency corresponds to the number of occurrences in the entire development dataset, as described in Table 3.1. These results are based on hourly data from the four TAHMO stations with no gaps during at least 2 full years within our study period (no missing data for at least 66% of the considered period 2018-2020): Daffiama (TA00251), Pusiga (TA00264), Bongo (TA00254), Kpandai (TA00259).

Figure 3.3d shows a progressively erratic diurnal cycle of convection during the rainy season starting from May due to the strengthening of the African Easterly Jet. A stronger African Easterly Jet consequently enhances horizontal moisture transport (visible in WV 7.3 µm data) and the formation of large mesoscale convective systems that propagate overnight and result in large accumulated rainfall values. This pattern peaks in early September where almost 1000 mm falls during nighttime. Morning hours (6 a.m.-12 p.m.) [-20]have generally the least rainfall accumulation, as well as fewer precipitation events (Figure 3.3e). Stable atmospheric conditions are more often found around this time of the day.

On average, northern Ghana is more often under the influence of the hot and arid North Easterly trade wind, which blows air that comes from the Sahara desert, usually carrying a considerable amount of dust, while the southern part of the country receives more maritime influx through moist SW winds.

Table 3.1: Development dataset distribution in training, validation, and test datasets. The validation and test datasets contained sequences from 2020 and were created using a dry/rain ratio computed from all 2020 data to simulate a realistic distribution.

Dataset	Year	Dry Samples	Rain Samples	Total n_Samples	Ratio Dry/Rain
Training	2018, 2019, 2020	4218	1055	5273	4.0
Validation	2020	6627	235	6862	28.2
Test	2020	6627	235	6862	28.2

3.3.3. DATA PREPROCESSING

Sparse ground training data pose a challenge to any ML-based rainfall retrieval model. The methodology described in this section presents a way to overcome the lack of dense ground data by using an image to point approach such as described in [66]. For this purpose, TIR and WV images were cropped to create a matrix of 32 × 32 pixels (96 × 96 km) with the TAHMO station located in a central 2 × 2 pixels square. The spatial resolution of the model corresponds to the pixel size, which is approximately 3.1 km [66]. Cropped images were then grouped to form 3 h (12-image) sequences. The chosen temporal resolution is in line with the rainfall duration pattern of this area. Integrity of the sequences was mandatory: if any sequence included missing data, it was discarded from the process.

Hourly TAHMO ground measurements were accumulated into 3 h intervals to match the temporal scale of the input sequences. A threshold of 1 mm/3 h was selected to discriminate between rain and no-rain sequences. We based our choice of threshold on the short and intense nature of rainfall events in our study region, with most events lasting no more than 3 h. It is recognized that there are different possible and reasonable definitions, but 1mm/3hours was also used in our previous work developing the first version of this model, making a direct comparison more consistent [66]. We aggregated 30-minute resolution IMERG data in a similar fashion for comparison.

To include temporal information about the satellite observations, we mapped the MSG data timestamp onto a circle to represent its cyclical nature. Particularly, from the timestamp, we extracted the month number, from 0 to 11, and hour of the day, from 0 to 21, due to the sequences being 3 h in length. We performed the mapping by converting these two variables into two two-dimensional arrays using sine and cosine transformations. In this way we avoided jump discontinuities from 11 pm to midnight and December to January. Equations (3.1) and (3.1) provide the timestamp encoding, where X is the time variable in question.

$$X_{sin} = sin(\frac{2\pi \times X}{max(X)}) \tag{3.1}$$

$$X_{cos} = cos(\frac{2\pi \times X}{max(X)}) \tag{3.2}$$

The development dataset is highly skewed, as is the rainfall binary classification problem. This means that the number of no-rain sequences is much larger than rain sequences. To deal with this imbalance, we followed the methodology of [66] and used a hybrid approach of data resampling and weighted loss function.

The dataset was split in such a way that the rain sequences in the training dataset were oversampled with a ratio of 4:1 dry/rain, while both validation and test datasets had a ratio of 28.2:1 dry/rain, representative of the full 2020 data. The training dataset contained sequences from 2018, 2019, and 2020, while the validation and test datasets only had sequences from 2020. The dataset distribution was based on the minority class, i.e., rain samples, divided following an approximate 70-15-15 (training-validation-test) ratio. The dry samples were selected randomly using the corresponding dry/rain ratios (Table 3.1).

3.3.4. SATELLITE DATA ANALYSIS

To study the differences between the TIR and WV spectral channels of SEVIRI and their complementarity, satellite data were analyzed using pixel analysis. We followed a top-down approach comparing data from the two channels from the larger synoptic scale over the entirety of West Africa-20° W to 20° E-to the smaller scale (mesoscale) using already cropped MSG images from relevant sequences used for model validation.

The aim of the larger-scale comparison was to visualize the water vapor exclusion of low-level nonconvective features hidden by the West African monsoon during the rainy season. For each SEVIRI TIR channel, the relationship between observed radiance R and the equivalent brightness temperature T_b is given by EUMETSAT and expressed in Equation (3.4). In this relation, R is the observed radiances in mW m^{-2} sr^{-1} (cm^{-1})$^{-1}$, Tb is the equivalent brightness temperature in K, ν_c is a central wavenumber of the spectral channel in cm^{-1}, c_1 and c_2 are constants with values $c_1 = 2hc_2$, $c_2 = \frac{hc}{\kappa}$, where h is Planck's constant, c is the speed of light, and κ is the Boltzmann constant. The central wavenumber ν and the so-called band c correction coefficients A and B were determined by EUMETSAT from a nonlinear regression of a precalculated lookup table using the Planck function for the different thermal infrared SEVIRI channels and are provided on EUMETSAT's website [83].

$$T_b = [\frac{c_2 v_c}{log(1 + c_1 v_c^3/R)} - B]/A \tag{3.3}$$

We analyzed sequences at the smaller scale using gray-level histograms of the normalized pixel values, positively related to equivalent brightness temperature. Because temperature is not constant with height, if the atmosphere is conditionally unstable, there is a negative temperature lapse rate Γ between the Earth surface and a layer at $height = Z$ that can be simplified using the relation expressed in Equation (3.4), where T is the absolute temperature and z the altitude.

$$\Gamma = -\frac{dT}{dz} \tag{3.4}$$

In raw satellite imagery, pixel radiances with values approaching the unity are bright pixels, and they translate into absorption at lower levels of the atmosphere, i.e., the effective layer is located at low levels, which corresponds to higher temperatures. Darker pixels have values closer to 0, which indicate colder temperatures of the effective layer, and therefore, its location will be at a higher altitude. Meaningful events for evaluation were selected manually based on (1) the misclassified probabilistic output values of the models, so that events for which one or both models misclassified a sequence but the combination of both corrected the classification were selected, and (2) the WV mean pixel value being at least a standard deviation away from TIR mean value.

3.3.5. MODEL DEVELOPMENT

We built our model on RainRunner [66]. We expanded the input layer to feed two different streams of twelve 32 × 32 × 1 matrices for a total of 24 input images, with one stream per each SEVIRI channel (TIR and WV).

We increased the number of nodes from 8 to 16, following the increase in the number of input images (from 12 to 24). Figure 3.4 illustrates a condensed diagram of the bispectral model structure. The inputs of WV and TIR are convoluted separately in order to learn information individually from each channel. The output of the convolution and pooling layers is a 2-dimensional (8 × 8 × 1) single tensor generated from each image of the sequence, i.e., 2 convolutions are applied in series. The tensors are then flattened and concatenated before being fed to a multilayer perceptron. The timestamp (month and time of the day) is added directly into the fully connected layer after preprocessing along with the 2D tensors from the convolutional layers. The model has 11,019,197 learnable parameters. The batch size was set to 64 and the learning rate was fixed to 0.0001. The number of passes trough the training dataset was fixed at 300 epochs with an early stopping callback set to 50 to halt the training in case the model was overfitting. The function for the dense layer(s) is a rectified linear function (ReLu), while the output layer function is a logistic function, or sigmoid, which returns a probabilistic output between 0 and 1, where 1 represents 100% rain and 0 is 100% dry. A decision boundary line at 0.5 is used for the classifier to make a distinction between the two classes. Lastly, a weighted loss function was applied to deal with the imbalanced dataset, where dry sequences have 0.2 and rain sequences 0.8 coefficients, which reflected the ratio of dry/rain sequences of the training dataset.

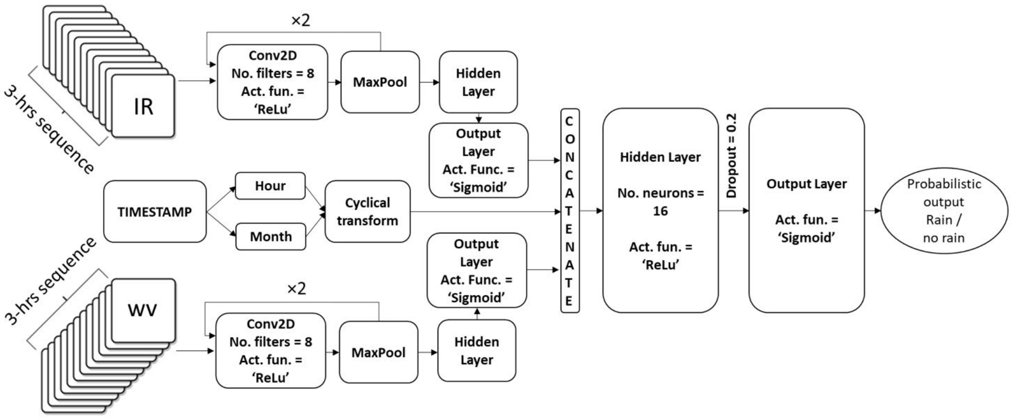


Figure 3.4: Schematic overview of the proposed bispectral (WV + TIR) RainRunner architecture.

3.3.6. PERFORMANCE EVALUATION AND ASSESSMENT OF DATA CONTRIBUTION

In order to assess the individual contributions of water vapor and timestamp, we conducted an ablation study in which we evaluated four models with different inputs but similar architecture/hyperparameters: (1) TIR data only; (2) WV data only; (3) TIR and WV data combined; and (4) TIR and WV data together with the observation timestamp. For a robust comparison, we applied an ensemble average to 10 runs of each model, so as to reduce the variance of the predictions. We evaluated model performance using a set of categorical metrics based on the contingency table, represented in Figure 3.5. These are Accuracy (Equation (3.5)), Probability of Detection (POD, Equation (3.6)), Success Ratio and its complimentary False Alarm Ratio (SR and FAR, Equation (3.7)), Frequency Bias (FBias, Equation (3.8)), F1 score (Equation (3.9)), and Critical Success Index (CSI, Equation (3.10)). POD, SR, CSI, and F1score range between 0 and 1, with 1 being the optimal value. FBias can adopt values from 0 to $+\infty$, with the target value being 1. If FBias is below 1, the model is underforecasting the event; if it is above 1, it is overforecasting it. F1score represents the harmonic mean of SR and POD and is especially valuable for imbalanced problems. Hence, the best averaged models were ranked according to F1 score. Accuracy, POD, SR, FAR, FBias, and CSI are performance metrics commonly used in meteorology for dichotomous forecast verification [84–86]. F1 score is widely used in the Deep Learning field and is especially useful to evaluate highly skewed binary classification problems.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{3.5}$$

$$POD = \frac{TP}{TP+FN} \tag{3.6}$$

$$SR = 1 - FAR = 1 - \frac{FP}{TP+FP} \tag{3.7}$$

$$FBias = \frac{POD}{SR} \tag{3.8}$$

$$F1score = \frac{2 \times SR \times POD}{SR+POD} \tag{3.9}$$

$$CSI = \frac{1}{\frac{1}{SR} + \frac{1}{POD} - 1} \tag{3.10}$$

Evaluation of model performance and of the contribution of each model input was also performed by miscassification analysis, that is,

		Observed: Rain	Observed: No rain
Predicted	Rain	Hits (True Positives, TP)	False alarms (False Positives, FP)
	No rain	Misses (False Negatives, FN)	Correct negatives (True Negatives, TN)

Figure 3.5: Contingency table for the binary rainfall classification problem.

analysis of the distribution of misclassified sequences through the day and across different months, seasons, TAHMO stations, and rain intensities or categories. Rain categories were defined according to the Glossary of Meteorology of the American Meteorological Society (AMS, https://glossary.ametsoc.org/wiki/Rain (accessed on 1 June 2021)), except the "very light rain" category, which was introduced in [66] for a more detailed results analysis, and is as follows:

- Very light rain: 1 mm/3h < RR < 1 mm/h;
- Light rain: 1 mm/h < RR < 2.5 mm/h;
- Moderate rain: 2.5 mm/h < RR < 7.6 mm/h;
- Heavy rain: RR > 7.6 mm/h.

3.4. RESULTS

3.4.1. MODEL PERFORMANCE ON THE INDEPENDENT TEST DATASET

Figure 3.6 displays the contingency tables of the four models evaluated here, the best single run of the model with TIR, WV, and the timestamp as input, and of IMERG Final Run for comparison. Initially, the models that use WV and TIR alone performed similarly, with the TIR model missing a slightly lower number of rain events and the WV model showing less false alarms (false positives). Combining the two channels leads to fewer misclassified dry and rainy sequences. The number of false alarms decreases further when the timestamp is included into the model. On the other hand, IMERG has considerably less misses (false negatives), which can be explained by the model making use of a constellation of LEO PMW satellites, able to sense rainfall more directly than TIR sensors. The best single run of the model with all inputs presents the overall lowest number of false alarms (229), at the expense of a high number of misses (78), which corresponds to a third of the all rainy sequences.

(a) RainRunner TIR

Predicted \ Observed	Rain	No rain
Rain	167	352
No rain	68	6031

(b) RainRunner WV

Predicted \ Observed	Rain	No rain
Rain	164	333
No rain	71	6050

(c) RainRunner TIR+WV

Predicted \ Observed	Rain	No rain
Rain	173	315
No rain	62	6068

(d) RainRunner TIR+WV+Timestamp

Predicted \ Observed	Rain	No rain
Rain	173	255
No rain	62	6128

(e) Best single TIR+WV+Timestamp run

Predicted \ Observed	Rain	No rain
Rain	157	229
No rain	78	6154

(f) IMERG Final Run

Predicted \ Observed	Rain	No rain
Rain	191	316
No rain	44	6067

Figure 3.6: Contingency tables on the independent test dataset of ensemble averaged RainRunner model results using different inputs: (**a**) TIR only, (**b**) WV only, (**c**) TIR + WV, (**d**) TIR + WV + Time, (**e**) the single best run and (**f**) results for the benchmark model, and IMERG-Final.

For better visualization, the categorical metrics are also represented in the Roebber performance diagram [84] in Figure 3.7, where all IMERG products are plotted as reference models. In this diagram, a perfect forecast would be in the top-right corner, with POD, SR, and CSI equal to one.

IMERG-Final has the highest number of hits (true positives). As a consequence, it also has the highest POD of all models, although it has a Fbias well above 2, which means it is severely overdetecting rainfall. The IMERG Early and Late Run products have similar Fbias, yet a lower POD and SR. IMERG Early has comparable performance to the WV_TIR model, while it is outperformed by the TIR_WV + Timestamp which achieves lower Fbias at the same short latency time. The benefit of adding WV and timestamp is noticeable in this diagram, as it progressively leads to a higher success ratio (SR) as well as a lower Fbias, reaching the lowest FBias of all models ($1.5 < Fbias < 2.0$).

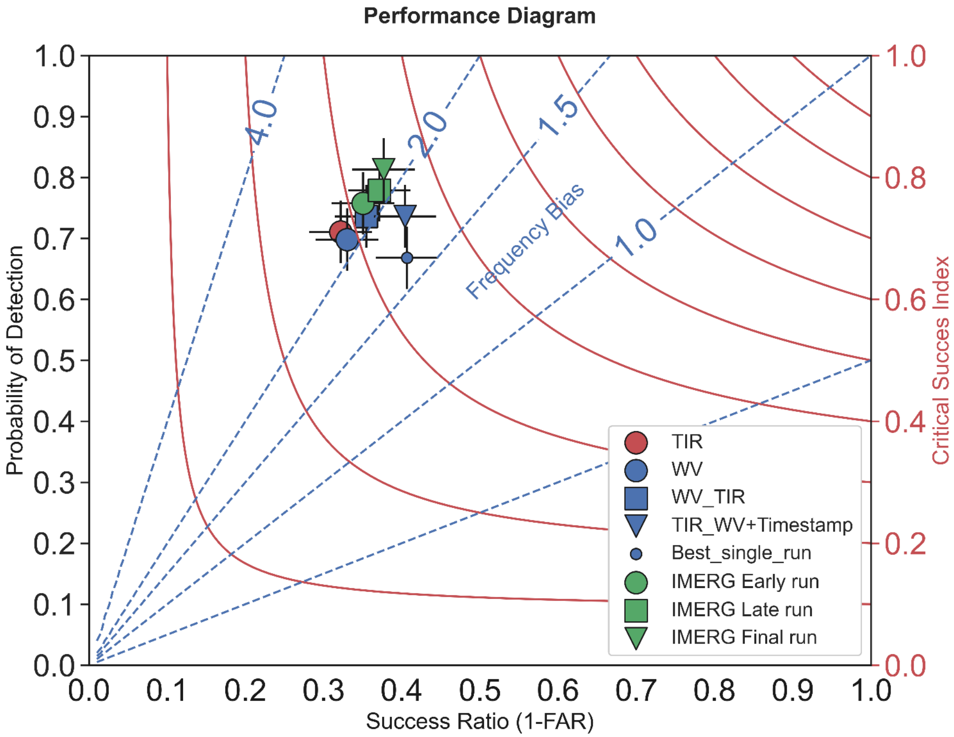


Figure 3.7: Roebber performance diagram on the test dataset.

3.4.2. MISCLASSIFICATION ANALYSIS

Figure 3.8 shows the distribution of misclassified sequences among different factors, i.e., time of the day, month, season, station, and rain category. The northernmost stations overall have fewer misclassified sequences compared with those more to the south of our study region. Overall, the combination of WV and TIR with the timestamp results in the least number of misclassifications of our developed models. The addition of the timestamp is particularly valuable during the dry season. The rainy season (boreal summer) shows a poorer performance than the dry season for all models. It is worth mentioning that IMERG has the highest number of incorrectly classified sequences during the second half of the rainy season (from July to October), highlighting the fact that the influence of the African Easterly Jet on rainfall patterns is a true challenge, even for the most advanced models. The WV model, whose strongest advantage is the correct depiction of convective motions, shows the most misclassifications for light and very light (stratiform) rain detection.

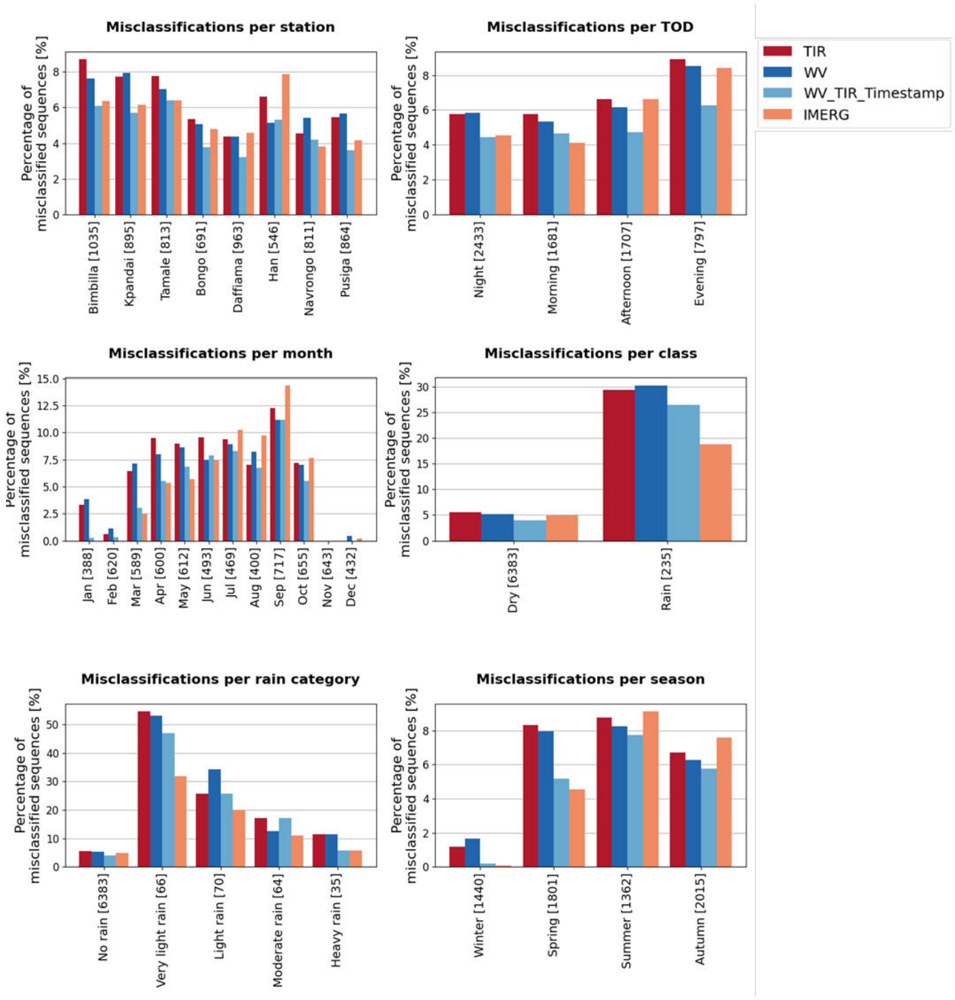


Figure 3.8: Misclassification analysis on the independent test dataset based on station, time of the day, month, class, rain category, and season.

Figure 3.9 illustrates the contribution of the WV and the timestamp information in the model by comparing the probabilistic output of the combined model + timestamp with RainRunner TIR-only (10.8 µm). The addition of the number of the month makes the predictions for the trimester December-January-February (DJF) much lower, with values close to 0. Concretely, while the mean probabilistic output of the model using TIR alone was 0.14, it decreased to 0.005 when incorporating the timestamp. On the other hand, dry intervals during

July-August-September (JAS) are still the most difficult to classify for both models. Results suggest that the addition of the time of the day is especially beneficial during the early rainy season, when the African Easterly Jet is not yet offsetting the diurnal convective cycle and rainfall is still occurring during late afternoon hours.

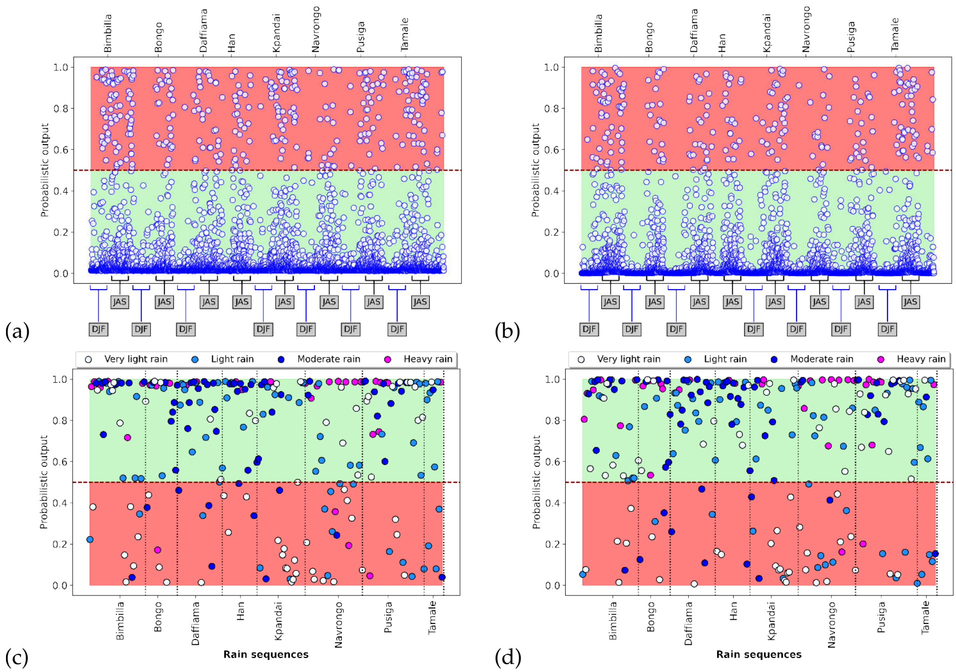


Figure 3.9: Comparison of the ensemble probabilistic output on the test dataset for dry (**a**,**b**) and rainy (**c**,**d**) sequences: The classification threshold applied for classification is 0.5, as indicated in the plots; green color indicates the truth: dry for (**a** and **b**) (output < 0.5) and rain for (**c** and **d**) (output > 0.5). Subplots (**a** and **c**) correspond to the TIR-only ensemble, while (**b** and **d**) correspond to the TIR + WV + Timestamp ensemble. Subfigures (**a**,**b**) present the seasons as acronyms, where JAS stands for July-August-September, i.e, peak of the rainy season, and DJF stands for December-January-February, i.e., the midst of the dry season.

Figure 3.9c,d shows how TIR-only predictions of rainy sequences are closer to unity than the model combined with timestamp. This is particularly true for some rain events that occurred during the shoulder season (March/April or October/November) and obtained a lower probabilistic output with the model using timestamp.

Four heavy rainfall events were misclassified by the model using TIR alone, while only two heavy events were misclassified by the combined model. This is probably due to the ability of WV data to capture strong convective motions associated to heavy rainfall.

3.4.3. PIXEL ANALYSIS COMPARISON

Satellite images over large areas are useful to understand the differences between the TIR and the two WV channels. Figure 3.10 shows a snapshot of West African atmospheric dynamics on 23 July 2020 at noon using equivalent temperature brightness units. Midday is the time at which the solar heating cycle is at its peak and early convection is visible. The image retrieved at 10.8 µm shows information not always related to rainfall, such as many low-level clouds spread across the whole region. Where the sky is clear, the brightness temperature is an indicator of the land surface temperature (the dark red area on the upper part of the figure is near the Sahara Desert). Areas of intense convection (dark blue) are highlighted in water vapor imagery. The softer red shade shown in 7.3 µm is clearly the top of the West African Monsoon layer that acts as threshold level for this channel, hiding low-level clouds. Above this level, the African Easterly Jet transports moisture eastwards and promotes slanted convection. The largest sensitivity range for channel 5 (6.2 µm) is around 350 hPa, which makes this channel completely blind to the West African Monsoon as well as to most of its associated lower-level features. It is still a useful channel to locate deep convective motions that take place in the upper troposphere, where the average temperature is around 240K.

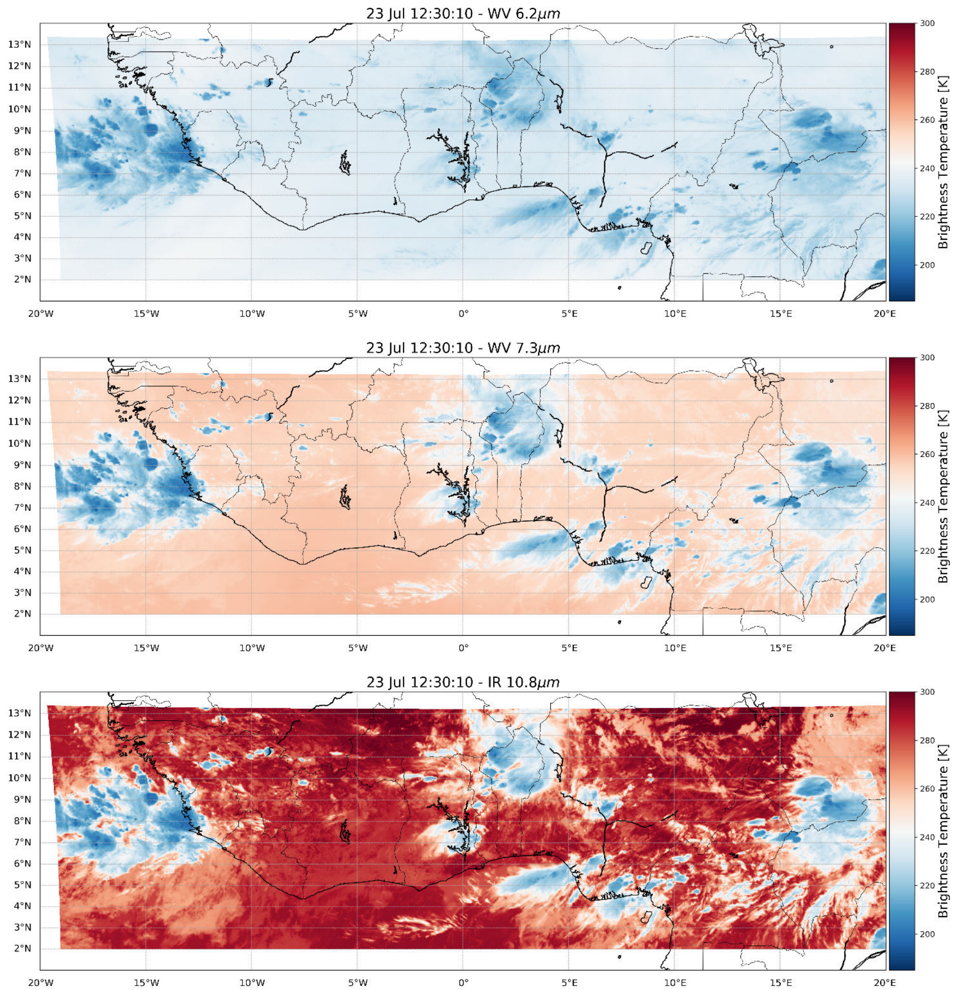


Figure 3.10: Data from the three considered SEVIRI infrared channels over West Africa on July 23 2020 at noon: from top to bottom: WV 6.2 μm (Channel 5); WV 7.3 μm (Channel 6); and IR 10.8 μm (Channel 9).

Figure 3.11 displays some of the analyzed misclassified sequences where the bispectral approach proved to be useful for the model and reflected some insight into atmospheric dynamics. The output of the four evaluated models for these sequences is presented together with the corresponding ground truth in Table 3.2. In Figure 3.11, the images on the left side were selected from entire sequences for being illustrative of the atmospheric event at hand. On the right hand side, the gray-level histogram shows the pixel distribution of each corresponding sequence.

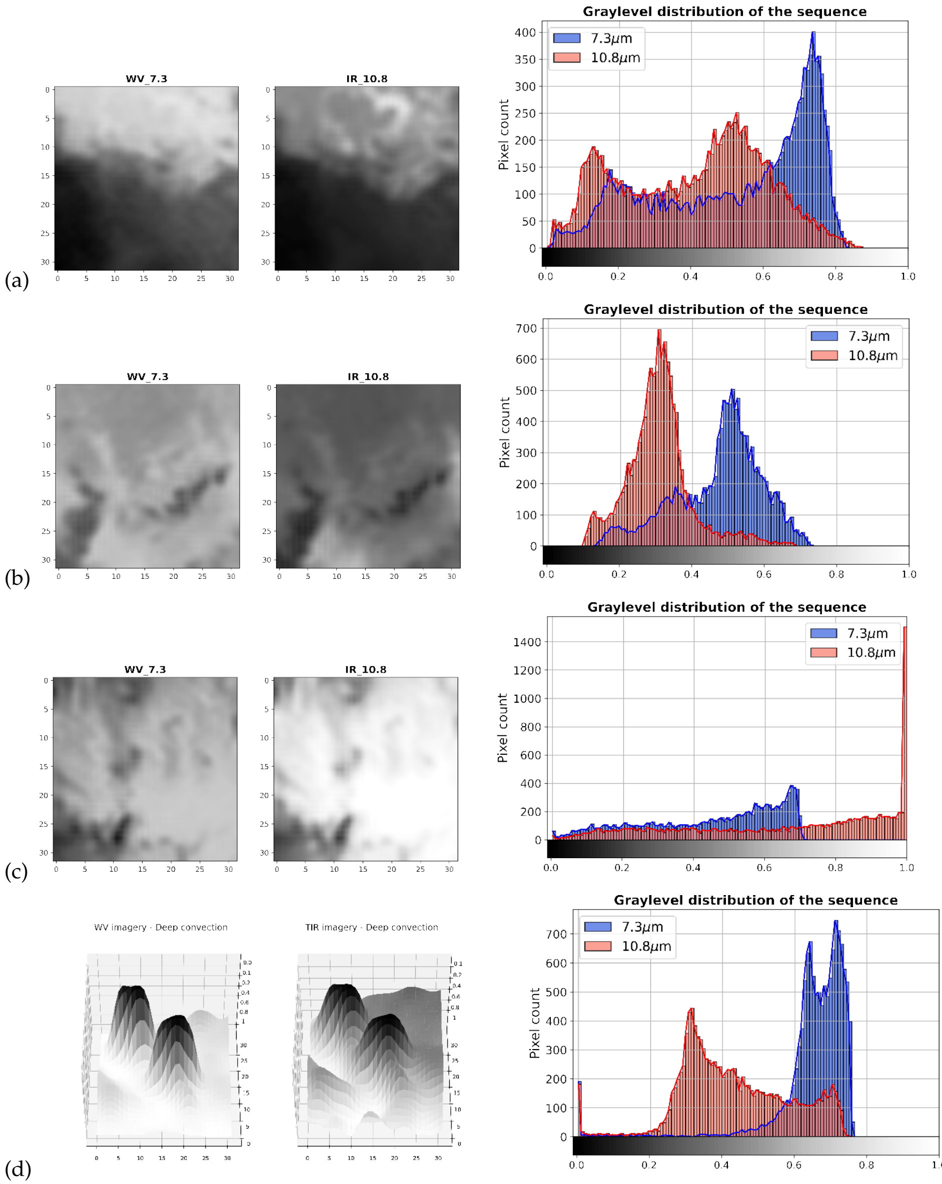


Figure 3.11: Pixel analysis of relevant atmospheric events: (**a**) dry intrusion from the North, (**b**) dry slot, (**c**) low-level moisture detected in WV, and (**d**) 3D deep convective motions of a heavy precipitation event as seen in WV and TIR imagery.

From top to bottom, Figure 3.11a shows a clear dry intrusion. Dry intrusions happen when a tropical system advects air from a dry source, generally right after a precipitation event. They are visible as a sharp gradient in WV imagery but are difficult to locate in TIR imagery, because warmer clouds linger for a longer period of time.

A dry slot is seen in Figure 3.11b. Dry slots can be a consequence of dry intrusions, or they might happen along the transition zone between convective and stratiform rain in larger mesoscale convective systems.

In these two cases, while the model using TIR data alone misclassified the sequence as rainy, the addition of WV allowed to correct it.

Figure 3.11c is a dry sequence from January 2020 (dry season) that was misclassified by WV as rainy. However, TIR data show that there were no rain-bearing clouds at that moment. This can happen when an anomalous low-level moist southerly circulation peaks up during certain days of the dry season, while, at higher levels, dry air is present. In this situation, the 7.3 µm channel retrieves water vapor content from lower levels resulting in incorrect predictions.

Figure 3.11d is a 3D surface plot of the 2D TIR and WV data, aimed at better showing convective motions of a violent rain event as seen from both channels. The Z-axis corresponds to the pixel values. The gray-level of each pixel in WV imagery gives information about the layer depth and clearly shows where strong convection occurs.

As for the gray-level histograms, two distinct peaks are observable in each histogram. That is, the two channels generate an asymmetric bimodal pixel distribution at different brightness temperatures. In the case of WV imagery, the peak is an indication of the most frequently occurring height of the effective layer during the sequence.

Table 3.2: Predicted probabilities from each ensemble model for the selected events in Figure 3.11.

Event	Ground Truth	TIR	WV	TIR + WV	TIR + WV + Timestamp
(a) Kpandai, 30.09.2020, 18h	0	0.60	0.18	0.48	0.47
(b) Bimbilla, 27.05.2020, 9h	0	0.51	0.10	0.42	0.21
(c) Tamale, 23.01.2020, 18h	0	0.42	0.64	0.32	0.14
(d) Pusiga, 27.05.2020, 12h	1	0.04	0.16	0.45	0.20

3.5. DISCUSSION

This study proposed a Deep Learning approach to tackle the challenge of rainfall detection in the Sudanian savanna of West Africa by using bispectral MSG data, i.e., TIR and WV data, as well as temporal information. WV data proved to be useful in detecting the midlevel African Easterly Jet, a main driver of rainfall dynamics in this area. This jet creates a thermodynamic environment favorable for deep convection, observed in WV data without the contamination from low-level clouds observed in TIR data. Furthermore, results show the complementarity of the two MSG channels in scenarios where a monospectral approach would result in misclassifications (Table 3.2). WV allows to reduce the number of false alarms and increase the success ratio in cases where dry air masses-dry slots and dry intrusions-in between tropical systems, missed in TIR data, suppress rainfall (Figure 3.11a,b and Appendix). While TIR data alone would detect the rain-bearing clouds and misclassify these events as rain, the addition of WV data allows to correct the classification. Another scenario in which WV and TIR results are complementary for correct rainfall binary classification is when there is low-level moisture with no rain-bearing clouds. Although WV data alone would misclassify these cases as rain, TIR data are able to correct them for the absence of rain-bearing clouds. For certain events that are more difficult to identify, the gray-level histogram can be helpful to distinguish dry from wet conditions during dry intrusions and dry slots, indicated by the the distance between the mode of the WV and TIR pixel distributions. Of the three scenarios-dry intrusions, dry slots, and low-level moisture-dry intrusions are the most challenging because of the sharp gradient present in the image (Figure 3.11a). This sharp gradient can lead to TIR data alone misclassifying rainfall with high certainty, making it difficult for the bispectral model to capture the correct development of the dry air advection into the rainfall area.

Incorporating temporal information further allows the model to learn regional seasonal and diurnal rainfall patterns. Its contribution is most evident during the dry season, when the model correctly expects mostly dry sequences. This is most advantageous in scenarios with low-level moisture during the dry season, when adding timestamp information reduces rainfall misclassification (Figure 3.11c and Appendix **??**). In fact, the misclassification analysis shows that the model based only on WV data achieves the lowest performance among all models during the dry season. This is because of the variable height of the effective layer. During the dry season, there is very dry air higher up in the atmosphere, and the satellite sensor might detect some anomalous low-level moist currents that are not correlated with rainfall and that might be misclassified as rain. However, these misclassifications can be corrected with the addition of TIR and temporal information. Because most of the analyzed dry slots and dry intrusions events (tables in the Appendix **??**) take place during the early or late rainy season, when the atmosphere is more dynamic, the timestamp contribution is unclear. In this scenario, adding timestamp information reduced the chances of misclassification in 50% of the analyzed cases.

The flipside of including WV data is that it fails to retrieve stratiform rainfall. Stratiform or warm rain is the precipitation that falls from low-level clouds and is usually associated with light rainfall events.

However, its relationship with low-level clouds remains very uncertain, since the presence of such clouds is only sporadically linked to rainfall [87]. The model using WV alone is the worst performing model for very light and light rain (Figure 3.8), likely to be found in stratiform clouds. Depending on the application, this insensitivity to low-level clouds might be a strength or a weakness. More than 80% of the rainfall in tropical inland areas comes from mesoscale convective systems (MCs), and in fact, the presence of low clouds or high clouds such as thin-iced cirrus leads to an overforecast of precipitation in models that only make use of TIR data, which can be seen in Figures 3.6 and 3.7, Table 3.2, and the Annex. The adoption of Channel 5 (WV 6.2 µm as opposed to the used 7.3 µm) would focus the model even more on deep convective events that are strictly related to heavy rainfall events, since it only detects upper-level WV structures (Figure 3.10). However, no information on stratiform rain and shallow convection can be extracted from this channel, so it would result in more missed events. Looking at the Roebber diagram in Figure 3.7, a main drawback of our approach is the low POD, which might be partly explained by our model missing these kinds of light rainfall events.

Along the same line, the discrepancy in misclassified sequences between northern and more southern stations (Figure 3.8) is in agreement with the literature, and it is most likely due to a progressively higher availability of moisture towards the coast, which leads to a slight increase in rain from warm clouds [88].

Our approach can provide the basis to develop a full alternative solution to the established Cold Cloud Duration (CCD) method. This method is a cloud indexing statistical approach applied to the TIR channel to distinguish convective rain clouds from nonrain low clouds. It assumes a positive linear relationship between cloud tops and rainfall to find an optimal temperature threshold for a certain area [89]. However, because of the complexities of convective rainfall, both the temperature threshold and the linear regression relationship depend on local characteristics of the area under consideration. Even if the region of interest is divided into many calibration subareas, the results exhibit several discontinuities in the rainfall estimates. Additionally, each calibration area requires many ground measurements. At the moment, West African gauge coverage is far from sufficient to make this method a reliable option. The strength of this method relies on its simple approach to achieve reliable results at very low temporal resolutions (POD: 0.69, SR: 0.75, BIAS: 0.9 for wet dekadals detection) [62]. The combination of the TIR and WV channels automatically excludes nonconvective features within the whole region of interest. Furthermore, the temporal resolution is higher than for TAMSAT (3hrs vs daily), which is very beneficial in a convective precipitation context. Similarly to the CCD-based CHIRPS and TAMSAT, the model developed in this study is specifically designed for equatorial Africa. The addition of WV data is expected to be less effective in detecting rainfall outside the tropics, where convective rainfall is less dominant. Different factors play a role in rainfall formation in midlatitudes, in particular frontal systems.

Finally, an important advantage of the model is the short latency time of 3 h, as compared with the 3.5 months latency of IMERG Final Run and the 12 h latency of IMERG Late Run. Only IMERG Early Run has a comparable latency time, i.e., 4 h. Precipitation estimates have an important operational value and are essential for crop models and applications such as flood and drought monitoring-for which timeliness is essential in an operational setting.

A promising direction for further development is to transform binary rainfall detection into rainfall estimation. However, geostationary (GEO) IR images have the limitation of providing only indirect rainfall estimates. Passive Microwave Sensors (PMWs) remain the most direct satellite observations for rainfall retrieval, capable of retrieving the rainfall rate by receiving the backscattered signal of hydrometeors. Therefore, the addition of PMW estimates could prove beneficial. As a starting point, rainfall estimates derived from GEO IR imagery could be locally adjusted whenever a PMW observation is available for that region, although post-processing calibration is required to account for grid mismatch [90]. Another advised future development is to increase the temporal and spatial resolution of the model. Certain rainfall events are so highly localized in space and time that the current scale (i.e., 3 h 96 km × 96 km sequences) is too coarse for their detection. As an example, the heavy rain event in Pusiga in May (Figure 1) was incorrectly classified by all models including IMERG Final Run. A well-defined small dark blob in WV imagery appears only at the end of the sequence, while the previous images contained mostly bright pixels that made the sequence easily misclassified as dry. Moreover, a higher temporal resolution would lead to fewer incomplete sequences, which would increase the size of the development dataset. Because most rainfall in this area is attributed to localized pockets of rapid moist air ascent, which are sometimes not larger than a few kilometers, reducing the area of the cropped MSG images could also be beneficial for their detection. On this matter, the new Meteosat Third Generation, for which the first satellite was launched in December 2022, is set to deliver higher spatial (2 km) and temporal (10 min) resolution [60]. These new data could potentially allow the WV channel to detect smaller-scale rising air motions even with the current input shape.

The combination of multiple SEVIRI channels to enhance low-level features by applying a temperature brightness difference between relevant channels might improve the detection of warm rainfall. However, it is likely that precipitation will be more overdetected unless a better relation between the two variables is defined through a $T_b - RR$

relationship. On the other hand, the adoption of the other WV channel 6.2 µm may bring more reliable results on the detection of heavy rainfall events, which account for most of the accumulated rainfall on the ground.

3.6. CONCLUSIONS

This work shows that a DL model is able to tackle rainfall detection in regions where sparse rain gauge networks and erratic precipitation patterns pose a challenge to existing rainfall estimation methods. The incorporation of water vapor information into the model is noticeable and results in a reduced number of false alarms. The true value of WV data for rainfall detection lies in its capacity to detect dry air intrusions into tropical easterly waves, which is of particular interest for regions close to the Sahara Desert. We also show how dry intrusions and dry slots result in false positives using only TIR data, which might be the reason why TIR rainfall products tend to overdetect rainfall. This can be corrected with the addition of WV data. However, using WV data alone can also result in false positives in scenarios with low-level moisture that occur most often during the dry season and that can be corrected by TIR data. This points to the complementarity of WV and TIR data for satellite rainfall estimation in Southern West Africa. Another new input to the model from the original TIR-only version [66] is the temporal information related to date and time. Results reveal that while the addition of temporal information is beneficial in scenarios with anomalous low-level moisture during the dry season, it does not have a clear effect during the rainy season. Finally, our approach allows to decrease false alarms and reach a lower FBias than the much more complex state-of-the-art IMERG Final Run (FBias < 2.0).

4

APPLICATION OF THE RAINRUNNER MODELS TO THE SUDANIAN SAVANNA OF WEST AFRICA

4.1. INTRODUCTION

In the last years Deep Learning (DL) has proven to be a powerful tool for weather and climate modeling, often surpassing purely physical models and other Machine Learning (ML) methods [91–93]. The success of DL has been enabled by a combination of efficient learning algorithms and a vast parametric space, which makes DL models much more complex than other ML models. Unfortunately, as models become more complex, they also become less transparent, creating a trade-off between model explainability and performance [94]. In fact, some known challenges of DL include limitations in generalization and interpretability. Firstly, data-driven models might learn spurious non-physical correlations, due to the high complexity of the physical processes, that might overfit the training dataset and not perform well when applied to other contexts. Secondly, the complex architectures of DL make its decision-making process opaque and difficult to understand, which is known as the "black-box" nature of DL [95, 96].

To address the "black-box" problem, Explainable Artificial Intelligence (XAI) aims to enhance the interpretability and transparency of ML models. XAI encompasses a range of techniques to establish clearer connections between the inputs and outputs of the ML models, enabling users to understand the reasoning behind model decisions [94, 97]. By understanding a ML model, users can gain trust in it as well as acquire new learnings about physical processes being modeled, such as possible causal relationships between input and output variables [94]. Thus, DL can be used to broaden our knowledge about physical processes. In climate models and projections, coupling a DL model with causal discovery has been shown to improve model performance and provide insights into the physical drivers of atmospheric processes [96].

When applied to the complex rainfall processes in West Africa, dominated by the West African Monsoon system [98], XAI could provide new insights into the relationships between different atmospheric variables and rainfall. The West African Monsoon system is a large-scale circulation characterized by reversal of low-level winds transporting moisture inland from the Atlantic Ocean [98]. Furthermore, rainfall dynamics in Southern West Africa, including the Sudanian Savanna present diverse diurnal and seasonal cycles, and complex relationships between meteorological variables such as atmospheric moisture, ocean and land temperatures, wind and cloud top temperatures [99–102].

The ablation study performed in Chapter 3, that evaluated the contribution of thermal infrared (TIR), water vapor (WV) and temporal information to rainfall detection (binary rain/no-rain classification) was a step in this direction. The results showed that the main effect of combining TIR and WV is to reduce false precipitation detection. During the dry season, TIR allowed to correct some misclassification derived from non-precipitation low-level moisture observed in WV.

During the rainy season, WV corrected TIR precipitation over-detection by (1) inhibiting non-precipitating low-level features and (2) detecting dry air masses advected from the Sahara Desert and that create discountinuities in precipitation events.

In this chapter we evaluate the generalization capability of the four RainRunner models introduced in the previous chapters, and trained over Northern Ghana, to the wider Sudanian Savanna region of West Africa. These are: a model only using TIR data, one only using WV data, one combining TIR and WV data and one based on these two kinds of data as well as temporal information.

Particularly, we address the questions of (1) whether the models trained over Northern Ghana achieve similar performances across the Sudanian Savanna, (2) whether the findings derived in Chapter 3 about the contribution of the different information types (i.e. temperature, water vapor and time of the day and of the year) to satellite rainfall retrieval hold true across the wider region, and (3) what can this analysis teach us about the poor performance of existing satellite rainfall products, and possible improvements.

Ultimately, the final goal remains to improve the quality of rainfall information in West Africa, as a necessary step for food and economic safety in the region.

4.2. DATA AND STUDY AREA

4.2.1. GROUNDTRUTH DATA: TAHMO STATION DATA

The original groundtruth dataset considered for this study was composed of 36 TAHMO stations in the Sudanian Savanna region, depicted in Figure 4.1, color-coded per country. For better visualization, the color and marker scheme will be maintained through the remaining of this chapter. Figure 4.2 shows their data availability during the study period.

In order to have a robust performance evaluation we filtered out of the dataset the stations with large data gaps. Particularly, only stations with data available for at least 50% of the two halves of the rainy season (i.e., March to June and July to October) were considered. To form the groundtruth dataset we accumulated hourly TAHMO data into 3-hour intervals. Table 4.1 describes the characteristics of the selected 14 stations, as well as their distribution of 3-hour rainy and dry samples or intervals.

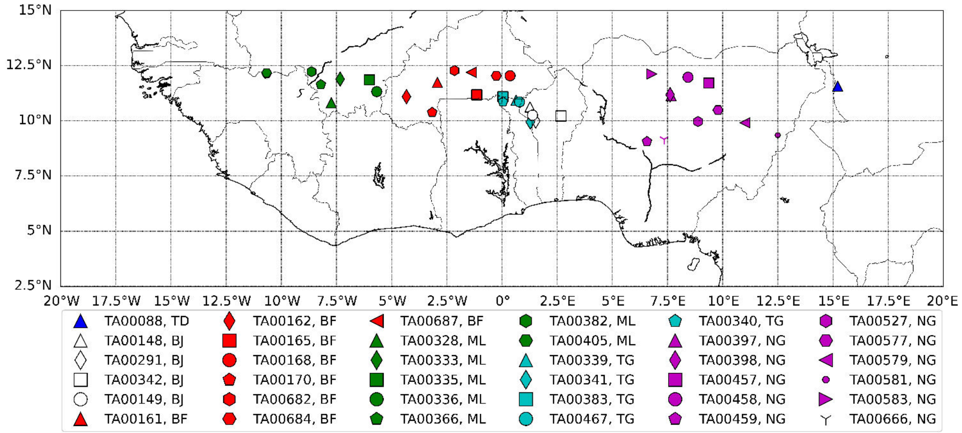


Figure 4.1: TAHMO stations in the Sudanian Savanna region of West Africa included in this study.

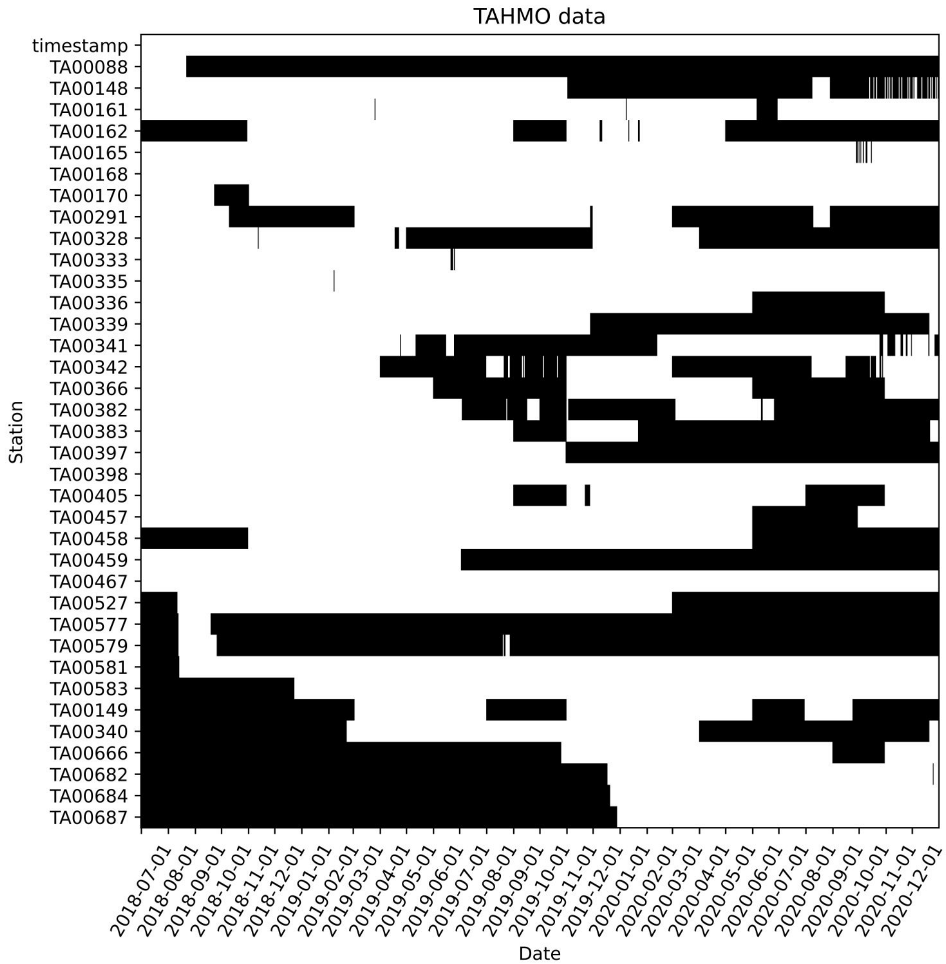


Figure 4.2: Data availability (white) of TAHMO stations in the Sudanian Savanna included in this research during the study period.

Station code	Country	Latitude [°N]	Longitude [°E]	Elevation (m)	Percentage of study period recorded [%]	Start recording (within study period)	End recording (within study period)	Total no. intervals	No. rainy intervals	No. dry intervals
TA00161	Burkina Faso (BF)	11.74	-2.93	271.7	98.6	01-07-2018	31-12-2020	7033	270	6763
TA00165	BF	11.18	-1.15	322	97.9	01-07-2018	31-12-2020	7168	283	6885
TA00168	BF	12.05	0.36	314.9	99.98	01-07-2018	31-12-2020	7224	229	6995
TA00170	BF	10.39	-3.17	289	90.26	01-07-2018	31-12-2020	6821	288	6533
TA00333	Mali (ML)	11.89	-7.34	381	88.39	01-07-2018	31-12-2020	7209	329	6880
TA00335	ML	11.85	-6.02	353	89.13	01-07-2018	31-12-2020	7104	315	6789
TA00336	ML	11.32	-5.68	396	55.58	01-07-2018	31-12-2020	6014	231	5783
TA00341	Togo (TG)	9.95	1.28	343.4	82.24	01-07-2018	31-12-2020	4920	250	4670
TA00398	Nigeria (NG)	11.18	7.62	691	99.7	01-07-2018	31-12-2020	4325	222	4103
TA00405	ML	12.16	-10.67	449.108	82.33	01-07-2018	31-12-2020	3760	188	3572
TA00457	NG	11.71	9.37	441	92.64	01-07-2018	31-12-2020	4323	112	4211
TA00467	TG	10.86	0.79	145	95.87	01-07-2018	31-12-2020	7225	207	7018
TA00581	NG	9.35	12.5	220	80.68	13-08-2018	31-12-2020	6887	208	6679
TA00583	NG	12.12	6.79	492	50.86	23-12-2018	31-12-2020	5865	187	5678
Total samples								85878	3319	82559

Table 4.1: Selected TAHMO stations in the Sudanian Savanna dataset, with their characteristics.

The stations are distributed across Mali, Burkina Faso, Togo, Benin and Nigeria and located at different altitudes, as depicted in Figure 4.3.

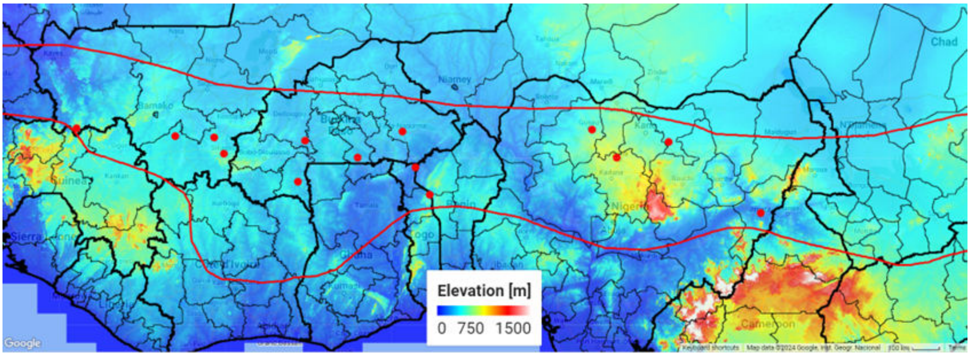


Figure 4.3: Digital Elevation Model of West Africa, with the Sudanian Savanna region marked in red, and location of the selected TAHMO stations.

4.2.2. MODEL INPUT DATA: MSG TIR AND WV DATA

In line with the previous chapters, 15-min level 1.5 data from channels 6 (7.3 µm, WV) and 9 (10.8 µm, TIR) of the SEVIRI instrument, onboard the MSG satellite, were used as input to the model. Examples of the data from both channels are represented in Figure 4.4. Both images were captured simultaneously on the 21st of August 2021 at 17:30, in the peak of the rainy season in the Sudanian Savanna. Cold clouds are observable above West Africa in both images, whereas due to its effective layer (i.e., the layer of the atmosphere that most contributes to the signal captured by a sensor in a particular wavelength, and thus the layer observable in an image) being at higher altitudes, channel 6 filters out some low-level features observed with channel 9.

4.2.3. REFERENCE DATA: RAINRUNNER MODELS AND IMERG

In order to evaluate the generalization capability of the RainRunner models developed previously in this thesis, we compare the performance that they achieved on the stations they were trained on, in Northern Ghana, with the performance achieved on the previously unseen Sudanian Savanna stations.

Additionally, we employ IMERG data as an external reference, in its Early, Late and Final runs. Because of the development of IMERG during the writing of this dissertation, here we utilize IMERG v07, that was fully released between 2023 and 2024 [103]. IMERG v07 incorporated several changes to improve rainfall retrieval performance as compared to the v06 version, used in previous Chapters. These changes include among others intercalibration of passive microwave (PMW) estimates to correct

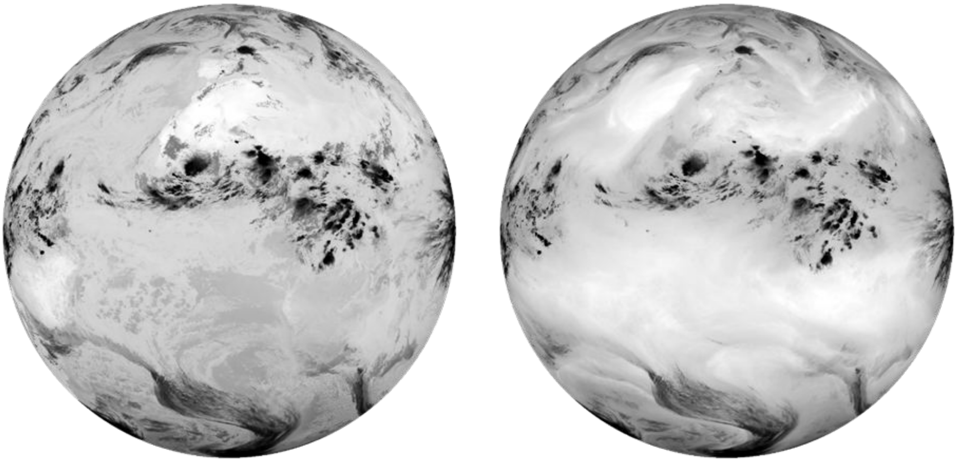


Figure 4.4: Images from the two MSG channels used in this study, on the 21st of August 2021 at 17:30 UCT. Left: Channel 9, 10.8 µm; right: Channel 6, 7.3 µm.

biases and improved retrieval algorithms for both PMW and TIR data. For TIR, PERSIANN-CCS was substituted for Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Dynamic Infrared Rain Rate (PDIR-Now), that dynamically shifting relationships between the IR brightness temperatures [103]. Over Africa, PDIR-NOW has been reported to importantly overtaken PERSIANN-CCS, with a decreased false alarm ratio (FAR) and bias [104].

4.3. METHODOLOGY

4.3.1. QUALITY CONTROL OF THE TAHMO RAINFALL DATASET

A first step for a reliable performance evaluation was to evaluate the data of the 14 selected TAHMO stations to exclude possible faulty stations, that would have derived in untrustworthy results. This included:

1. **Outlier detection** through a “high-level” evaluation of the stations’ rainfall observations, to identify stations reporting unrealistic values.

2. **Consistency evaluation** by comparison of rainfall records from stations very close to each other. Although due to the highly regional small-scale rainfall processes rainfall observations can be different between close-by stations, the large-scale rainfall patterns should be similar.

4.3.2. PREPROCESSING OF MODEL INPUT DATA

The preprocessing of data into a format appropriate for input into the model followed the same methodology as in previous chapters. Both TIR and WV images were cropped to 32 $\times$ 32 pixels (96 $\times$ 96 km) images with the TAHMO stations located in a 2 $\times$ 2 pixels central square. Then, these images were combined in 3-hour, i.e. 12 images, sequences. Lastly, we assigned to each sequence its corresponding temporal information: the hour of the day and the month of the year of the first image in the sequence.

4.3.3. PROCESSING OF REFERENCE DATA

In order to be comparable to model output, IMERG data needed to be processed into a similar format. For this, 30-min IMERG Early, Late and Final Runs by first accumulating them into 3-hour intervals and later classified in rain/no-rain. A threshold of 1mm/3hr was used for this classification.

4.3.4. MODEL IMPLEMENTATION

To study the generalization capability of the four models, i.e., (1) TIR data only; (2) WV data only; (3) TIR and WV data combined; and (4) TIR and WV data together with the observation timestamp in the wider Sudanian Savanna region we re-ran the ensemble models on the 14 stations considered here. This is, we ran the same 10 pre-trained model runs from Chapter 3 over each station, and averaged the results for an improved robustness. It is important to remember that the output of the RainRunner models is one single rain / no-rain value corresponding to a point target (rain gauge) within the 2 $\times$ 2 pixels central square of the input images (32 x 32 pixels).

4.3.5. PERFORMANCE EVALUATION

We evaluated the performance of the models by comparison to the corresponding groundtruth data: 3-hour TAHMO data from the 12 selected stations. We performed a global evaluation (i.e., considering all data) as well as a seasonal evaluation. The dry season was defined as November to February, included, and the rainy season was divided in two halves: March to June and July to October, included. As reference, we included IMERG data (Early, Late and Final runs) in the evaluation.

To evaluate the generalization capability of the models we compared their global and seasonal performance over the previously unseen stations across the Sudanian Savanna to the performance achieved on the Northern Ghana stations, used for training.

We used Probability of Detection (POD), Success Ratio (SR), Frequency Bias (FBias) and Critical Score Index (CSI) as performance metrics.

Based on these metrics, we represented model performances on Roebber diagrams.

The analysis evaluated both global model performance across the study region and individual station performance to compare models and regional differences.

4.4. RESULTS

4.4.1. TAHMO DATASET QUALITY CONTROL

The station outlier and consistency evaluation resulted in the decision to exclude the stations TA00457, in Nigeria, and TA00467, in Togo, ad these station did not pass the data quality control. Particularly, station TA00457 contained unrealistic daily rainfall values, with up to around 600 mm [Figure 4.5].

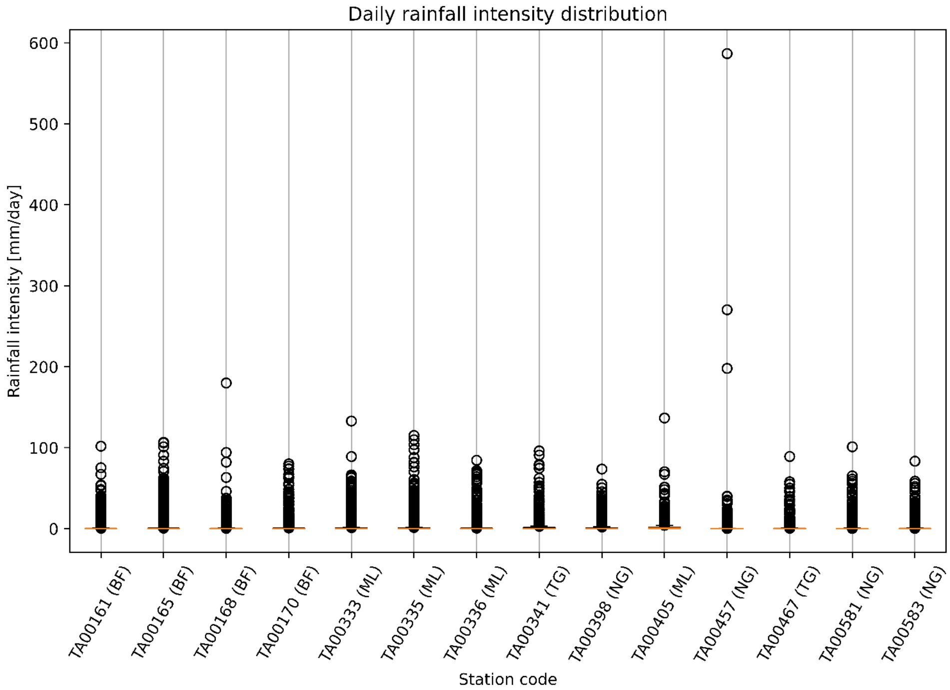


Figure 4.5: Daily rainfall values of the 14 TAHMO stations in the Sudanian Savanna dataset.

We compared the daily rainfall timeseries of station TA00467 with that of IMERG daily data and, for proximity as can be seen in Figure 4.1, to that of station TA00339, that due to data gaps was not considered in this evaluation. Figure 4.6 shows this comparison. As can observed in the figure, station TA00467 failed to capture most rainfall in the second half of the 2019 rainy season, from July 2019, while it was captured by both IMERG and station TA00339. Furthermore, the correlation between the daily rainfall time series of the two TAHMO stations was as low as 0.49. Comparing the time series from both stations with that of IMERG resulted in a correlation of 0.62 for station TA00467 and 0.73 for station TA00339.

Thus, stations TA00457 and TA00467 were excluded from the Sudanian Savanna dataset for the remaining analysis.

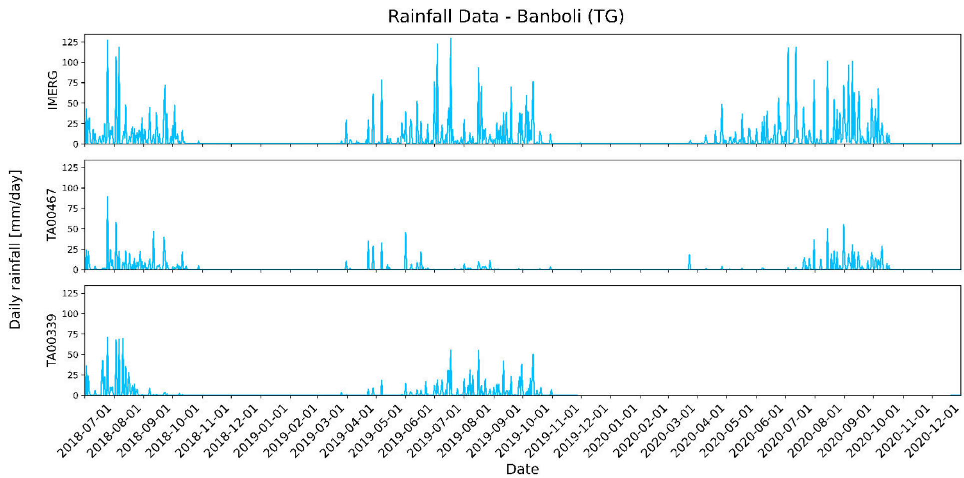


Figure 4.6: Daily rainfall timeseries of IMERG v6 and the TAHMO stations TA00467 and TA00339.

4.4.2. PERFORMANCE EVALUATION

Results of model performance evaluation are shown globally over the whole Sudanian Savanna dataset and for individual stations. Figure 4.7 shows the global performances of all four RainRunner models over the (previously unseen) Sudanian Savanna stations, as well as the corresponding IMERG Early, Late and Final runs. For reference, the performance of IMERG Final and of the RainRunner models on the Northern Ghana stations are also shown. Figure 4.8 shows global performances divided in the two halves of the rainy season (March to June and July to October, included). As a reminder, the training dataset of the models comprised data from July 2018 to December 2020 from 8 TAHMO stations in Northern Ghana, while the evaluation and test datasets only contained 2020 data. This is, while the reference RainRunner model performances correspond only with 2020 data, the results showed here for the Sudanian Savanna also include data from 2018 and 2019.

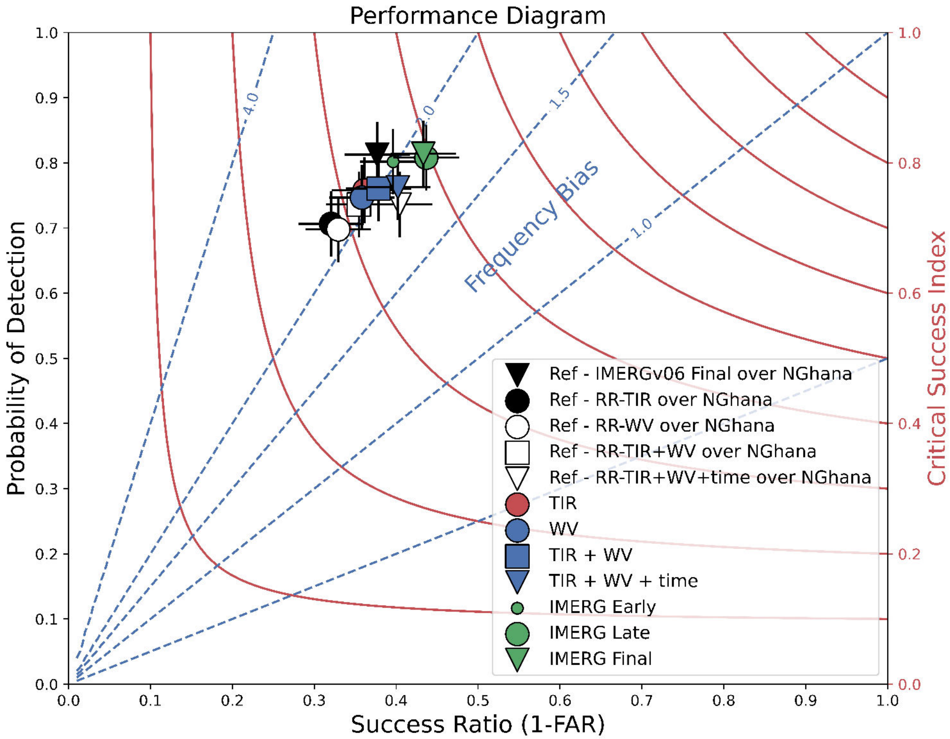


Figure 4.7: Performances of the four RainRunner models and IMERG Early, Late and Final runs on the Sudanian Savanna dataset, for the whole study period. For reference, the performance of the four RainRunner models and IMERG Final over Northern Ghana are also shown.

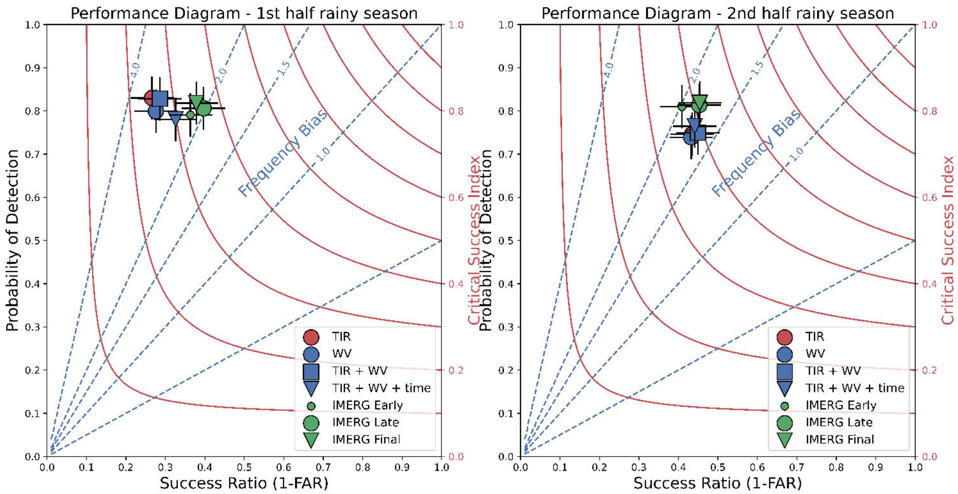


Figure 4.8: Performances of the four RainRunner models and IMERG Early, Late and Final runs on the Sudanian Savanna dataset, divided in the two halves of the rainy season.

As for individual stations, Figure 4.9 depicts the resulting Roebber performance diagrams per station. In this case, and due to the already cluttered Roebber diagrams, the performances over Northern Ghana are displayed as comprised in a green rectangle. This rectangle remains unmoved in all the performance diagrams in this chapter, to serve as a static reference to aid visual interpretation of the results.

It can be seen that the performances achieved in the 12 stations are very similar among them and to those of northern Ghana (green rectangle), with all stations simultaneously achieving higher POD and SR than the lower boundaries in northern Ghana. Furthermore, the performances slightly improve consistently (i.e. getting closer to the top-right corner of the diagram) when both the WV and the timestamp layers are added to the model. This is also observed in Table 4.2, which presents the performance metrics averaged over the 12 stations, for the 4 models.

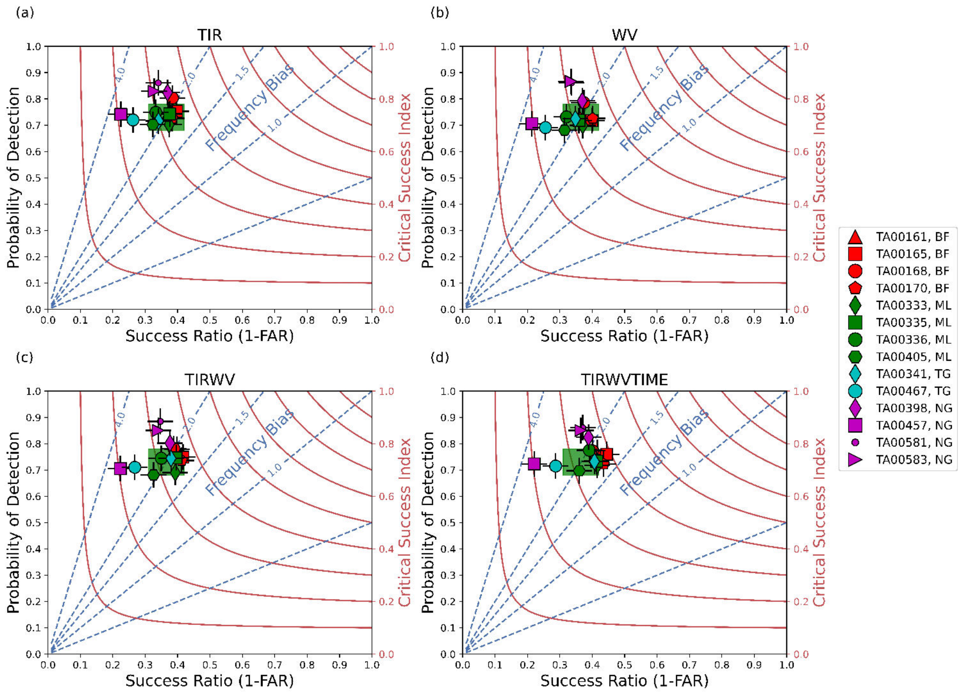


Figure 4.9: Performance diagrams of the RainRunner models on the Sudanian Savanna stations, for the entire study period (1/07/2018 to 31/12/2020). The green rectangle serves as reference and indicates the area comprised between the performances of the RainRunner models on the Northern Ghana stations. See Figure 3.7

Model	POD	SR	FBias	CSI
TIR	0.76	0.36	2.13	0.32
WV	0.75	0.36	2.12	0.32
TIR+WV	0.77	0.38	2.05	0.34
TIR+WV+time	0.77	0.40	1.93	0.36

Table 4.2: Average performance metrics of the four RainRunner models (TIR, WV, TIR+WV and TIR+WV+time) on the 12 TAHMO stations in the Sudanian Savanna dataset.

Figure 4.10 and Figure 4.11 show the performance diagrams for the two halves of the rainy season: March to June and July to October, included. Table 4.3 shows the performance metrics averaged over all stations for the two periods. During the first half of the rainy season, the added value of incorporating the WV and temporal layers in the model is significant in that false alarms from the TIR-only model are corrected, with the SR increasing from 0.26 to 0.32 and the FBias decreasing from 3.19 to 2.47. Here, the information layer with the largest effect on model performance is the time. In this part of the season, there are also larges differences in performance across the stations. On the other hand, the four models perform better in all metrics except POD during the second half of the rainy season, included the TIR model, that obtains a considerably higher SR (lower FAR) than during the first half of the rainy season, i.e., 0.43 compared to 0.26. The FBias of all models decreases to around 1.75. During these months, performances do not change significantly with additional layers, and except from a small increase in average POD, the addition of time does not improve performance. The model that uses a combination of TIR and WV data has the lowest FBias. Furthermore, there are smaller differences in performance across stations.

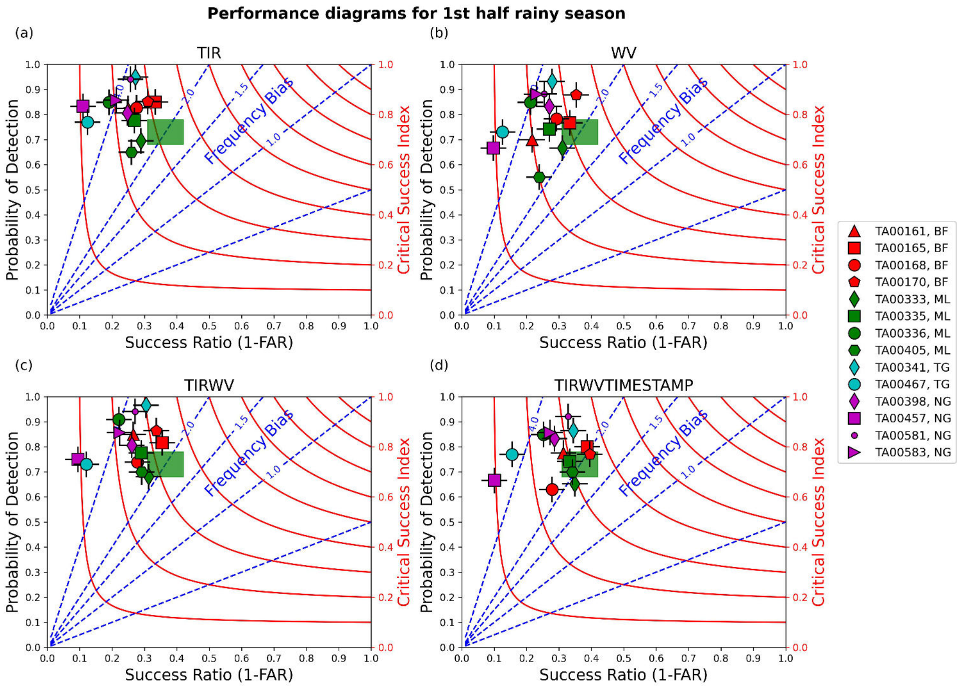


Figure 4.10: Performance diagrams during the first half of the rainy season. The green rectangle serves as reference and indicates the area comprised between the performances of the RainRunner models on the Northern Ghana stations. See Figure 3.7

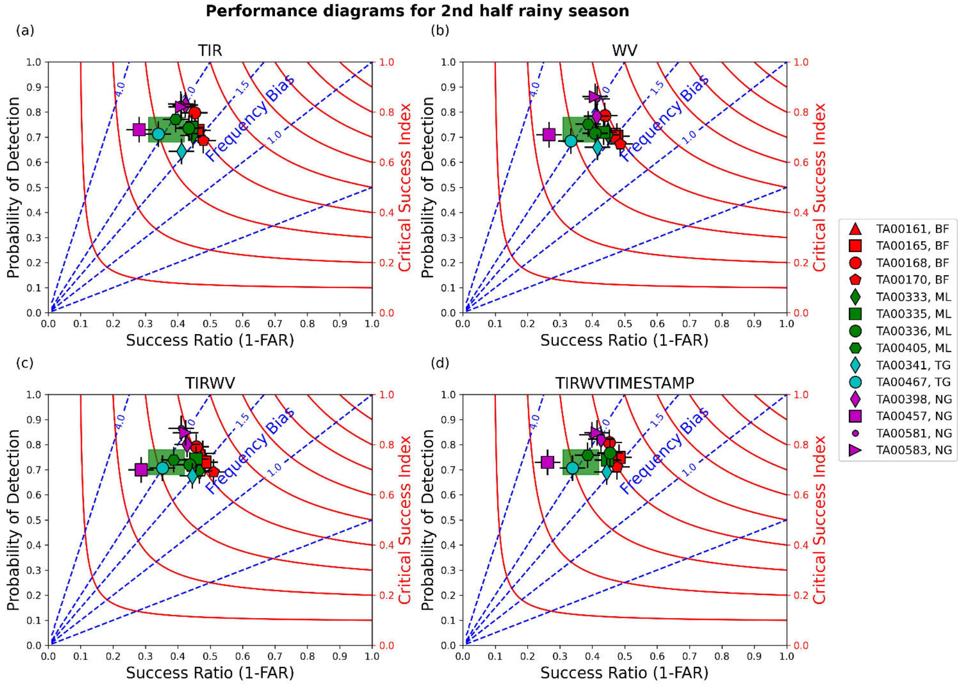


Figure 4.11: Performance diagram during the second half of the rainy season. The green rectangle serves as reference and indicates the area comprised between the performances of the RainRunner models on the Northern Ghana stations. See Figure 3.7

Model	POD (1st half)	POD (2nd half)	SR (1st half)	SR (2nd half)	FBias (1st half)	FBias (2nd half)	CSI (1st half)	CSI (2nd half)
TIR	0.82	0.75	0.26	0.43	3.19	1.74	0.25	0.38
WV	0.79	0.75	0.27	0.43	2.96	1.74	0.25	0.37
TIR+WV	0.83	0.76	0.28	0.45	2.97	1.70	0.27	0.39
TIR+WV+time	0.78	0.77	0.32	0.44	2.47	1.76	0.30	0.39

Table 4.3: Performance metrics averaged over all the considered TAHMO stations, for the four models developed, during the two halves of the rainy season (March to June and July to October).

4.5. DISCUSSION

The general performance of the models over the whole period shows high similarity to that of the stations in northern Ghana, pointing towards a good generalization capability of the model, i.e., it achieves a similar performance on data from unseen stations than the stations it was trained on. It is important to remember that these models were trained using data from only 8 stations in northern Ghana, during 2.5 years. Figure 4.7 shows that when combining TIR, WV and temporal information, RainRunner achieves a somewhat higher SR and lower FBias than IMERG Early. IMERG Late and IMERG Final are the best performing models in the overall study period, which is not striking due to the high complexity of the precipitation retrieval algorithms. However, Figure 4.8 shows that during the second half of the rainy season, all RainRunner models achieve a smaller FBias than all IMERG models, although the difference is small. A phenomena that is known to cause over-detection in both TIR and PMW-based satellite rainfall retrieval, and that might be partly the reason for the higher FBias of IMERG, is "virga" precipitation, i.e., precipitation that evaporates before reaching the ground [41, 54]. This kind of precipitation is estimated to account for over 50% of precipitation over the Sahara [41], and 10-15% in the rest of Africa [54]. Virga precipitation has been found to account for 30% - 50% of false PMW-detected precipitation events in arid regions [41].

Comparing the influence of different sources of information on both the global performance and the performance on individual stations, the combination of TIR, WV and temporal data is beneficial over the overall study period, similarly to the effect visible in Northern Ghana. Although POD decreases, SR and CSI increase, and FBias gets closer to 1. This indicates similar rainfall processes captured in TIR, WV and temporal data across the whole region.

As for the seasonal performance Figure 4.8, the combination of TIR and WV is beneficial in the two halves of the rainy season. The addition of temporal information has the largest impact during the first half of the rainy season, when it improves general performances (this is, favoring a decrease in FBias and FAR over an increase in POD). However, temporal information does not improve performance during the second half of the rainy season. Evaluating performances at the level of individual stations, a similar pattern can be found [Figure 4.10 and Figure 4.11]. We suggest that the explanation is in line with the findings from Chapter 3. The chapter concluded that the two main contributions of WV information as an addition to TIR information in satellite rainfall retrieval were:

1. **Inhibition of low-level features**. Due to an effective layer at a higher altitude, the part of the atmosphere observable in WV data is shallower and restricted to higher altitudes than for TIR, that at 10.8 μm is in an atmospheric window and therefore can pass through the atmosphere. This effect can be observed in Figure 4.4.

Because of this higher effective layer, WV data highlights areas of strong convection – higher cloud tops - , usually related to heavy rainfall. In this way it would be an alternative to the traditional Cold Cloud Duration (CCD) method, employed by other products such as TAMSAT and CHIRPS.

2. **Detection of dry air masses advected from the Sahara Desert**, that produce vertical discontinuities in rainfall events in the form of dry air intrusions. In such cases, rain-bearing clouds with cold tops identified in TIR imagery can produce rainfall that does not reach the ground. Including WV data allows to depict this phenomenon and correct for false alarms. This kind of events is disregarded by CCD methods.

Here, the WV inhibition of low-level features improves model performance in some cases. This might explain improvement in performance during the whole rainy season. Figure 4.12 to Figure 4.15 show examples of cases where the addition of WV information corrected a false alarms of the TIR-only model. The images show the 3-hour sequences formed by twelve 32 x 32 pixels images in the TIR and WV wavebands. The images are centered on different TAHMO stations, in this case in Burkina Faso and Mali. They depict what seems to be non-precipitating low-level features [Figure 4.12 and Figure 4.13] and dry air intrusions [Figure 4.14 and Figure 4.15]. Non-precipitating low-level features can be recognized by the differences in the amount and distribution of pixels of low reflectance (black) in the TIR and WV images. The TIR images show dark pixels (cold) throughout, whereas the WV images show fewer and localized dark pixels. The effective layer of WV data (the highest layer in the atmosphere that can be observed in the data) is higher than the non-precipitating low-level features and therefore filters them out. Dry air intrusions are seen as a continuous "blanket" of air that is also somewhat observable in the TIR imagery. Table 4.4 shows the output of the TIR-only, WV-only, and TIR+WV RainRunner models for these sequences. As a reminder, this output is a real number between 0 and 1, and a threshold of 0.5 is then used to classify sequences as dry (< 0.5) or rainy (≥ 0.5).

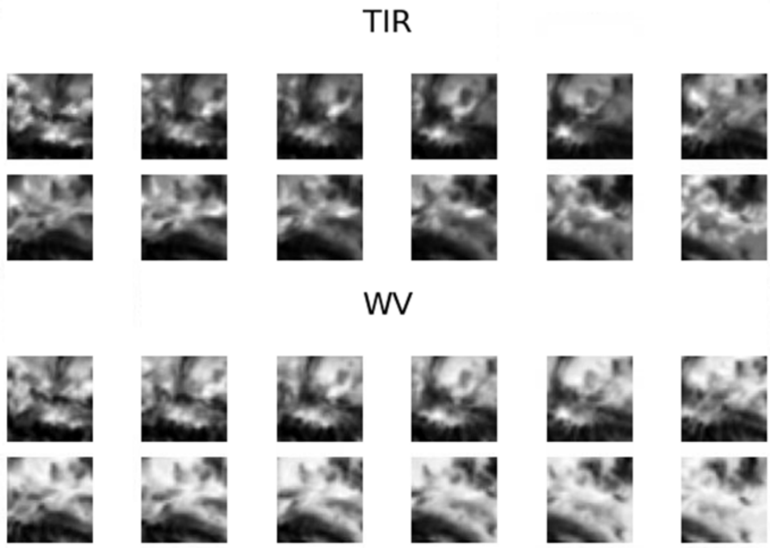


Figure 4.12: Example of a case when the addition of WV information corrected a false alarms of the TIR-only model. Station TA00168 on the 8th of June 2019.

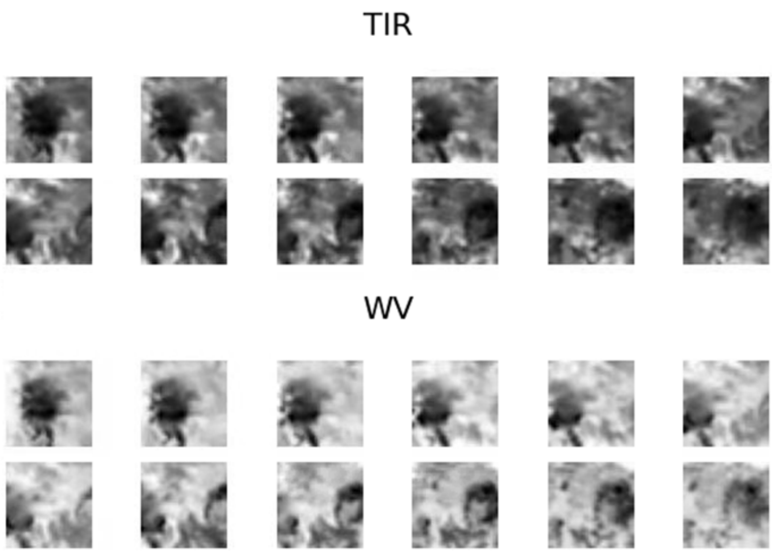


Figure 4.13: Example of a case when the addition of WV information corrected a false alarms of the TIR-only model. Station TA00170 on the 12th of September 2018.

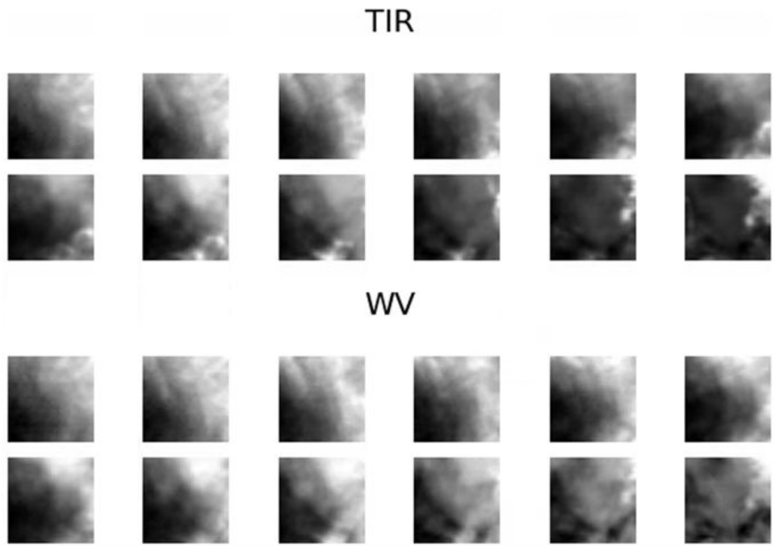


Figure 4.14: Example of a case when the addition of WV information corrected false alarms of the TIR-only model. Station TA00405 on the 20th of June 2019.

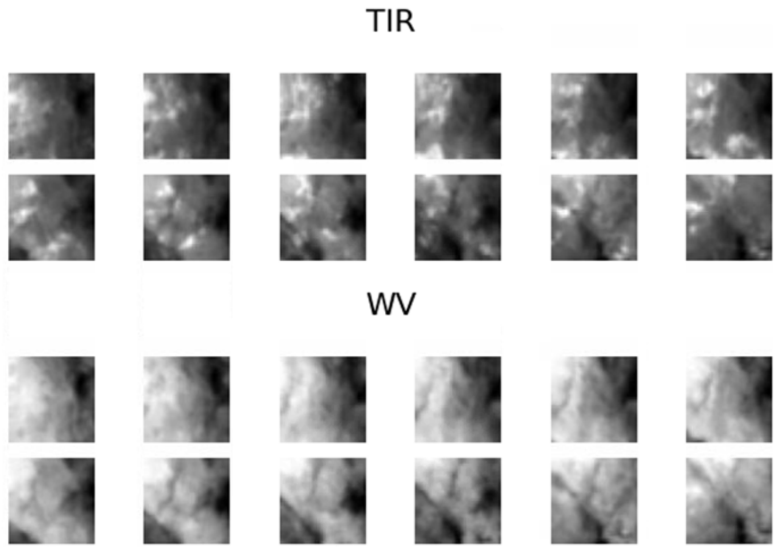


Figure 4.15: Example of a case when the addition of WV information corrected false alarms of the TIR-only model. Station TA00336 on the 20th of May 2019.

Fig.	**Station**	**Country**	**Date**	**TIR model output**	**WV model output**	**TIR+WV model output**
4.10	TA00168	Burkina Faso	08-06-2019	0.84	0.45	0.44
4.11	TA00170	Burkina Faso	12-09-2018	0.68	0.08	0.49
4.12	TA00405	Mali	20-06-2019	0.58	0.23	0.38
4.13	TA00336	Mali	20-05-2019	0.52	0.04	0.24

Table 4.4: Probabilistic output of the RainRunner models (TIR-only, WV-only and TIR+WV) associated with the sequences in Figures 4.10-4.13.

The different results between the two halves of the rainy season might find an explanation in the atmospheric processes that define West African climate and rainfall dynamics mechanisms, dominated by the West African Monsoon system [98], Figure 4.16. An in-depth analysis of the complexity of what elements within the monsoon system produce the seasonal cycle over equatorial Africa are beyond the scope of the thesis, but a number of potential factors have been pointed out in the literature [98, 99]. One of the phenomena that dictates rainfall processes in West Africa is the movement of the Intertropical Convergence Zone (ITCZ). Although the extent to which the progression of the ITCZ influences rainfall is debated in literature [105], it is referred to in the following as a reference to track the associated movement of the Sahara Thermal Low (STL) during the year. The development of the STL is induced by the temperature gradient between the colder Atlantic Ocean and the warmer Sahara and Sahel, and initiates the northward migration of the ITCZ and associated moisture transport [99].

Figure 4.17 shows rainfall distribution over West Africa in the first dekads of June to October of 2020. The figure depicts the progression first northwards and then southwards of the ITCZ and associated rainfall. Our study area is roughly comprised between the southern border of Senegal and the northern border of Cote d’Ivoire.

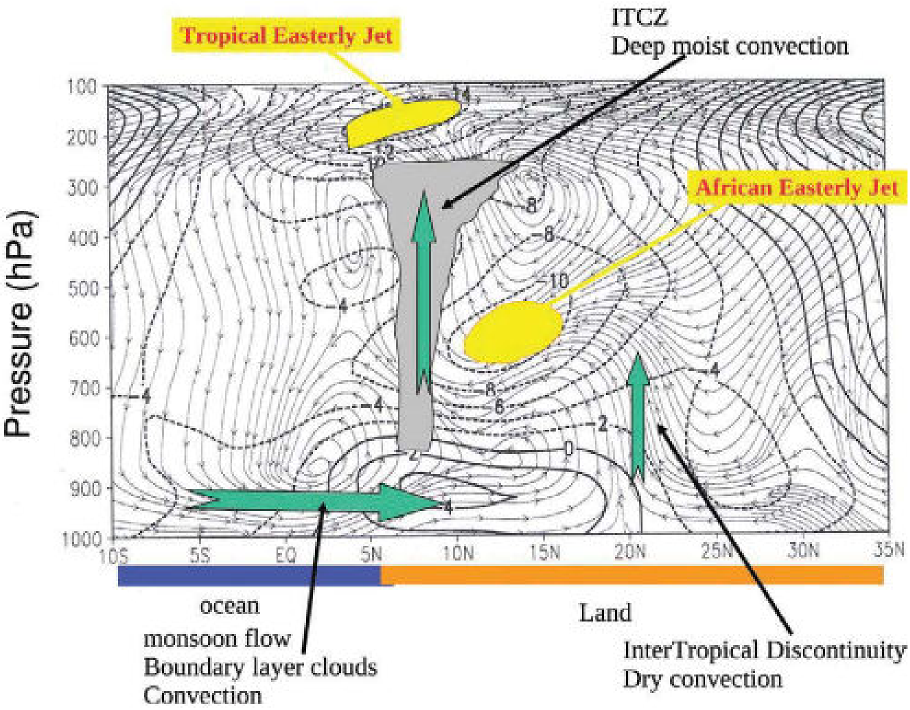


Figure 4.16: Mean Meridional Circulation (stream lines) and associated mean Zonal Wind (m/s in contours) over West Africa during the summer season [98].

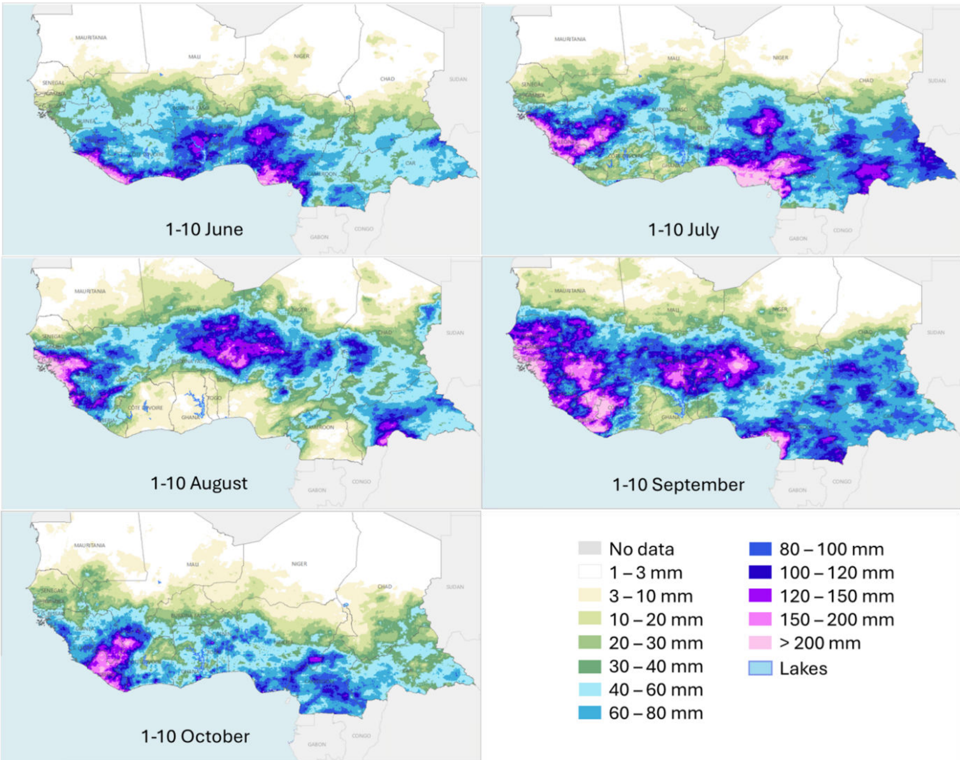


Figure 4.17: Rainfall distribution over West Africa in the first dekads of June to September 2020, adapted from [106–110].

Firstly, during the first half of the rainy season the ITCZ is at lower latitudes and therefore the Sahara Thermal Low (STL) reaches further south. This leads to conditions of convective inhibition (CIN) that suppress rainfall. The movement of the ITCZ over the Sudanian Savanna during the first months of the rainy season creates highly different atmospheric conditions within our study area. By July to September, the rainy season peaks over the Sudanian Savanna and the ITCZ reaches its northernmost position, reaching latitudes far beyond this region. During this period, the likelihood of dry air masses being advected from the Sahara Desert decreases.

Additionally, the seasonal and diurnal distribution of precipitation types in the Sudanian Savanna is closely related to the movements of the West African Monsoon (WAM) and the Intertropical Convergence Zone (ITCZ) [40]. The establishment of the WAM (northward and southward progression of the ITF) in the region at the beginning of the rainy season (March to June) and its retreat phase at the end of the season (September to October) favors conditions of greater instability and wind shear that promote the development of organized convection in the form of Mesoscale Convective System (MCS) type rain systems, such as Wide Convective Cores (WCCs). Deep Convection Cores (DCC) rain events reach their maximum frequency in May. On the other hand, when the ITCZ and the associated WAM rain belt are in their northernmost position (July and August), despite the high atmospheric moisture content there is a weaker wind shear, which is more conducive to the development of isolated and less organized convection compared to the transition phases. During this phase, there is less organized convection and isolated shallow convection or Isolated Shallow Echo (ISE) - weak convective rainfall or warm rain -rain events reach their maximum frequency [40]. Of these precipitation types, DCC precipitation shows the larger diurnal variability, with most events occurring during the late afternoon-early evening, after convection has built up during the day [40]. Although strong convective rainfall represents a small number of events in West Africa during the second half of the rainy season 3%, they contribute by close to 50% to the total rainfall [111].

These processes have multiple implications for satellite rainfall retrieval: With the STL over an area, the probability of TIR data detecting rain-bearing clouds whose rainfall is later suppressed by the dry and dusty air is higher. However, these masses of dry air are observed by WV imagery, enabling corrections for false alarms in rainfall detection. Furthermore, the African Easterly Jet Figure 4.16, situated at around 600 hPa, that corresponds to the level of maximum sensitivity range of the WV channel used here [112], is linked to Saharan dust transport [113] and might therefore amplify dry air and aerosol advection, and rainfall inhibition. This can explain the large added benefit of WV observations to model performance during the first half of the rainy season, and the better performance of the TIR model during the second half of the rainy season. This is, with the ITCZ in its northernmost position, the likelihood of CIN decreases and therefore the relationship between TIR data and rainfall on the ground is more direct, and additional sources of information barely change the performance of the model. The larger differences in performance across stations during the first half of the rainy season than during the second half might be due to the ITCZ travelling over our study area during the first months of the rainy season, creating large atmospheric differences at its pass. Moreover, a possible reason of the significant performance improvement introduced by temporal information during the first half of the rainy season, is the stronger diurnal cycle of DCC rain events, that dominate the first half of the rainy season. As for seasonal information (i.e., the number of the month), we suggest that the model might learn to expect more rainfall inhibition during the first part of the rainy season. Lastly, the weaker wind shear during the months of July and August will induce a smaller displacement of rainfall from the cloud (top) to the ground, which might improve TIR performance during the second half of the rainy season. Opposed to this, other research has shown the prevalence of mesoscale convective systems, in the form of squall lines in the second half of the rainy season [114, 115]. Because squall lines are larger in extent and have higher tops (deeper convection) than localized convection, they are more clearly visible in TIR data. Furthermore, squall lines have a weaker diurnal cycle. This paradigm could also be supported by our results, in the better performance of TIR during the second half of the rainy season, as well as the little impact of adding temporal information. Therefore, we can only conclude that the models are identifying a stronger temporal signature during March-June than during July-September, but the concrete mechanisms behind this effect are still unclear. More research is needed to determine which view is more correct but the value of going back and forth between ML generated results and physical reasoning is a very productive method.

Supporting these claims, Table 4.5 presents the number of rainy intervals, the number of intervals that were classified as rainy by the TIR-only model and their ratio. Interestingly, this seems to roughly follow the progression to the North and then back to the South of the ITCZ during the year. The considerably smaller general numbers from March to June are partly due to the fact that the study period is from the 1st of July 2018 to the 31st of December 2020, and therefore excludes the first half of the 2018 rainy season, added to the peak of the rainy season happening in the second half.

Month	Actual rain	Predicted rain	Actual/pred
March	35	161	0.22
April	86	375	0.23
May	207	674	0.31
June	260	625	0.42
July	671	1149	0.60
August	773	1131	0.69
September	624	1065	0.59
October	319	750	0.43

Table 4.5: Number of rainy intervals, intervals that were predicted as rain by the TIR-only model and their ratio. Our study area expands between 9.35°N and 12.16°N.

Finally, even if the addition of WV improves the performance of the models all over the Sudanian Savanna region of West Africa, they still obtain a high FAR. A potential reason for rainfall over-detection, as seen in literature [11], might be a higher atmospheric aerosol concentration than in other regions of the world. Therefore, a possible future avenue to improve model performance would be to add information related to aerosols. Chapter 5 will look into some of these avenues.

5

FUTURE PERSPECTIVES FOR SATELLITE RAINFALL RETRIEVAL IN WEST AFRICA

During the time that I spent carrying out this research, I started or devised various research avenues based on the insights I gained along the way. Some of them did not materialize due to time constraints, others would require separate research projects on their own. I would like to reflect on some of them in this dissertation so that future researchers can consider them, since I believe they will yield advances in the study of West African rainfall. To this end, this chapter summarizes some reflections for future work.

This chapter is organized by sections according to various topics that future research should consider. It begins by describing the unique atmospheric conditions of West Africa, which differ significantly from those in other regions. The chapter then suggests expanding the use of Earth Observation (EO) data beyond the types traditionally employed for satellite rainfall retrieval. It proposes leveraging data cubes and Deep Learning to process EO data more effectively. Additionally, it recommends Citizen Science as a valuable source of ground data. Finally, the chapter addresses the need of interdisciplinary collaboration to achieve a true breakthrough in satellite rainfall retrieval

5.1. DIFFERENT ATMOSPHERIC CONDITIONS IN WEST AFRICA

The starting hypotheses of this dissertation were that the poor performance of satellite rainfall products over West Africa is partly due to (1) **a drier atmosphere** and (2) **a higher atmospheric aerosol concentration** in West Africa than in other regions of the world [11]. According to Mc Collum (2000) [11], a drier atmosphere leads to higher cloud bases due to the limited moisture available at lower altitudes. As a result, precipitation is more likely to evaporate before reaching the ground. A higher aerosol concentration translates into many Cloud Condensation Nuclei (CNN), and therefore smaller droplets and inefficient rainfall processes. Furthermore, **rainfall patterns in this region are highly seasonal**, following the latitudinal movement of the ITCZ. **Therefore, if we aim to improve rainfall retrieval in this area, we must account for regional conditions.** In this dissertation, including both water vapor and temporal information in the RainRunner rainfall retrieval model (Chapter 3) resulted in better performances over a TIR-only model. WV data did not only depict deep convective motion, filtering out low-level features, but also highlighted dry air masses advected from the Sahara Desert, that suppress rainfall. The main contribution of temporal information was to correct false alarms during the dry season and the first months of the rainy season. Evaluating the performance of the models across the wider Sudanian Savanna (Chapter 4) pointed towards a link between the importance of adding

WV information to a TIR-only model and regional seasonal atmospheric dynamics, that would explain the difference performances in the two halves of the rainy season. This is in line with Chapter 3, since Saharan dry (and dusty) air moves northwards crossing our research area during the first half of the rainy season.

The next logical step from the initial hypotheses would be to add information about aerosols to capture (1) the small-scale interactions between rainfall and aerosols and (2) the large-scale movement of the ITCZ. Figure 5.1 shows monthly averages of the aerosol product AER_AI from the TROPOMI instrument onboard ESA's Sentinel 5P satellite, that contains the Aerosol Index (AI) – representative of the content of atmospheric aerosols, for 2019. The AI is based on wavelength-dependent changes in Rayleigh scattering in the ultraviolet (UV) region of the electromagnetic spectrum, where ozone absorption is very small. It is calculated based on spectral contrast in the UV spectral range, using the 354 nm and 388 nm wavelengths. Positive values (that correspond with the Absorbing Aerosol Index, AAI) signal the presence of UV-absorbing aerosols such as dust and smoke. Strong negative values might indicate the presence of non-absorbing aerosols. Because clouds result in near-zero values, this index can be used in their presence. It is useful to track absorbing aerosols such as desert dust, smoke from biomass burning and volcanic eruptions also in the presence of clouds [116].

The yearly progression in Figure 5.1 can be divided in the following stages:

From December to March, within the dry season in the Sudanian Savanna, the distribution of atmospheric aerosols is less organized. This corresponds to the *Harmattan* or *Harmattan (Dust) Haze* phenomenon experienced in sub-Sahel West Africa between mid-November and March, peaking in January and February, characterized by a widespread high density of atmospheric aerosols [117]. The main source of dust during this period is the Bodele depression, Chad, that can be observed in the figure. The Bodele depression is the largest global source of mineral aerosols and is estimated to produce around 50% of those originating in the Sahara [117, 118]. Rainfall and dust influence each other in that, on one side, rainfall in the dust source is linked to dust emission rates in the sub-Sahel region [117, 118]. On the other side, anecdotal evidence gathered from on-the-ground conversations claims that years with a dustier Harmattan are followed by heavier rainy seasons. In the interest of understanding better sub-Sahel African rainfall dynamics, beyond satellite rainfall retrieval, it would be interesting to evaluate whether there are indeed interconnections between atmospheric aerosol concentrations during the Harmattan months - potentially linked to previous rainfall in the main dust sources - and rainfall intensity during the following rainy season. The same kind of Earth Observation (EO)

5

data could be used for such analyses.

Later in the year, in April, aerosols become denser and more structured, pushed inland by southwesterly moist air blowing inland from the Atlantic Ocean, following the northwards progression of the ITCZ (that separates this moist air from the northeasterly Harmattan winds). From April to September, the dense aerosol area progresses roughly as the ITCZ: first moving northwards and then, albeit in a less structured manner, retreating southwards. During these months, the high concentration of atmospheric aerosols increases the abundance of Cloud Condensation Nuclei (CCN), leading to the formation of many small droplets. This derives in inefficient rainfall processes and can result in rainfall overestimation by TIR-based satellite rainfall retrieval methods. Passive microwave (PMW) satellite rainfall retrieval may also overestimate rainfall in these conditions, as the abundance of CNN enhances ice particle formation and lightning activity in convective systems, increasing PMW scattering (McCollum, 2000).

Finally, from October onwards the density of aerosols in the atmosphere decreases. and in December, aerosols disperse over a wider area.

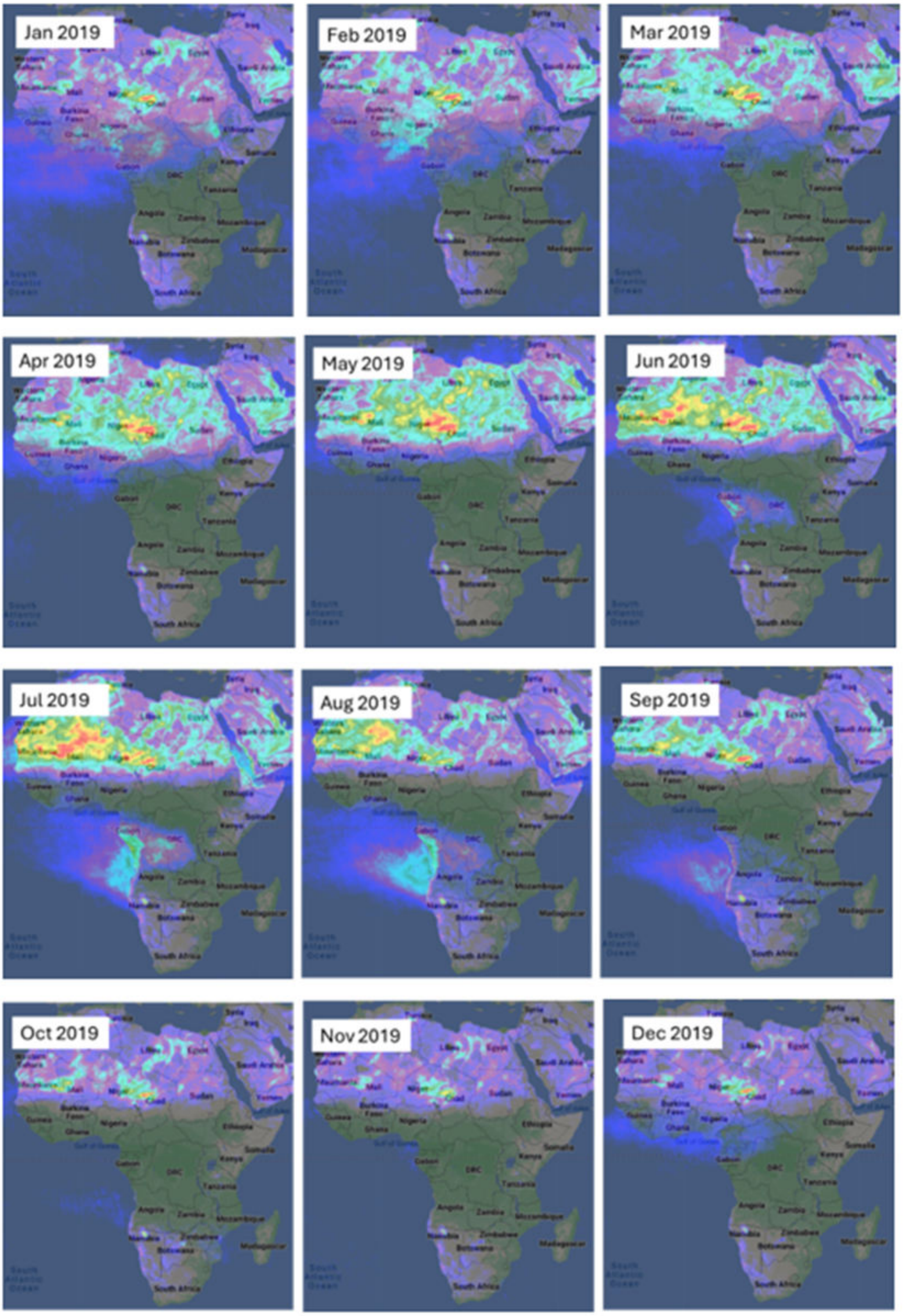


Figure 5.1: Monthly averages of the AER_AI product of the satellite Sentinel 5P, for 2019. The colour scale (ranging from dark blue to cyan, green, yellow, and red) represents aerosol presence, with green to red indicating a strong aerosol presence.

5.2. BEYOND THE TRADITIONAL USE OF EARTH OBSERVATION DATA FOR SATELLITE RAINFALL RETRIEVAL

Earth Observation (EO) satellites have been observing the planet since the launch of Tiros 1, the first weather satellite, in 1960. Continuously orbiting the Earth, these satellites provide a privileged, global perspective of our atmosphere, delivering Terabytes of data daily. **Despite this wealth of information, only a small portion of EO data is currently utilized for satellite rainfall retrieval, primarily thermal infrared (TIR) and passive microwave (PMW) data.**

Rainfall retrieval products for Africa, such as CHIRPS and TAMSAT, rely predominantly on TIR data from geostationary (GEO) satellites, with CHIRPS incorporating rain gauge data to enhance accuracy. Meanwhile, IMERG achieves superior performances than other products by incorporating PMW data from “as many as possible” low-Earth orbit (LEO) satellites, which detect hydrometeors in the atmosphere, alongside reanalysis and rain gauge data. However, these approaches focus on cloud-top temperature and localized hydrometeor information and do not account for broader interactions of the Earth system that influence rainfall.

However, rainfall is intricately linked to the Earth system, as it both influences and is influenced by various atmospheric, terrestrial, and hydrological processes. For instance, rainfall impacts soil conditions, vegetation health, and water availability, while factors such as wind patterns, land surface temperatures, topography and soil moisture play crucial roles in rainfall formation and distribution. Different conditions under the clouds can significantly affect the amount and distribution of rainfall reaching the ground, which is neglected in Cold Cloud Duration methods. Including information from other Earth system variables could enable data-driven models to capture the complex relationships that drive rainfall. Expanding the scope of EO data used in satellite rainfall retrieval beyond TIR and PMW would allow the model to learn from these interactions, potentially improving performance and providing a more comprehensive understanding of the system.

This dissertation demonstrates the value of integrating additional EO data to improve performance of the RainRunner rainfall retrieval model. By combining water vapor (WV) data with thermal infrared (TIR) data, the RainRunner model showed improved performance in binary rainfall classification. This supports the idea that including more variables related to rainfall can improve the performance of satellite rainfall retrieval. **It is foreseen that further inclusion of data from other bands, sensors and satellites will yield additional performance improvements.**

Future research should consider this idea and explore the inclusion of additional EO datasets that capture variables related to rainfall. The selection of these datasets should be guided by factors such as spatial and temporal resolution, data quality, and historical record length. For example, although often modern satellites offer high-resolution data, that can be useful to analyze small-scale weather patterns, they have the trade-off of shorter historical records on which to train data-driven models. Additionally, remote sensing sensors inherently have resolution trade-offs, across spatial, temporal, spectral and radiometric resolutions. Although it is beyond the scope of this dissertation to specify which datasets should be used, some data to be considered – and associated potentially useful products – are:

- **Soil Moisture:** Changes in soil moisture, as initially shown by Hasenauer et al. (2014) [119] and more recently by Mosaffa et al. (2023) [120], can provide indirect information about precipitation. Examples of EO products containing information on soil moisture are those from ESA's Sentinel 1 and SMOS satellites, and NASA's SMAP missions.

- **Land Surface Temperature (LST):** Although the relationship between LST and rainfall is complex, the two variables are undoubtedly interlinked and influence each other. For instance, higher land surface temperatures can increase evaporation, potentially causing precipitation to evaporate before reaching the ground, and drive increased atmospheric moisture. Conversely, when precipitation reaches the ground it cools it down. Some relevant EO sensors that produce LST datasets are the Sea and Land Surface Temperature Radiometer (SLSTR) onboard ESA's Sentinel-3, Terra and Aqua onboard NASA's MODIS, and the long-record USGS Landsat series.

- **Topography:** Elevation affects local weather patterns through processes such as orographic lifting, that can cause condensation in mountainous regions and affect regional rainfall distribution [121]. Therefore, incorporating Digital Elevation Models (DEM) could improve spatial rainfall estimates. Examples of such datasets are the Space Shuttle Radar Topography Mission (SRTM)'s DEM and the ASTER Global DEM.

- **Other parameters:** Multispectral imagery can provide information on vegetation status, cloud albedo, and surface irradiance, among others, all of which interact with rainfall dynamics. Examples of multispectral sensors to consider are the Multispectral Instrument (MSI) onboard Sentinel 2, and, particularly relevant in this dissertation, the SEVIRI instrument onboard Meteosat Second Generation (MSG). Although this dissertation used only two MSG's

SEVIRI channels, the instrument provides many other wavebands with the same spatial and temporal resolution, enabling seamless incorporation into RainRunner or other similar satellite rainfall retrieval models.

Lastly, the launch of the new generation of the Meteosat satellites deserves special attention, as a key advancement in European weather observation. Meteosat Third Generation (MTG) was launched in December 2022 and offers new possibilities for satellite rainfall retrieval thanks to its higher spatial and temporal resolution compared to MSG utilized in this research. Particularly, the temporal resolution was improved from 15 to 10 minutes and the spatial resolution, from 3 to 2 km. Furthermore, MTG carries a new set of sensors. These developments will undoubtedly yield better results in future research.

It is worth noting that this holistic or systemic view of the Earth aligns well with the evolution of the global climate sciences landscape. Particularly, this idea aligns with the principles of the digital twins of the Earth, a new concept that is rapidly gaining popularity, enabled by modern technologies and supercomputing facilities. They are digital representations of the Earth that intend to replicate our physical world with all its processes and uncertainties [122]. The popularity and relevance that digital twins of the Earth have gained during the last years - during the development of this dissertation – is reflected in the ambitious Destination Earth project (DestinE) of the European Commission, to be implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency (ESA). DestinE aims to create a digital twin of the Earth system to support decision making, particularly in the face of climate change [123].

5.3. DEEP LEARNING AND DATA CUBES TO HARVEST THE POWER OF EARTH OBSERVATION

The advancement of Deep Learning methods, that are highly data-intensive, could offer a pathway to fully utilize the vast amounts of EO data. This dissertation served as a showcase of the application of Deep Learning methods for rainfall retrieval in West Africa, and as a DL-based diagnosis of how the characteristic atmospheric conditions of West Africa affect the performance of satellite rainfall products. As such, the Deep Learning methods considered (CNNs and ConvLSTMs) are early DL methods and in no way the most advanced. DL is a rapidly changing field, with new, more powerful architectures being continuously developed. Therefore, the development of an operational model would require extensive research to design the most suitable DL architecture.

Two more complex DL architectures whose use has flourished during the development of this research and that have yielded promising results in satellite rainfall estimation and prediction are Trajectory Gated Recurrent Unit (TrajGRU) and Generative Adversarial Networks (GANs). While the first one was explicitly developed for rainfall nowcasting [124], GANs have been successfully applied also to rainfall estimation, such as in PrecipGan [125] and Sat2rain [126]. Following the development of the PERSIANN satellite rainfall products, the developer group at the University of California Irvine applied a modification of GAN, conditional GAN (cGAN) to estimate rainfall [127]. Very close to this research, both PrecipGAN and Sat2Rain combined multiple channels from geostationary satellites and elevation data to estimate rainfall. However, once again these models required a dense ground observation network for training, which is missing in the case of West Africa. Sat2rain was trained on radar data over Asia [126]. and PERSIANN-cGAN, on data from the GPM Ground Validation Data Archive over the United States [127].

However, pursuing a multi-satellite and multi-instrument approach poses the challenge of working with different orientations and spatial resolutions, among others. A solution to this can be inspired by EO data cubes. They are a proven concept by Geoscience Australia (GA), the Australian Science Industry (CSIRO) and a super-computer facility (NCI) and are a representation of EO data in which different data is organised as spatially aligned pixels ready for analysis [128]. Although in this application the data cubes would not contain analysis-ready data, the concept can be used instead as an efficient structure to organise all input satellite data for ingestion into the RainRunner model. In such an arrangement, different products could be thought of as different layers, and data could be collocated using the same geographic coordinates. A possible methodology to create such data cubes used in a promising early investigation during the course of this research is the combination of the open data cube libraries with the PostgreSQL database manager. This allows to access the stored data by location, date, dimensions and resolution of the desired output. Further, this dissertation addressed the lack of dense target data to train a DL model by considering point data from the centre of the input image, instead of gridded data. In this way, data cubes should also be centred on point measurements.

Figure 5.2(a) shows an example of an initial setup devised in early stages of this research. It contains ground swaths of multiple satellites (large green, red, yellow and blue geometries) overlayed on point rainfall measurements (dots) surrounded by squares of the shape of the desired input data to the model (small squares). Figure 5.2(b) represents the concept of a data cube, that could overcome this obstacle, centered on a rainfall measurement.

This methodology has the potential to be expanded to other applications: Organizing relevant EO data layers in a data cube to be input into a Deep Learning model to predict a target variable of interest, for its center.

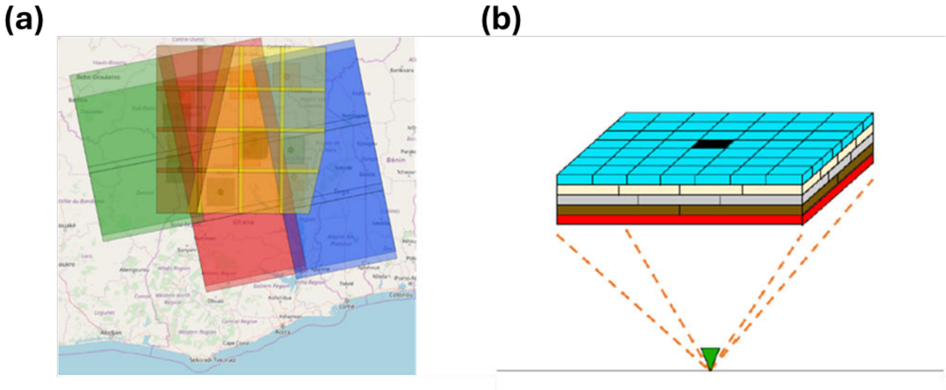


Figure 5.2: (a) Multi-satellite and multi-sensor experimental setup, with different satellite ground swaths and (b) data cube concept, containing multiple data layers, centred on one rainfall measurement.

5.4. CITIZEN SCIENCE AND AS A GROUND DATA SOURCE

One of the main challenges facing rainfall information in Sub-Saharan Africa is the sparse rain gauge distribution. Without dense ground data it becomes extremely challenging to develop or validate a satellite rainfall retrieval model, whether analytical or DL-based. The World Bank estimates that the overall cost of modernizing and maintaining of a full-fledged weather observation system is around $1.5 billion [129], which is highly limiting for many African nations.

Various innovative rainfall sensing methods could overcome this obstacle, such as opportunistic sensing through GPS sensors or telephone microwave links and Citizen Science (CS). CS has the added value of involving the general population in science and the tackling of the problem.

During this research we carried out a CS project named Schools and Satellites (SaS) in which 51 citizen scientists across the North of Ghana measured daily rainfall during a period of 2.5 years. Despite the high educational and societal value of SaS, the data gathered was inconsistent and not always reliable. Therefore, these data were deemed unfit for the training dataset, and were only used for model validation. Even then, several assumptions had to be made.

Citizen Science remains a powerful tool to gather ground observations, particularly relevant in data-scarce contexts such as Sub-Saharan Africa, and it must be explored. Some of the key learnings from SaS to design effective CS projects that produce scientifically valid data are:

- To ensure trustworthiness and high data quality, besides data quality control mechanisms, citizen scientists should be given extensive training and supervision.
- To avoid data gaps, every project should be adapted to the context of a particular region, including internet connectivity and available resources. Another strategy that can help avoid data gaps is to introduce redundancy in the design of the observation network, with more than once citizen scientist in virtually the same location. The necessary distance between redundant citizen scientists would be determined by the resolution of the model.
- To ensure harmony and avoid possible conflicts, as well as appropriateness of the project, cultural awareness and collaboration with local partners are imperative.
- It is necessary to work with engagement strategies throughout the project to keep citizen scientists engaged.
- Any project involving CS must have a dedicated team to ensure its success.

For a detailed explanation of SaS, the reader is directed to the Appendix.

Lastly, there are initiatives such as the school-to-school program of the Trans-African Hydro-Meteorological Observatory (TAHMO) [15] that are a step between a citizen science campaign and a high-cost fully-fledged observation network. TAHMO places weather stations in or near schools, integrating them in the curriculum. In this set-up, relatively rich schools subscribe to the program in a “buy one, pay two” arrangement, with one station placed in a relatively poor school in rural Africa. Both schools are then paired in a “sister schools” program through which they engage in conversation about climate and its importance in their respective societies. In this way, the project supports education and cultural exchange while ensuring that the weather stations are maintained and secured [15].

5.5. THE NEED OF INTERDISCIPLINARY COLLABORATION

Throughout the time that I have spent working on this research I have observed that academia often operates in silos, with researchers focused solely on their individual fields. However, as this Chapter has made clear, this is a problem that involves multiple disciplines. Expertise from Computer Science is needed to design optimal model architectures; Earth Observation informs the selection of suitable satellites and sensors; Atmospheric Science and Hydrology recommend relevant information to include in the models based on the physical sciences of rainfall processes; and, where Citizen Science is to be employed, Social Sciences are key in designing engaging and effective programs.

This need for collaboration is particularly needed in fast-moving fields like Deep Learning and Big Data, and especially to tackle one of the most urgent issues of our time: building resilience against climate change. **It is only through such interdisciplinary efforts that we can develop reliable rainfall information that supports agriculture, ecosystem management, and overall resilience in Sub-Saharan Africa. Only from such efforts will we be able to progress towards a climate-resilient and food- and economically-secure Africa.**

6

CONCLUSIONS

This Chapter summarizes the key findings of this thesis, discusses the advantages and limitations of the proposed methodologies and reflects on the scientific relevance of this work by highlighting its contributions to the scientific community.

6.1. MAIN FINDINGS

In this work, we developed four Deep Learning (DL) models for satellite rainfall retrieval called RainRunner. The models perform a binary classification (rain / no-rain) of 3-hour intervals based on Meteosat thermal infrared (TIR), water vapor (WV), a combination of TIR and WV data, and a combination of these and temporal data. The models were developed locally over the North of Ghana and applied across the Sudanian Savanna in West-Africa, achieving similar results. This work serves as a proof-of-concept for DL-based satellite rainfall retrieval in the data-scarce region of West Africa. Furthermore, performance evaluation in all stages of model development and testing gave us important insights into rainfall processes in the region, in particular related to relations between deep convection systems, non-precipitating low-level features and dry Saharan air – as depicted in EO imagery – and rainfall measured on the ground. These findings provide an explanation for the poor performance of TIR-only satellite rainfall products in the region, debunking existing methods based on Cold Cloud Duration (CCD). In the following we give an answer to the research questions formulated in Chapter 1:

RQ1. How can DL be exploited to improve satellite rainfall retrieval in data-scarce contexts?

In Chapter 2 we proposed a methodology to apply DL methods to data-scarce contexts, showing promising results. It overcame the challenge of data-scarcity for model training by replacing the gridded target training dataset with point data. Before feeding images to the model, they are cropped so as to be centered on a target point rainfall measurement. Once trained, the output of the model is not a gridded rainfall map but a point rainfall estimation (in this case binary classification rain / no-rain). Therefore, the model does not need dense gridded data but only point data.

The chapter presented two models - one based on CNN and one combining CNN and ConvLSTM. Both models obtained performances comparable to state-of-the-art satellite rainfall products. This is a remarkable result given that they are based on simple DL architectures, developed with a small training dataset—observations from 8 stations over 2.5 years, with 20.4% data gaps—and only using TIR data. Both models consistently outperformed PERSIANN-CCS in our test dataset. The RainRunner models even outperformed the much more complex IMERG product during the second half of the rainy season.

As for model architecture, the CNN-based model marginally outperformed the model combining CNN and ConvLSTM yet with a faster training process, even if at the expense of a higher number of training parameters (4.7 times more parameters).

Above all, the high performance that the models were able to achieve despite the significant challenge of data scarcity showed their high efficiency and, ultimately, the potential of DL to model rainfall in regions with low data availability.

RQ2. What role can water vapor observations and temporal information, added to thermal infrared information, play in satellite rainfall retrieval?

Chapter 3 continued the development of the (CNN-based) RainRunner model by incorporating WV and temporal data into the model. Then, Chapter 4 tested the model on 12 other stations in the wider Sudanian Savanna, resulting in similar performances and findings coherent with Chapter 3. These chapters served as an example of how meteorology knowledge can be used to analyze the results of a DL model, and explain them from a physical perspective.

The combination of TIR with WV data was beneficial in all instances – for the north of Ghana and the wider savanna region. This indicates similar rainfall processes captured. The role that water vapor information plays in satellite rainfall retrieval - particularly in the Sudanian Savanna region - can be summarized in the two following points:

1. **Inhibition of low-level features.** The effective layer for WV data - that is, the atmospheric layer that contributes most to the signal at the corresponding wavelength - is located at a higher altitude than that of TIR data, which can penetrate through the whole depth of the atmosphere without being absorbed. As a result, the part of the atmosphere observable in WV data is shallower and limited to higher altitudes. Therefore, WV data highlights areas of strong convection – higher cloud tops - , usually related to heavy rainfall. This effect can be used to discard low-level features that do not produce rainfall. In this way it constitutes an alternative to the traditional Cold Cloud Duration (CCD) method, employed by other products such as TAMSAT and CHIRPS.

2. **Detection of dry air masses advected from the Sahara Desert**, that produce vertical discontinuities in rainfall events. The results of Chapter 4 show that water vapor adds greater value during the first half of the rainy season (March to June). This can be explained by dry air being advected into the region from the Sahara Desert during this season. This is part of a larger atmospheric process that plays a crucial role in West African climate and rainfall dynamics during the year, associated with converging trade winds and low pressure systems, and often loosely referred

to as the movement of the ITCZ. During the first months of the rainy season the Sahara Thermal Low (STL) reaches further south, creating convective inhibition (CIN) and suppressing rainfall. This dry air is observed in WV data, and therefore its CIN-effect is corrected for in rainfall estimation. Indeed, during this phase the model based only on TIR data shows a higher False Alarm Ratio. During the second half of the rainy season, the occurrence of dry air suppressing rainfall is less probable than during the previous months. In these instances, TIR data has a more direct relationship to rainfall recorded on the ground and adding WV data does not alter the model performance significantly.

Temporal information also proved more valuable during the first than during the second half of the rainy season. We suggest that this might be explained by the predominant rainfall mechanisms in these two periods: DCC rain events, with a strong diurnal cycle, are most frequent during the first half of the rainy season. As for seasonal information (i.e., the number of the month), we suggest that the model might learn to expect more rainfall inhibition during the first part of the rainy season. Furthermore, the weaker wind shear during the months of July and August will induce a smaller displacement of rainfall from the cloud (top) to the ground, which might improve TIR performance during the second half of the rainy season.

Lastly, the added benefit of incorporating information sources is smaller during the second half of the rainy season. This might be due to the absence of convective inhibition during this phase, that makes the relationship between rainfall and TIR data more direct.

RQ3. Can a DL satellite rainfall retrieval model developed for the north of Ghana be extrapolated to the wider Sudanian Savanna bioclimatic region?

This research question was addressed in Chapter 4, were we applied the model developed using data from 8 stations in the north of Ghana to 12 other stations in the Sudanian Savanna. The model achieved similar performances, showing a good generalization capability. Comparing the performances of the RainRunner models across stations during the two halves of the rainy season shows larger differences between stations during the first than during the second halves of the rainy season. We suggest that the explanation behind this is the larger spatial variability in atmospheric conditions across the whole Sudanian Savanna. During this period, the model combining TIR, WV and temporal information shows a greater generalization ability than those using only one (TIR or WV) or two (TIR and WV) sources of information.

6.2. ADVANTAGES AND LIMITATIONS

This dissertation served as a showcase for the application of Deep Learning methods for rainfall retrieval in West Africa, and as a DL-based diagnosis of how satellite observations capture the characteristic atmospheric conditions of West Africa and how this affects the performance of existing satellite rainfall products.

One advantage of DL satellite rainfall retrieval as discussed in this dissertation is the high efficiency of DL to learn relationships between variables, as compared to physics-based methods. This is demonstrated by the performances achieved by RainRunner, comparable to state-of-the-art models and even surpassing IMERG during the second half of the rainy season, both in Northern Ghana and in the broader Sudanian Savanna, in terms of Frequency Bias (FBias).

Furthermore, with our methodology we showed how meteorology knowledge can be used to analyze the results obtained with a DL model, and to get insights into physical processes. In this line and as proposed in Chapter 5, collaboration between the Deep Learning and Meteorology fields can lead to important advancements in the field of satellite rainfall retrieval.

This thesis also shows that by incorporating WV observations into a TIR-based rainfall retrieval model, the rainfall over-detection typical of TIR methods can be minimized. Across the whole Sudanian Savanna, WV data was able to correct for non-precipitating low-level features as well as to identify dry Saharan air suppressing rainfall. However, even if the addition of WV improved the performance of the models, they still obtained a high FAR. Among others, potential reasons for this, as seen in the literature (Mc Collum, 2000), might be a higher concentration of atmospheric aerosols, higher clouds tops and a high land surface temperature that causes rainfall to evaporate before reaching the ground. Therefore, possible future avenues to improve model performance would be to add other layers with satellite atmospheric aerosols and land surface temperature data.

It is also important to bear in mind that the DL methods considered here are early DL methods and in no way the most advanced. Therefore, an exhaustive investigation of the most appropriate DL architecture should be carried out should RainRunner be operationalized.

6.3. SCIENTIFIC CONTRIBUTION OF THIS DISSERTATION

The findings from this dissertation advance satellite rainfall retrieval in the Sudanian Savanna region of West Africa in various ways:

1. Using simple DL architectures, RainRunner **demonstrates the great potential of DL methods for satellite rainfall retrieval in data-scarce regions.**

2. By using DL as a diagnostic tool, this work identifies relationships between non-precipitating low-level features, dry Saharan air and strong convective areas, with rainfall measured on the ground. In this way, **it debunks the CCD method and explains TIR rainfall over-detection** in two ways: (1) TIR-based methods are often contaminated with low-level features that do not produce rainfall and (2) convective clouds observed in TIR imagery might not produce rainfall or, it might not reach the ground because of convective inhibition and rainfall suppression produced by dry Saharan air.

3. **It provides evidence in support of regional models over global models.** Particularly, we suggest that regionally training a DL rainfall model can result in better performances than global models, especially in areas with complex, highly region-specific meteorological characteristics, such as the Sudanian Savanna region of West Africa.

BIBLIOGRAPHY

[1] G. G. Tappan, W. M. Cushing, S. E. Cotillon, J. A. Hutchinson, B. Pengra, I. Alfari, E. Botoni, A. Soulé, and S. M. Herrmann. *Landscapes of West Africa: A window on a changing world*. 2016. doi: 10.5066/F7N014QZ.

[2] S. Nicholson. "The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability". In: *ISRN Meteorology* 2013 (Jan. 2013). doi: 10.1155/2013/453521.

[3] C. M. Dunning, E. C. L. Black, and R. P. Allan. "The onset and cessation of seasonal rainfall over Africa". In: *Journal of Geophysical Research* 121.19 (2016), pp. 11405–11424. doi: 10.1002/2016JD025428.

[4] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen. "The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes". In: *Scientific Data* 2 (2015), pp. 1–21. doi: 10.1038/sdata.2015.66.

[5] E. Tarnavsky, D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and F. Kayitakire. "Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present". In: *Journal of Applied Meteorology and Climatology* 53.12 (2014), pp. 2805–2822. doi: 10.1175/JAMC-D-14-0016.1.

[6] F. Satge, D. Defrance, B. Sultan, M.-P. Bonnet, F. Seyler, N. Rouché, F. Pierron, and J.-E. Paturel. "Evaluation of 23 gridded precipitation datasets across West Africa". In: *Journal of Hydrology* 581 (Nov. 2019), p. 124412. doi: 10.1016/j.jhydrol.2019.124412.

[7] C. Le Coz and N. van De Giesen. "Comparison of Rainfall Products over Sub-Saharan Africa". In: *Journal of Hydrometeorology* 21 (2020), pp. 553–596. doi: 10.1175/JHM-D-18-0256.1.

[8] P. Nguyen, M. Ombadi, S. Sorooshian, K. Hsu, A. AghaKouchak, D. Braithwaite, H. Ashouri, and A. R. Thorstensen. "The PERSIANN family of global satellite precipitation data: a review and evaluation of products". In: *Hydrology and Earth System Sciences* 22 (Nov. 2018), pp. 5801–5816. doi: 10.5194/hess-22-5801-2018.

[9] R. Pradhan, Y. Markonis, M. Vargas Godoy, A. Villalba Pradas, K. Andreadis, E. Nikolopoulos, S. M. Papalexiou, A. Rahim, F. Tapiador, and M. Hanel. “Review of GPM IMERG performance: A global perspective”. In: *Remote Sensing of Environment* 268 (Nov. 2021). doi: 10.1016/j.rse.2021.112754.

[10] A. K. Dezfuli, C. M. Ichoku, G. J. Huffman, K. I. Mohr, J. S. Selker, N. van de Giesen, R. Hochreutener, and F. O. Annor. “Validation of IMERG precipitation in Africa”. In: *Journal of Hydrometeorology* 18.10 (2017), pp. 2817–2825. doi: 10.1175/JHM-D-17-0139.1.

[11] J. R. McCollum, A. Gruber, and M. B. Ba. “Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa”. In: *Journal of Applied Meteorology* 39.5 (2000), pp. 666–679. doi: 10.1175/1520-0450-39.5.666.

[12] K. L. Hsu, X. Gao, S. Sorooshian, and H. V. Gupta. “Precipitation estimation from remotely sensed information using artificial neural networks”. In: *Journal of Applied Meteorology* 36.9 (1997), pp. 1176–1190. doi: 10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

[13] F. Tapiador, C. Kidd, K. Hsu, and F. Marzano. “Neural networks in satellite rainfall estimation”. In: *Meteorological Applications - METEOROL APPL* 11 (Mar. 2004), pp. 83–91. doi: 10.1017/S1350482704001173.

[14] C. Shen. “A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists”. In: *Water Resources Research* 54.11 (2018), pp. 8558–8593. doi: 10.1029/2018WR022643.

[15] N. Van De Giesen, R. Hut, and J. Selker. “The Trans-African Hydro-Meteorological Observatory (TAHMO)”. In: *WIREs Water* 1 (2014), pp. 341–348. doi: 10.1002/wat2.1034.

[16] G. Huffman, D. Bolvin, D. Braithwaite, K. Hsu, and R. Joyce. *Algorithm Theoretical Basis Document (ATBD) NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG)*. 29 pages. 2013.

[17] Y. Hong, K. L. Hsu, S. Sorooshian, and X. Gao. “Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System”. In: *Journal of Applied Meteorology* 43 (2004), pp. 1834–1852. doi: 10.1175/JAM2173.1.

[18] M. Estébanez-Camarena, R. Taormina, N. van de Giesen, and M.-C. Ten Veldhuis. “The Potential of Deep Learning for Satellite Rainfall Detection over Data-Scarce Regions, the West African Savanna”. In: *Remote Sensing* 15 (Apr. 2023). doi: 10.3390/rs15071922.

[19] M. Mortimore. *Adapting to Drought. Farmers, Famine and Desertification in West Africa*. Cambridge, UK: Cambridge University Press, 1989.

[20] C. Agnew. “Spatial aspects of drought in the Sahel”. In: *Arid. Environ.* 18 (1990), pp. 279–293. doi: 10.1016/S0140-1963(18)30838-3.

[21] B. Sultan and M. Gaetani. “Agriculture in West Africa in the Twenty-First Century: Climate Change and Impacts Scenarios, and Potential for Adaptation”. In: *Front. Plant Sci.* 7 (2016), pp. 1–20. doi: 10.3389/fpls.2016.01262.

[22] United Nations Department of Economic and Social Affairs, Population Division. *World Population Prospects 2022: Summary of Results*. Tech. rep. UN DESA/POP/2022/TR/NO. 3. New York, NY, USA, 2022.

[23] F. Song, L. Leung, J. Lu, L. Dong, W. Zhou, B. Harrop, and Y. Qian. “Emergence of seasonal delay of tropical rainfall during 1979–2019”. In: *Nat. Clim. Chang.* 11 (2021), pp. 605–612. doi: 10.1038/s41558-021-01066-x.

[24] S. Shukla, G. Husak, W. Turner, F. Davenport, C. Funk, L. Harrison, and N. Krell. “A slow rainy season onset is a reliable harbinger of drought in most food insecure regions in Sub-Saharan Africa”. In: *PLoS ONE* 16 (2021), e0242883. doi: 10.1371/journal.pone.0242883.

[25] G. Naumann, L. Alfieri, K. Wyser, L. Mentaschi, R. Betts, H. Carrao, J. Spinoni, J. Vogt, and L. Feyen. “Global changes in drought conditions under different levels of warming”. In: *Geophys. Res. Lett.* 45 (2018), pp. 3285–3296. doi: 10.1002/2017GL076521.

[26] F. Clapp. *Estimating Monthly Precipitation from Satellite Data*. Tech. rep. Silver Spring, MD, USA: NMC Office Note 46, 1970.

[27] I. Flitcroft, J. Milford, and G. Dugdale. “Relating Point to Area Average Rainfall in Semiarid West Africa and the Implications for Rainfall Estimates Derived from Satellite Data”. In: *J. Appl. Meteor. Climatol.* 28 (1989), pp. 252–266. doi: https://doi.org/10.1175/1520-0450(1989)028<0252:RPTAAR>2.0.CO;2.

[28] Y. Arnaud, M. Desbois, and A. Gioda. “Towards a Rainfall Estimation Using Meteosat over Africa”. In: *Hydraulics Hydrology of Arid Lands*. Reprinted from Proceedings of the Int’l. Symposium HY & IR Div.IASCE, San Diego, CA, USA, 30 July–2 August. 1990.

[29] M. Ba, R. Frouin, and S. Nicholson. “Satellite-Derived Interannual Variability of West African Rainfall during 1983–1988”. In: *J. Appl. Meteor. Climatol.* 34 (1995), pp. 411–431. doi: https://doi.org/10.1175/1520-0450(1995)034<0411:SDIVOW>2.0.CO;2.

[30] W. Atiah, L. Amekudzi, J. Aryee, K. Preko, and S. Danuor. “Validation of Satellite and Merged Rainfall Data over Ghana, West Africa”. In: *Atmosphere* 11 (2020), p. 859. doi: https://doi.org/10.3390/ATMOS11080859.

[31] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, *et al.* “Skillful Precipitation Nowcasting Using Deep Generative Models of Radar”. In: *Nature* 597 (2021), pp. 672–677. doi: 10.1038/s41586-021-03854-z.

[32] Q. Miao, B. Pan, H. Wang, K. Hsu, and S. Sorooshian. “Improving Monsoon Precipitation Prediction Using Combined Convolutional and Long Short Term Memory Neural Network”. In: *Water* 11.5 (2019), p. 977. doi: 10.3390/w11050977.

[33] Y. Tao, X. Gao, A. Ihler, K. Hsu, and S. Sorooshian. “Deep Neural Networks for Precipitation Estimation from Remotely Sensed Information”. In: *2016 IEEE Congress on Evolutionary Computation (CEC)*. Piscataway, NJ, USA: IEEE Press, 2016, pp. 1349–1355. doi: 10.1109/CEC.2016.7743945.

[34] Y. Qiangqiang, S. Huanfeng, L. Tongwen, L. Zhiwei, L. Shuwen, J. Yun, X. Hongzhang, T. Weiwei, Y. Qianqian, W. Jiwen, *et al.* “Deep learning in environmental remote sensing: Achievements and challenges”. In: *Remote Sens. Environ.* 241 (2020), p. 111716. doi: 10.1016/j.rse.2020.111716.

[35] J. Schmid. “The SEVIRI Instrument”. In: *Assembly* (2000), pp. 1–10.

[36] EUMETSAT. *MSG Level 1.5 Image Data Format Description*. 2017.

[37] W. B. Group. *Third Ghana Economic Update: Agriculture as an Engine of Growth and Jobs Creation*. 2018.

[38] C. McSweeney, M. New, and G. Lizcano. *UNDP General Climate Change Country Profiles: Ghana*. New York, NY, USA, 2008.

[39] G. Climatology. *Ghana Meteorological Agency*. https://www.meteo.gov.gh/gmet/climatology/. Accessed: 2021-10-19.

[40] M. Maranan, A. Fink, and P. Knippertz. “Rainfall Types over Southern West Africa: Objective Identification, Climatology and Synoptic Environment”. In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 1628–1648. doi: 10.1002/qj.3345.

[41] B. Geerts and T. Dejene. “Regional and Diurnal Variability of the Vertical Structure of Precipitation Systems in Africa Based on Spaceborne Radar Data”. In: *J. Clim.* 18 (2005), pp. 893–916. doi: 10.1175/JCLI-3316.1.

[42] A. Moraux, S. Dewitte, B. Cornelis, and A. Munteanu. “A Deep Learning Multimodal Method for Precipitation Estimation”. In: *Remote Sens.* 13.16 (2021), p. 3278. doi: 10.3390/rs13163278.

[43] H. Wu, Q. Yang, J. Liu, and G. Wang. “A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China”. In: *J. Hydrol.* 584 (2020), p. 124664. doi: 10.1016/j.jhydrol.2020.124664.

[44] J. Johnson and T. Khoshgoftaar. “Survey on Deep Learning with Class Imbalance”. In: *J. Big Data* 6 (2019), p. 27. doi: 10.1186/s40537-019-0192-5.

[45] B. Pan, K. Hsu, A. AghaKouchak, and S. Sorooshian. “Improving Precipitation Estimation Using Convolutional Neural Network”. In: *Water Resour. Res.* 55 (2019), pp. 2301–2321. doi: 10.1029/2018WR024090.

[46] S. Hochreiter and J. Schmidhuber. “Long Short-term Memory”. In: *Neural Comput.* 9 (1997), pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735.

[47] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”. In: *Adv. Neural Inf. Process. Syst.* 2015, pp. 802–810.

[48] J. Yu, Z. Wang, V. Vasudevan, L. Yeung, M. Seyedhosseini, and Y. Wu. *Coca: Contrastive captioners are image-text foundation models*. https://arxiv.org/abs/2205.01917. arXiv:2205.01917. 2022.

[49] K. Salman, M. Naseer, M. Hayat, S. Zamir, F. Khan, and M. Shah. “Transformers in vision: A survey”. In: *ACM Comput. Surv.* 54 (2022), pp. 1–41. doi: 10.1145/3505244.

[50] M. Schuster and K. Paliwal. “Bidirectional recurrent neural networks”. In: *IEEE Trans. Signal Process.* 45 (1997), pp. 2673–2681. doi: 10.1109/78.650093.

[51] P. Roebber. “Visualizing Multiple Measures of Forecast Quality”. In: *Weather Forecast.* 24 (2009), pp. 601–608. doi: 10.1175/2008WAF2222159.1.

6

[52] D. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*. San Diego, CA, USA, 2015, pp. 1–15.

[53] S. Raschka. *Model evaluation, model selection, and algorithm selection in machine learning*. arXiv:1811.12808. 2018. doi: 10.48550/arXiv.1811.12808.

[54] Y. Wang, Y. You, and M. Kulie. “Global Virga Precipitation Distribution Derived From Three Spaceborne Radars and Its Contribution to the False Radiometer Precipitation Detection”. In: *Geophys. Res. Lett.* 45 (2018), pp. 4446–4455. doi: 10.1029/2018GL077891.

[55] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. *A ConvNet for the 2020s*. https://arxiv.org/abs/2201.03545. arXiv:2201.03545. 2022. doi: 10.48550/arXiv.2201.03545.

[56] M. Estébanez-Camarena, F. Curzi, R. Taormina, N. van de Giesen, and M.-C. ten Veldhuis. “The Role of Water Vapor Observations in Satellite Rainfall Detection Highlighted by a Deep Learning Approach”. In: *Atmosphere* 14.6 (2023), p. 974. doi: 10.3390/atmos14060974.

[57] FAO. *Ghana at a glance*. (accessed on 1 December 2022). url: https://www.fao.org/ghana/fao-in-ghana/ghana-at-a-glance/en/#:~:text=Agriculture%5C%20contributes%5C%20to%5C%2054%5C%20%5C%25%5C%20of,food%5C%20needs%5C%20of%5C%20the%5C%20country..

[58] P. Bechtold. *Convective Precipitation*. (accessed on 1 December 2022). 2019. url: https://confluence.ecmwf.int/display/FUG/9.6+Convective+Precipitation#id-9.6ConvectivePrecipitation-ConvectivePrecipita

[59] C. L. Coz and N. van de Giesen. “Comparison of Rainfall Products over Sub-Saharan Africa”. In: *Journal of Hydrometeorology* 21.4 (2020), pp. 553–596. doi: 10.1175/JHM-D-18-0256.1. url: https://journals.ametsoc.org/view/journals/hydr/21/4/jhm-d-18-0256.1.xml.

[60] ESA. *Meteosat Third Generation takes major step towards its first launch*. (accessed on 1 December 2022). 2021. url: https://www.eumetsat.int/meteosat-third-generation.

[61] C. Shen, X. Chen, and E. Laloy. “Editorial: Broadening the Use of Machine Learning in Hydrology”. In: *Frontiers in Water* 3 (2021). issn: 2624-9375. doi: 10.3389/frwa.2021.681023. url: https://www.frontiersin.org/article/10.3389/frwa.2021.681023.

[62] E. Tarnavsky, D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and F. Kayitakire. “Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present”. In: *Journal of Applied Meteorology and Climatology* 53 (12 2014), pp. 2805–2822. issn: 15588432. doi: 10.1175/JAMC-D-14-0016.1.

[63] G. Dybkjær. “A simple self-calibrating cold cloud duration technique applied in West Africa and Bangladesh”. In: *Geografisk Tidsskrift-Danish Journal of Geography* 103.1 (2003), pp. 83–97. doi: 10.1080/00167223.2003.10649482. eprint: https://doi.org/10.1080/00167223.2003.10649482. url: https://doi.org/10.1080/00167223.2003.10649482.

[64] F. Satgé, D. Defrance, B. Sultan, M.-p. Bonnet, N. Rouche, F. Pierron, J.-e. Paturel, F. Satgé, D. Defrance, B. Sultan, M.-p. Bonnet, and F. Seyler. “Evaluation of 23 gridded precipitation datasets across West Africa To cite this version : HAL Id : hal-02626156”. In: *Journal of Hydrology* 581 (2020). Evaluation of 23 gridded precipitation datasets reliability across West Africa through comparison with rain gauges, at a daily and monthly scale. 4-years period (2000-2003), p. 124412.

[65] W. A. Atiah, L. K. Amekudzi, J. N. A. Aryee, K. Preko, and S. K. Danuor. “Validation of satellite and merged rainfall data over Ghana, West Africa”. In: *Atmosphere* 11 (8 2020). Evaluation of 23 gridded precipitation datasets reliability across West Africa through comparison with rain gauges, at a daily and monthly scale. 4-years period (2000-2003). issn: 20734433. doi: 10.3390/ATMOS11080859.

[66] M. Estébanez-Camarena, R. Taormina, N. van de Giesen, and M.-C. ten Veldhuis. “The Potential of Deep Learning for Satellite Rainfall Detection over Data-Scarce Regions, the West African Savanna”. In: *Remote Sensing* 15.7 (2023). issn: 2072-4292. doi: 10.3390/rs15071922. url: https://www.mdpi.com/2072-4292/15/7/1922.

[67] M. Maranan, A. H. Fink, and P. Knippertzand. “Rainfall Types over Southern West Africa: Objective Identification, Climatology and Synoptic Environment”. In: *Quarterly Journal of the Royal Meteorological Society* 144 (July 2018), pp. 1628–1648. doi: 10.1002/qj.3345.

[68] C. Kidd and G. Huffman. “Global precipitation measurement”. In: *Meteorological Applications* 18 (Aug. 2011), pp. 334–353. doi: 10.1002/met.284.

6

[69] P. M. Aoki. “CNNs for Precipitation Estimation from Geostationary Satellite Imagery”. In: (June 2017), pp. 1–9. url: http://cs231n.stanford.edu/reports/2017/pdfs/557.pdf.

[70] A. Moraux, S. Dewitte, B. Cornelis, and A. Munteanu. “Deep learning for precipitation estimation from satellite and rain gauges measurements”. In: *Remote Sensing* 11 (21 2019), pp. 1–22. issn: 20724292. doi: 10.3390/rs11212463.

[71] G. Barbero, U. Moisello, and S. Todeschini. “Evaluation of the Areal Reduction Factor in an Urban Area through Rainfall Records of Limited Length: A Case Study”. In: *Journal of Hydrologic Engineering* 19 (Nov. 2014), pp. 05014016–1. doi: 10.1061/(ASCE)HE.1943-5584.0001022.

[72] L. Tomassini, D. Parker, A. Stirling, C. Bain, C. Senior, and S. Milton. “The interaction between moist diabatic processes and the atmospheric circulation in African Easterly Wave propagation: The interaction between moist processes and circulation in AEWs”. In: *Quarterly Journal of the Royal Meteorological Society* 143 (Oct. 2017). doi: 10.1002/qj.3173.

6

[73] L. Tomassini. “The Interaction between Moist Convection and the Atmospheric Circulation in the Tropics”. In: *Bulletin of the American Meteorological Society* 101.8 (2020), E1378–E1396. doi: 10.1175/BAMS-D-19-0180.1. url: https://journals.ametsoc.org/view/journals/bams/101/8/bamsD190180.xml.

[74] G. J. Berry and C. Thorncroft. “Case Study of an Intense African Easterly Wave”. In: *Monthly Weather Review* 133.4 (2005), pp. 752–766. doi: 10.1175/MWR2884.1. url: https://journals.ametsoc.org/view/journals/mwre/133/4/mwr2884.1.xml.

[75] ESA. “MSG’s SEVIRI Instrument”. In: *ESA EUMETSAT* (2016), p. 33. url: https://www.esa.int/esapub/bulletin/bullet111/chapter4_bul111.pdf.

[76] N. Selami, G. Sèze, M. Gaetani, J.-Y. Grandpeix, C. Flamant, J. Cuesta, and B. Noureddine. “Cloud Cover over the Sahara during the Summer and Associated Circulation Features”. In: *Atmosphere* 12 (Mar. 2021), p. 428. doi: 10.3390/atmos12040428.

[77] N. van de Giesen, R. Hut, and J. Selker. “The Trans-African Hydro-Meteorological Observatory (TAHMO)”. In: *Wiley Interdisciplinary Reviews: Water* 1 (June 2014). doi: 10.1002/wat2.1034.

[78] U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf. "GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle". In: *Theoretical and Applied Climatology* (115 2014), pp. 15–40. doi: 10.1007/s00704-013-0860-x.

[79] J. Tan, G. J. Huffman, D. T. Bolvin, and E. J. Nelkin. "IMERG V06: Changes to the Morphing Algorithm". In: *Journal of Atmospheric and Oceanic Technology* 36.12 (2019), pp. 2471–2482. doi: 10.1175/JTECH-D-19-0114.1. url: https://journals.ametsoc.org/view/journals/atot/36/12/jtech-d-19-0114.1.xml.

[80] P. Windmeijer and W. Andriesse. "Inland valleys in West Africa : an agro-ecological characterization of rice - growing environments". In: 52 (1993). url: https://edepot.wur.nl/73431.

[81] S. Nicholson. "The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability". In: *ISRN Meteorology* 2013 (Jan. 2013). doi: 10.1155/2013/453521.

[82] P. Knippertz and A. H. Fink. "Dry-Season Precipitation in Tropical West Africa and Its Relation to Forcing from the Extratropics". In: *Monthly Weather Review* 136.9 (2008), pp. 3579–3596. doi: 10.1175/2008MWR2295.1. url: https://journals.ametsoc.org/view/journals/mwre/136/9/2008mwr2295.1.xml.

[83] J. Schmetz, P. Pili, S. Tjemkes, and Diete. "SEVIRI CALIBRATION". In: *Bulletin of the American Meteorological Society* (July 2002). url: https://www-cdn.eumetsat.int/files/2020-04/pdf_sci_bams0702_msg-calib.pdf.

[84] P. J. Roebber. "Visualizing Multiple Measures of Forecast Quality". In: *Weather and Forecasting* 24.2 (2009), pp. 601–608. doi: 10.1175/2008WAF2222159.1. url: https://journals.ametsoc.org/view/journals/wefo/24/2/2008waf2222159_1.xml.

[85] A. Karagiannidis, K. Lagouvardos, and V. Kotroni. "The use of lightning data and Meteosat Infrared imagery for the nowcasting of lightning activity". In: *Atmospheric Research* 168 (Sept. 2015). doi: 10.1016/j.atmosres.2015.08.011.

[86] E. D. W. Mitchell, S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas. "The National Severe Storms Laboratory Tornado Detection Algorithm". In: *Weather and Forecasting* 13.2 (1998), pp. 352–366. doi: https://doi.org/10.1175/1520-0434(1998)013<0352:TNSSLT>2.0.CO;2. url: https://journals.ametsoc.org/view/

6

journals/wefo/13/2/1520-0434_1998_013_0352_tnsslt_2_0_co_2.xml.

[87] C. Liu and E. J. Zipser. ""Warm Rain" in the Tropics: Seasonal and Regional Distributions Based on 9 yr of TRMM Data". In: *Journal of Climate* 22.3 (2009), pp. 767–779. doi: 10.1175/2008JCLI2641.1. url: https://journals.ametsoc.org/view/journals/clim/22/3/2008jcli2641.1.xml.

[88] I. Reinares Martínez, J.-P. Chaboureau, and J. Handwerker. "Warm Rain in Southern West Africa: A Case Study at Savè". In: *Atmosphere* 11 (Mar. 2020), p. 298. doi: 10.3390/atmos11030298.

[89] C. Domenikiotis, M. Spiliotopoulos, E. Galakou, and N. Dalezios. "ASSESSMENT OF THE COLD CLOUD DURATION (CCD) METHODOLOGY FOR RAINFALL ESTIMATION IN CENTRAL GREECE". In: Nov. 2003.

[90] K. Hsu, N. Karbalee, and D. Braithwaite. "Improving PERSIANN-CCS Using Passive Microwave Rainfall Estimation". In: Apr. 2020, pp. 375–391. isbn: 978-3-030-24567-2. doi: 10.1007/978-3-030-24568-9_21.

[91] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang. *Machine Learning Methods in Weather and Climate Applications: A Survey*. Nov. 2023. doi: 10.3390/app132112019.

[92] B. Bochenek and Z. Ustrnul. "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives". In: *Atmosphere* 13.2 (Feb. 2022). issn: 20734433. doi: 10.3390/atmos13020180.

[93] J. Shi, A. Shirali, B. Jin, S. Zhou, W. Hu, R. Rangaraj, S. Wang, J. Han, Z. Wang, U. Lall, Y. Wu, L. Bobadilla, and G. Narasimhan. "Deep Learning and Foundation Models for Weather Prediction: A Survey". In: (Jan. 2025). url: http://arxiv.org/abs/2501.06907.

[94] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". In: *Information Fusion* 58 (June 2020), pp. 82–115. issn: 15662535. doi: 10.1016/j.inffus.2019.12.012.

[95] V. Eyring, W. D. Collins, P. Gentine, E. A. Barnes, M. Barreiro, T. Beucler, M. Bocquet, C. S. Bretherton, H. M. Christensen, K. Dagon, D. J. Gagne, D. Hall, D. Hammerling, S. Hoyer, F. Iglesias-Suarez, I. Lopez-Gomez, M. C. McGraw, G. A. Meehl, M. J. Molina, C. Monteleoni, J. Mueller, M. S. Pritchard, D. Rolnick, J. Runge, P. Stier, O. Watt-Meyer, K. Weigel, R. Yu, and L. Zanna. "Pushing the frontiers in climate modelling and analysis with machine learning".

6

In: *Nature Climate Change* 14.9 (Sept. 2024), pp. 916–928. issn: 17586798. doi: 10.1038/s41558-024-02095-y.

[96] F. Iglesias-Suarez, P. Gentine, B. Solino-Fernandez, T. Beucler, M. Pritchard, J. Runge, and V. Eyring. “Causally-Informed Deep Learning to Improve Climate Models and Projections”. In: *Journal of Geophysical Research: Atmospheres* 129.4 (Feb. 2024). issn: 21698996. doi: 10.1029/2023JD039202.

[97] A. Abusitta, M. Q. Li, and B. C. Fung. “Survey on Explainable AI: Techniques, challenges and open issues”. In: *Expert Systems with Applications* 255 (Dec. 2024). issn: 09574174. doi: 10.1016/j.eswa.2024.124710.

[98] M. B., I. Diallo, and J. S. “West African Monsoon in State-of-the-Science Regional Climate Models”. In: *Climate Variability - Regional and Thematic Patterns*. InTech, Aug. 2013. doi: 10.5772/55140.

[99] S. Kothe, D. Lüthi, and B. Ahrens. “Analysis of the West African Monsoon system in the regional climate model COSMO-CLM”. In: *International Journal of Climatology* 34.2 (Feb. 2014), pp. 481–493. issn: 08998418. doi: 10.1002/joc.3702.

[100] E. K. Vizy and K. H. Cook. “Understanding the summertime diurnal cycle of precipitation over sub-Saharan West Africa: regions with daytime rainfall peaks in the absence of significant topographic features”. In: *Climate Dynamics* 52.5-6 (Mar. 2019), pp. 2903–2922. issn: 14320894. doi: 10.1007/s00382-018-4315-z.

[101] C. Klein, F. Nkrumah, C. M. Taylor, and E. A. Adefisan. “Seasonality and trends of drivers of mesoscale convective systems in Southern West Africa”. In: *Journal of Climate* 34.1 (Jan. 2021), pp. 71–87. issn: 08948755. doi: 10.1175/JCLI-D-20-0194.1.

[102] M. A. Janiga and C. D. Thorncroft. “Convection over tropical Africa and the East Atlantic during the West African monsoon: Regional and diurnal variability”. In: *Journal of Climate* 27.11 (2014), pp. 4189–4208. issn: 08948755. doi: 10.1175/JCLI-D-13-00449.1.

[103] G. J. Huffman, D. T. Bolvin, R. Joyce, O. A. Kelley, E. J. Nelkin, A. Portier, E. F. Stocker, J. Tan, D. C. Watters, and B. J. West. “IMERG V07 Release Notes Changes from V06 to V07”. In: (2024). url: https://gpm.nasa.gov/data/news.

[104] P. Nguyen, M. Ombadi, V. A. Gorooh, E. J. Shearer, M. Sadeghi, S. Sorooshian, K. Hsu, D. Bolvin, and M. F. Ralph. “Persiann dynamic infrared–rain rate (PDIR-now): A near-real-time, quasi-global satellite precipitation dataset”. In: *Journal of Hydrometeorology* 21.12

6

(2020), pp. 2893–2906. issn: 15257541. doi: 10.1175/JHM-D-20-0177.1.

[105] S. Nicholson. “The ITCZ and the Seasonal Cycle over Equatorial Africa”. In: *Bulletin of the American Meteorological Society* 99 (Feb. 2018), pp. 337–348. doi: 10.1175/BAMS-D-16-0287.1.

[106] World Food Programme. *West Africa Seasonal Monitor. 2020 Season. 1-10 June 2020*. Relief Web. [Online PDF] Available at: https://reliefweb. 06a3-3744-8a8d-f50e44e24b6b/WFP-0000116889.pdf. 2020.

[107] World Food Programme. *West Africa Seasonal Monitor. 2020 Season. 1-10 July 2020*. [Online PDF] Available at: https://reliefweb.int/attachme 74fb-340c-83d5-513d04c12fc0/WFP-0000117903.pdf. 2020.

[108] World Food Programme. *West Africa Seasonal Monitor. 2020 Season. 1-10 August 2020*. Relief Web. [Online PDF] Available at: https://reliefweb.int/attachments/94db5dbb-1566-30eb-946a-805b6b6704 0000118580.pdf. 2020.

[109] World Food Programme. *West Africa Seasonal Monitor. 2020 Season. 1-10 September 2020*. Relief Web. [Online PDF] Available at: https://reliefweb.int/attachments/8e6eee22-c13b-393a-be8f-461bb87e398 0000119159.pdf. 2020.

[110] World Food Programme. *West Africa Seasonal Monitor. 2020 Season. 1-10 October 2020*. Food Security Cluster. [Online PDF] Available at: https://fscluster.org/sites/default/files/documents/wafricaseason202 2020.

[111] A. K. Dezfuli, C. M. Ichoku, K. I. Mohr, and G. J. Huffman. “Precipitation characteristics in West and East Africa from satellite and in situ observations”. In: *Journal of Hydrometeorology* 18.6 (June 2017), pp. 1799–1805. issn: 15257541. doi: 10.1175/JHM-D-17-0068.1. url: www.ametsoc.org/PUBSReuseLicenses.

[112] C. G. Georgiev, P. Santurette, F. Dupont, and P. Brunel. *QUANTITATIVE EVALUATION OF 6.2μm, 7.3μm, 8.7μm METEOSAT CHANNELS RESPONSE TO TROPOSPHERIC MOISTURE DISTRIBUTION*. Tech. rep.

[113] E. Bercos-Hickey, T. R. Nathan, and S. H. Chen. “On the relationship between the african easterly jet, saharan mineral dust aerosols, and west african precipitation”. In: *Journal of Climate* 33.9 (May 2020), pp. 3533–3546. issn: 08948755. doi: 10.1175/JCLI-D-18-0661.1.

[114] A. H. Fink and A. Reiner. “Spatiotemporal variability of the relation between African Easterly Waves and West African Squall Lines in 1998 and 1999”. In: *Journal of Geophysical Research D: Atmospheres* 108.11 (2003), pp. 1–17. issn: 01480227. doi: 10.1029/2002jd002816.

[115] A. H. Fink, D. G. Vincent, and V. Ermert. “Rainfall types in the West African Sudanian zone during the summer monsoon 2002”. In: *Monthly Weather Review* 134.8 (2006), pp. 2143–2164. issn: 00270644. doi: 10.1175/MWR3182.1.

[116] D. Stein Zweers. *TROPOMI ATBD of the UV aerosol index*. Royal Netherlands Meteorological Institute. [PDF ONLINE] Available at: https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-UV-Aerosol-Index.pdf. 2022.

[117] A. C. Anuforom. “Spatial distribution and temporal variability of Harmattan dust haze in sub-Sahel West Africa”. In: *Atmospheric Environment* 41.39 (2007), pp. 9079–9090. doi: 10.1016/j.atmosenv.2007.08.003.

[118] R. Washington, C. Bouet, G. Cautenet, and I. Tegen. “Dust as a tipping element: The Bodélé Depression, Chad”. In: *Proceedings of the National Academy of Sciences* 106.49 (2009), pp. 20564–20571. doi: 10.1073/pnas.0711850106.

[119] S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani. “Rainfall from satellite soil moisture data”. In: *Journal of Geophysical Research: Atmospheres* (2014), pp. 5128–5141. doi: 10.1002/2014JD021489.

[120] H. Mosaffa, P. Filippucci, and C. e. a. Massari. “SM2RAIN-Climate, a monthly global long-term rainfall dataset for climatological studies”. In: *Sci Data* 10 (2023), p. 749. doi: 10.1038/s41597-023-02654-6.

[121] J. Houze R. A. “Orographic effects on precipitating clouds”. In: *Rev. Geophys.* 50 (2012), RG1001. doi: 10.1029/2011RG000365.

[122] X. Li, M. Feng, Y. Ran, Y. Su, F. Liu, C. Huang, H. Shen, Q. Xiao, J. Su, S. Yuan, and H. Guo. “Big data in Earth system science and progress towards a digital twin”. In: *Nature Reviews Earth & Environment* 4.5 (2023), pp. 319–332. doi: 10.1038/s43017-023-00409-w.

[123] J. Hoffmann, P. Bauer, I. Sandu, N. Wedi, T. Geenen, and D. Thiemert. “Destination Earth – A digital twin in support of climate services”. In: *Climate Services* 30 (2023), p. 100394. doi: 10.1016/j.cliser.2023.100394.

[124] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. *Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model*. 2017. arXiv: 1706.03458 [cs.CV]. url: https://arxiv.org/abs/1706.03458.

[125] C. Wang, G. Tang, and P. Gentine. “PrecipGAN: Merging microwave and infrared data for satellite precipitation estimation using generative adversarial network”. In: *Geophysical Research Letters* 48.3 (2021). doi: 10.1029/2020GL092032.

[126] H. Sakaino and A. Higuchi. “Sat2rain: Multiple satellite images to rainfall amounts conversion by improved GAN”. In: *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 1484–1491. doi: 10.1109/ICMLA55696.2022.00233.

[127] N. Hayatbini, B. Kong, K. .-. Hsu, P. Nguyen, S. Sorooshian, G. Stephens, C. Fowlkes, R. Nemani, and S. Ganguly. “Conditional Generative Adversarial Networks (cGANs) for Near Real-Time Precipitation Estimation from Multispectral GOES-16 Satellite Imageries—PERSI cGAN”. In: *Remote Sensing* 11.19 (2019), p. 2193. doi: 10.3390/rs11192193.

[128] T. Dhu, B. Dunn, B. Lewis, L. Lymburner, N. Mueller, E. Telfer, and C. Phillips. “Digital earth Australia – unlocking new value from earth observation data”. In: *Big Earth Data* 1.1–2 (2017), pp. 64–74. doi: 10.1080/20964471.2017.1402490.

6

[129] A. Tzachor, C. E. Richards, M. Gudoshava, P. Nying’uro, H. Misiani, J. G. Ongoma, Y. Yair, Y. Mulugetta, and A. T. Gaye. *How to reduce Africa’s undue exposure to climate risks*. Nature. Comment, Nature, Available at: https://doi.org/10.1038/d41586-023-02547-w. 2023. url: https://doi.org/10.1038/d41586-023-02547-w.

[130] S. de Vries, M. Estébanez-Camarena, N. van de Giesen, M.-C. ten Veldhuis, K. Duah, J. Davids, P. Silwal, R. Prajapati, and F. Annor. “Leveraging Citizen Science for Sustainable Development Education and Water Security in Northern Ghana”. In: *Youth and Water Security in Africa*. UNESCO, 2022.

[131] M. Estébanez Camarena and P. Martinez. “PyrSat – Prevention and Response to Wildfires with an Intelligent Earth Observation CubeSat”. In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018, pp. 1–5.

[132] M. Estébanez Camarena, L. Feetham, A. F. Scannapieco, and N. Aouf. “FPGA-based Multi-Sensor Relative Navigation in Space: Preliminary Analysis in the Framework of the I3DS H2020 Project”. In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018.

[133] M. Gbenga Ogungbuyi, P. Martinez, and M. Estébanez Camarena. “Spatiotemporal Investigations of Oil Ground Spills and MODIS Fire Products in Near Real-Time”. In: *Proceedings of the 69th In-*

ternational Astronautical Congress. Bremen, Germany, Oct. 2018, pp. 1–5.

[134] A. F. Scannapieco, L. Feetham, M. Estébanez Camarena, and N. Aouf. “Space-oriented Navigation Solutions with Integrated Sensor-Suite: The I3DS H2020 Project”. In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018.

[135] E. de Miguel, E. Prado, M. Estébanez, A. Martin, and M. Gonzalez. “Direct Geolocation of Satellite Images with the EO-CFI Libraries”. In: *Living Planet Symposium, Proceedings of the Conference*. Ed. by L. Ouwehand. Prague, Czech Republic: ESA-SP Volume 740, May 2016, p. 410. isbn: 978-92-9221-305-3.

[136] M. Estébanez Camarena. “Prediction of Science Fiction that Came True”. In: *Outer Space and Popular Culture*. Ed. by U. Editor. Springer, 2019, pp. 129–144.

[137] S. Anih, G. Badela, T. Campbell, G. Cromhout, M. Estébanez Camarena, L. W. Y. Feng, T. Hugbo, P. H. Khwambala, K. Konar, D. Lindgren, S. Madlanga, R. Maharaj, A. K. Nath, M. G. Ogungbuyi, Z. Pandey, E. Pieterse, and M. Tanner. “Space Resource Utilization: A View from an Emerging Space Faring Nation”. In: *Space Resource Utilization*. Ed. by A. Froehlich. Cham: Springer International Publishing, 2018. isbn: 978-3-319-66968-7.

LIST OF TABLES

LIST OF FIGURES

ANNEX A

Schools and Satellites (SaS) CSEOL Citizen Science Earth Observation Lab

This chapter contains parts of deliverable 4.1, "WP1 - Ground-truth data collection." and deliverable 4.2, "Report WP2 - Algorithm developement", submitted by the Schools and Satellites team, led by the author, to the Citizen Science Earth Observation Lab (CSEOL).

Report WP1 Schools and Satellites

DELIVERABLE #4.1

ADAPTED CONTENT – ONLY SELECTED CONTENT HAS BEEN KEPT IN THIS VERSION.

FOR MORE INFORMATION ON THE SCHOOLS AND SATELLITES PROJECT, CONTACT THE AUTHOR.

Schools and Satellites

Work Package 1: Ground-truth data collection

Objectives

This work package had the following objectives:

O1. Expand the rainfall measurement network in Ghana, by starting in the pilot area. ☑

O2. Introduce an educational module for schoolchildren in Ghana on climate change, the water cycle and earth observation possibilities, that can be added to the regular school curriculum of Ghana. ☑

Results

CS data collection

Designing and developing the CS data collection in the pilot area

For farmers and schoolteachers in the northern regions of Ghana to participate in the SaS project, we have developed instruction materials:

- To make you own #SchoolsandSatellites rain gauge, we recommend watching this video and using the manual 'How to make your own rain gauge'. This pdf also explains what to do.
- In the Rainfall Measurement Syllabus, we explain all the necessary steps to take rainfall measurements yourself or with your students.
- And for teachers, we have made a Teachers Manual that gives a lesson plan to teach their students about rainfall patterns, climate change and taking rainfall measurements.
- To understand how to make your own rainfall graphs, please us this instruction sheet.

The teachers manual is based on a research among the teachers and students interviewed during the fieldtrip in February 2020, and on the Educational Curriculum for Climate Change for Junior High Schools in Ghana.

To collect the data, an African server was opened by S4W, and an African ODK S4W form created. Errors that occurred in the pre-pilot have after that been solved by providing different smartphones to the citizen scientists.

We have produced 50 low-cost soda bottle rain gauges in the summer of 2020.

Roll out of the designed CS data collection

In corona times and with corona measures in place, we performed one on one instructing of the teachers joining the project. At those moments, teachers also received the S4W rain gauges, a SaS t-shirt, a smartphone with the ODK app already installed, and help with installation of the holder for the rain gauge.

This way, we immediately ensured that the installed S4W rain gauges location was suitable concerning rainfall measurement standards.

Furthermore, we have held a survey among the observers of the SaS Citizen Observatory. Of the 44 new observers, 15 responded to the survey. We held a lottery between the respondents, where they could win a raincoat and boots. We held this lottery to motivate them to fill in the survey and to create a fun element into it all.

Validation of the CS measurements

The first validation step is to view all measurements and compare them with the produced photos of the measurements and check on GPS and time of measurement. This will ensure that most of the errors can be corrected.

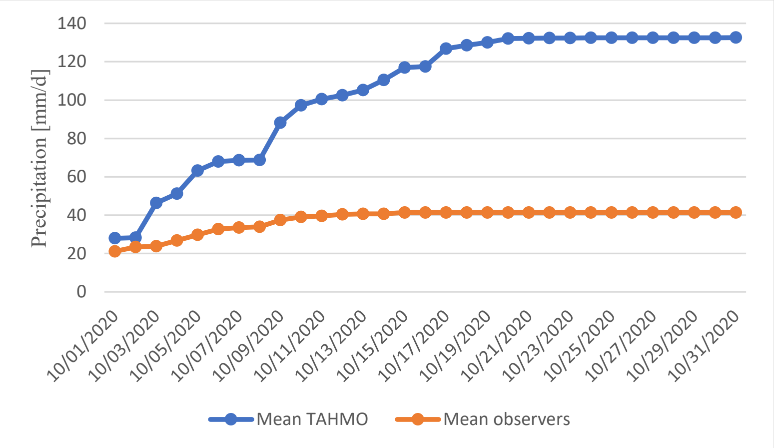
This process is taken up by Smartphones4Water. After correction, a continuation of the error is prevented, by giving feedback to the citizen scientists. See Appendix F for an example of such a feedback input.

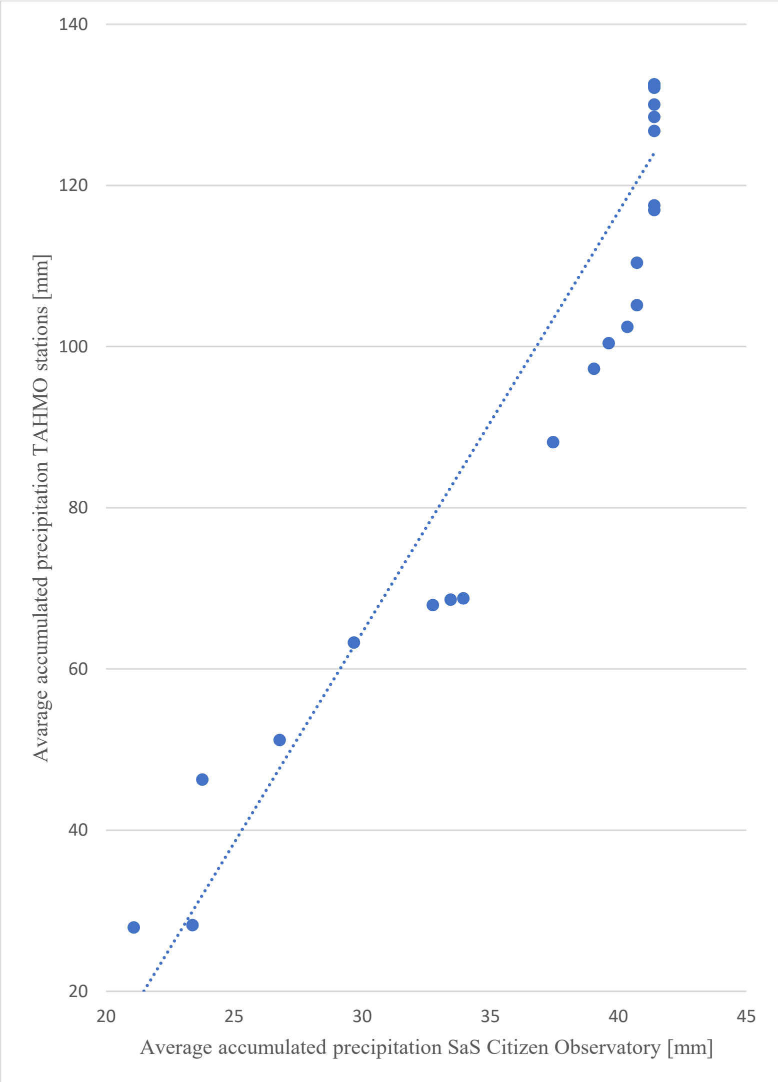
The second validation step is to compare the measurements with each other. Even though we expect a different value over space, outliers can be spotted and removed.

The python script developed by Bilal Abou Hashish also functions for this purpose: to have some insight in the data and to detect outliers. From the script, it can be inferred that the citizen scientists largely are good at taking measurements. By the script, bar plots of any desired timeframe can be plotted, as well as specifying the citizen scientists or clusters to include in the plot.

The third validation step is to compare with the TAHMO weather stations installed in the area.

A first comparison has been made with the data collected in September and October 2020 by the SaS Citizen Observatory (CO) and the TAHMO stations. Since the TAHMO stations are spread over the complete north and the CO is for now only spread over 4 districts, the accumulated rainfall is quite different between the two networks. However, when we create a double mass curve of the two networks, the networks seem to be proportional. Of course, this must be checked again after more data has been collected by the SaS CO.




140
120
100
80
60
40
20
20
25
30
35
40
45
Avarage accumulated precipitation TAHMO stations [mm]
Average accumulated precipitation SaS Citizen Observatory [mm]

Appendices

Appendix D: Roll out of the Citizen Science network

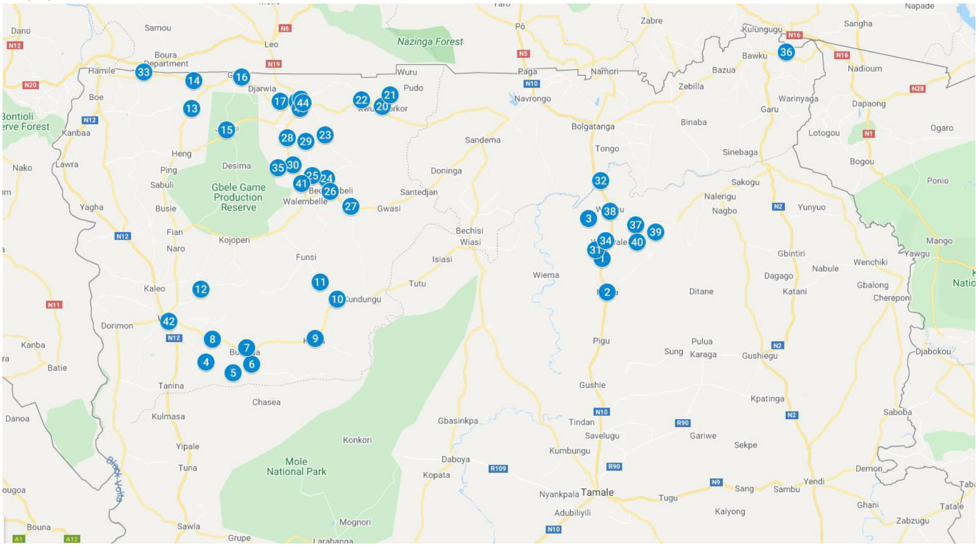


Figure 1 Locations of SaS manual rain gauges. Can also be found on this map.

SCHOOLS AND SATELLITES INSTALLATIONS

Schools and Satellites project seeks to create a large number of rainfall stations across the northern part of Ghana. This is to help look into the changing pattern of rainfall in this part of the country. There has been a change of the onset of the rainy season and the peak period for the raining season. This instability in respect of the rainfall has had grave effects on the productivity of farmers.

As part of the educational component of the schools and satellites project, rainfall stations were to be installed in Junior High Schools. The project seeks to give the school child a hand-on experience on how to take rainfall measurements, plot graphs with the data and share their analysis with their colleagues. All these are being done under the guidance and supervision of the designated teachers in the respective schools involved in the project.

The innovative rain gauges installed are made from 1.5L Fanta bottle.

The installation of the Schools and Satellites (SaS) rain gauges began on the 10th of August 2020. It all looked impossible at first in view of the global pandemic (COVID-19) which has brought a lot of restrictions including shutting down of schools.

The initial plan had to be changed completely to meet the prevailing conditions in the country due to the COVID-19. The new plan was to engage with the education directorate to assist us with potential schools for the installations. Again, the head teacher or the teacher of the selected school had to be resident in the community where the school is located. This was to ensure that the measurements were done on daily basis.

The education offices were consulted to assist with the list of schools. We finally, got four (4) education offices giving us a list to work with. These were West Mamprusi District (North East Region), Wa East, Sissala East and Sissala West (Upper West Region). The number of schools visited included West Mamprusi 10, Wa East 10, Sissala East 17 and Sissala West 6. One was also installed at Gbewaa College of Education in the Upper East region. This makes it a total of 44 schools or sites where rain gauges were installed.

LOGISTICS

The SaS project was to provide all the needed logistics to enable the teachers carry out the task. The teachers were also trained on how to take measurements as well as use the education materials co-developed with SaS team for their lessons. The teachers were then to train their students when school finally re-opens. The logistics made available to the selected schools and assigned teachers included:

1. Plastic rain gauge made from Fanta bottle
2. A mobile phone (Itel A56) installed with the ODK app and educational materials
3. A branded gear with the name of the project
4. A sticker to indicate the school is a SaS school

The education offices and their selected schools welcome us and the project. Most of them expressed great optimism about the project. They have made commitment to ensure that they take care of all that have been handed over to them.

The project also received the endorsement of the assigned (head) teachers. Some of the teachers also engage in farming and therefore see this as an opportunity to monitor the rainfall in their respective communities. They acknowledge the impact it will have on their decision making as far as their farm work is concerned. They promised to also keep hard copies of the daily records.

IMPACT OF SaS

The selected schools for the project are all located in farming communities. From the interactions that went on with the teachers, rainfall data will be useful to the community. Farmers still depend on the traditional knowledge and assumptions for the onset of the rains and the peak period. This on many occasions fail and they lose their investment (crops).

The project has prospect of influencing decisions of farmers in the communities where these innovative rain gauges have been installed. Hence, saving them of their investments. The teachers have been encouraged to share the outcome of this research work with their communities.

Though we did not install on time to cover the beginning of the raining period, we should have some data coming in.

Report WP2 Algorithm development

SCHOOLS AND SATELLITES - DELIVERABLE #4.2

Schools and Satellites

Mónica Estébanez Camarena
TU Delft

ADAPTED CONTENT – ONLY SELECTED CONTENT HAS BEEN KEPT IN THIS VERSION.

FOR MORE INFORMATION ON THE SCHOOLS AND SATELLITES PROJECT, CONTACT THE AUTHOR.

Work Package 2: Algorithm development

Introduction

The aim of this work package was to create a satellite rainfall product able to estimate rainfall without the need of data from the ground, using Deep Learning (DL) and with Earth Observation (EO) data as input.

The model developed, RainRunner, is able to detect rainfall on the ground making use of satellite data only, with a 0.03 degrees spatial resolution and a 3-hour temporal resolution.

The proof-of-concept model achieves similar performances to the well-established Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System (PERSIANN-CCS), developed at the University of California Irvine, and the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), developed by NASA. The main advantage of this model is that it is fully data-driven, uses a simpler algorithm, is tailored to regional rainfall characteristics (in this case of West Africa), and runs in quasi real-time, i.e., it can be applied as soon as GEO IR images become available.

Finally, the model has been tested on the citizen science-based rainfall dataset collected by SaS citizen scientists. On these data, the model outperforms the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) and the Tropical Application of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT) products and has a similar performance to PERSIANN-CCS and IMERG.

Another goal set in the project proposal of using all Sentinel satellites, Meteosat Second Generation (MSG) and a DEM was not met. Instead, the current version of RainRunner uses only MSG data. There are reasons to believe adding the remaining layers would improve the performance of the model. Strategies to add further layers to the model are outlined in this document.

To the best knowledge of the project team, this is the first work in which a DL-based rainfall estimation model is locally trained over Africa. Additionally, a methodology to train DL models when accurate high-density data is missing on the ground has been developed, in which EO images are linked to point-based ground measurements to train the DL model. Potential applications of this include using Citizen Science data to complement station data and create large datasets for EO-DL problems.

RainRunner could set a stage towards better rainfall information in areas of the world where it is currently missing, ultimately contributing to climate adaption in these areas. The methodology developed in this project, i.e. such combination of Deep Learning, Citizen Science and Earth Observation, could potentially be extrapolated to other real-world problems.

This report describes the work developed in this work package, limitations, outcomes and suggestions for future developments.

Results

Citizen Science data

Data taken by 38 citizen scientists in 2019 and 2020 have been used to further test the model. In this case, since the CS data has a daily resolution, the performance of TAMSAT and CHIRPS was also evaluated against the RainRunner algorithm [Figure 8]. For RainRunner, IMERG and PERSIANN-CCS, running at 3-hour resolution, first predictions were made on each of the eight 3-hour intervals in each day. A day was only classified as dry when all eight intervals were classified as dry. Otherwise, the day in question would be classified as rain. Figure 8 shows the performance of all models on the CS dataset. The overall performance is higher compared to Figure 7, as expected as a result of the daily instead of 3-hour resolution. Again, IMERG seems to be leading but very close to RainRunner and with somewhat lower scores for PERSIANN-CCS. In this case, CHIRPS and TAMSAT show a clear disadvantage with substantially lower performance scores.

The rainfall product currently being used by the Ghana Meteorological Agency is CHIRPS. Therefore, and looking at the much higher performance, RainRuner can introduce a clear advantage in this region. Furthermore, the high performance achieved by the two RainRunner models in 38 previously unseen locations show the generalization capability of the models.

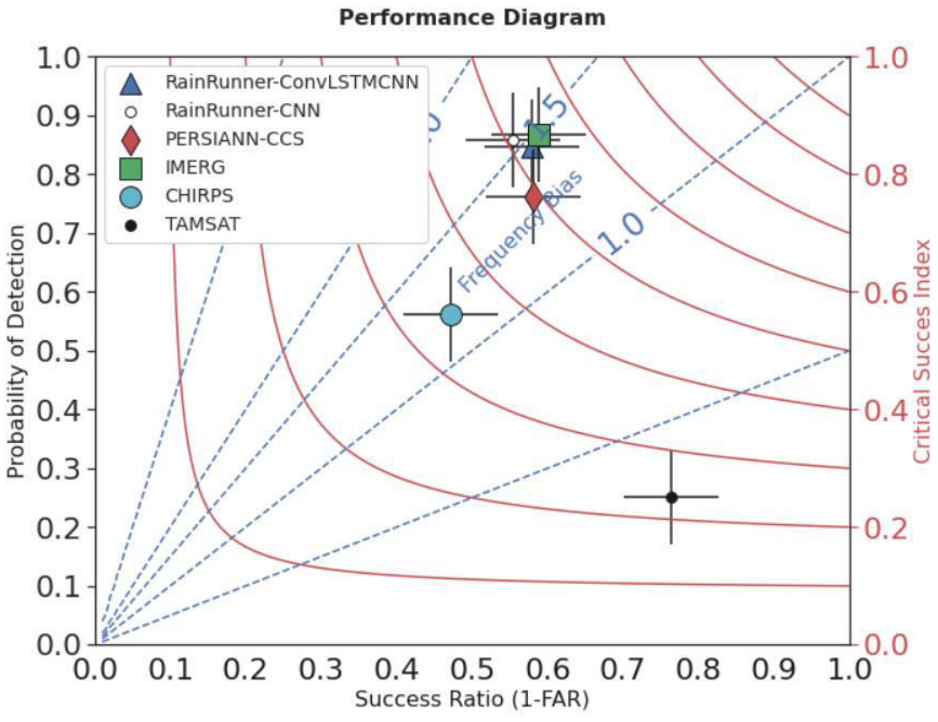


Figure 1: Performance diagram on the citizen science data.

Lessons learned from Citizen science data collection campaign

Possibly the main challenge with the CS data recording was that the citizen scientists recorded rainfall data at different times of the day. Partly this was due to the fact that in-person workshops could not be organized before the full CS campaign in 2020 due to COVID-19. The number of citizen scientists increased during the campaign from 6 to 51, and some of the citizen scientists did not fully understand that measurements were always meant to be taken at the same time, ideally in the morning before school. It was also challenging for the citizen scientists to get to the rain gauges, located in the schools and often far away from the teachers' residence and of difficult access, even more so with schools closed because of the pandemic. Another reported reason for a challenging access to the rain gauge were extremely heavy storms at the recording time. Lastly, it was often difficult to upload the recording via the Smartphones4Water (S4W) app due to a poor internet connection or to the S4W app requesting the GPS coordinates, and the phones not being able to provide them due to the need of an internet connection. The GPS issue of the S4W forms was tackled in 2021, but the internet poor connectivity was still an important limitation. In the cases where the citizen scientists took measurements at a time different than the morning, sometimes this was just after a heavy storm, wanting to report that rain event.

Working with these non-uniform data poses challenges, since we are unsure about the timing of the reported rainfall, and this introduces uncertainty that the performance of the models is correctly evaluated.

As a way to tackle this issue, the definition of a day was made considering the distribution of recording times [Figure 9]. The RainRunner 3-hour intervals were also considered, so that in our definition of a day the start and end corresponded with the start and end of the intervals (i.e. 0 AM, 3 AM, 6 AM, 9 AM, 12 PM, 3 PM, 6 PM, 9 PM). Because most measurements were taken in the morning and the heaviest rainfall happens in the evening [Figure 10], it was decided to define a day as from 12 PM to 12 PM. Then, all CS measurements were assigned to 12 PM of that day, corresponding to rainfall fallen (or not) between 12 PM of the previous day and 12 PM of the current day. In this way, we avoid errors such as evening rainfall in a day where it did not rain in the morning being missed if the citizen scientist recorded "dry". However, it is important to note that there might be errors if a citizen scientist records an evening rainfall event just after it happens, since that measurement will be assigned to that day, but it would have corresponded to the next one.

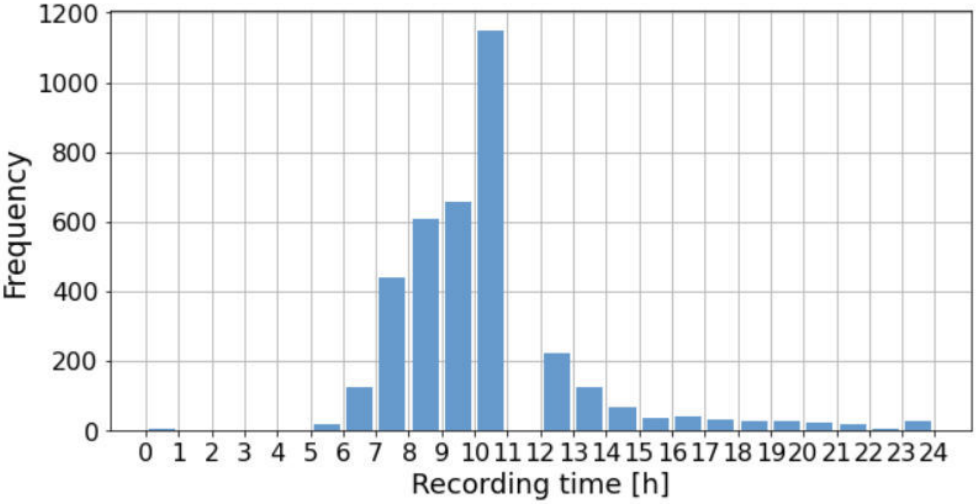


Figure 2: Distribution of citizen scientist recordings in time.

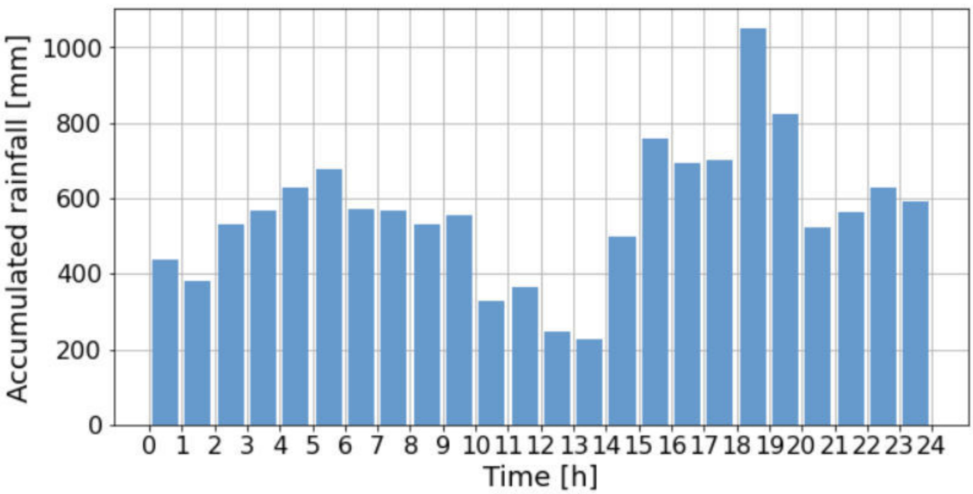


Figure 3: Total accumulated rainfall through the day, in 2019 and 2020 and at the 8 TAHMO stations.

ANNEX C

Supplementary material to Chapter 3.

.

This chapter has been published in Atmosphere 14(6), 974.as an Annex to *The Role of Water Vapor Observations in Satellite Rainfall Detection Highlighted by a Deep Learning Approach*. (2023). Authors: Estébanez-Camarena, M., Curzi, F., Taormina, R., van de Giesen, N., and ten Veldhuis, M. -C. doi: 10.3390/atmos14060974, [56]

0.1. PIXEL ANALYSIS

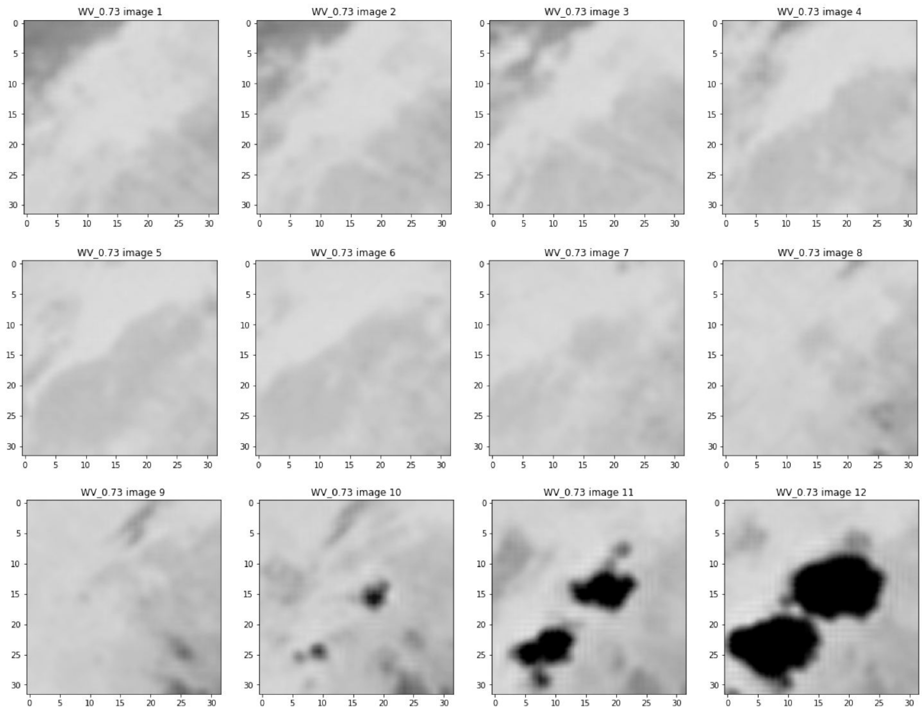


Figure 1: Example of a misclassified rain sequence in WV imagery due to coarse temporal resolution. Pusiga—May 2020.

0.2. DRY SLOTS

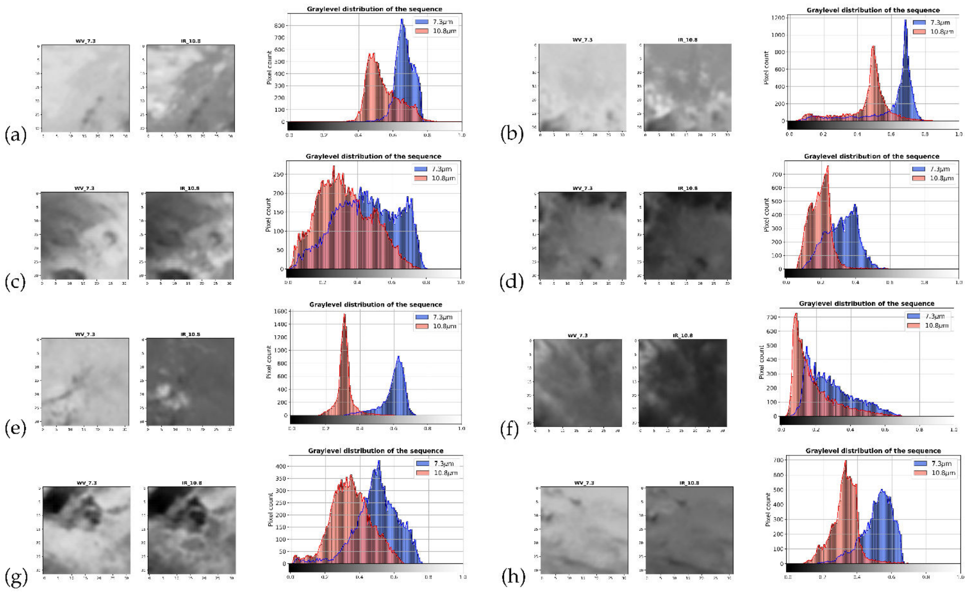


Figure 2: Dry slots observed at TAHMO locations. The characteristics of each dry slot are presented in Table 1.

Event	Ground Truth	TIR	WV	TIR + WV	TIR + WV + Timestamp
(a) Bimbilla, 2020.09.07, 09 h	0	0.51	0.03	0.007	0.17
(b) Bimbilla, 2020.09.12, 06 h	0	0.56	0.54	0.23	0.39
(c) Han, 2020.10.01, 15 h	0	0.73	0.34	0.42	0.33
(d) Bongo, 2020.04.10, 15 h	0	0.66	0.28	0.49	0.38
(e) Bongo, 2020.05.09, 12 h	0	0.41	0.11	0.32	0.35
(f) Daffiama, 2020.05.15, 15 h	0	0.74	0.28	0.47	0.25
(g) Tamale, 2020.06.14, 12 h	0	0.52	0.09	0.29	0.23
(h) Navrongo, 2020.06.20, 15 h	0	0.45	0.04	0.35	0.50

Table 1: Location, date and time of the dry slots depicted in Figure 2, together with the corresponding grountruth (rain = 1/ no-rain = 0) and resulting probabilistic output of the different models.

0.3. DRY INTRUSIONS

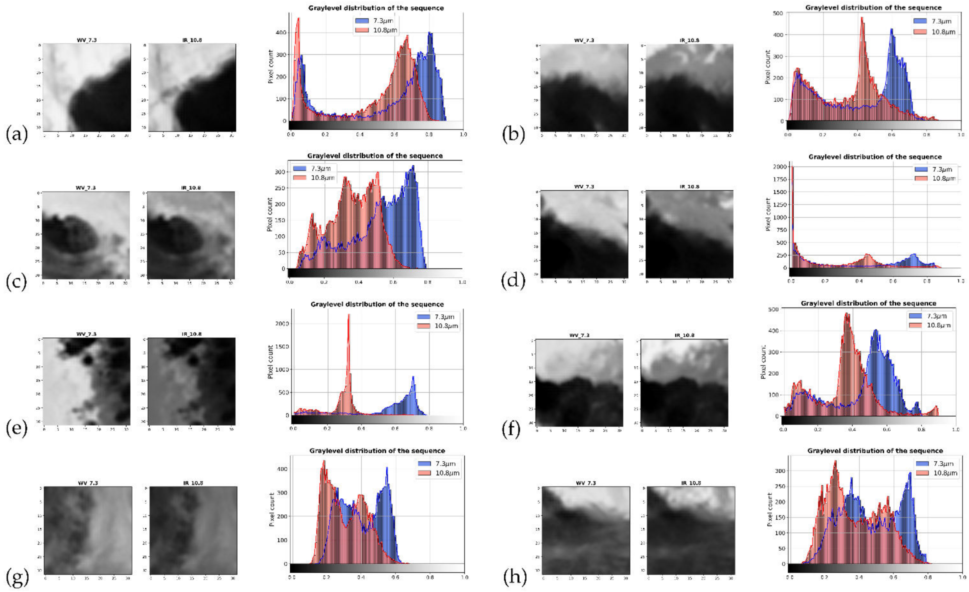


Figure 3: Dry intrusions observed at TAHMO locations. The characteristics of each dry intrusion are presented in Table 2.

Event	Ground Truth	TIR	WV	TIR + WV	TIR + WV + Timestamp
(a) Bimbilla, 2020.03.22, 15 h	0	0.57	0.14	0.67	0.17
(b) Bimbilla, 2020.05.06, 21 h	0	0.92	0.45	0.64	0.23
(c) Bimbilla, 2020.07.26, 12 h	1	0.51	0.26	0.75	0.78
(d) Bimbilla, 2020.09.30, 15 h	0	0.77	0.69	0.46	0.59
(e) Navrongo, 2020.05.17, 12 h	0	0.81	0.41	0.69	0.48
(f) Pusiga, 2020.05.06, 00 h	0	0.50	0.34	0.24	0.29
(g) Pusiga, 2020.07.15, 03 h	0	0.51	0.37	0.46	0.54
(h) Bongo, 2020.09.25, 12 h	0	0.62	0.27	0.44	0.44

Table 2: Location, date and time of the dry intrusions depicted in Figure 3, together with the corresponding grountruth (rain = 1/ no-rain = 0) and resulting probabilistic output of the different models.

0.4. LOW-LEVEL MOISTURE

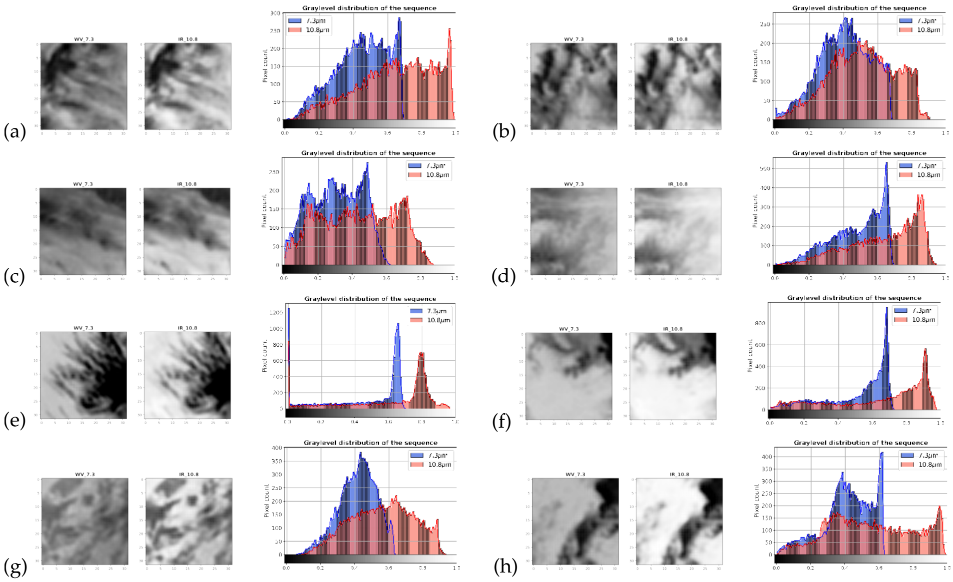


Figure 4: Low-level moisture events observed at TAHMO locations during dry season. The characteristics of each low-level moisture event are presented in Table 3.

Event	**Ground Truth**	**TIR**	**WV**	**TIR + WV**	**TIR + WV + Timestamp**
(a) Bimbilla, 2020.02.11, 18 h	0	0.42	0.83	0.28	0.30
(b) Bimbilla, 2020.12.22, 12 h	0	0.41	0.65	0.37	0.18
(c) Daffiama, 2020.01.22, 21 h	0	0.28	0.54	0.20	0.02
(d) Daffiama, 2020.01.23, 06 h	0	0.22	0.66	0.14	<0.01
(e) Kpandai, 2020.01.25, 00 h	0	0.43	0.58	0.18	0.02
(f) Kpandai, 2020.10.21, 00 h	0	0.30	0.63	0.37	0.23
(g) Navrongo, 2020.02.12, 00 h	0	0.07	0.50	0.24	<0.01
(h) Pusiga, 2020.02.11, 18 h	0	0.08	0.55	0.24	0.05

Table 3: Location, date and time of the low-level moisture events depicted in Figure 4, together with the corresponding grountruth (rain = 1/ no-rain = 0) and resulting probabilistic output of the different models.

LIST OF PUBLICATIONS

1. PUBLICATIONS

1. M. Estébanez-Camarena, R. Taormina, N. van de Giesen, and M.-C. Ten Veldhuis. "The Potential of Deep Learning for Satellite Rainfall Detection over Data-Scarce Regions, the West African Savanna". In: *Remote Sensing* 15 (Apr. 2023). doi: 10.3390/rs15071922
2. M. Estébanez-Camarena, F. Curzi, R. Taormina, N. van de Giesen, and M.-C. ten Veldhuis. "The Role of Water Vapor Observations in Satellite Rainfall Detection Highlighted by a Deep Learning Approach". In: *Atmosphere* 14.6 (2023), p. 974. doi: 10.3390/atmos14060974
3. S. de Vries, M. Estébanez-Camarena, N. van de Giesen, M.-C. ten Veldhuis, K. Duah, J. Davids, P. Silwal, R. Prajapati, and F. Annor. "Leveraging Citizen Science for Sustainable Development Education and Water Security in Northern Ghana". In: *Youth and Water Security in Africa*. UNESCO, 2022

2. CONFERENCE PROCEEDINGS

1. M. Estébanez Camarena and P. Martinez. "PyrSat – Prevention and Response to Wildfires with an Intelligent Earth Observation CubeSat". In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018, pp. 1–5
2. M. Estébanez Camarena, L. Feetham, A. F. Scannapieco, and N. Aouf. "FPGA-based Multi-Sensor Relative Navigation in Space: Preliminary Analysis in the Framework of the I3DS H2020 Project". In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018
3. M. Gbenga Ogungbuyi, P. Martinez, and M. Estébanez Camarena. "Spatiotemporal Investigations of Oil Ground Spills and MODIS Fire Products in Near Real-Time". In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018, pp. 1–5
4. A. F. Scannapieco, L. Feetham, M. Estébanez Camarena, and N. Aouf. "Space-oriented Navigation Solutions with Integrated Sensor-Suite: The I3DS H2020 Project". In: *Proceedings of the 69th International Astronautical Congress*. Bremen, Germany, Oct. 2018
5. E. de Miguel, E. Prado, M. Estébanez, A. Martin, and M. Gonzalez. "Direct Geolocation of Satellite Images with the EO-CFI Libraries". In: *Living Planet Symposium, Proceedings of the Conference*. Ed. by L. Ouwehand. Prague, Czech Republic: ESA-SP Volume 740, May 2016, p. 410. isbn: 978-92-9221-305-3

6. Additionally, the work in this dissertation has been presented by the author in multiple national and international conferences, including the ESA Phi-week (ESRIN, Frascati, Italy, 2019 and online, 2020, video presentation: https://www.youtube.com/watch?v=oPMDqXpke8g); the AGU Fall Meeting (online, 2020), the EGU General Assembly (online, 2020 and in Vienna, Austria, 2022), the GEO Week (Accra, Ghana, 2022; **keynote presentation** during the *Young Women in GIS* panel discussion) and the EMS Annual Meeting (Bratislava, Slovakia, 2023).

3. CONTRIBUTIONS TO BOOKS

1. M. Estébanez Camarena. “Prediction of Science Fiction that Came True”. In: *Outer Space and Popular Culture*. Ed. by U. Editor. Springer, 2019, pp. 129–144
2. S. Anih, G. Badela, T. Campbell, G. Cromhout, M. Estébanez Camarena, L. W. Y. Feng, T. Hugbo, P. H. Khwambala, K. Konar, D. Lindgren, S. Madlanga, R. Maharaj, A. K. Nath, M. G. Ogungbuyi, Z. Pandey, E. Pieterse, and M. Tanner. “Space Resource Utilization: A View from an Emerging Space Faring Nation”. In: *Space Resource Utilization*. Ed. by A. Froehlich. Cham: Springer International Publishing, 2018. isbn: 978-3-319-66968-7

ACKNOWLEDGEMENTS

I have been lucky to have the support and guidance of many people in my PhD, to whom I want to express my heartfelt gratitude.

First and foremost I would like to thank my supervisors, that have accompanied me during the last six years. To Nick for giving me the chance to pursue this PhD and therefore a door to my career in providing climate services in Africa, I will be always grateful. Thank you for your inspiration. To Marie-Claire, for your unwavering direction, support and interesting conversations, about science and life. You both are great scientists and great people, and it has been a priviledge to work with you. To both, thank you for your patience and your continuous encouragement. I could not have done this without you.

To my Schools and Satellites team: Sandra, Jeff, Frank, Kwame, Mark, Halilu, and others. Thank you for helping make SaS a success. Those 2.5 years were my favorite part of my PhD. Thank you for sharing them with me. To all the SaS citizen scientists - teachers that involved their students and farmers – thank you for all your work and engagement. You were really the ambassadors of this adventure.

To Riccardo, thank you for all your guidance with Deep Learning! You have been a large part of this work. To Fabio, your enthusiasm in your research and in improving RainRunner and its interpretation gave me the push I needed to go on through a hard PhD time.

Thank you also to my friends in Ghana, that gave me a more personal view to this work. To Paakwesi and his family, thank you for giving me a home in Tamale. To Ziro, besides our friendship, thank you for showing me the traditional culture from the Dagomba people in Northern Ghana - including your drumming and dance classes after long days of work. To Kwame (again), who over the years became a good friend, for his continuous support and his friendship. That first trip to Tamale with you will remain a very sweet (though accidented) memory. You gave me a great introduction to Ghana. Thank you also to Bebs, Harriet and Dan for always make me feel home in Ghana.

To my friends from TU Delft, from the office, from the beers at PSOR and from the department – Juan, Jerom, Ties, Paul, Sergio, Boran, Jessica, Robert, Mario, Dengxiao, Juan Carlo, Luuk, Bas, Gijs, Julia, Judith, Tessel, and the rest. To my two paranymphs: David, it has been such a pleasure to work shoulder with shoulder with you during these years, in our PhDs and everywhere, I have gained another brother. Su, thanks for walking next to me during these years and for being always a constant. I really appreciate all of you.

Thanks to the TU Delft Global Initiative for creating a community of researchers for impact. I am proud of keeping flying the flag of “Science for the benefit of people. All people. Worldwide.”

I am lucky to continue in the same line of work after my PhD, and my professional path is already intertwined with that of some of these people. For others, I hope that our paths cross again. In any case, thank you for all we have shared these years.

To Weather Impact - Stefan, Paula, Joanne, Tamara, Joep, Lorenzo, Shashwat, Bob, Janina and Glenn, thank you for giving me the space and understanding I needed to finish this PhD. But most importantly, thank you for the chance to keep working for food and economic safety in Africa and the world. I could not have landed in a better place and with a better team.

Thanks to Peter Martinez and the SpaceLab in South Africa - special mention to Brooks - for being the turning point that took me here.

To my neighbors in the Pionierskwartier and my fellow activists and change-makers, thank you. Living and working with you strengthened my conviction that another world is possible, and that we can change the world. It is this determination that has taken me here.

To Hans and Mathilde, for your warm support and opening your doors to me, and for doing your best me Dutch - even if my busy schedule made it not always easy.

To my dance students and friends, dancing with you has given me a much needed balance and joy even through my PhD years. Thank you.

To my Delft friends outside of uni: Michelle, Gaby, Lau, Roge, Ted, Diana, Valeria, Rachele and Fran. Thank you for making my every day better. Thanks to my childhood friends - Joaquín, Gabi, Goso, Luis, Gómez and Miguel - for always being there, also in the distance, I carry you always with me.

Thanks also to everyone that I might not have mentioned but that have also been in my life these years.

To Randy, all the love, strength and peace that you give me have been essential to close my PhD. I love you and I am really grateful for you.

Lastly, thanks to my family - Mamá, Edu, Car, Maria José, Mariano, Maria José (hija) and Clara - for your love, support, encouragement and for always believing in me, and making sure I did too. You are the behind-the-scenes of all my successes.

And of course, let’s not forget my little Marli. She will not read this, but she is also key in it all.

Like Maya Angelou said, “I have had so many rainbows in my clouds”, and for that, I am thankful.

CURRICULUM VITæ

Mónica Estébanez Camarena

18-01-1992	Born in Madrid, Spain.

EDUCATION

2015	Bachelor of Science (BSc) Aerospace Engineering Technical University of Madrid (UPM), Spain
2017	Masters of Philosophy (MPhil) in Space Studies University of Cape Town (UCT), South Africa
2018	Masters of Science (MSc) by Research in Space Navigation University of Cranfield, United Kingdom

AWARDS

2017	Best Proposal Presented by an Individual for the 2017 ESA's Sentinel Small Sat (S^3)*Challenge*
2019	Funding recipient for the Schools and Satellites (SaS) project, from the Citizen Science Earth Observation Lab (CSEOL)