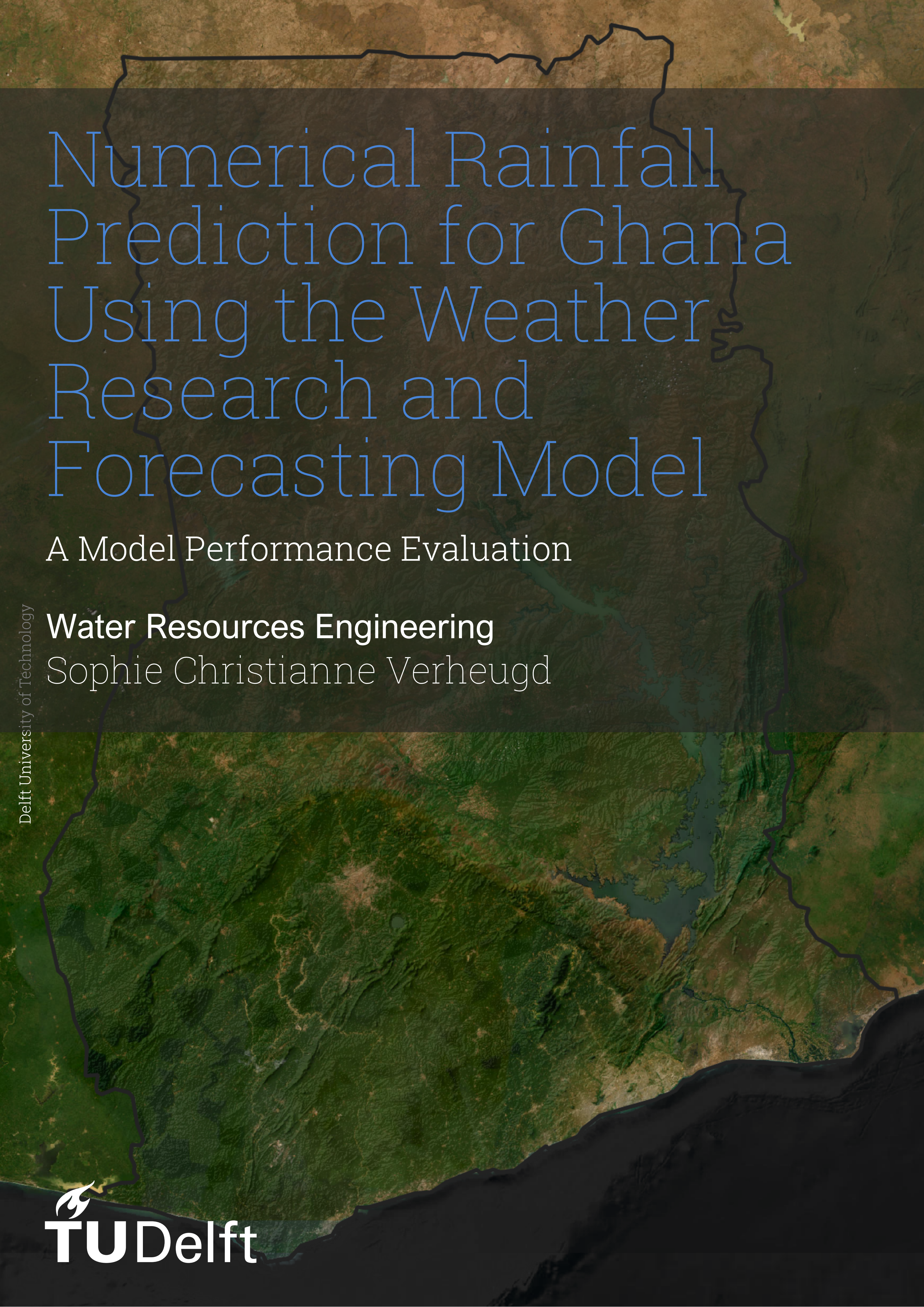


Numerical Rainfall Prediction for Ghana Using the Weather Research and Forecasting Model

A Model Performance Evaluation

Water Resources Engineering

Sophie Christianne Verheugd

Delft University of Technology

Numerical Rainfall Prediction for Ghana Using the Weather Research and Forecasting Model

A Model Performance Evaluation

by

Sophie Christianne Verheugd

Student Name	Student Number
Sophie Verheugd	4674553

First supervisor:	Prof.dr.ir N.C. van de Giesen
Second supervisor:	Prof.dr.ir R. Uijlenhoet
Third supervisor:	Dr. A.A. Nuijens
External supervisor:	Dr. S.Ligtenberg & MSc. L. Occelli
Project Duration:	March, 2024 - September, 2024
Faculty:	Faculty of Civil Engineering and Geosciences, Delft

Cover: World Imagery from the arcgismapserver (Modified)
Style: TU Delft Report Style, with modifications by Daan Zwaneveld

TUDelft

Preface

This thesis serves as the final project for my Master's in Environmental Engineering at the Civil Engineering Faculty of the Technical University in Delft. After six years of studying, I'm approaching the end of my academic journey. I'm grateful for the opportunity to dedicate my last six months to a project that I enjoyed and have borne much passion for. To finally being able to apply all the knowledge and skills I've acquired over the years to this final work has been incredibly fulfilling, and I hope it will have a positive impact.

The success of this research would not have been possible without the support of my colleagues at Weather Impact, where I completed my graduation internship. Working alongside you at the office in Amersfoort made the project progress much more smoothly. The coffee breaks and lunches were a refreshing change and I looked forward to coming to the office every day. Besides this I would like to express my special gratitude to Lorenzo for our weekly meetings, during which I could share my doubts and uncertainties about the research, which you took with great patience and understanding. The way you handled both my challenges and your own work was inspiring to me and consistently put me back on the right path, leaving me relieved and motivated after every meeting. In addition, I want to thank Stefan for the opportunity you gave me to work at the company. The little moments in which you asked how everything was going and if I was still enjoying it meant a lot for me so thank you for that.

Furthermore, I would like to thank my main supervisor, Nick, for the reassuring update meetings, and my other two supervisors, Remko and Louise, for their occasional help and input to my research. I would also like to express my thanks to Agostino Meroni, whose extensive knowledge in the field of WRF helped me a lot in defining my experiments.

Finally, I would like to thank my family, boyfriend and friends for their great support, listening ears, and very welcome distractions.

Sophie Christianne Verheugd
Delft, September 2024

Summary

Agricultural businesses in Ghana heavily depend on rain-fed farming, but they are facing increasing challenges due to climate change and population growth. Accurate weather forecasts providing useful information on the onset of the rainy season are therefore of utmost importance. Unfortunately, the country still lacks reliable weather information and precise forecasts.

This study aims to enhance and expand weather forecasting research in Ghana by running the regional Weather Research and Forecasting (WRF) model over Ghana and verifying its performance in forecasting the local onset of West Africa's annual rainy season. Three experimental setups were conducted. The first experimental setup tested four different planetary boundary layer (PBL) schemes. Results showed that the schemes performed differently across the various agro-ecological zones of Ghana, highlighting the influence of spatial context on the model's performance and the choice of PBL scheme.

Further analysis within a smaller domain in the Forest zone revealed promising insights into accurately capturing categorized precipitation. In the third and final experiment, the WRF model was compared to the ECMWF's operational forecast, demonstrating WRF's capability of detecting local variations in rainfall, including heavier precipitation amounts, which the ECMWF model was not capable of detecting. This comparison underscored the added value of the WRF model over global models.

Whether the WRF model can serve as an accurate weather prediction model to forecast the local onset of the rainy season in Ghana depends not only on its performance but also on the chosen definition of the onset. However, the results highlighted its qualities and robustness for multiple definitions, indicating a high potential for this purpose.

Contents

Preface i

Summary ii

1 Introduction 1
- 1.1 Problem Description . . . 1
- 1.2 Weather Prediction . . . 2
  - 1.2.1 Numerical Weather Prediction Models . . . 2
  - 1.2.2 Regional NWP Models in Africa . . . 2
  - 1.2.3 Regional NWP Models and the West African Monsoon . . . 2
- 1.3 Climatology of Ghana . . . 2
  - 1.3.1 The West African Monsoon (WAM) . . . 2
  - 1.3.2 Onset of the Rainy Season: Definitions . . . 3
- 1.4 Research Objective . . . 4
- 1.5 Structure of the Study . . . 4

2 Parametrization Schemes in WRF 6
- 2.1 Overview of WRF Parametrization Schemes . . . 6
- 2.2 Previous Studies . . . 7
- 2.3 Parametrization Schemes in this Study . . . 8

3 Experimental Setup 1 9
- 3.1 Model Configuration . . . 10
  - 3.1.1 Parametrization Schemes . . . 10
  - 3.1.2 Time and Spatial Domain . . . 10
- 3.2 Validation Method . . . 11
  - 3.2.1 Validation Data . . . 11
  - 3.2.2 Metrics . . . 12
- 3.3 Results . . . 14
  - 3.3.1 WRF Output . . . 14
  - 3.3.2 Point Validation . . . 16
  - 3.3.3 Summary . . . 21
- 3.4 Conclusion . . . 23

4 Experimental Setup 2 24
- 4.1 Model Configuration . . . 25
  - 4.1.1 Time and Spatial Domain . . . 25
- 4.2 Validation Method . . . 25
  - 4.2.1 SEEPS . . . 25
  - 4.2.2 Interpolation of TAHMO Data . . . 27
- 4.3 Results . . . 29
  - 4.3.1 Accumulated Precipitation . . . 29
  - 4.3.2 Daily Precipitation . . . 29
  - 4.3.3 Summary . . . 33
- 4.4 Conclusion . . . 33

5 Experimental Setup 3 35
- 5.1 Model Configuration . . . 36
  - 5.1.1 Time and Spatial Domain . . . 36
- 5.2 Validation Method . . . 37
  - 5.2.1 Validation Data . . . 37

5.2.2 Metrics . . . 37
5.3 Results . . . 39
5.3.1 Probability Density Function . . . 39
5.3.2 Heatmap of Daily Precipitation . . . 39
5.3.3 SEEPS . . . 40
5.3.4 Final Metric . . . 42
5.3.5 Summary . . . 42
5.4 Conclusion . . . 43

6 Discussion 44

7 Conclusion and Recommendations 46
7.0.1 Conclusion . . . 46
7.0.2 Recommendations . . . 47

References 48

A Source Code 51
A.1 Point Validation . . . 51
A.2 IDW Interpolation Function . . . 53
A.3 Automating WRF . . . 53
A.3.1 Bash script . . . 53
A.4 SEEPS . . . 55

B Figures 58
B.1 IDW Interpolation Experimental Setup 2 . . . 58
B.2 SEEPS Experimental Setup 2 . . . 59
B.3 Extra Results Experimental Setup 2 . . . 61

List of Figures

1.1 Visualization of the WAM: average monthly precipitation and wind data ERA5 from the years 2000 to 2023. 3
1.2 Methodology 5

2.1 Interactions between the different WRF physics schemes, copied from the WRF users guide [15] 6

3.1 Sketch Experimental Setup 1 9
3.2 Spatial domain Experimental Setup 1 11
3.3 Validation data Experimental Setup 1 12
3.4 Accumulated precipitation at leading time D+10 for the model time period of 2018-04-26 to 2018-05-06 14
3.5 Accumulated precipitation at leading time D+10 for the model time period of 2020-04-11 to 2020-04-21 14
3.6 General statistics of all ten datasets. 15
3.7 Accumulated precipitation (mm) over time, averaged over each zone, for all 8 simulations and observations. 16
3.8 Normalized Mean Absolute Error (MAE) of daily precipitation (mm/d) for the 2018 simulations at the selected TAHMO locations. 17
3.9 Normalized Mean Absolute Error (MAE) of daily precipitation (mm/d) for the 2020 simulations at the selected TAHMO locations. 18
3.10 TA00011 and TA00012 precipitation time series from 2018-04-27 to 2018-05-05 18
3.11 Mean absolute error [mm/d] of daily rainfall against Time and Space Correlation [-] for the 2018 simulations. Blue = YSU, Green= MYJ, Red = ACM2, Purple = MYNN2. . . . 19
3.12 Mean absolute error [mm/d] of daily rainfall against Time and Space Correlation [-] for the 2020 simulations. Blue = YSU, Green= MYJ, Red = ACM2, Purple = MYNN2. . . . 19

4.1 Sketch Experimental Setup 2 24
4.2 Model domain and validation domain for the second set of experiments displayed on a Digital Elevation Model (DEM) map, downloaded from [27] 26
4.3 Cumulative distribution for 24-h precipitation at station TA00047 for the period of 2018-2023. 27
4.4 Interpolated TAHMO data for 5-day Accumulated Precipitation (mm), using the IDW method. 28
4.5 The spatial distribution of the Mean Bias Error (MBE) of the 10 simulations for each PBL scheme, along with the accumulated precipitation timeseries averaged over the whole validation domain. 29
4.6 SEEPS, with a t1 value of 0.25 mm. The numbers inside each grid cell depict the rainfall amount, ranging from 0 (no rainfall on all simulation days) to 100 (heavy rainfall on all simulation days). The spatial average is the mean SEEPS skill score for the entire validation domain. 30
4.7 The MAE of YSU over the whole validation domain, where all observed and modeled daily rainfall above threshold ’T’ is masked. 32
4.8 The MAE with all observed and modeled daily rainfall above 9 mm/d masked for PBL schemes ACM2, MYNN2 and MYJ. 32
4.9 The mean absolute error (MAE) and mean bias error (MBE) of all four PBL schemes, averaged across the validation domain, and plotted against the threshold values ’T’ where daily rainfall above is masked. 32

5.1 Sketch Experimental Setup 3 . . . 35
5.2 Model domain and validation domain for the final set of experiments displayed on a Digital Elevation Model (DEM) map, downloaded from [27]. . . . 36
5.3 Logarithmic-scale Probability Density Function (PDF) of daily precipitation (mm/day) for all five ECMWF and WRF simulations, with observed data over the same time periods. 39
5.4 Heatmap of the aggregated daily precipitation of ECMWF and WRF compared to the observed daily rainfall. . . . 40
5.5 SEEPS skill score for different leading times (D+), with associated error fractions and spatial averages of both models. . . . 41
5.6 Final Metric: All normalized confusion matrix variables plotted against 6-day average rainfall (mm/d) threshold values. . . . 42

B.1 IDW-interpolation output of daily precipitation (mm/d) of TAHMO for experimental setup 2 58
B.2 Precipitation time series at the four selected TAHMO stations.The gray band shows periods with no data values. . . . 59
B.3 Cumulative distribution for 24-h precipitation at station TA00047 for the period of 2018-2023 with t1=1mm . . . 60
B.4 SEEPS, with a t1 value of 1 mm. The numbers inside each grid cell depict the rainfall amount, ranging from 0 (no rainfall on all simulation days) to 100 (heavy rainfall on all simulation days). The spatial average is the mean SEEPS skill score for the entire validation domain. . . . 60

List of Tables

3.1 Model Configuration Scheme 10
3.2 2018 Simulations, Statistical Summary 21
3.3 2020 Simulations, Statistical Summary 21
3.4 Experimental Setup 1: Recommended usage of each PBL scheme per zone. 22

4.1 SEEPS Error Fractions for Different Configurations and Precipitation Categories 31

B.1 Evolution of the MBE and MAE against threshold values, with all daily rainfall above the threshold masked. 61

1

Introduction

1.1. Problem Description

Economic growth in West African countries relies heavily on agriculture. According to a study by the Organization for Economic Co-operation and Development (OECD), the food economy accounts for 66% of total employment, with 78% of this economy rooted in agriculture [1]. Despite its critical role, agriculture in these countries faces increasing challenges. Population growth demands higher production rates, which businesses struggle to meet. Additionally, heavier droughts due to climate change, declining soil fertility, and over dependence on rain-fed farming—combined with a lack of technology and weather forecasting information—add significant uncertainty to the sector.

Ghana, one of the countries in West Africa, is facing these same challenges. While its agricultural employment may be lower than that of neighboring countries like Burkina Faso or Togo, agriculture remains the largest industry in Ghana. The 2017/18 Ghana Census of Agriculture (National Report) showed that agriculture contributed more than one-fifth (21.2%) of the GDP in 2017 [2]. Recognizing agriculture's importance to national growth, Ghana has included it in its development agenda, aiming to improve food and nutrition security, drive economic growth, and reduce poverty.

However, many challenges must be addressed to achieve these goals. As previously mentioned, agricultural businesses in West Africa rely heavily on rain-fed farming. In Ghana, only 0.3% of farmland is irrigated [3], making rainfall a crucial factor. The rainy season, known as the West African Monsoon (WAM), provides the majority of the region's total rainfall, typically occurring between late February and September. The timing of planting must align with this onset to optimize yield. Therefore, information on the onset of the rainy season (ORS) and weather forecasts is critical.

Unfortunately, African countries still suffer from a lack of reliable weather information, largely due to the extremely low number of weather stations across the continent. The Trans-African Hydro-Meteorological Observatory (TAHMO) project [4] aims to address this gap by developing a network of 20,000 weather stations across Sub-Saharan Africa. Since the project's inception, over 500 stations have been installed, including in Ghana. This development opens new opportunities for climate research and numerical weather prediction modeling. The newly available data provide higher-resolution ground-based measurements, which can be used to train and validate weather models, improving forecasting accuracy. These advancements in weather forecasting have the potential to mitigate future agricultural challenges and enhance businesses in West Africa. This study aims to enhance and expand weather forecasting research in Ghana, with the goal of providing more accurate predictions for the rainy season onset.

1.2. Weather Prediction

1.2.1. Numerical Weather Prediction Models

Numerical weather prediction (NWP) models are crucial tools to obtain future atmospheric conditions and are globally used in operational forecasting [5]. With physical equations, they can predict the weather based on current weather conditions. Global models like the European Centre for Medium-Range Weather Forecasts (ECMWF) and the The Global Forecast System of the National Centers for Environmental Prediction (NCEP-GFS), which are some of the most used resources in weather and climate forecasting and provide increasingly available data around the world, are also used in West African countries. Performance assessment on these models has already been done in some previous research [6, 7, 8], although validation of these regions is still relatively rare compared to other parts of the world. This is because ground validation sites and access to the data are limited in Africa, which is needed to make accurate comparisons. The studies showed biases and spatial variability between the models, but overall good performances on larger timescales. However, on shorter timescales, the models perform quite poorly in predicting precipitation. This indicates room for improvement or implementation of more detailed, high-resolution, regional models.

1.2.2. Regional NWP Models in Africa

In contrast to global models that provide forecasts for the entire earth, regional NWP models focus on a specific region or domain. With their higher resolution grid-spacing they simulate atmospheric conditions on a more local scale and in this way can provide more fine-scale details. Short to medium-range forecasts for localized weather events can best be executed with them. Unfortunately, the current situation in West-African countries shows very limited availability in regional numerical weather prediction models [5, 6, 7] which is a problem that spreads out over the whole continent. A previous study [5] surveyed 15 Southern African Development Community (SADC) countries to assess their capability in performing NWP and climate modeling activities. The results showed a great disparity in terms of numerical weather predictions, with only 60% either testing or operationally running a numerical weather prediction model. The lack of NWP availability is mostly due to some challenges associated with good model performance: to obtain accurate results, the models require high-resolution observational weather data and a lot of computational power for performing the numerical calculations, which most of the (West)-African countries do not have [4, 5, 6, 7]. Of this deficient quantity of models, the Weather Research and Forecasting (WRF) model is mostly used throughout Africa and its performance is already assessed before [9, 10, 11].

1.2.3. Regional NWP Models and the West African Monsoon

Several researchers, by testing various parametrization schemes in WRF, have tried to improve the performance of simulating the West African Monsoon (WAM) and thus the rainy season [9, 10, 11]. The growing number of scientific and modeling communities that have been trying to adequately understand the WAM have provided some important new insights and advancements in the optimization of the regional climate model. However, with the WAM being a complex large-scale system that is influenced by many multi-scale atmospheric components [9], it makes it very challenging to accurately simulate its variable spatial and temporal extent. This leaves room for more optimization research in the WRF model, building on already existing findings on the sensitivity and performance of different parametrization schemes. In this way, our knowledge of WAM and its underlying processes will be expanded which in turn gives new insights on the onset of the rainy season; the specific focus of this study.

1.3. Climatology of Ghana

1.3.1. The West African Monsoon (WAM)

Ghana has a tropical monsoonal climate, characterized by distinct dry and wet seasons associated with the large-scale system known as the West African Monsoon (WAM). The WAM involves the migration of a zonal rainfall band from the Guinea Coast to the dry northern Sahel region and back [12]. This tropical rain belt, also known as the Intertropical Convergence Zone (ITCZ), is marked by intense convection and heavy precipitation.

The migration of the ITCZ results from the rapid heating of the land surface compared to the adjacent ocean in early summer. This creates a north-south pressure gradient at low levels, with high pressure near the coast and low pressure toward the northern Sahara, causing winds to transmit from south to north (Figure 1.1). At higher levels, the situation is reversed: warm air rises into the atmosphere in the northern part, creating high pressure there and low pressure near the coast, causing winds to move away from the land. This results in a circulation cell that brings moist winds from the coast towards the continent at lower levels and away from the continent at higher levels [12].

At the front where these moisture-laden winds meet the dry 'Harmattan' winds from the north, an inter-tropical front (ITF) forms. As the heat low deepens, this front penetrates even further north, and the ITCZ, located about 400 km south of the ITF, follows its migration [13]. Due to the up and downward movement of this rain belt, Ghana experiences two distinct rainfall regimes: a bi-modal pattern in the south (coastal and forest zones), experiencing two distinct rainy seasons within a year, and a uni-modal pattern in the transition and Savannah zones [14].

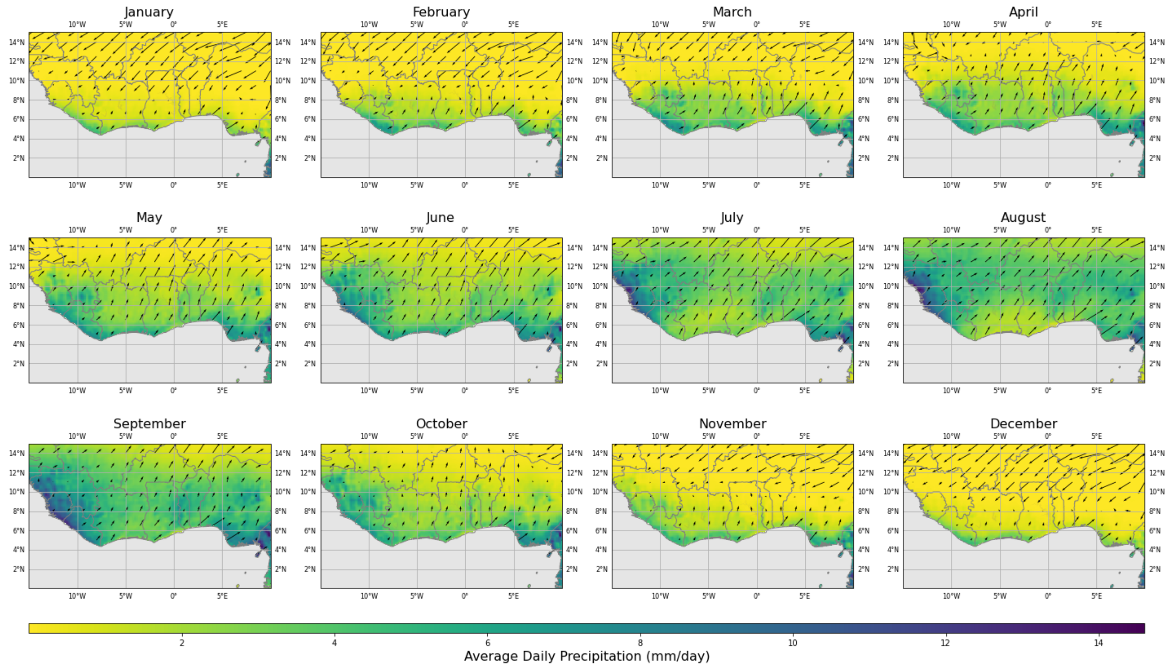


Figure 1.1: Visualization of the WAM: average daily precipitation and wind data ERA5 from the years 2000 to 2023.

1.3.2. Onset of the Rainy Season: Definitions

While the driving factors of the WAM and its progression over West Africa are generally understood, the factors defining the onset of the monsoon are more complex and open to interpretation. A study by Fitzpatrick et al. [13] defined 20 WAM onset definitions. The onset is classified by local or regional definitions, depending on different parameters across various spatial and temporal scales. These definitions are relevant for different types of end users and serve various purposes. The regional onset of the rainy season is often marked by a sudden northward shift or jump of the ITCZ from the coastal region to the Sahel. This definition predominantly focuses on large-scale dynamics over the West African region, which is beyond the scope of a single nation. In contrast, local onset definitions are constrained to single grid cells or ground-based measurements, making them more usable for local stakeholders, such as farmers. Notably, the research showed that the timing of the local rainy season onset and the ITCZ shift, or regional onset, are weakly correlated. This underscores the importance of distinguishing between the two phenomena and focusing on the appropriate one based on your specific objectives.

The data used to indicate the regional onset varies from precipitation to outgoing longwave radiation, zonal winds, and mean sea level pressure, mostly derived from satellite data and weather models. For local onset definitions, only local precipitation amounts meeting a specific threshold are used, which can be modified depending on the stakeholder's needs (e.g., different crops). The reliance on rainfall at individual locations results in low spatial and temporal consistency of the onset, as convective precipitation is characterized by localized intense rainfall. Given the agronomic focus of this research, the localized onset definitions are preferred and most suitable for the associated stakeholders. Therefore, considering the various local onset definitions proposed by Fitzpatrick et al., which all use thresholds of precipitation within a certain number of days, accurate prediction of rainfall becomes critically important.

1.4. Research Objective

This research will extend the analyses on the performance of the Weather Research and Forecasting (WRF) model in Ghana, by assessing the outputs of different parametrization schemes. It aims to accurately predict the precipitation in Ghana on relatively short timescales, from which the output can be used to indicate the onset of the rainy season and in this way help agricultural businesses in their decision-making processes. The following research question is defined:

"What is the performance of the Weather Research and Forecasting (WRF) model in forecasting the local onset of West Africa's annual rainy season in Ghana?"

To address this question, a variety of model experiments will be conducted. Initially, design decisions and filtering will be applied to the model to gain an overall understanding of its performance and determine the appropriate calibration. Ultimately, a comparison with ECMWF IFS data will be made to demonstrate the potential added value of WRF. The following sub-questions are formulated to guide the research:

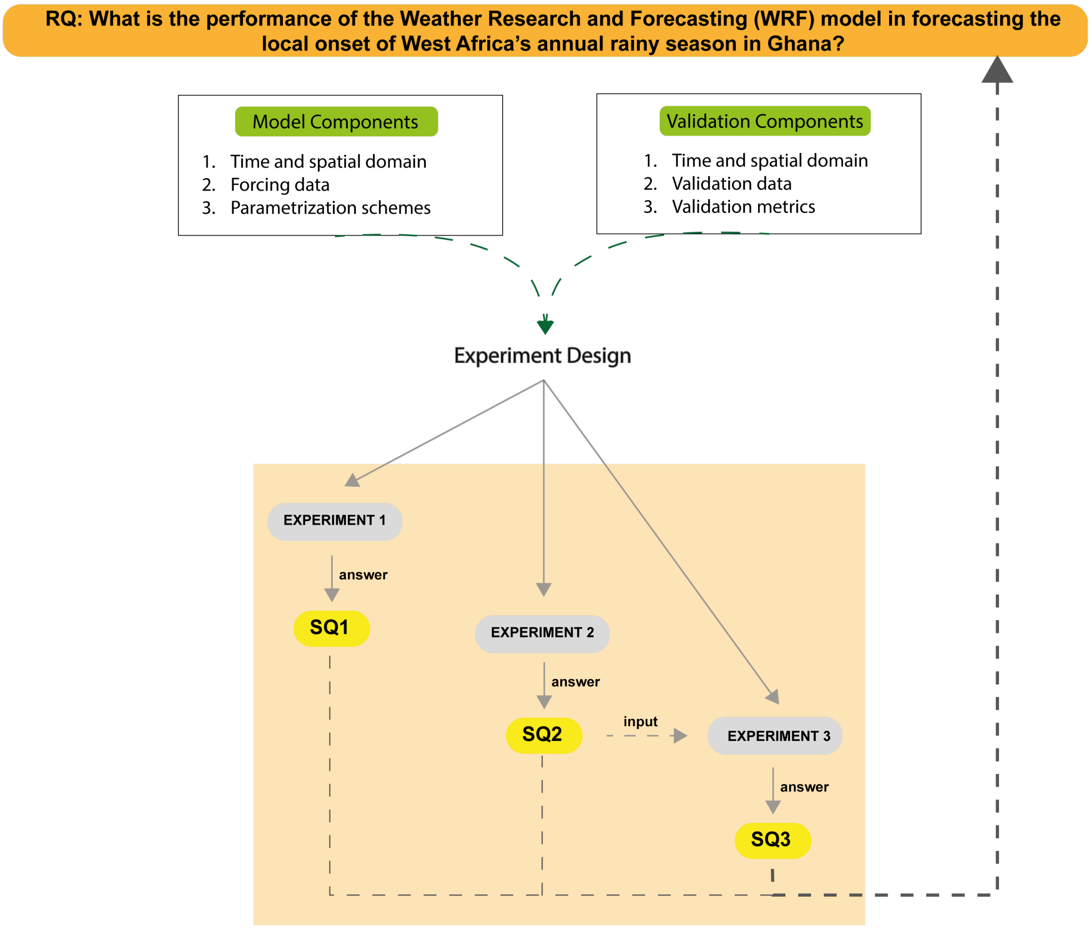
1. How do the selected combinations of physics schemes of the WRF model perform in forecasting precipitation around the onset of the rainy season in: a. the whole country of Ghana b. the different agro-ecological zones?
2. Which model configuration is the optimal choice considering the context of this study?
3. How does the selected model configuration perform compared to the precipitation forecasts of ECMWF IFS?

1.5. Structure of the Study

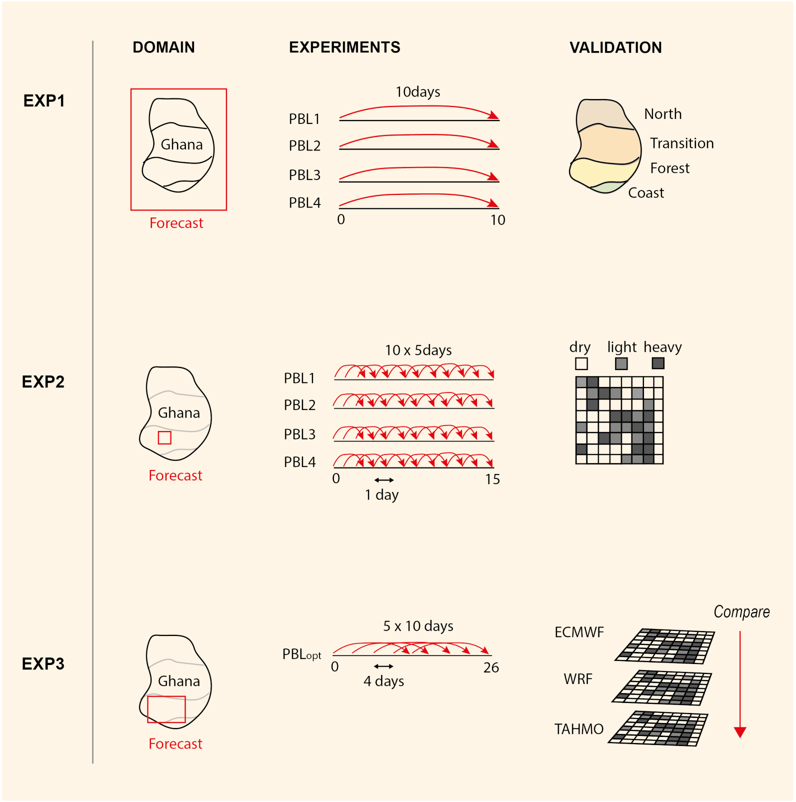
This study involves three experimental setups designed to test the performance of the WRF model. Each setup corresponds to a specific research sub-question, with Experimental Setup 1 addressing Sub-question 1 (SQ1), Experimental Setup 2 addressing Sub-question 2 (SQ2), and Experimental Setup 3 addressing Sub-question 3 (SQ3). The methodology flowchart, depicted in Figure 1.2a, illustrates the overall research process and how each setup contributes to answering the main research question. As input for the model components, a literature review on WRF will be conducted first. After that, each experimental setup is detailed in its own chapter, which includes:

- Description of the model configuration, including temporal and spatial domains
- Validation metrics used
- Results obtained
- A summary and conclusion that link the findings to the respective sub-question

General doubts and uncertainties encountered during the research will be highlighted in the ”Discussion” section. Finally, the research will be summarized in the ”Conclusion and Recommendations” chapter, including an answer to the main research question. It will also provide general suggestions and ideas for further research, and highlight potential operational applications of the model.



(a) Flowchart Methodology



(b) Sketches of the Experimental Setups

Figure 1.2: Methodology

2

Parametrization Schemes in WRF

2.1. Overview of WRF Parametrization Schemes

WRF-ARW is built out of multiple physics schemes responsible for different components of the physical processes in the Earth's atmosphere, interacting with each other during simulation [15]. These parametrization schemes are methods used to represent processes that occur on smaller scales than the model's grid resolution and therefore cannot be directly resolved. These so-called sub-grid processes, including cumulus convection, boundary layer turbulence, land surface interaction, and cloud microphysics, are represented in the model by empirical or theoretical relationships involving variables that the model can resolve. In this way, the smaller processes are estimated and provide additional tendencies for the resolved variables (which are at the grid scale or larger). The interactions between the schemes are visualized in the figure below, coming from the WRF user guide.

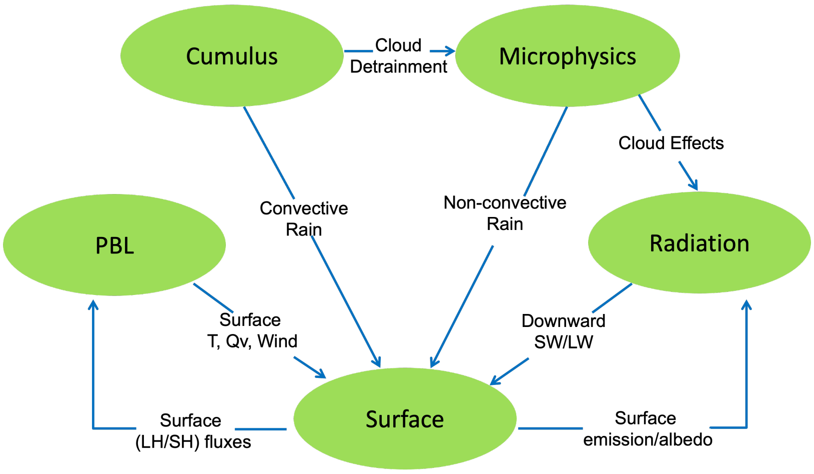


Figure 2.1: Interactions between the different WRF physics schemes, copied from the WRF users guide [15].

Various parametrization techniques have been developed and applied in atmospheric and climate modeling, each with its strengths and weaknesses. Sensitivity and performance analyses on these parametrization schemes in the WRF model in West or other parts of Africa, have been increasingly studied throughout the last couple of years [6, 7, 9, 16, 17, 18, 19] showing high dependency on the choice of implemented physics scheme. These studies form the literary basis for this particular study which is focused on Ghana only. A brief summary of three of these studies is provided below.

2.2. Previous Studies

Gbode et al. [9] investigated the sensitivity of the WRF model to numerous physical schemes, focusing on the large-scale simulation of the WAM. The aim was to identify optimal combinations for long-term regional climate simulation over the monsoon region. The applied domain covered the entire West African region with a grid resolution of 20 km, and the initial and lateral boundary conditions were taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim). For the time simulation, a 2-month regime of the WAM from August to September 2007 (a 'normal' monsoon year) was used, during which the WAM was fully developed in August. A total of 27 runs were produced, testing 2 Microphysic schemes (MP), 6 Cumulus (CU) Convection schemes, and 3 Planetary Boundary Layer (PBL) schemes (Figure 2.1). These three parametrization schemes play an important role in atmospheric moisture and heat distribution and thus are critical factors in simulating and predicting the WAM dynamical system. The results showed noticeable spatial systematic errors in all simulations. However, combinations with nTDK, BMJ, and nSAS (CU) showed low absolute errors, and high correlation statistics. Simulations with KF and the old GF (CU) performed rather poorly. This contradicts the research of Flaounas et al. [17], who concluded that combinations with KF, notably combined with the PBL scheme MYJ, seemed best to represent the WAM. The results also showed that, in general, the model simulated the diurnal cycles of surface temperature more adequately than those of precipitation, suggesting that any form of evaluation is subjective to the variable of interest.

The study of Agyeman et al. [16] also focused on long-term seasonal precipitation simulation by WRF, but on Ghana specifically. The goal was to determine a suitable combination of physics schemes for simulating seasonal precipitation over Ghana, taking into consideration timing, location, and evolution. The study used ERA-interim as forcing data and simulated 8 months (from April to November) for one wet year (2008) and one dry year (2001). Using double-nesting, the outer domain covered the whole of West Africa with a resolution of 50km, while the inner domain covered Ghana with a resolution of 10km. Again the Microphysics, Cumulus and Planetary Boundary Layer schemes were tested, containing: WSM6 & WSM5 (MP), BMJ & GD (CU), YSU & ACM2 (PBL) . The performance was evaluated looking at spatial and temporal evaluation statistics over Ghana and its four agro-ecological zones. The study showed clear differences in the performance of the parametrization combinations between the different zones (some for instance performing very well in the north, but bad in the south), highlighting their sensitivity for different geographical and atmospheric settings. As an example, combinations with ACM2 provided good estimates for the western part of the transitional zone and parts of the forest zone (unlike YSU), while some of these same configurations underestimated rain for the northern zone. As a whole, the schemes overestimated precipitation for the coastal zone and underestimated in the northern zones. The combination of WSM6 (MP), GD (CU) and ACM2 (PBL) simulated the best temporal and spatial patterns with the least bias for both years and thus was recommended for Ghana.

Meroni et al. [19] studied the sensitivity of the Weather Research and Forecasting (WRF) model on a more localized scale by simulating three heavy rainfall events in Sub-Saharan Africa. The research focused on the influence of the Microphysical (MP) and Planetary Boundary Layer (PBL) schemes. It was found that in terms of forecasting heavy rainfall location and rainfall structure, the PBL scheme has a greater impact than the MP scheme, which is in line with earlier findings by Flaounas et al. (2011). For the initial conditions and 3-hourly boundary conditions, the Integrated Forecast System (IFS) of ECMWF at a 0.125° grid spacing was used. Three two-way nested domains were established for each case study, with grid spacings of 13.5, 4.5, and 1.5 km. These case studies were conducted in Kenya, Uganda, and South Africa. Validation was performed using ground-based sensors and satellite estimates, specifically IMERG and CMORPH, which are considered reliable satellite products for Sub-Saharan Africa [20]. The PBL schemes tested were YSU, MYJ, and ACM2, while the MP schemes tested were WSM6 and THOM. The choices were based on the results of previous studies, as well as the WRF tropical suite, which is a combination of schemes recommended in the WRF model community to use properly in tropical environments. The WRF tropical suite includes the WSM6 for the MP, NTD for CU, RRTMG for longwave and shortwave radiation (RAD), YSU for PBL, RMM5 for the Surface Layer (SL), and Unified-NOAH as the Land Surface Model (LSM).

The results showed equally well performances between the WSM6 and THOM scheme, with slightly better results from WSM6, as THOM was found to systematically overestimate. However, there were notable differences in the performance outcomes of the PBL schemes. The MYJ scheme showed poor capabilities in simulating rainfall compared to YSU and ACM2. Meroni explained that MYJ is the only local PBL scheme tested, meaning that it does not explicitly include the vertical transport by large eddies which is an important factor in convective rainfall dynamics, thus resulted in large biases in the simulations. Instead, YSU and ACM2 are non-local schemes that do implement vertical transport by large eddies, and therefore produce more qualified simulation outcomes. The recommendation of ACM2 is in line with the suggestion made by Agyeman et al. [16].

Most of the former research kept their radiation and surface physics correspondingly constant. Meroni [19], Flaounas [17], and Gbode [9] et al. all kept their long wave radiation scheme on the Rapid Radiative Transfer Model or newer version (RRTGM), also confirmed by Noble et al. [18]. As for the land surface model, the unified Noah model is most commonly used [9, 16, 19, 18] in combination with the MM5 surface physics scheme [19, 16]. Supported by previous research and because of their lower impact, these schemes will also be held constant in this study.

2.3. Parametrization Schemes in this Study

The selection of parametrization schemes for this study is based on the literature review presented above. It was chosen to focus only on the PBL schemes and keep the other parametrizations constant. This decision is informed by previous studies indicating that some CU and MP schemes consistently outperform others, while the performance of PBL schemes varies significantly, especially between different landscapes. The New Tiedtke (CU) and WSM6 (MP) schemes have consistently demonstrated good performance in many reviewed studies [9, 16, 21]. Tiedtke is also used in other globally recognized weather models, such as the Integrated Forecasting System (IFS), developed by ECMWF, and the Icosahedral Nonhydrostatic (ICON) model, developed by the Max Planck Institute for Meteorology and the German Weather Service [7].

The four PBL schemes analyzed in this study are ACM2, MYNN2, YSU and MYJ. A more detailed description of these schemes, along with an indication of which PBL is applied in each experimental setup, is provided in the following sections.

3

Experimental Setup 1

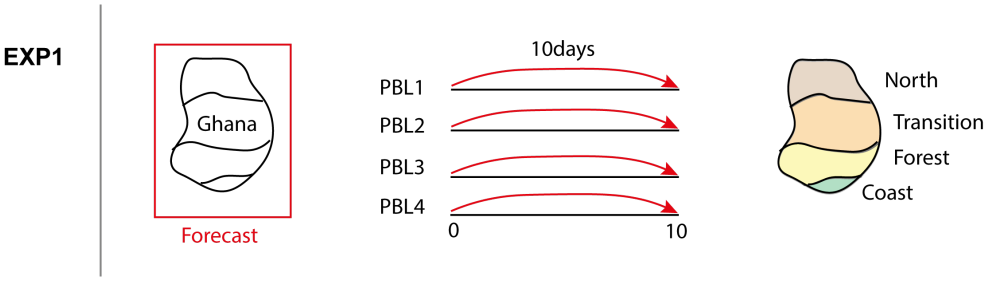


Figure 3.1: Sketch Experimental Setup 1

This chapter addresses sub-question 1 of this research: *"How do the selected combinations of physics schemes of the WRF model perform in forecasting precipitation around the onset of the rainy season in: a. the whole country of Ghana b. the different agro-ecological zones?"* This will be done by testing various WRF model configurations across the entire country of Ghana. By evaluating the different outputs using the provided validation metrics, both for the specific zones as for the country as a whole, we can obtain an initial indication of the model's performance.

3.1. Model Configuration

3.1.1. Parametrization Schemes

As mentioned previously, this study focuses on four distinct PBL schemes: ACM2, MYNN2, YSU and MYJ. These were selected based on a thorough analysis of literature, filtering out poorly performing schemes and focusing on those that presented good but varying results. Table 3.1 provides an overview of all the parametrization schemes used in this study. All PBL schemes are tested with the other parametrization schemes held constant, except for the Surface Layer scheme in combination with MYJ as the PBL scheme, since MYJ is not a possible combination with MM5.

Table 3.1: Model Configuration Scheme

	Microphysics (MP)	Cumulus Convection (CU)	Radiation (RD)	Land Surface (LS)	Surface Layer (SL)	Planetary B.Layer (PBL)
1	WSM6	nTDK	RRTMG	UN	old MM5	ACM2
2	WSM6	nTDK	RRTMG	UN	old MM5	MYNN2
3	WSM6	nTDK	RRTMG	UN	old MM5	YSU
4	WSM6	nTDK	RRTMG	UN	ES	MYJ

WSM6 = WSM6; nTDK= new Tiedtke; RRTMG = Rapid Radiative Transfer Model; UN = Unified Noah; old MM5 = old MM5 Monin-Obukhov; ES = Eta Similarity; ACM2 = ACM2; MYNN2 = Mellor-Yamada-Nakanishi and Niino Level 2.5; YSU = YSU; MYJ = Mellor-Yamada-Janjic

All PBL schemes are briefly described below. A more advanced description of the schemes and its interaction with the WRF model can be found in the technical note of Skamarock et al. [22].

The ACM2 scheme [23] is a first-order hybrid scheme that combines non-local upward mixing with local downward mixing, allowing it to represent both supergrid-scale and subgrid-scale components of turbulent transport. This scheme is an enhancement of the explicit non-local ACM1 scheme. Using a first-order eddy-diffusion component enhances its effectiveness in modeling local mixing processes.

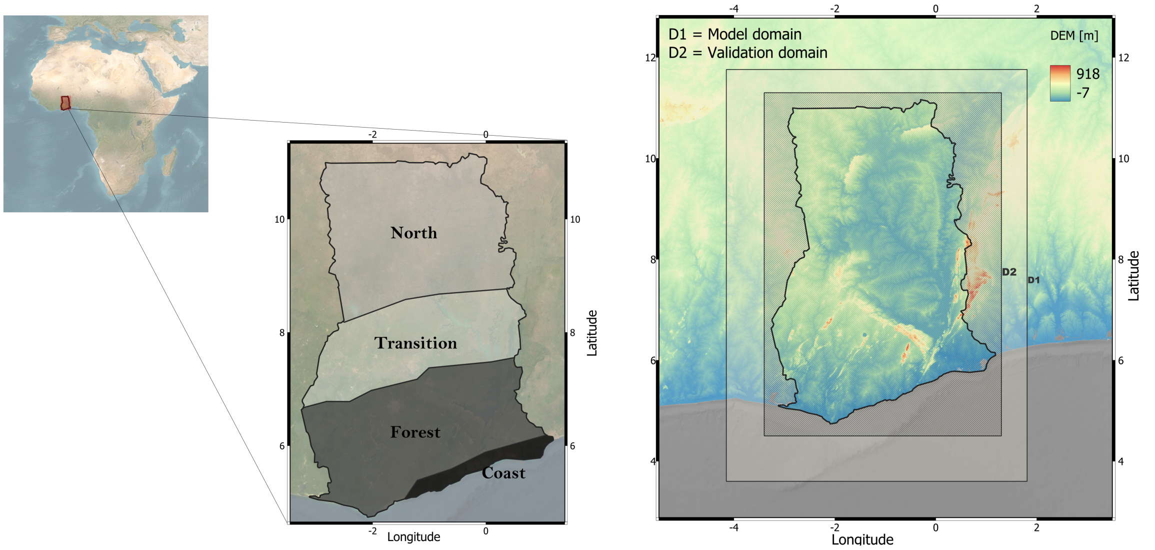
The MYNN2 scheme [24] is an improved version of the Mellor-Yamada 1.5-order local parameterization scheme (described below), meaning it also makes use of the turbulent kinetic energy (TKE) equation to parameterize turbulence. It enhances numerical stability and accurately represents turbulence dynamics. The scheme predicts subgrid TKE terms, which are essential for capturing the effects of unresolved turbulence on larger-scale atmospheric processes.

The YSU scheme [25], which is one of the most widely used PBL schemes in the WRF community, is a non-local first-order scheme. It uses a counter-gradient term in the eddy-diffusion equation, which accounts for the transport of heat and momentum by large eddies. The top of the boundary layer is defined using a critical bulk Richardson number.

The MYJ scheme [26] is an 1.5-order scheme with local vertical mixing and is particularly notable for its use of turbulent kinetic energy (TKE) to represent turbulent processes. As a local scheme MYJ only takes local gradients into account and does not account for the influence of larger eddies, making it particularly effective in stable and weakly unstable boundary layers.

3.1.2. Time and Spatial Domain

The model domain is focused on Ghana and spans 4.21°W to 1.86°E longitude and 3.53°N to 11.87°N latitude with a resolution of 9 x 9 km. The initial and boundary conditions are derived from ECMWF IFS operational analysis data with a grid sizing of 0.125 degrees. The boundary conditions are applied every 6 hours. While WRF is designed for high-resolution weather forecasting, the reduced resolution of 9km is applied uniformly across the entire domain. To account for potential large gradients, a relaxation zone is utilized. By running the model for a larger domain than the validation domain, with boundaries at 3.4°W to 1.3°E longitude and 4.5°N to 11.3°N latitude, and nudging the boundary conditions gradually, the relaxation zone helps to mitigate these gradients (Figure 3.2b). This zone is on average 10 grid cells wide. As shown in the map, the decision was made to expand the buffer zone in the southwest. In other words, the validation domain is positioned downstream with respect to the



(a) Ghana in the context of Africa and it's agro-economical zones.

(b) Model domain and validation domain for the first set of experiments displayed on a Digital Elevation Model (DEM) map, downloaded from [27].

Figure 3.2: Spatial domain Experimental Setup 1

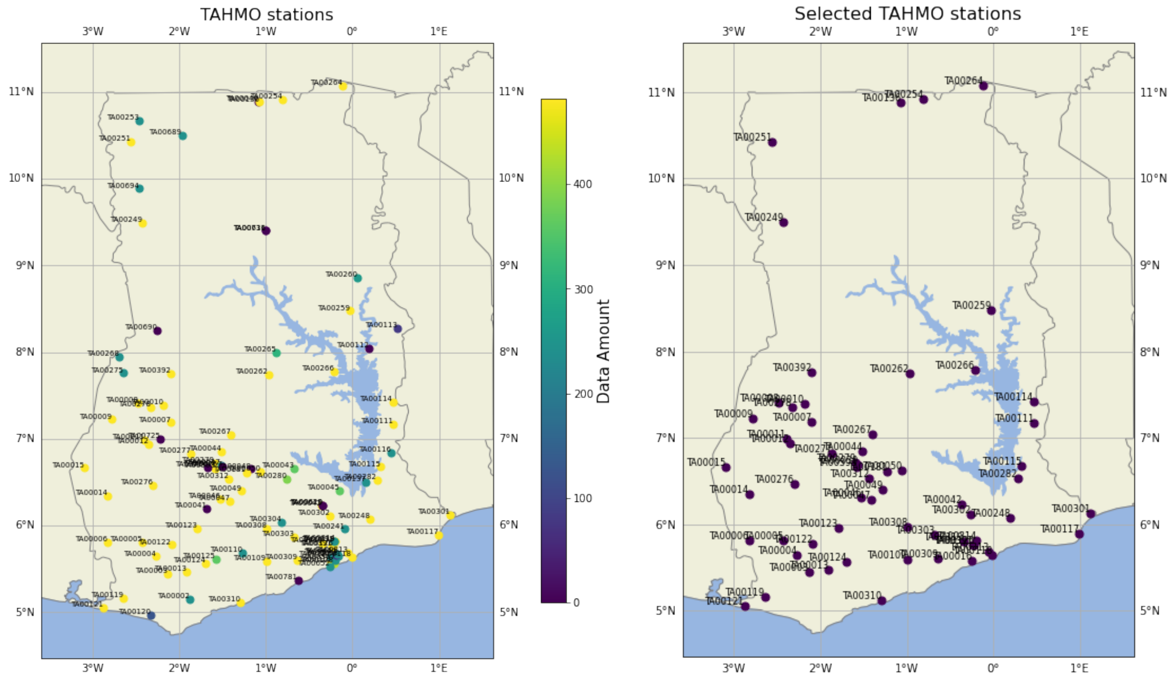
lower-level climatological wind, originating from south-west direction during the monsoon period [17]. Since the weather systems are propagating from south-west to northeast, extending the buffer zone upstream allows the systems to develop sufficiently before reaching the area of interest.

The validation domain focuses on Ghana as a whole and its different agro-ecological zones defined by the Ghana Meteorological Agency (GMet) classification: the North, Transition, Forest, and Coastal zone (Figure 3.2a) [14]. These zones are specified by their distinct meteorological factors and climatic conditions. Their divergent characteristics can have great effect on the model's performance, so the decision was made to validate them separately. The model's time domain was chosen based on the regional onset analyses coming from L.Occelli's study [28]. The model will be run for the onset dates of two specific years, 2018 and 2020. One of these years, 2020, has an average onset date that aligns with the region's climatological onset date, while the other year, 2018, deviates from this norm. To minimize the amount of runs, the onset date of the northern area was considered, as this ensures that the entire country experiences rainfall. As mentioned in the introduction, local onset definitions are primarily based on a few consecutive days of rainfall (at least 6 days). Given the objective to generate practical tools for local onset prediction, and the need for the model to eventually run operationally, it was decided to incorporate a 10-day forecast for this first set of experiments. This range is restricted by the use of the ECMWF 10-day operational forecast data, which cannot be extended. Taking these considerations into account, the model is run for the periods of April 26, 2018, to May 6, 2018, and April 11, 2020, to April 21, 2020. Throughout the rest of the report, these two simulations will be referred to simply as Simulations 2018 and 2020.

3.2. Validation Method

3.2.1. Validation Data

Concerning the validation dataset, the ground-based Trans-African Hydro-Meteorological Observatory [4] network is used. Figure 3.3a shows all TAHMO stations located throughout the whole country, and their associated amount of available hourly precipitation data for the associated time domain specified above. To ensure high-quality of validation data, a threshold of 95% of available data was chosen to select useful stations for the verification process (Figure 3.3b). Looking at the map it becomes clear that the northern part of Ghana has a very low density of weather stations, thus a sparse amount of validation opportunities. This needs to be taken into account when looking at the statistical outcomes. To validate



(a) Data amount of the TAHMO stations with a maximum possibility of 480 data counts for the time period of 2018-04-26 to 2018-05-06 and 2020-04-11 to 2020-04-21.

(b) Selected stations after using a threshold of 95% of available data within the time domain.

Figure 3.3: Validation data Experimental Setup 1

the agro-ecological zones separately, the stations are divided into the four zones accordingly.

3.2.2. Metrics

Both the temporal and spatial performance of the WRF model are evaluated. Since most local onset definitions rely on specific rainfall thresholds, precipitation amounts are initially assessed using continuous verification indices. To evaluate the model on a small scale, each model grid cell overlapping with the weather stations is extracted for point validation by comparing their values.

The following metrics are defined: the Mean Bias Error (MBE), the Mean Absolute Error (MAE), the Normalized MAE, the Root Mean Squared Error (RMSE) and the Pearson Correlation coefficient (r). The equations are shown below.

$$\text{MBE}_s = \frac{\sum_{i=1}^{n}(M_{i,s} - O_{i,s})}{n} \tag{3.1}$$

$$\text{MAE}_s = \frac{\sum_{i=1}^{n}|M_{i,s} - O_{i,s}|}{n} \tag{3.2}$$

$$\text{MAE}_{\text{norm}} = \frac{\text{MAE}}{O_{\text{mean}}} \tag{3.3}$$

$$r_s = \frac{\text{cov}(M, O)}{\sigma_O \sigma_M} \tag{3.4}$$

$$\text{RMSE}_s = \sqrt{\frac{\sum_{i=1}^{n}(M_{i,s} - O_{i,s})^2}{n}} \tag{3.5}$$

In the temporal domain:

- $M_{i,s}$ = Model value at station s, at time step i

- $O_{i,s}$ = Observation value at station s, at time step i
- n = Number of time steps
- σ_O = Standard deviation of observation in the time domain
- σ_M = Standard deviation of model in the time domain
- cov(M, O) = Covariance of model and observations
- $Statistic_s$ = Statistic at station s

In the spatial domain:

- $M_{i,s}$ = Model value at station i, at time step s
- $O_{i,s}$ = Observation value at station i, at time step s
- n = Number of stations
- σ_O = Standard deviation of observation in the space domain
- σ_M = Standard deviation of model in the space domain
- cov(M, O) = Covariance of model and observations
- $Statistic_s$ = Statistic at time step s

At each station location, the difference between the model grid value and observation will be calculated first. This error value is then used to calculate the statistical metrics in the spatial and temporal domains. Once the temporal statistics for each station location and the spatial statistics for each day are known, the aggregated regional values will be calculated by taking the average statistical value of all the associated stations in that region.

The source code of these metrics can be found in Appendix A.1.

3.3. Results

3.3.1. WRF Output

10-day Accumulated Precipitation

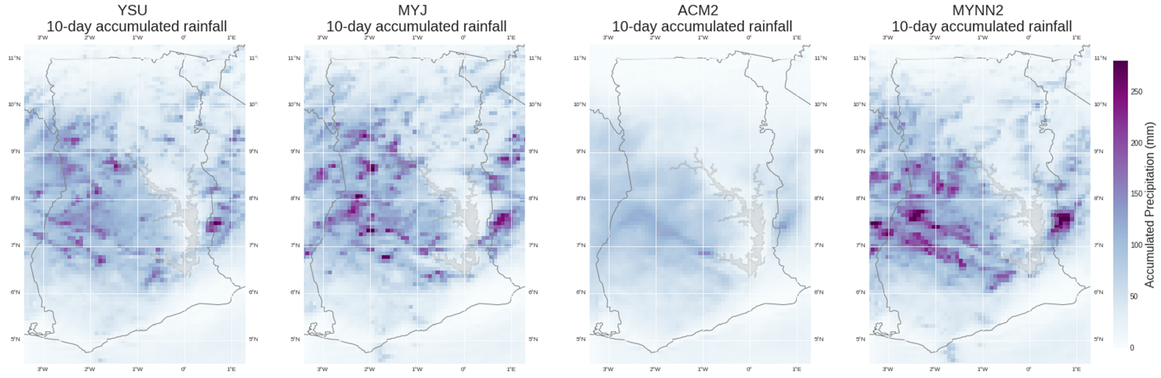


Figure 3.4: Accumulated precipitation at leading time D+10 for the model time period of 2018-04-26 to 2018-05-06

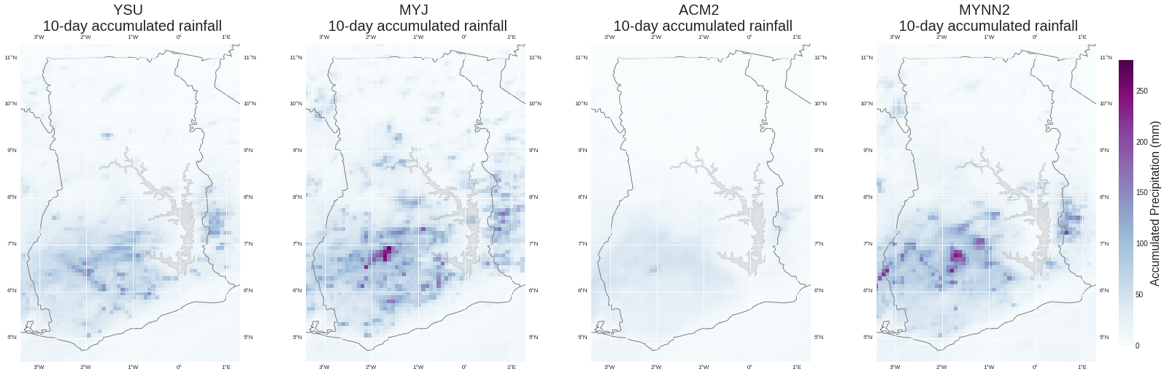


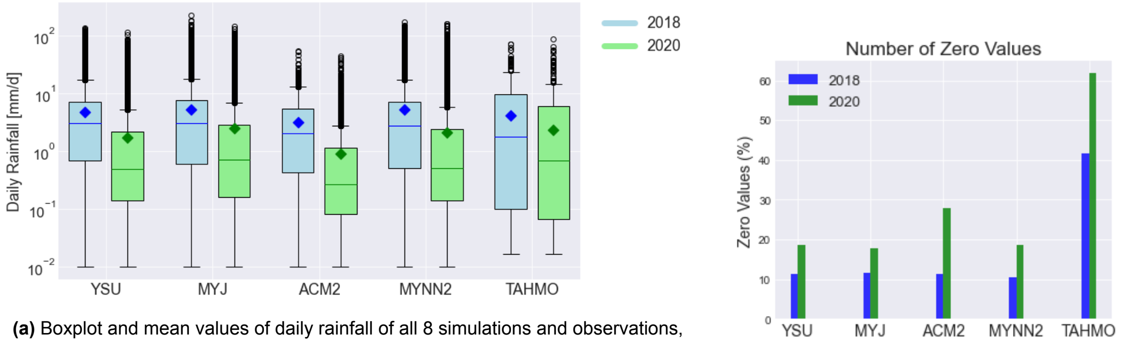
Figure 3.5: Accumulated precipitation at leading time D+10 for the model time period of 2020-04-11 to 2020-04-21

Figures 3.4 and 3.5 depict the accumulated precipitation up to the 10th forecast day (D+10) from the 2018 and 2020 WRF simulations. A notable observation is the significant variation in precipitation magnitude among different schemes. For example, ACM2 shows maximum values of 100 mm, whereas MYJ and MYNN2 reach a maximum of 280 mm. Another noticeable aspect is the rainfall pattern of each scheme: ACM2 generates more widespread rainfall, while other schemes produce localized precipitation clouds with heavy rainfall peaks.

The maps also illustrate that Ghana experienced considerably more rainfall in late April and early May 2018 compared to April 2020. Additionally, rainfall is already prevalent in the north, whereas in 2020 it remains relatively low.

A consistent finding across all eight simulations is the minimal rainfall over Lake Volta. Since Lake Volta is a large water body, it has different thermal properties compared to the surrounding land. The lake's surface may remain cooler during the day, potentially inhibiting convective processes.

Boxplot of Daily Rainfall



(a) Boxplot and mean values of daily rainfall of all 8 simulations and observations, containing: The interquartile range (the box itself); The lower quartile (low edge) and upper quartile (upper edge); The median (line splitting the box); The whiskers (1.5* IQR); The mean (diamonds) and the outliers (black dots).

(b) Percentage of zero values in the total datasets of daily precipitation.

Figure 3.6: General statistics of all ten datasets.

To gain a general understanding of the characteristics of all model configurations and observational data, a boxplot of the daily precipitation aggregated over the whole country is utilized (Figure 3.6a). This involves all grid points within the country's borders for the model simulations and all weather stations for the observations. The decision to include all grid points, rather than subtracting only those at the station location, is aimed at preserving all valuable information. The boxes represent the interquartile range (IQR), which contains the middle 50% of the data. The lower edge of the box corresponds to the 25th percentile (lower quartile), while the upper edge represents the 75th percentile (upper quartile). The line splitting the box indicates the median. The whiskers extend to a maximum of 1.5 times the IQR from the bottom and top of the box. The diamonds represent the mean of the complete datasets, and the black dots indicate all outliers.

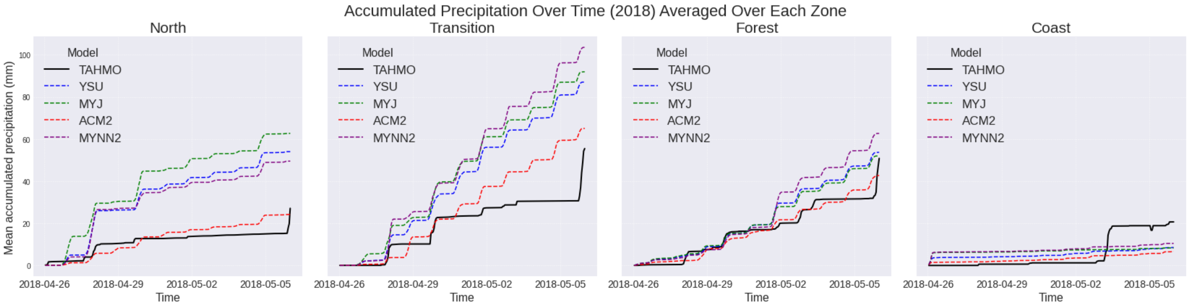
Given that the dataset contains a high concentration of low precipitation values as well as some very high rainfall amounts, a logarithmic scale is applied to account for this log-normal distribution. Zero values are masked from the boxplot but included in the mean calculation. Consequently, the diamonds reflect the true mean of the entire dataset, while the boxplot represents only the active data. The percentage of zero values is plotted separately in the graph next to it (Figure 3.6b).

The skewed nature of the datasets becomes apparent, with the mean significantly higher than the median due to the influence of high precipitation outliers. This difference would have been even bigger if the boxplot had included the zero values, further lowering the median. Additionally, the presence of numerous high precipitation outliers, and no low precipitation outliers, further underscores this skewed distribution. Moreover, the whiskers of the TAHMO datasets do not reach as low as those of the model outputs, likely due to the limited capability of the weather stations to capture low precipitation amounts.

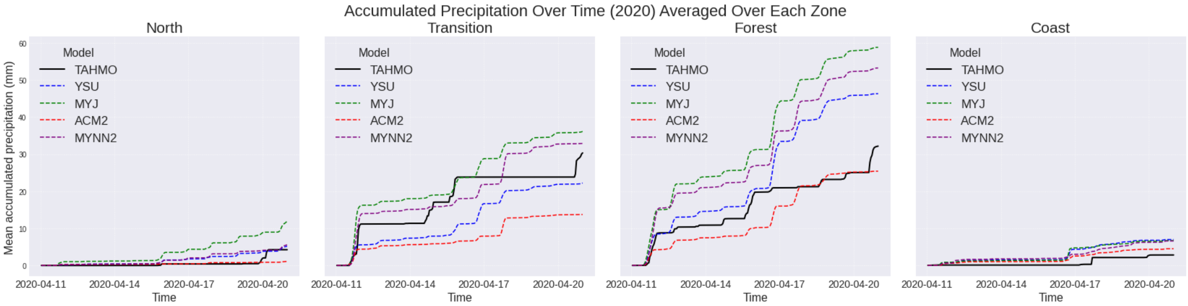
The boxplots of 2020 are centered around lower precipitation amounts compared to 2018, suggesting less precipitation, which supports previous observations. ACM2 exhibits the lowest median and mean values for both years (except for the median of TAHMO in 2018), reinforcing earlier findings and indicating a potential underestimation of rainfall. The IQR of TAHMO is much wider compared to the model outputs, suggesting a broader distribution of the observed rainfall. Lastly, YSU, MYJ, and MYNN2 schemes exhibit more outliers compared to ACM2 and TAHMO, particularly in 2018.

Figure 3.6b shows that TAHMO has a significantly higher percentage of zero values than the model outputs. This is partly due to the previously mentioned limited capability of TAHMO to capture low precipitation amounts, whereas the models can easily produce small amounts (which are not zero). The ACM2 model has the second highest number of zero values in 2020, while for 2018 the amount of zero values are relatively similar across all models.

Accumulated Precipitation over Time



(a) Accumulated precipitation (mm) over time, averaged over each zone for model time period of 2018-04-26 to 2018-05-06



(b) Accumulated precipitation (mm) over time, averaged over each zone for model time period of 2020-04-11 to 2020-04-21

Figure 3.7: Accumulated precipitation (mm) over time, averaged over each zone, for all 8 simulations and observations.

Figure 3.7 visualizes the accumulated precipitation (mm) over time, averaged over each zone, for all 8 simulations and observations. There are some clear agreements between model and observations, along with some discrepancies. In 2018, all schemes show high overestimation, especially in the North and the Transition zone. However, all schemes perform relatively well in the Forest zone. In the coastal region, there is a minimal steady increase in rainfall, with TAHMO being the only one experiencing an abrupt increase at the end. Interestingly, the timing of the increase in rainfall is reasonably consistent across all datasets, all exhibiting a stepped curve. A closer examination of the datasets reveals that rainfall typically occurs in the late afternoon or early evening, while it remains relatively dry at night and in the morning, particularly in the Forest and Transition zones.

For 2020, TAHMO's daily mean precipitation is the second highest among the datasets, just below MYJ (Figure 3.6a). However, as shown in Figure 3.7b, TAHMO does not consistently show high accumulated values over time. In particular, in the Forest zone, YSU, MYNN2, and MYJ report significantly higher values than TAHMO. Additionally, in the Coastal zone, all model simulations show higher precipitation levels than those observed (even though these values are relatively low).

This discrepancy can be attributed to the distribution patterns observed in the accumulated rainfall maps (Figure 3.5) and the high number of outliers in the model simulations. The model simulations produce more sporadic, very high precipitation values at specific grid cells, which skews the total accumulated precipitation upwards. In contrast, TAHMO provides more consistent and evenly distributed precipitation events over time, leading to fewer extreme values and outliers. As a result, this consistency produces higher daily means but lower accumulated precipitation levels.

3.3.2. Point Validation

Mean Absolute Error: Daily Precipitation

To better visualize the discrepancies between observations and model simulations, the normalized Mean Absolute Error (MAE) is plotted at each station location (Figure 3.8 and Figure 3.9). The MAE

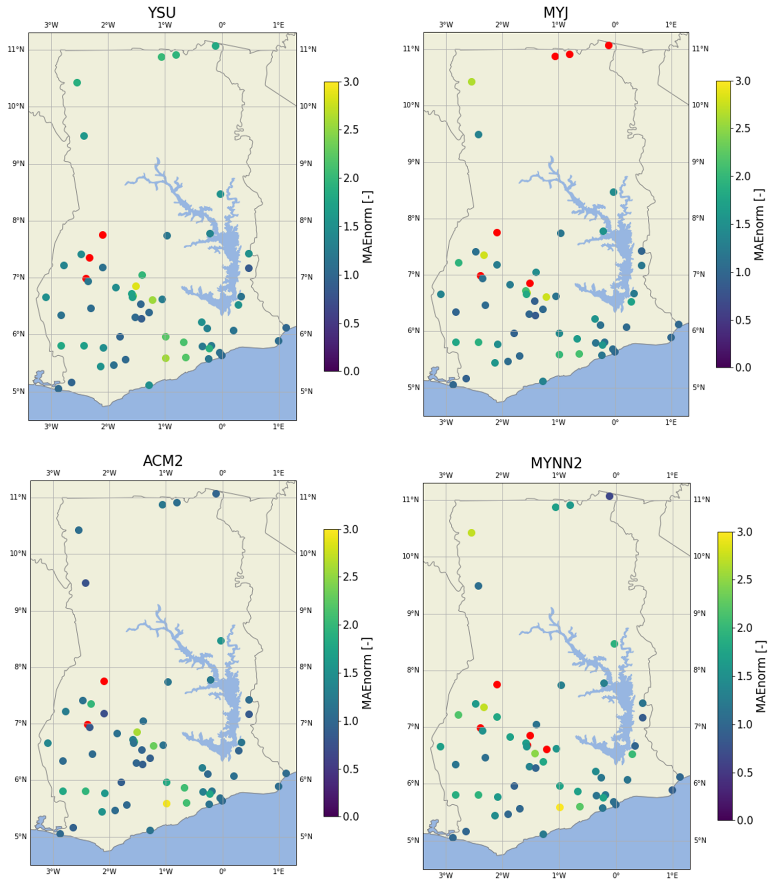


Figure 3.8: Normalized Mean Absolute Error (MAE) of daily precipitation (mm/d) for the 2018 simulations at the selected TAHMO locations.

represents the average of the absolute errors between observed and modeled daily precipitation and is normalized by dividing it by the average daily precipitation value from the TAHMO data at the corresponding station. The stations depicted in red are stations with MAE values above 3, meaning the model error is three times as high as the average daily precipitation value, measured over those ten days. Stations where the observed 10-day average daily precipitation value is zero, making it not possible to normalize the error values, are masked and not shown in the map.

The outliers (red dots) are spread out over the whole country and not specifically characteristic of one region. The year of 2020 shows more outliers and worse results than the year 2018. Besides that, in both years, MYNN2 and MYJ show more outliers and higher normalized MAE values than ACM2 who is performing relatively well. What is also noticeable is that all PBL schemes perform well at some specific station locations and bad at others. For instance at stations TA00011 (lat: 6.98, lon:-2.39) WRF is constantly performing poorly, while at station location TA00617 (lat: 6.69, lon:-1.52) it is performing well in each run.

A closer examination of station TA00011 reveals an interesting discrepancy when compared to the nearby station TA00012. Despite their proximity, TA00012 consistently performs well in the 2018 simulations, while TA00011 performs poorly. Plotting the time series for both stations (Figure 3.10) reveals that TA00011 does not show any precipitation peaks, whereas TA00012 records significant rainfall. This abnormal discrepancy suggests potential measurement failures or other external factors negatively impacting the data from station TA00011.

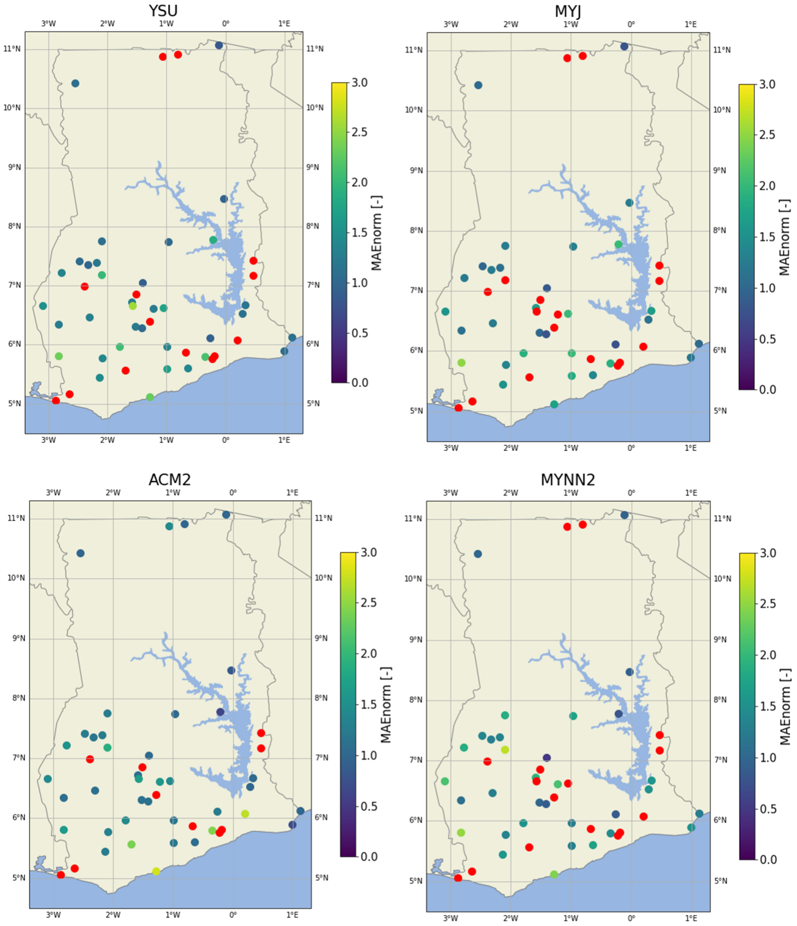


Figure 3.9: Normalized Mean Absolute Error (MAE) of daily precipitation (mm/d) for the 2020 simulations at the selected TAHMO locations.

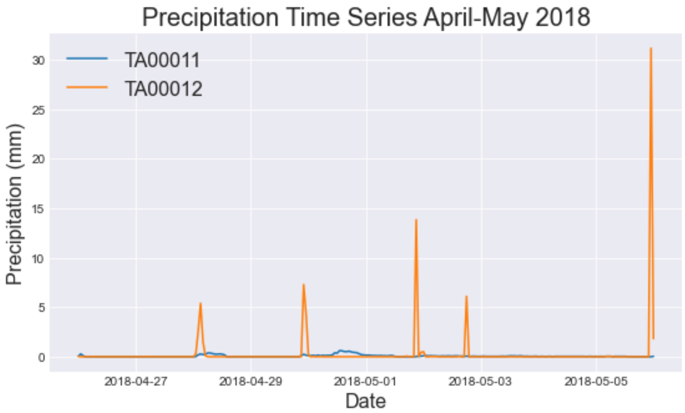


Figure 3.10: TA00011 and TA00012 precipitation time series from 2018-04-27 to 2018-05-05

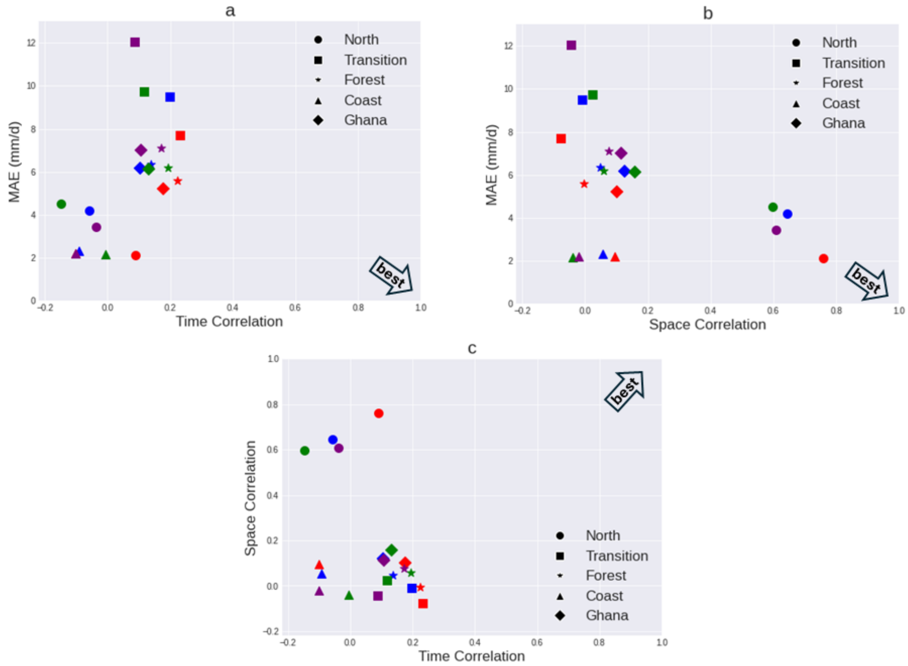


Figure 3.11: Mean absolute error [mm/d] of daily rainfall against Time and Space Correlation [-] for the 2018 simulations. Blue = YSU, Green= MYJ, Red = ACM2, Purple = MYNN2.

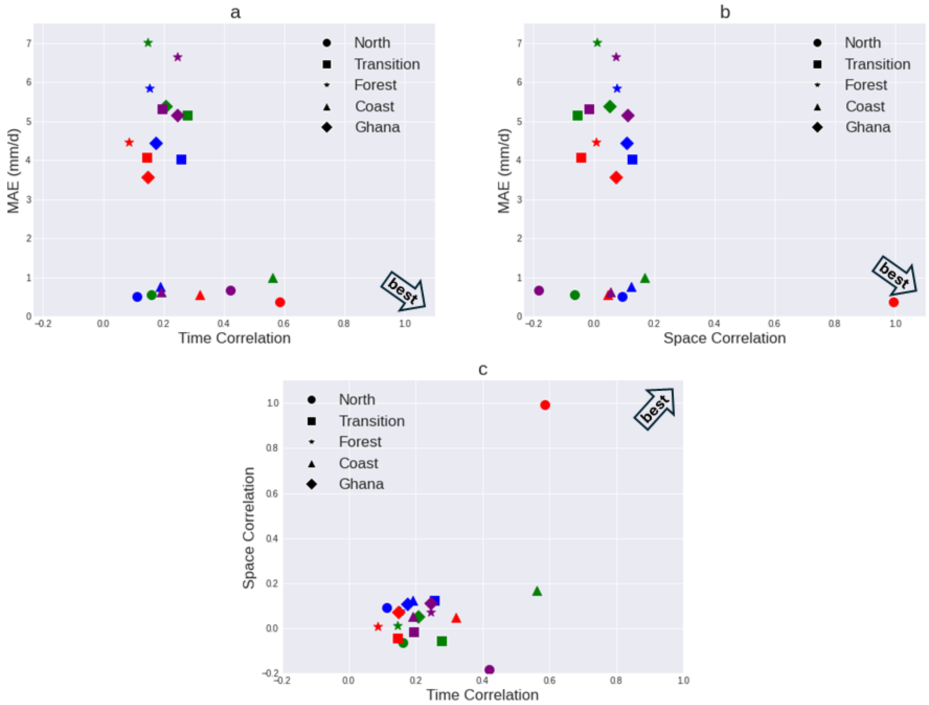


Figure 3.12: Mean absolute error [mm/d] of daily rainfall against Time and Space Correlation [-] for the 2020 simulations. Blue = YSU, Green= MYJ, Red = ACM2, Purple = MYNN2.

Temporal and Spatial Correlation: Daily Precipitation

Now that the relative bias of the various simulations at each station location is indicated, the scale can be aggregated to regional and national scale. Figures 3.11 and Figure 3.12 plot the regionalized mean absolute errors of daily precipitation against their corresponding temporal and spatial correlation coefficients. These coefficients represent the linear relationship between modeled and observed daily precipitation values in both the temporal and spatial domains. In the temporal domain, this involves calculating the correlation of simulated daily precipitation over a 10-day period for each station location, and then averaging these correlations across the stations within each zone to obtain a zonal time correlation. In the spatial domain, the correlation of the spatial distribution of rainfall across the associated zones is calculated for each day, and the average of these daily correlations over the 10-day period is used to produce a single spatial correlation value for each zone.

Starting with an examination of the results from the 2018 simulations, one observation is the negative correlations in the northern and coastal regions. The negative correlation in the Coast is likely directly linked to the previously identified underestimation in Figure 3.7a, where TAHMO shows a sudden increase in rainfall towards the end of the simulation period, which is not captured by the model outputs. The northern zone performs very well in terms of space correlations with values between 0.6 and 0.8.

The transition zone is showing the highest MAE values compared to the other regions, ranging from 12 mm/d with MYNN2 and 8 mm/d with ACM2. ACM2 outperforms the other schemes in terms of absolute error values and is also showing the best results in terms of time correlation. However, for spatial correlation, the other schemes perform slightly better. Overall, ACM2 appears to be the best-performing scheme in the 2018 simulations.

Upon reviewing the scatter plot data from 2020, the PBL schemes exhibit both similarities and differences. In terms of similarities, all PBL schemes show weak spatial correlation coefficients, mostly positive, ranging between 0 and 0.2. However, some schemes, MYNN and MYJ, exhibit negative correlations for the Northern region. This contrasts with ACM2, which shows a perfect spatial correlation of 1 for the same region. The accuracy of these correlations is questionable due to the small amount of stations in this region, making it statistically difficult to verify. The temporal correlation is slightly better, varying from 0 to 0.6. However, most schemes still fall between 0.1 and 0.3.

Regarding the MAE, there seems to be a clear division between the Transition and Forest zones, with errors between 4.5 and 7 mm/d, and the Coastal and Northern zones, with errors below 1 mm/day. The low error values can be explained by the very low precipitation amounts in these regions, making discrepancies less noticeable.

Examining the regions and PBL schemes individually, ACM2 again outperforms all the other PBL schemes in terms of MAE but performs the poorest in spatial and temporal correlation in the rainy zones (Forest and Transition). In contrast, the MYJ scheme exhibits the highest errors but shows relatively good correlations compared to other schemes. MYJ performs particularly well in the Coastal region. For the Northern zone, ACM2 performs exceptionally well, with correlation coefficients of 1 and 0.4 and low absolute errors. For the zones that are experiencing more rain, namely the Transition and Forest zone, YSU and MYNN2 seem to be the best-performing schemes in terms of correlation coefficients, respectively. On average, considering both correlations and error values, YSU performs the best for Ghana as a whole.

3.3.3. Summary

Table 3.2: 2018 Simulations, Statistical Summary

PBL	Zone	MBE [mm/d]	RMSE [mm/d]	MAE [mm/d]	r_temporal [-]	r_spatial [-]
YSU	North	1.079	10.955	4.192	-0.06	0.64
	Transition	4.030	13.924	9.505	0.20	-0.01
	Forest	0.390	10.565	6.330	0.14	0.05
	Coast	-1.331	6.306	2.313	-0.09	0.06
	Ghana	0.870	10.845	6.202	0.10	0.12
MYJ	North	1.102	7.998	4.510	-0.15	0.60
	Transition	4.037	14.601	9.733	0.12	0.02
	Forest	0.358	10.854	6.188	0.19	0.06
	Coast	-1.572	6.293	2.164	-0.01	-0.04
	Ghana	0.822	10.968	6.162	0.13	0.16
ACM2	North	-1.429	5.825	2.124	0.09	0.76
	Transition	2.003	12.064	7.686	0.23	-0.08
	Forest	-0.375	9.451	5.596	0.22	0.00
	Coast	-1.389	6.232	2.213	-0.10	0.09
	Ghana	-0.169	9.398	5.250	0.18	0.10
MYNN2	North	0.045	7.286	3.428	-0.04	0.61
	Transition	6.928	18.030	12.048	0.09	-0.04
	Forest	1.912	12.672	7.094	0.17	0.08
	Coast	-1.481	6.269	2.186	-0.10	-0.02
	Ghana	2.213	12.863	7.044	0.11	0.11

Table 3.3: 2020 Simulations, Statistical Summary

PBL	Zone	MBE [mm/d]	RMSE [mm/d]	MAE [mm/d]	r_temporal [-]	r_spatial [-]
YSU	North	-0.213	1.927	0.522	0.11	0.09
	Transition	0.253	7.262	4.023	0.26	0.13
	Forest	1.746	11.522	5.856	0.15	0.08
	Coast	0.312	1.832	0.756	0.19	0.12
	Ghana	1.138	9.563	4.443	0.17	0.11
MYJ	North	-0.113	1.854	0.565	0.16	-0.06
	Transition	1.576	12.253	5.151	0.28	-0.06
	Forest	3.349	14.056	7.028	0.15	0.01
	Coast	0.609	3.387	1.005	0.56	0.17
	Ghana	2.402	12.226	5.397	0.21	0.05
ACM2	North	-0.370	1.856	0.383	0.59	0.99
	Transition	-0.417	8.108	4.078	0.14	-0.04
	Forest	-0.349	9.915	4.462	0.09	0.01
	Coast	0.129	1.684	0.564	0.32	0.05
	Ghana	-0.301	8.518	3.562	0.15	0.07
MYNN2	North	-0.041	2.407	0.670	0.42	-0.18
	Transition	1.354	8.947	5.323	0.19	-0.02
	Forest	2.985	12.880	6.652	0.25	0.07
	Coast	0.214	1.671	0.627	0.19	0.05
	Ghana	2.094	10.803	5.156	0.25	0.11

From this initial set of experiments, it is immediately apparent that the parametrization schemes cannot be evaluated through a singular, straightforward assessment. However, some general observations can be drawn: When examining the magnitude of the different simulations, along with their corresponding statistical features and accumulated rainfall maps, it is evident that the MYJ and MYNN2 schemes produce more localized intense rainfall patterns with high maximum values. In contrast, the ACM2 scheme exhibits a more homogeneous, widespread rainfall distribution without localized extremes, and the YSU scheme represents a combination of both characteristics. The boxplot of the observed rainfall data revealed a significantly broader distribution compared to all model outputs. The error and correlation calculations, summarized in Tables 3.2 and 3.3, reveal the following insights:

- The ACM2 scheme generally underestimates rainfall, as previously observed, resulting in negative MBE values (Tables 3.2 and 3.3). However, this also leads to the lowest error values (RMSE and MAE), thus the best-performing scheme in terms of error metrics.
- All other schemes are generally overestimating, especially in the Transition and Forest zones. This constant overestimation of the model corresponds with the findings of Agyeman's study [16].
- The MYJ scheme shows the highest errors in 2020, while the MYNN2 scheme exhibits the highest errors in 2018.
- The YSU scheme performs moderately, with its error metrics falling between the other evaluated schemes.
- All PBL schemes underestimate the precipitation in the North in the 2020 simulation and in the Coast in the 2018 simulation (Tables 3.2 and 3.3).

The overall higher magnitudes and resulting higher errors of the MYJ and MYNN2 schemes, contrasted with the lower precipitation amounts and errors of the ACM2 scheme, are further supported by the findings of Moya-Alvarez et al. [29]. Their study, which evaluated the sensitivity of 12 PBL schemes in precipitation forecasts in the Andes of Peru, highlighted clear differences in boundary layer stabilities among the schemes and their impact on precipitation amounts. ACM2 was identified as one of the most stable schemes, resulting in lower rainfall amounts, whereas MYJ, one of the most unstable schemes, led to significantly higher precipitation. The non-local nature of ACM2, which includes large eddies and not only local grid-point interactions, allows for efficient mixing, resulting in smoother temperature and moisture gradients and a more stable environment. Increased stability reduces the likelihood of convective activity, thereby decreasing precipitation. This contrasts with local schemes like MYJ and MYNN2, where the localized nature leads to larger gradients and more unstable conditions.

Regarding spatial and temporal correlation coefficients, the following conclusions can be drawn:

- With some exceptions, most schemes perform suboptimally, fluctuating between 0 and 0.3. The ACM2 scheme performs the worst in both time and space correlations, except for the time correlation in the 2018 simulation, where it performs the best.
- No single PBL scheme consistently outperforms the others in terms of correlation. The performance fluctuates significantly between the years and the zones.
- The northern region shows positive and negative outliers in terms of correlation, but the reliability of these results is questionable due to the low number of stations in this region.

Taking all results and statistical findings into account, a general usage recommendation is provided for the different zones and parametrization schemes. In this scoring matrix, the PBLs are rated relative to each other, indicating which scheme performs better or worse in comparison to the others. However, this rating is rather subjectively done by considering all the statistical scores provided in Tables 3.2 and 3.3 and the Figures shown in section 3.3. On average, ACM2 is demonstrating the best performance in simulating rainfall for the whole country, followed by YSU. This is in line with with the suggestions of Agyeman et al. and Meroni et al. [16, 19]. The lower rainfall produced by the ACM2 scheme appears to be particularly suitable for the Northern and Coastal regions, which, during the analyzed time frame, also experience less rainfall compared to other regions. For the Transition zone, YSU is the most suitable PBL scheme, while for the Forest zone, it is still unclear which PBL scheme is most suitable to use.

	North	**Transition**	**Forest**	**Coast**	**Ghana**
YSU	Average	Best	Average	Average	Good
MYJ	Average	Average	Average	Average	Average
ACM2	Best	Average	Average	Best	Best
MYNN2	Average	Worst	Average	Average	Average

Table 3.4: Experimental Setup 1: Recommended usage of each PBL scheme per zone.

3.4. Conclusion

This chapter aimed to address sub-question 1: ”How do the selected combinations of physics schemes of the WRF model perform in forecasting precipitation around onset of the rainy season in: a. the whole country of Ghana b. the different agro-ecological zones?”

The findings indicate that the performance of the model configurations varies across the different zones and the validation metrics used. The model shows higher errors in the Transition and Forest zone, compared to the Northern and Coastal zone and relatively weak correlations in general. On average, the YSU and ACM2 schemes perform the best, while MYNN2 and MYJ show higher errors despite their temporal and spatial correlation outperforming that of ACM2.

Some general questions arises after these first set of results. At first, the uncertainty in the applied validation method. The metrics use point data which limits our ability to capture the complete distribution of rainfall, thereby omitting a significant portion of the WRF model output. This raises concerns about the spatial distribution accuracy of the models, given that only a few grid points are considered. It is possible that the rain clouds are slightly misaligned with the selected grid points, which might not necessarily indicate poor model performance, but is shown that way with these metrics.

Besides this, the high error values warrant further investigation. Are these errors for instance due to a consistent overestimation that could be corrected with a simple adjustment factor? Alternatively, do the MYNN2 and MYJ schemes predominantly overestimate heavier rainfall peaks, and would their performance improve if we only considered daily rainfall up to a certain maximum?

To address these questions and issues a new set of experiments is defined, focusing on a smaller domain where stations are closely spaced, allowing for interpolation to the WRF grid scale. Additionally, employing categorical evaluations can mitigate the magnitudinal errors mentioned earlier.

Reflecting on the initial results, there is still much to learn about the Forest zone. Different schemes yield varying statistical performances, and no single simulation consistently outperforms the others. Given that this area is densely filled with agricultural land [30] further investigation is warranted.

Furthermore, the current reliance on only two simulations per scheme may not provide a comprehensive understanding. Initial conditions could have a significant impact, or this specific time period might be an outlier. To enhance the robustness of the findings, automating WRF to conduct more simulations for each PBL scheme would be beneficial.

4

Experimental Setup 2

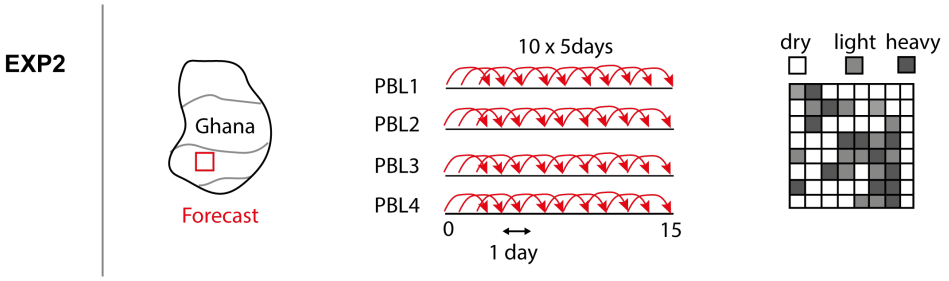


Figure 4.1: Sketch Experimental Setup 2

This chapter delves into the second experimental setup of this research. The focus here is on a smaller area within the Forest zone, selected for its denser observational data. This experimental setup aims to address the second sub-question of this research: *”Which model configuration is the optimal choice considering the context of this study?”*

4.1. Model Configuration

4.1.1. Time and Spatial Domain

In light of the findings and uncertainties from the initial experiments, the decision has been made to concentrate the second phase of experiments on a small area within the Forest zone. Further investigation in this zone is warranted due to its high density of weather stations, offering the most reliable observational data, and extensive coverage of agricultural grounds. The selected model domain exhibited the highest density of reliable stations within the zone. In the initial experiments, all planetary boundary layer (PBL) schemes performed similarly in this zone, with no clear standout. Therefore, a re-evaluation of the sensitivity of all PBL schemes is planned, starting from scratch.

The time focus of this set of experiments is the period surrounding the onset date of 2018, which for the southern and middle parts of Ghana was around February 19th.

The chosen domain takes into account the proximity of the weather stations to one another, allowing for data interpolation to the WRF grid. Moreover, the stations have provided reliable data within the selected timeframe.

Additionally, all simulations are conducted at a convective-resolving scale of 3km resolution, rather than the 9km scale, allowing for the Cumulus Convection scheme (CU) to be switched off [15]. Convective processes typically occur at scales of a few kilometers. When the resolution is smaller or comparable to the convective scale, the model is able to resolve the physical processes without the need for parametrization schemes. Switching off the CU scheme, despite controversy and uncertainty within the climate modeling community regarding the resolution threshold at which convective processes are fully resolved without an added-value of parametrization [15, 31], can potentially lead to improved model performance in terms of the magnitude, spatial, and temporal distribution of the rainfall [32, 33, 34].

Running a 5-day simulation over this domain only takes a few hours, allowing for the possibility of automating the model to perform multiple simulations in immediate succession. The automating bash script can be found in the Appendix A.3. This change allows for the validation of a PBL scheme to be based on multiple simulations, rather than just one. Specifically, 10 simulations of a 5-day forecast were run for each PBL, covering the period from 12-02-2018 to 26-02-2018. The lead times from the initial conditions (IC) are denoted as D+1 (up to 24 hours after IC) to D+5 (up to 120 hours after IC). The individual forecast days are referred to as forecast-day (FD) 1 to 5.

4.2. Validation Method

4.2.1. SEEPS

For the validation method, both continuous and categorical verification indices are utilized. The continuous indices include the Mean Bias Error (MBE) and the Mean Absolute Error (MAE), as written out earlier in this research. In the categorical verification field, there are various statistical methods available. Commonly used metrics in weather prediction, like the false alarm ratio (FAR) or the probability of detection (POD), are all based on contingency table practices and involve true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) [6, 21].

Rodwell et al. (2010) introduced the Stable Equitable Error in Probability Space (SEEPS) score, which utilizes different thresholds depending on the precipitation climatology of the region of interest [35, 36]. This refined method may be considered advantageous compared to the traditional metrics mentioned above. It is designed to be equitable, equally rewarding all forecast improvements, and aims to be as insensitive as possible to sampling uncertainty.

The error metric is based on three categories: dry, light, and heavy precipitation. The first threshold (t1) is set to a value of 0.25 mm/day, as described in [36], while the second threshold, distinguishing light from heavy rain, is based on the cumulative distribution of 24-hour precipitation at a given location. The observed probability of dry weather (p1), associated with threshold t1, divides the distribution into dry and wet days. The division of wet days is based on the assumption that light precipitation occurs twice as often as heavy precipitation, resulting in a ratio of 2:1 between p2 (light) and p3 (heavy).

For this particular research, the cumulative distribution is based on 5 years of TAHMO data for station

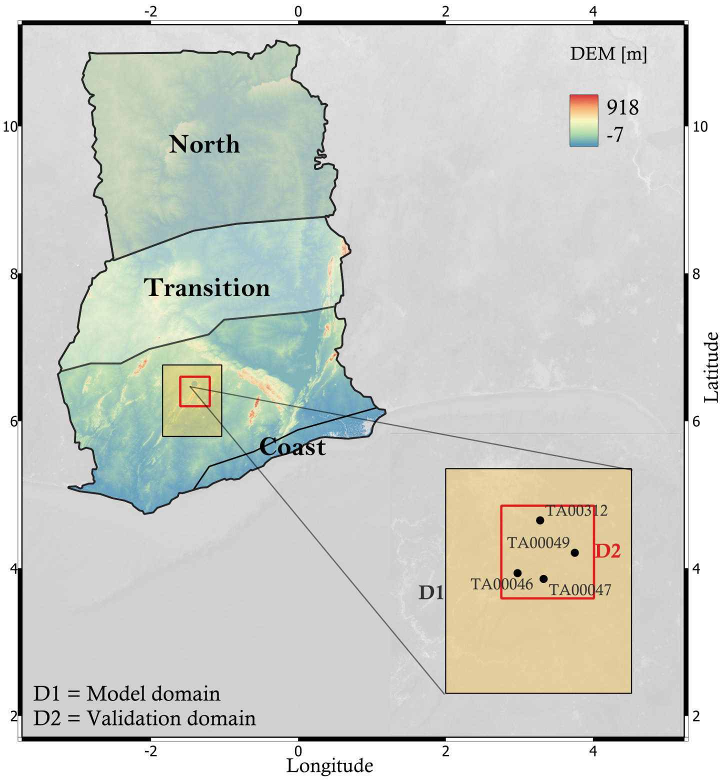


Figure 4.2: Model domain and validation domain for the second set of experiments displayed on a Digital Elevation Model (DEM) map, downloaded from [27]

TA00047, located within the validation domain as shown in Figure 4.2. This station was chosen as it appeared to have the least amount of data gaps for the longest possible time period (Figure B.2b in Appendix B.2).

The cumulative distribution together with the associating thresholds and probabilities are shown in Figure 4.3

The scoring matrix of SEEPS, rewritten only in terms of p1, is derived from the study by Haiden et al. (2012) [36] and is depicted below.

$$S = \frac{1}{2}\begin{pmatrix} 0 & \frac{1}{1-p_1} & \frac{4}{1-p_1} \\ \frac{1}{p_1} & 0 & \frac{3}{1-p_1} \\ \frac{1}{p_1}+\frac{3}{2+p_1} & \frac{3}{2+p_1} & 0 \end{pmatrix} \tag{4.1}$$

The observed precipitation categories run from left to right, while the forecast categories run from top to bottom. For example, if light precipitation is observed but heavy rain is forecasted, the error value will be calculated as 3/(2 + p1). Additionally, if dry weather is observed but heavy rain is forecasted, the error will be calculated as 1/p1 + 3/(2 + p1), which is even bigger. This indicates that SEEPS is an error-oriented score. To convert it to a skill score, the term 1 - SEEPS is used. Considering a sufficient amount of comparative values, the values are expected to be between 0 (bad) and 1 (good).

Since this study has an agricultural purpose, the choice of the initial threshold (t1) was reconsidered. In the agricultural sector, it is customary to use a threshold value of 1 mm instead of 0.25 mm to differentiate between dry and rainy conditions. The CDF and associated probability and threshold values can be found in Appendix B.2 Figure B.3.

All simulation days will be evaluated and no distinction between the different leading times (D+1 to D+5)

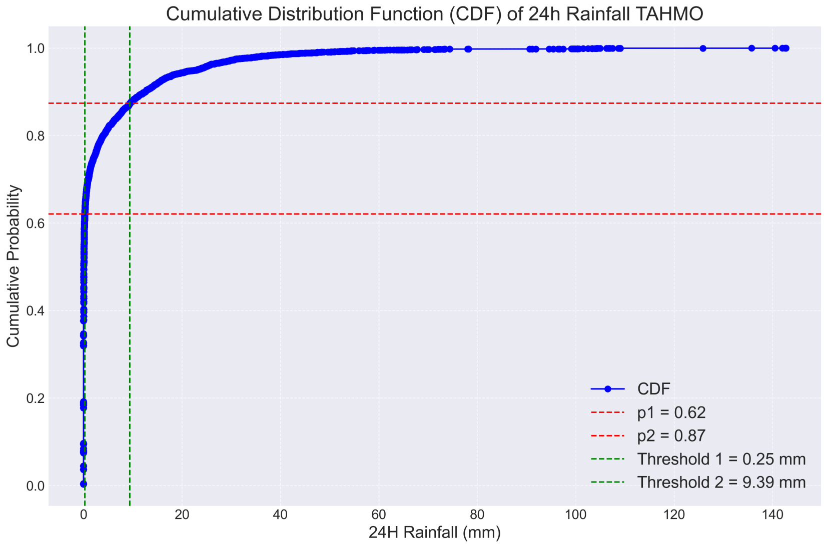


Figure 4.3: Cumulative distribution for 24-h precipitation at station TA00047 for the period of 2018-2023.

will be made yet. The SEEPS score at each grid point will then be the average of all the simulation days. The calculation of SEEPS can be found in Appendix A.4.

4.2.2. Interpolation of TAHMO Data

To ensure accurate comparison, the TAHMO data is interpolated to the WRF grid to expand its scale and increase the utility of the stations' data by filling in gaps and providing a more comprehensive dataset. The inverse distance weighting (IDW) method is used for this interpolation, which is a broadly applied application in spatial interpolation and a recommended technique within modelling and validation of precipitation data [37]. The basic idea behind the method is that the interpolated points are the most affected by the nearest observational point and least by the points with the most distance. The Inverse Distance Weighting (IDW) interpolation method is defined by the following equation:

$$Z(\mathbf{x}_0) = \frac{\sum_{i=1}^{N} \frac{Z(\mathbf{x}_i)}{d(\mathbf{x}_0,\mathbf{x}_i)^p}}{\sum_{i=1}^{N} \frac{1}{d(\mathbf{x}_0,\mathbf{x}_i)^p}} \tag{4.2}$$

where:

- $Z(\mathbf{x}_0)$ is the interpolated value at the location $\mathbf{x}_0$,
- N is the total number of known data points,
- $Z(\mathbf{x}_i)$ is the known value at the i-th data point,
- $d(\mathbf{x}_0, \mathbf{x}_i)$ is the distance between the interpolation point $\mathbf{x}_0$ and the i-th known data point $\mathbf{x}_i$,
- p is the power parameter that controls the influence of the distances.

The interpolation technique is applied both for observed 5-day accumulated rainfall (Figure 4.4) as for the daily rainfall (Appendix B.1, Figure B.1). The source code with additional explanation can be found in Appendix A.2. The code introduces an additional feature to the equation that accounts for a potential maximum distance. However, this maximum distance is not utilized because the stations are sufficiently close to one another. The accuracy of the results heavily depends on the selection of the power parameter p. Since the optimal value for p can vary depending on the specific context, and given that the four stations are situated close to each other, allowing for a lower power parameter, the default value of 2 has been selected for this implementation.

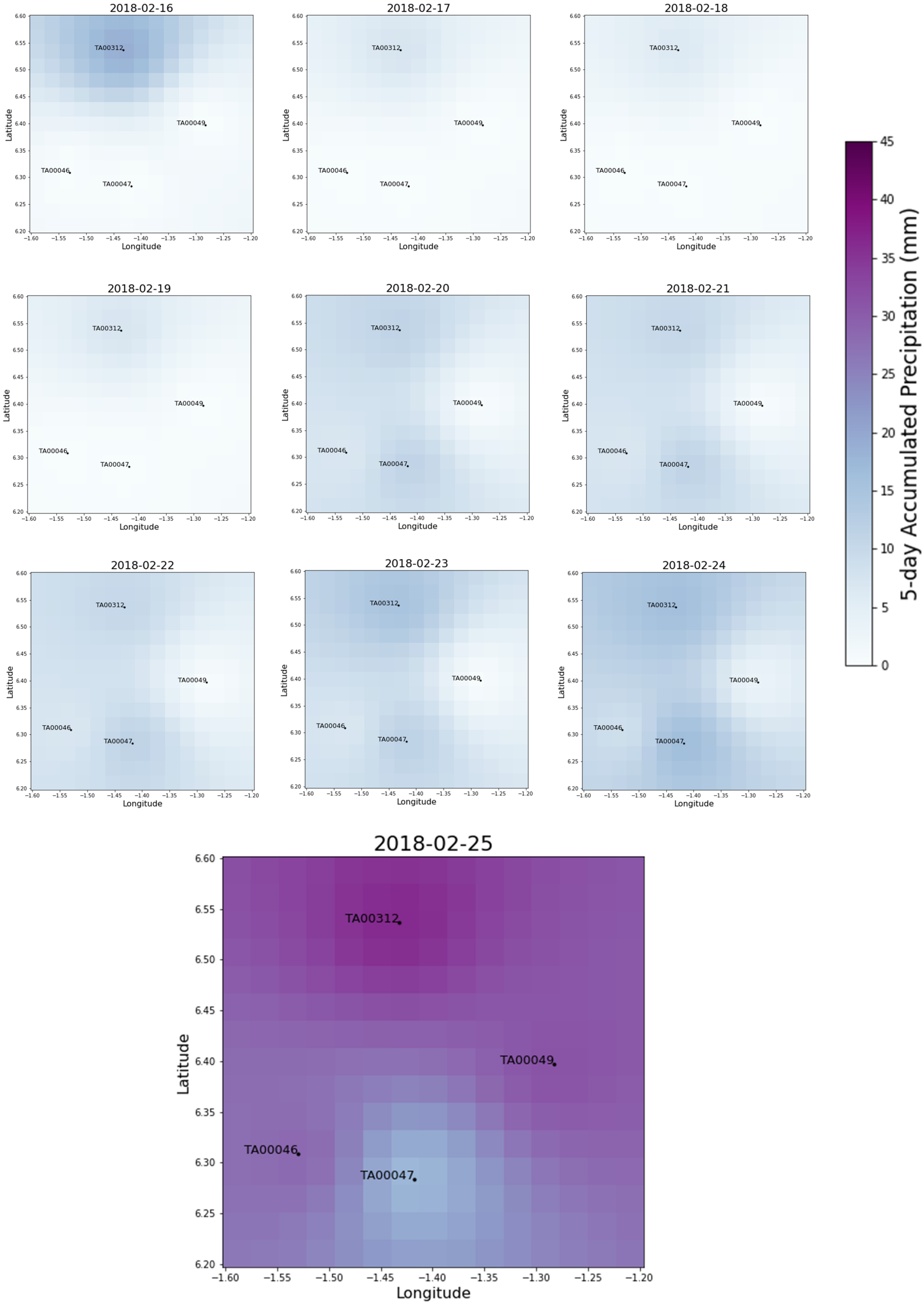


Figure 4.4: Interpolated TAHMO data for 5-day Accumulated Precipitation (mm), using the IDW method.

4.3. Results

4.3.1. Accumulated Precipitation

Firstly, the 5-day accumulated rainfall is analyzed, which represents the total precipitation generated in each 5-day simulation. The accumulated rainfall on the last forecast day (leading time D+5) is compared with the corresponding 5-day accumulated rainfall of the interpolated TAHMO data.

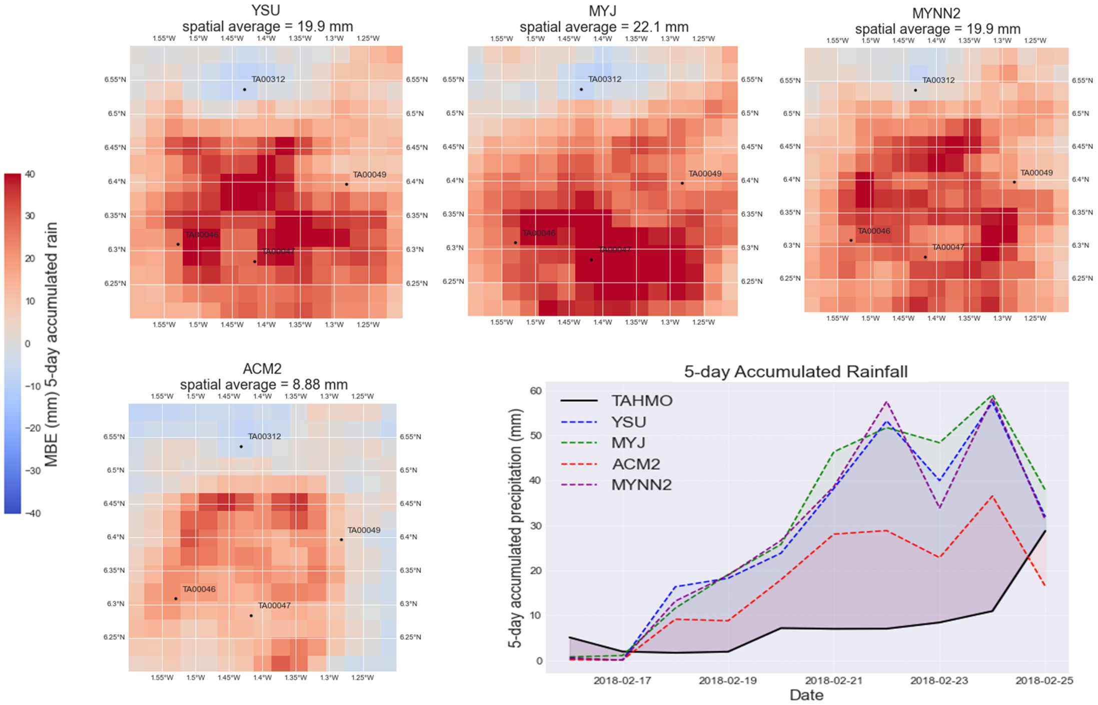


Figure 4.5: The spatial distribution of the Mean Bias Error (MBE) of the 10 simulations for each PBL scheme, along with the accumulated precipitation timeseries averaged over the whole validation domain.

Figure 4.5 shows the Mean Bias Error (MBE) of 5-day accumulated rainfall for the 10 simulations of each PBL scheme. Additionally, it displays the 5-day accumulated precipitation time series averaged across the entire validation domain. The time series indicate that all schemes perform rather poorly, with the start of the increase being too early and the magnitude of the increase being too large. This corresponds to the high MBE values displayed in the spatial maps. Notably, the YSU, MYJ, and MYNN2 schemes tend to overestimate significantly, while the ACM2 scheme performs better, though it underestimates around the borders of the domain.

Upon examining the locations of the four stations, all schemes appear to perform better around station TA00312 compared to the other locations. The highest overestimations are observed between the other three TAHMO stations. However, these overestimations are not distributed in a manner that would clearly question the accuracy of the interpolation technique.

4.3.2. Daily Precipitation

SEEPS

To evaluate daily rainfall, the Skill score of SEEPS (Stable Equitable Error in Probability Space) is calculated alongside its associated error counts for each rain category. Figure 4.6 illustrates this Skill score for each PBL scheme when a t1 threshold value of 0.25 is applied. The numbers inside each grid cell represent, roughly speaking, the rainfall amount: 0 indicates no rainfall across all simulation days, and 100 signifies heavy rainfall on all simulation days. These numbers are the sum of the observed categories (dry: 0, light: 1, heavy: 2) in the cell across all simulation days.

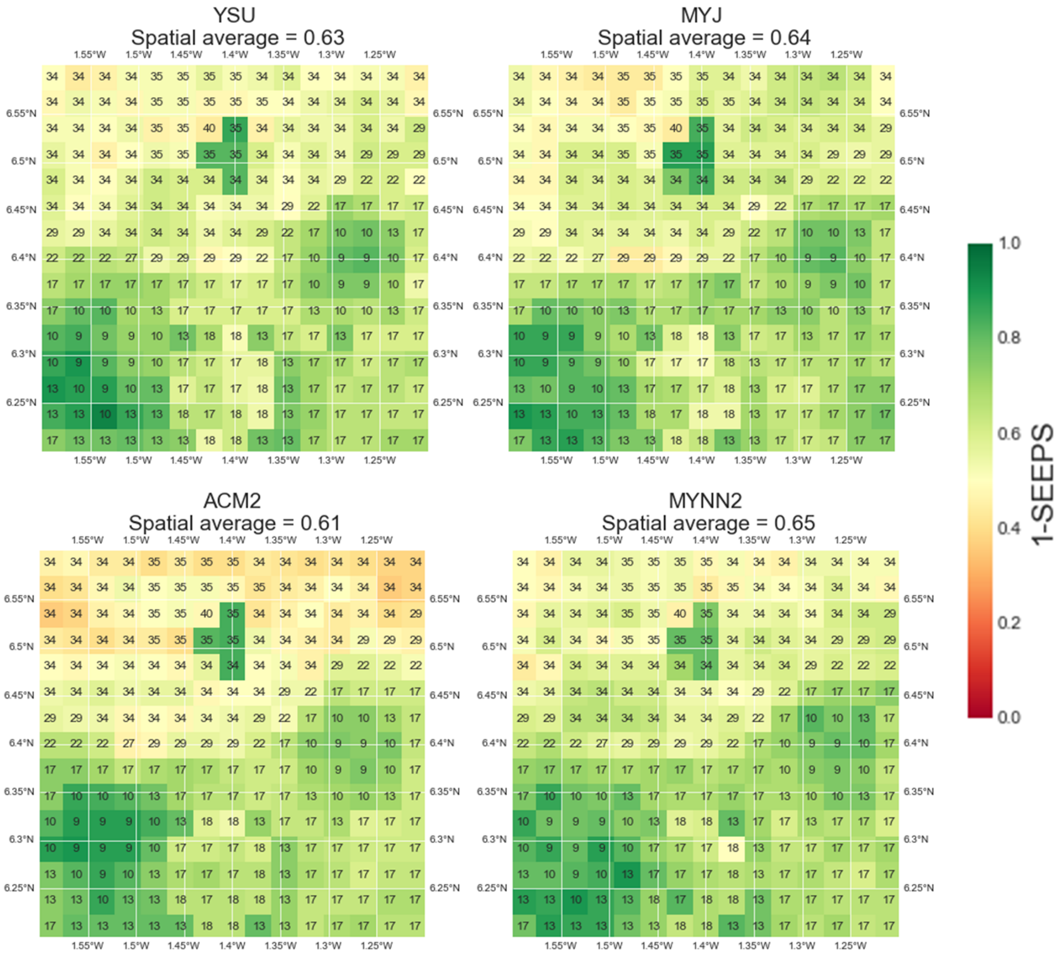


Figure 4.6: SEEPS, with a t1 value of 0.25 mm. The numbers inside each grid cell depict the rainfall amount, ranging from 0 (no rainfall on all simulation days) to 100 (heavy rainfall on all simulation days). The spatial average is the mean SEEPS skill score for the entire validation domain.

Several key points emerge from this analysis. Firstly, all four configurations demonstrate relatively good Skill scores, with most grid cells showing a Skill score of 0.6 or higher, resulting in a spatial average also above 0.6. Interestingly, all configurations perform better in the southern part of the domain compared to the northern part. This spatial distribution correlates with the rainfall amounts: grid cells with high SEEPS Skill scores exhibit the lowest rainfall amounts, whereas grid cells with lower SEEPS scores experience higher rainfall amounts. This indicates that all schemes are relatively better at predicting dry weather or small amounts of rainfall. Notably, the ACM2 scheme performs the worst in the north among all PBL schemes. One exception is the four green grid cells around 6.5N and 1.4W, which show consistently high scores across all PBL schemes. These grid points are in close proximity to the TA00312 station, but do not overlap its exact location. A closer look at the land surface reveals that these grid points are located above Lake Bosomtwe. The differing surface conditions, compared to the surrounding area, provide a steady and consistent moisture source, which could lead to more predictable atmospheric conditions that are easier for WRF to capture. This, in turn, results in better SEEPS scores.

However, when observing the t1=1 result maps (Appendix B.2 Figure B.4) the green cross largely disappears, suggesting that many rainfall observations at these grid points are near the threshold values. When the threshold is set to t1=0.25, more rainfall is classified as light/heavy, whereas with a threshold of t1=1, more rainfall is classified as dry/light. This shift results in lower SEEPS scores at these four points, indicating a sensitivity of SEEPS scores to the chosen threshold.

This tendency to perform worse with more rainfall is further evidenced by the error counts of the three rain categories in Table 4.1. The total counts reflect the frequency of each category across the entire domain (i.e., all simulation days and all grid cells combined). Below that, the error counts represent how many times the configuration incorrectly predicted each category. For both t1 threshold values, all PBL schemes exhibit a lower error fraction in the dry category than in the light and heavy categories. ACM2, in particular, excels at predicting dry weather but performs the worst in the light and heavy rain categories. This is why ACM2's average score improves significantly when t1 is set to 1, as there are substantially more "dry" counts (8442) compared to light (2584) and heavy counts (224), giving the dry category more weight.

Table 4.1: SEEPS Error Fractions for Different Configurations and Precipitation Categories

(a) t1 value of $0.25\,\mathrm{mm}$

	YSU			**MYJ**			**ACM2**			**MYNN2**		
	Dry	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**
Total counts	6417	4511	322	6417	4511	322	6417	4511	322	6417	4511	322
Error counts	3008	2885	224	3407	2786	231	1606	3548	246	2965	2890	197
Error fraction	0.469	0.640	0.696	0.531	0.618	0.717	0.250	0.787	0.764	0.462	0.641	0.612
Mean (1-SEEPS)		0.63			0.64			0.61			0.65	

(b) t1 value of $1\,\mathrm{mm}$

	YSU			**MYJ**			**ACM2**			**MYNN2**		
	Dry	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**	**Dry**	**Light**	**Heavy**
Total counts	8442	2584	224	8442	2584	224	8442	2584	224	8442	2584	224
Error counts	3296	1697	177	3683	1638	179	1726	2075	198	3242	1716	171
Error fraction	0.39	0.66	0.79	0.44	0.63	0.80	0.20	0.80	0.88	0.38	0.66	0.76
Mean (1-SEEPS)		0.71			0.72			0.73			0.73	

The MYJ scheme appears to be the least effective for forecasting dry weather but performs better in the light rainfall category. YSU generally shows the lowest performance but maintains a fair distribution across different categories. MYNN2 also demonstrates a balanced performance distribution and has the highest average score overall.

Considering the agricultural context of this study, it can be argued that mainly Table b (with related Figure B.4 in Appendix B.2) is of particular interest. In this table, ACM2 improves from being the worst-scoring PBL scheme to being among the best, alongside MYNN2. However, considering all categories as equally important in this study — since the identification of the transition from no rain to rain is of most importance and not only the prediction of dry weather — MYNN2 appears to be the best-performing scheme for this particular metric.

MAE with Moving Threshold

To better understand how the error values evolve, a moving threshold is applied to the model output. All rainfall values, both observed and modeled, above the threshold are masked, so the error metrics are only applied to rainfall values below the threshold. This is done for thresholds ranging from 0 up to 35 mm/day (and infinity, with no threshold applied). Figure 4.7 depicts this evolution of the absolute error of daily rainfall for the YSU simulations. The lower rainfall amounts, reflected in the maps associated with the initial threshold values, especially give rise to errors at the borders and northern part of the simulations. After a threshold of 15 mm, the absolute error rises more significantly in the lower center of the domain.

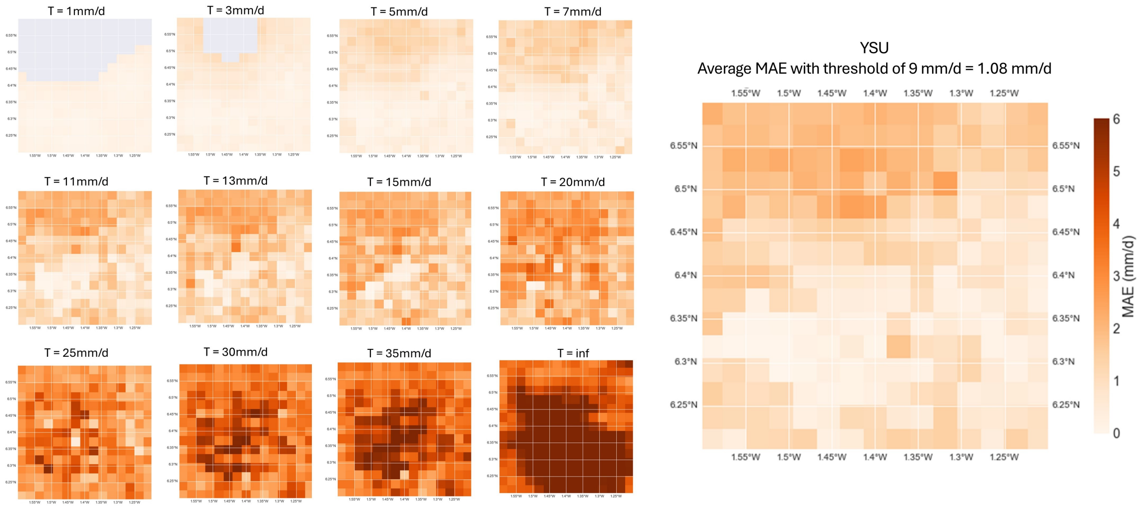


Figure 4.7: The MAE of YSU over the whole validation domain, where all observed and modeled daily rainfall above threshold 'T' is masked.

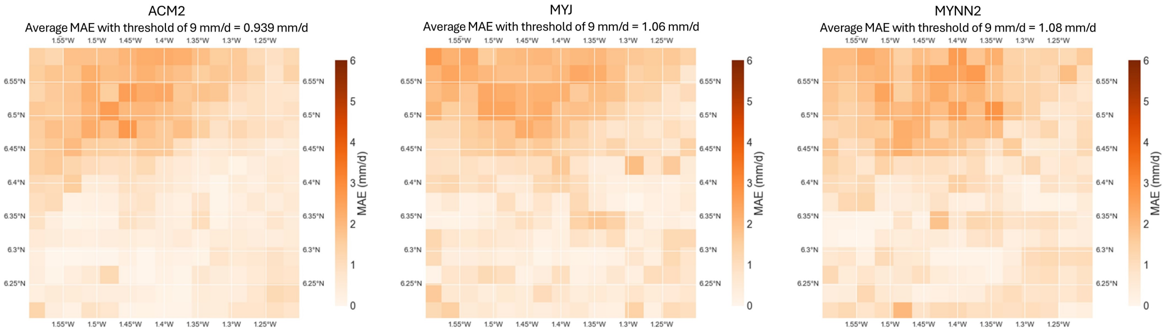


Figure 4.8: The MAE with all observed and modeled daily rainfall above 9 mm/d masked for PBL schemes ACM2, MYNN2 and MYJ.

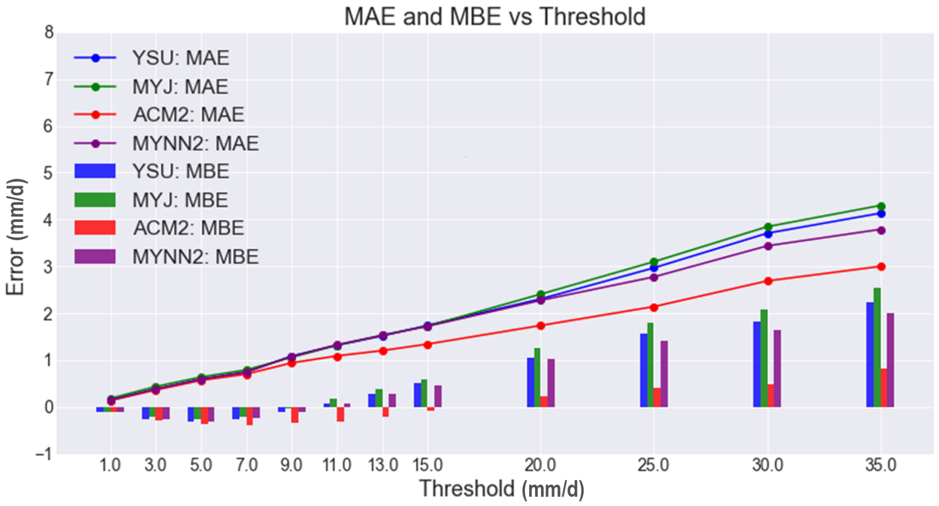


Figure 4.9: The mean absolute error (MAE) and mean bias error (MBE) of all four PBL schemes, averaged across the validation domain, and plotted against the threshold values 'T' where daily rainfall above is masked.

Examining the MAE map specifically for rainfall up to 9 mm. This value is close to the second threshold value of the SEEPS method, so you would expect to see similar patterns in the map, which is indeed the case: In the northern part of the domain, where SEEPS scored lower values for all PBL schemes, the 9 mm MAE map actually shows higher errors, which support each other's results (since higher errors would mean lower SEEPS skill scores). The same holds for the other PBL schemes (Figure 4.8).

Interestingly, this method reveals that the average MAE values of all PBL schemes do not differ significantly from each other when rainfall above 9 mm is masked. ACM2 still shows the lowest values, but the other schemes are only 0.1 mm higher, which is minimal. They diverge more significantly when heavy rainfall is included.

This trend is further highlighted when both errors are plotted against the thresholds, as shown in Figure 4.9 (up to a 35mm threshold). Several points stand out: first, the MAE values start to deviate from each other from the 9 mm threshold onwards, as previously described. More interestingly, the error values for these lower precipitation values are primarily due to underestimation, as indicated by the MBE error bars below zero. This suggests that the absolute error values at T=9 mm, visualized in Figures 4.7 and 4.8, and the lower SEEPS scores shown in Figure 4.6 in the northern part of the domain, are mostly due to underestimation. This is probably as a results of the higher precipitation amounts in this part of the domain.

The transition from underestimation to overestimation occurs between the 9 to 11 mm threshold for MYNN2, YSU, and MYJ, which again corresponds closely to the second SEEPS threshold, encompassing all three categories but excluding heavy rain events. In other words, for rainfall up to a value of 9/11 mm, the MBE is lowest for these PBL schemes, indicating the most balanced output. This outcome differs for ACM2, where the switch only happens after the 15 mm threshold, meaning ACM2 underestimates a larger portion of the total rainfall.

4.3.3. Summary

Summarizing the results, it can be stated that the models still perform quite poorly in terms of Mean Bias Error (MBE) and Mean Absolute Error (MAE) if no threshold is applied to mask certain rainfall amounts. Among the PBL schemes, ACM2 consistently showed the lowest biases, while MYJ performed the worst.

However, the categorical metric SEEPS provided very promising results, indicating that all schemes are effective in distinguishing between dry, light, and heavy precipitation. Using a threshold (t1) of 0.25, the MYNN2 scheme outperformed all others. When this threshold was adjusted to 1mm, both MYNN2 and ACM2 showed the highest scores, with MYNN2 being the most balanced in its error counts.

Considering these performance differences, deeper analyses were conducted to understand the source of these discrepancies. By implementing a moving threshold on the rainfall values, it became clear that as heavier rainfall amounts were excluded from the data, the error values rapidly decreased and converged. The divergence in performance primarily began after a threshold of 9mm. At this threshold, all PBL schemes except ACM2 generally transitioned from underestimation to overestimation. For ACM2 simulations, this transition occurred after a threshold value of 15mm.

4.4. Conclusion

This chapter aims to answer the second sub-question of this research: *"Which model configuration is the optimal choice considering the context of this study?"* By focusing on a smaller domain within the Forest area, which contains a substantial amount of agricultural land, and by increasing the number of simulations to enhance the robustness of the validation, the objective is to select the most suitable PBL scheme for furthering this research.

Considering these new set of results, it can be stated that the MYJ scheme generally performs the worst. The choice of which PBL scheme to use for the final set of simulations lies between ACM2 and MYNN2. ACM2 consistently shows the lowest error values, as observed in the first set of experiments, while MYNN2 performs best in terms of SEEPS score. Both PBL schemes appear adequate for implementation in the final simulations.

The decision hinges on whether to prioritize the lowest quantitative errors and accurate prediction of dry weather, albeit with the risk of underestimating rainfall amounts, or to focus on accurately predicting light to heavy rainfall amounts and the triggering of 'rain', with a potential risk of overestimating (for no-rain cases), which could result in some false alarms. Each scheme has its own benefits and drawbacks, offering different advantages in predicting the onset of the rainy season.

Given the agricultural context and the previously defined local onset definitions, which use various thresholds to define the onset, the model's performance heavily depends on the specific definition used. What is certain is that the transition from dry weather to rain is of utmost importance. Therefore, the decision was made to conduct the final set of experiments using the MYNN2 scheme. This scheme appeared to be the best at predicting multiple rainfall types (both light and heavy), making it more robust towards multiple definitions, despite resulting in some higher quantitative errors. It is important to note that this choice is somewhat subjective and not the only suitable option.

One aspect not yet captured in these validations, but worth examining, is the model's performance across separate forecast days (FD) and lead times (D+). Currently, all forecast days are combined and rated equally. However, analyzing each day individually can illuminate the evolution of the model's performance over its lead times, providing valuable insights into the model's strengths and weaknesses.

5

Experimental Setup 3

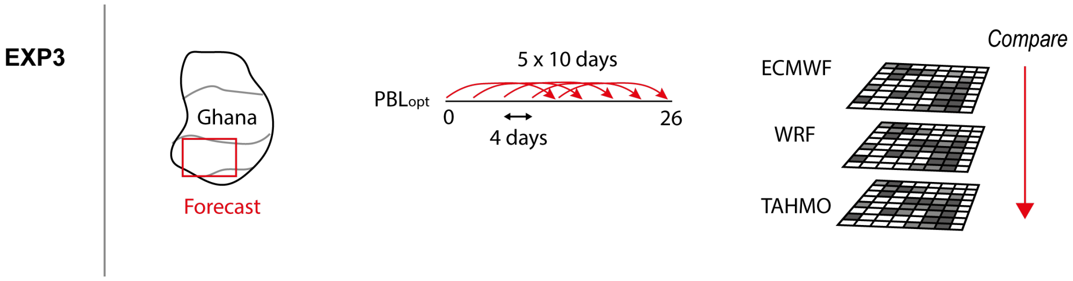


Figure 5.1: Sketch Experimental Setup 3

This chapter presents the final experimental setup of this research, addressing research sub-question 3: *"How does the selected model configuration perform compared to the precipitation forecasts of ECMWF IFS?"* This comparison will be made by validating both WRF and ECMWF across a large domain in the Forest zone.

5.1. Model Configuration

5.1.1. Time and Spatial Domain

The last model configuration focuses on a large domain over the agricultural sites in the forest zone, covering an area of 300 by 306 km, with a grid spacing of 3 km. The Forest domain was selected again due to its higher density of TAHMO stations and its extensive agricultural land coverage [30]. This larger area allows for more validation points, resulting in more accurate final validation. The model's boundary conditions, updated every six hours, and initial conditions are again forced by the ECMWF Integrated Forecast System with a resolution of 0.125 degrees. For the PBL scheme, the MYNN2 scheme was chosen as it showed the most promising results in the previous model experiments. Given the convection-solving resolution of 3 km, the cumulus convection scheme is again turned off.

Five simulations of 10 days each were executed, covering the period from February 10, 00:00h to March 8, 00:00h, 2018, with the onset occurring around February 19, according to L. Occelli [28]. This means the initial conditions (IC) of each simulation were shifted four days forward compared to the previous simulation:

- Simulation 1: February 10 - February 20
- Simulation 2: February 14 - February 24
- Simulation 3: February 18 - February 28
- Simulation 4: February 22 - March 4
- Simulation 5: February 26 - March 8

The longer time period allows us to capture more rainfall events, resulting in a more robust final validation. This validation will be conducted in only a part of the model domain, visualized in Figure 5.2, to account for enough relaxation zone.

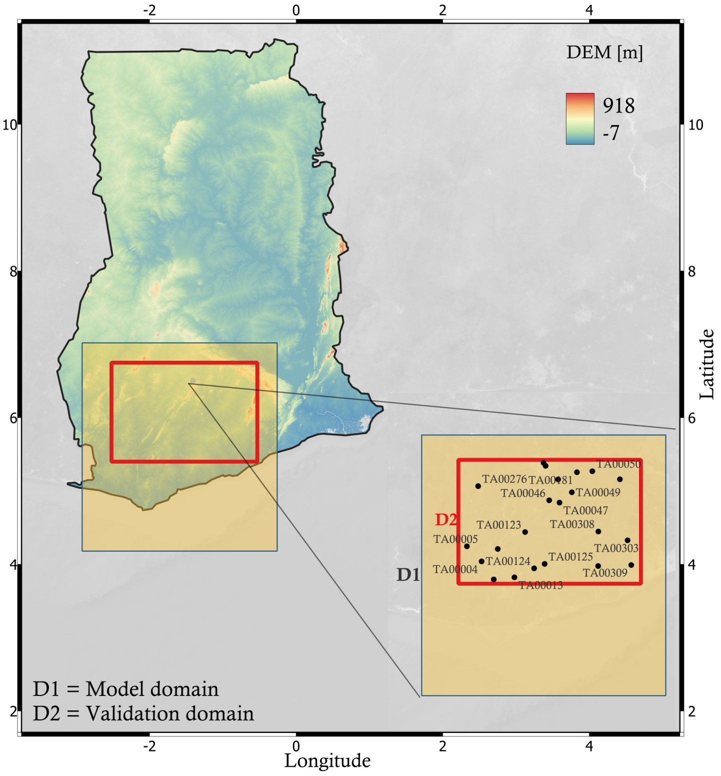


Figure 5.2: Model domain and validation domain for the final set of experiments displayed on a Digital Elevation Model (DEM) map, downloaded from [27].

5.2. Validation Method

5.2.1. Validation Data

This time, the research will validate both the WRF model output and the ECMWF IFS precipitation data against the TAHMO data, allowing for a direct comparison between the models. It is to note that the WRF model is not entirely independent of ECMWF, as it uses ECMWF data to force the model, also known as dynamicaL downscaling. In this context, this phase of the research investigates whether the dynamical downscaling of WRF adds value to precipitation forecasts or not.

The TAHMO data consists of interpolated data from 22 stations, which were selected based on the amount of available data during the simulation period, applying a threshold of 95%. The stations were interpolated using the Inverse Distance Weighting method, as in experimental setup 2, but this time with a power coefficient of 3. This adjustment was made because the validation domain covers a larger area, resulting in greater distances between stations. Using a higher power value ensures that the interpolation gives more weight to the nearest known points, providing more accurate estimations. A lower power parameter might produce a surface that is too smooth and not representative of the actual conditions.

5.2.2. Metrics

Only daily precipitation is validated this time. This will include some general analyses, such as the probability density function of all three datasets to compare their distributional shape, and a statistical heatmap that distinguishes between the various forecast days (FD) from the first day (FD1) to the tenth day (FD10). This approach reveals how the performance of both models evolves as the forecast time period increases, with the hypothesis that performance will be better for shorter lead times. This distinction will also be applied in a SEEPS validation. This time, only the t1 threshold value of 1mm will be considered, with the associated t2 value of 12.59mm.

The 0.125° grid data from ECMWF is downscaled to the WRF resolution of 3 km. The reason for downscaling ECMWF rather than upscaling WRF is to preserve the high-resolution qualities of the data. Upscaling WRF would aggregate the data, potentially resulting in the loss of important information.

Design Final Metric

A concluding metric has been designed, aligned with the primary objective of this research. Referring to the onset definitions by Fitzpatrick et al. [13], all local onset definitions utilize daily precipitation as input. The study investigated three local onset definitions:

- Yamada et al. (2013) [38]: Date when 6-day average rainfall exceeds 2 mm/day.
- Marteau et al. (2009) [39]: First rainy day (> 1 mm) of two consecutive rainy days with total rainfall > 20 mm and no 7-day period with total rainfall less than 5 mm in the following 20 days.
- Omotosho et al. (2000) [40]: The first 3 or 4 rainy days (> 10 mm) with no more than 7 days between them.

Although all definitions use daily rainfall as input, they differ in their onset criteria. Marteau's definition focuses on the continuation of rainfall, Omotosho's depends on heavy rainfall, and Yamada's uses the start of local rainfall as the identifier. The study found Marteau's and Yamada's definitions to be the most useful for local farmers and likely correlated.

However, these are just three of the many existing definitions. Amekudzi et al. [14] used the cumulative percentages of mean annual rainfall and rainy days to define the onset and cessation dates of the rainy season by identifying points of maximum positive and negative curvature, respectively. Gbangou et al. [41] tested two different definitions: one using a given percentage of cumulative seasonal rainfall and an agronomic onset definition based on absolute rainfall values, defined as the date when the average 4-day rainfall exceeds 10 mm, starting from March 1.

One common aspect of these definitions is that they all use rainfall amounts over a certain number of consecutive days. The threshold and density of the rain vary, aligning with different purposes.

The final metric of this paper utilizes the definition of Yamada et al., which is based on a 6-day average rainfall, but now with a moving threshold ranging from 2 mm/day to 12 mm/day instead of a fixed 2 mm/-

day. This approach allows for greater flexibility and adaptability to other possible definitions, rather than relying solely on this specific one, which is also depended on quite a low rainfall amount. Additionally, using a moving threshold demonstrates the model's robustness to different indicators. Besides that, this method was selected because it can be applied to 10-day forecasts. To be clear, the definition is not used to determine the onset of 2018, but rather as a threshold value at each grid cell of the models and observations. The binary output (threshold met: yes or no) is then validated using a confusion matrix, which shows true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives are considered the most important metric, as they indicate how often the model accurately crosses the threshold, which could hypothetically represent an onset threshold.

5.3. Results

5.3.1. Probability Density Function

Figure 5.3 illustrates the probability density function of daily precipitation (mm/day) on a logarithmic scale for five simulations of ECMWF and WRF, along with observed values from TAHMO over the same period. Due to the application of the logarithmic scale, all values below 0.1 mm/day are set to 0.1 mm/day, as it is not possible to take the log of zero.

Several key observations can be made. First, ECMWF consistently exhibits the highest peaks at very low precipitation amounts across all simulations, indicating a high frequency of days with minimal rainfall. In contrast, WRF shows lower densities at the low precipitation amounts, with a broader distribution extending towards higher precipitations amounts, suggesting more variability in rainfall. TAHMO presents the lowest and broadest peaks, indicating the most even distribution of daily precipitation values among the three datasets. This comparison highlights that ECMWF's rainfall distribution is not well-aligned with observed values, while WRF more closely resembles the observations, albeit not perfectly. ECMWF predicts more frequent occurrences of very low precipitation compared to TAHMO and WRF.

Additionally, the tails of the distributions differ significantly among the three datasets. The WRF model exhibits a tail that extends above 10^2 mm/day, indicating that it occasionally predicts more extreme precipitation values. In contrast, ECMWF's tail only extends slightly above 10^1 mm/day, suggesting no prediction of extreme precipitation. TAHMO's tail reaches values between the two, approaching but not surpassing the 10^2 mm/day mark.

Furthermore, all three datasets show the highest peaks in simulations 1 and 2, compared to other simulations, indicating that the initial part of the temporal coverage experiences less rainfall than the later days.

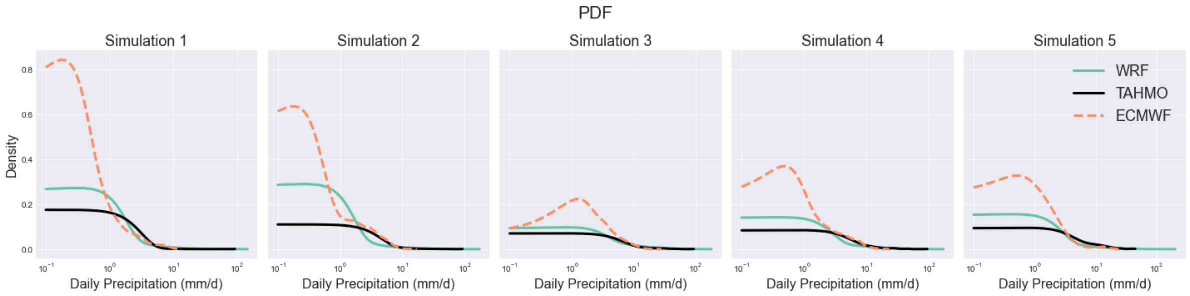


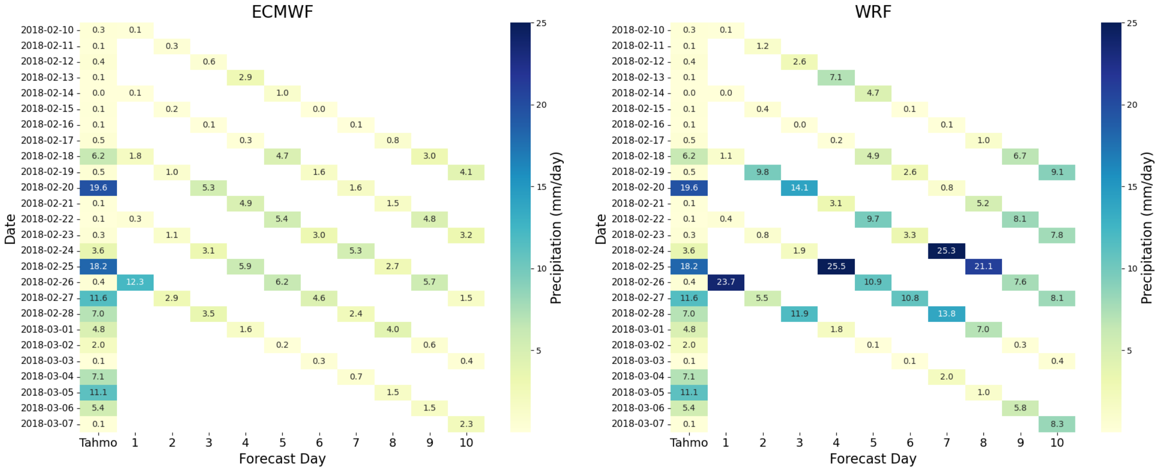
Figure 5.3: Logarithmic-scale Probability Density Function (PDF) of daily precipitation (mm/day) for all five ECMWF and WRF simulations, with observed data over the same time periods.

5.3.2. Heatmap of Daily Precipitation

The low precipitation amounts predicted by ECMWF are further evidenced by Figure 5.4. This heatmap displays the aggregated daily precipitation values across the entire domain for each date (y-axis) and forecast day (FD) (x-axis). Specifically, the upper diagonal line corresponds to simulation 1, covering the period from February 10th (FD1) to February 19th (FD10) (ending at 00:00 on February 20th, with FD10 representing the daily precipitation for February 19th from midnight to midnight). Similarly, the lower diagonal line represents simulation 5, which spans from February 26th to March 7th.

Not yet looking in detail, what stands out are the different color palettes. WRF shows a variation of color ranging from light yellow (very low precipitation) to dark blue (heavy precipitation), while ECMWF only displays light yellow to light green colors, indicating low to moderate precipitation. The observational column, however, also shows a more diverse color palette with low to heavy rainfall, which is more aligned with WRF. Examining the rows representing different days allows for a detailed analysis of the timing and accuracy of the forecasts, particularly on days with more rainfall.

- 2018-02-10 to 2018-02-17: Low precipitation amounts are measured and similarly predicted by both models. However, WRF FD4 and FD5 slightly overestimated around February 13th and 14th, while ECMWF showed less overestimation.



(a) Averaged daily precipitation over the entire model domain for the five ECMWF simulations, plotted separately for each forecast day.

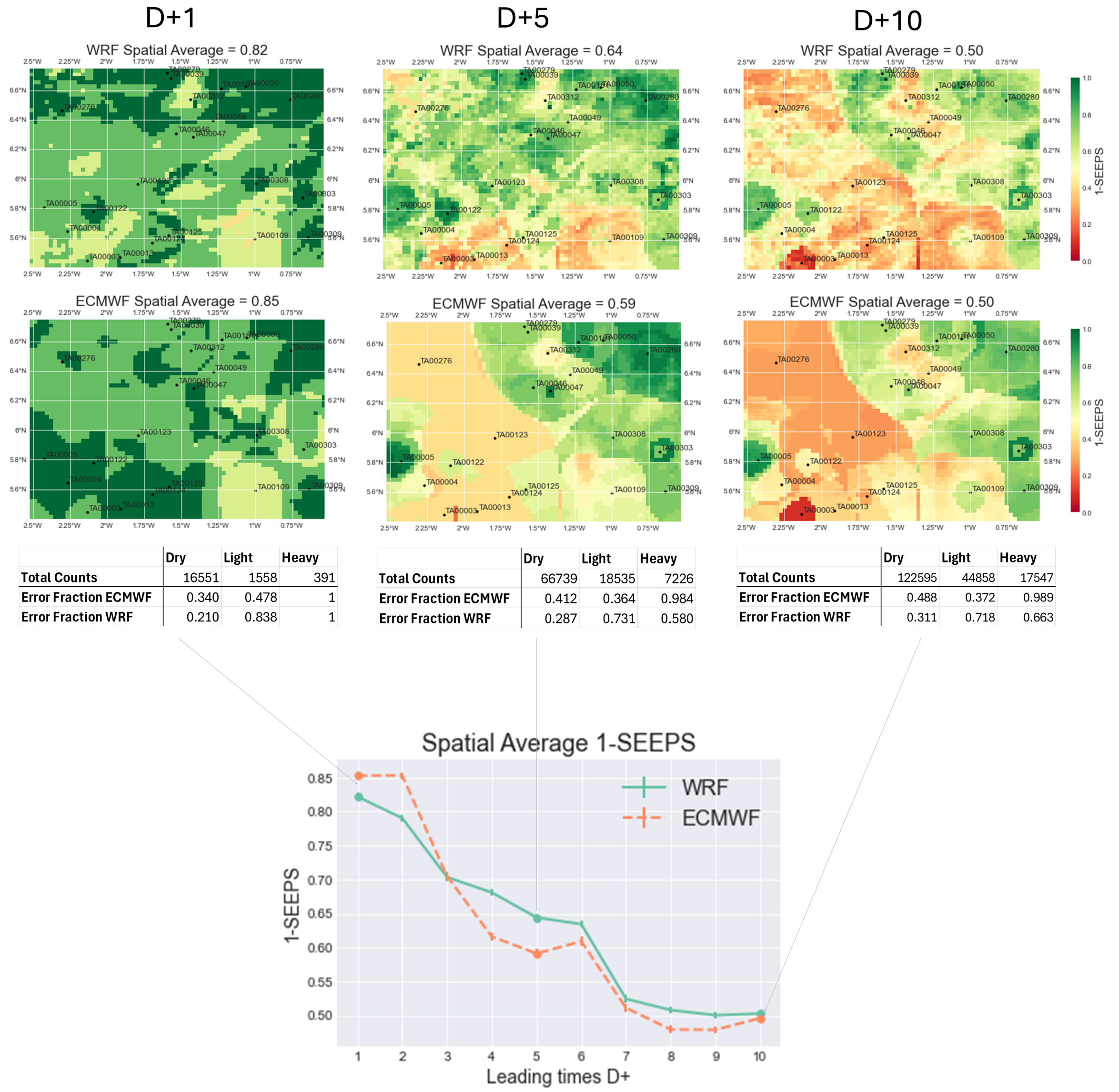
(b) Averaged daily precipitation over the entire model domain for the five WRF simulations, plotted separately for each forecast day.

Figure 5.4: Heatmap of the aggregated daily precipitation of ECMWF and WRF compared to the observed daily rainfall.

- 2018-02-18: Both models made accurate predictions at FD5 and FD9, with WRF performing slightly better. Interestingly, FD1 from both models was less accurate.
- 2018-02-20: WRF FD3 accurately captured the high precipitation amounts, whereas ECMWF FD3 did not. Both model's FD7 did not capture the peak.
- 2018-02-21 to 2018-02-23: WRF's earlier forecast days (FD1 to FD3) captured lower precipitation amounts well, but later forecast days tended to overestimate. ECMWF showed similar trends but with less overestimation.
- 2018-02-25: The precipitation peak was captured by WRF but not by ECMWF. However, WRF also showed significant overestimation the day before and after this date, suggesting timing issues or an inability to capture a sudden decrease. ECMWF showed moderate rainfall around this date. The biggest overestimation by both models occurred at FD1 on February 26th, just after the rainfall peak.
- 2018-02-27 to 2018-02-28: Heavy to moderate rainfall was captured well by WRF for all forecast days, but not by ECMWF. WRF slightly overestimated on February 28th, while ECMWF underestimated.
- 2018-03-01 to 2018-03-07: Low to moderate rainfall was measured, with one peak around March 5th. However, both models (from FD5 onward) mostly showed low precipitation amounts, indicating underestimation.

5.3.3. SEEPS

To further extend the lead time analysis, the SEEPS skill score of daily precipitation for all different lead times (D+) is calculated. Specifically, the SEEPS skill score for D+1 uses 5 days as comparable input, while D+10 uses 50 days (5 forecasts of 10 days each) as comparable input. Figure 5.5 depicts three of the ten spatial distributions of the 1-SEEPS score over the validation domain, specifically for lead times 1, 5, and 10. The total counts of the different categories are shown beneath the maps, along with the error fractions for ECMWF and WRF. To clearly illustrate the evolution of the skill for both models, the spatial average 1-SEEPS score is plotted against the lead times.



	Dry	Light	Heavy
Total Counts	16551	1558	391
Error Fraction ECMWF	0.340	0.478	1
Error Fraction WRF	0.210	0.838	1

	Dry	Light	Heavy
Total Counts	66739	18535	7226
Error Fraction ECMWF	0.412	0.364	0.984
Error Fraction WRF	0.287	0.731	0.580

	Dry	Light	Heavy
Total Counts	122595	44858	17547
Error Fraction ECMWF	0.488	0.372	0.989
Error Fraction WRF	0.311	0.718	0.663

Figure 5.5: SEEPS skill score for different leading times (D+), with associated error fractions and spatial averages of both models.

As shown, the skill score decreases as the lead time increases, which is expected. Both models start with a skill score between 0.82 and 0.85 for D+1, decreasing to around 0.5 by forecast day D+10. This trend is also visible in the spatial maps, with D+1 showing much more green areas compared to later days. The graph also shows that ECMWF is on average performing better up to leading time D+2 while after that WRF shows higher performance values.

The maps reveal that the spatial pattern of ECMWF quickly aligns with the interpolation patterns. This is particularly evident in the top left corner, where a distinct dividing line emerges between stations TA00276 and TA00123 on the left and a group of other stations on the right. This pattern is not observed for WRF at D+5 but starts to appear at D+10. However, this is not the only spatial border exhibiting the interpolation pattern. At D+10 for ECMWF, almost all borders seem to result from data interpolation.

This occurs because ECMWF operates at a lower resolution and is interpolated to fit the grid. Consequently, large parts of the domain reflect average values from ECMWF output, leading to minimal variation between grid cells. Therefore, the spatial patterns quickly resemble those of TAHMO. In contrast, WRF provides output for each grid cell with more detail, resulting in unique spatial patterns.

What becomes clear from the error fraction depicted in the tables, is that WRF outperforms ECMWF in categorizing dry and heavy rain, but ECMWF outperforms WRF in classifying light precipitation.

5.3.4. Final Metric

The final graph (Figure 5.6) illustrates the changes in the normalized confusion matrix variables - True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) - with respect to the 6-day average rainfall threshold. The normalization involves dividing the TP and FN counts by the total observed positive counts, and the TN and FP by the total observed negative counts. This means that the sum of TP and FN, and TN and FP, always add up to one. A normalized score of 1 indicates that all observations are correctly forecasted (a perfect score), while a score of 0 means none of the observations are modeled correctly.

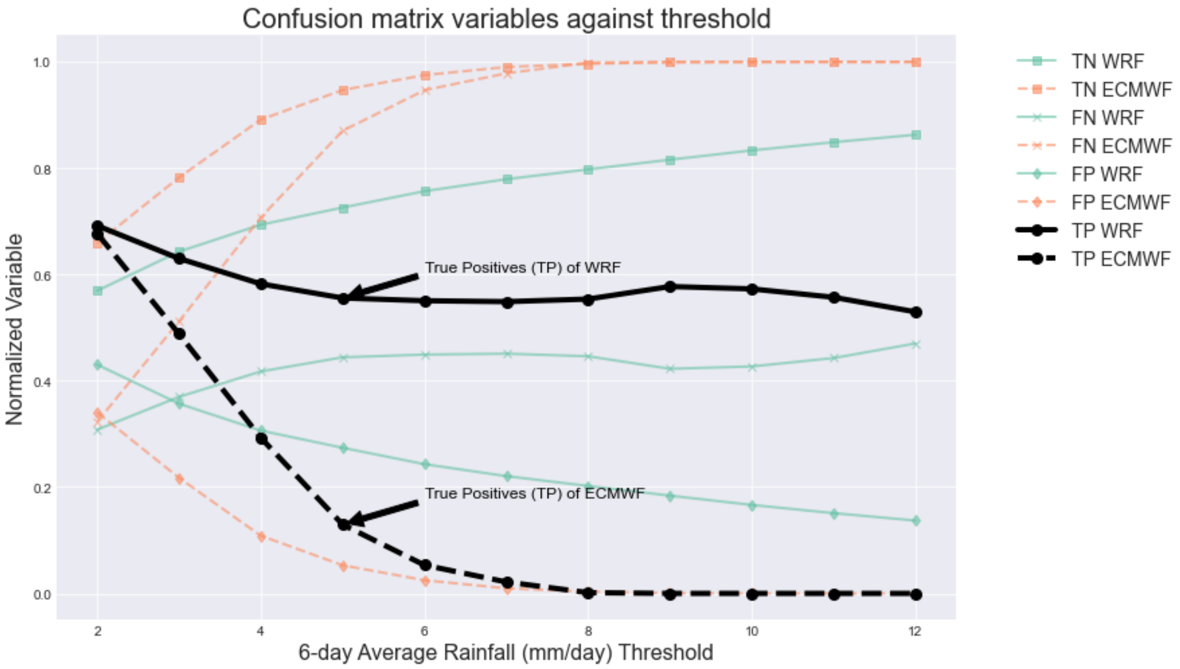


Figure 5.6: Final Metric: All normalized confusion matrix variables plotted against 6-day average rainfall (mm/d) threshold values.

The most notable observation is the consistent performance of WRF in predicting positive outcomes across various thresholds, as indicated by the relatively constant true positive (TP) rate. In contrast, ECMWF experiences a rapid decline as the threshold increases. Both models start with similar scores, WRF at 0.69 and ECMWF at 0.67. However, as the threshold increases, ECMWF's TP score drops to 0 at around an 8mm threshold, indicating its inability to predict a 6-day average rainfall of 8 mm/day. On the other hand, WRF consistently forecasts correctly across all thresholds, even up to a 6-day average of 12 mm/day with a TP score of 0.53. This suggests that ECMWF struggles to predict heavier rainfall, reinforcing earlier indications.

Looking at true negative (TN) predictions—instances where the models correctly identified non-threshold events—ECMWF consistently outperforms WRF across all thresholds. From a threshold of 7mm/day, ECMWF achieves a perfect score, while WRF's scores increase but never reach 1. This indicates that WRF produces some false positives, indicating a constant overestimation.

5.3.5. Summary

To summarize all notable validation metrics. Firstly, it is evident that ECMWF struggles to accurately predict variations in rainfall, particularly heavier precipitation amounts. The probability density function (pdf) and daily precipitation heatmap display a skewed distribution with a high frequency of low rainfall amounts for ECMWF, whereas WRF and TAHMO exhibit a more diverse and broader distribution.

Additionally, the spatial analyses indicated by the SEEPS maps showed that ECMWF results in less detailed spatial variability, potentially masking local variations and specific details. However, this could also be an advantage for ECMWF in certain scenarios, considering that station data is interpolated and may not accurately represent local phenomena. In contrast, WRF captures these local details, potentially leading to higher errors. Furthermore, the SEEPS score decreases for both models as the leading times increase. However, the differences in sample sizes between leading times may also influence these results.

In the final metric, it is evident that ECMWF is less robust when it comes to different thresholds for 6-day average rainfall. It performs well in correctly predicting lower amounts, but it quickly drops to a True Positive score of 0 when thresholds are increased. On the other hand, WRF consistently performs well across all thresholds.

5.4. Conclusion

This final experimental setup aims to address the last sub-question of this research: ”How does the selected model configuration perform compared to the precipitation forecasts of ECMWF IFS?” This comparison involves examining both models’ outputs against the interpolated TAHMO data over a wide area within the Forest zone.

The validation process revealed strengths and weaknesses in both models. ECMWF’s performance was weak in capturing spatial and temporal variability in rainfall, with a skewed distribution towards low rainfall amounts, making it unable to predict heavier precipitation. On the other hand, it demonstrated excellent performance in predicting light precipitation amounts and showed relatively good SEEPS scores. Meanwhile, WRF consistently overestimated, leading to more false positive values in the final validation metric. Nevertheless, WRF displayed promising results in almost all metrics, accurately capturing different types of rainfall at a finer scale, a quality that ECMWF lacked. This highlights the added value of WRF as a downscaling model.

6

Discussion

Several uncertainties appeared during this research, which will be discussed in this chapter. At first, the uncertainty in the reliability of the observational data. Despite applying a selection procedure to eliminate stations with low data amounts, some station locations consistently performed poorly, as was indicated in Chapter 3 in Figure 3.8. This suggests potential errors in the data itself, such as recording zero values when there is actually precipitation. However, this is a problem that is hard to verify. Comparing stations that are very close to each other or by visualizing precipitation time series in more detail, can provide helpful insights in the correctness of the data but still does not give a 100% guarantee.

One other common challenge in meteorology is the difficulty of accurately verifying local rainfall events. High-resolution models such as WRF provide detailed rainfall information on a very small scale (in this research 3km). However, this level of detail requires a dense network of observational data to accurately capture these phenomena. As mentioned in Chapter 2, it is possible that rain clouds may not align perfectly with the observation stations, leading to inaccuracies even after interpolation. Since the stations only represent a small portion of the actual conditions, the interpolated data may not accurately reflect what is happening between the stations. This can result in instances where intense localized rainfall occurs between stations, leading to artificially high errors in the interpolated data.

Additionally to these common uncertainties, some doubts on the implemented metrics are described below:

- Concluding table experimental setup 1 (Table 3.4): The concluding table of experimental setup 1, where the PBL schemes are rated in comparison to each other for each zone and the entire country, is somewhat subjective. The rating was based on considering all metrics equally. However, it could be questioned whether some metrics should be weighted more heavily than others. Additionally, since the rating was done in relation to each other, the overall performance was not clearly demonstrated, and its general usefulness was not highlighted.
- Model domain in experimental setup 2 (Figure 4.2): The model domain implemented in experimental setup 2 was relatively small, consisting of around 30x40 grid points. Although this allowed for multiple simulations to be conducted without requiring extensive computational time, there was a concern regarding the potential impact of the forcing data on the model output. It is necessary for the model domain to be sufficiently large to develop its own dynamics. A domain that is too small could potentially disrupt this process, leading to biased results. However, the outputs exhibited considerable variances among different configurations, indicating that the model generated its own dynamics and (partly) mitigated this issue.
- SEEPS experimental setup 3 (Figure 5.5): In the last experimental setup, the Stable Equitable Error in Probability Space (SEEPS) was calculated for various lead times, leading to significant variations in sample sizes between D+1 and D+10. This disparity could have significantly impacted the results. For example, the D+1 sample might not encompass the same range of weather predictions as the D+10 sample with 50 validation pairs, potentially affecting the stability of the

SEEPS scores. This uncertainty prompted a consideration to solely focus on individual forecast days (FD). This approach revealed consistently high scores (ranging from 0.89 to 0.59, ending at F10 with scores of 0.77 for WRF and 0.88 for ECMWF). However, since all scores were based on only 5 days, this method was not preferred as it could question the representativeness of the score. It also raises doubts about the statistical accuracy of the scores for the initial lead times presented in this paper (e.g., D+1, and D+2). Additionally, in experimental setups 2 and 3, there was an overlap in simulation time periods, resulting in the use of the same dates for verification multiple times. However, since these dates belonged to different simulations, they can still be considered independent from each other.

Furthermore, all experimental setups were limited by a relatively low number of simulations, due to time constraints and limited computational power. This limitation led to less statistically robust results. If there had been the opportunity to increase the number of simulations, encompassing a wider time range and capturing more weather phenomena, the results would have been more reliable. Additionally, a larger dataset would have made the aggregated and average values more meaningful.

As a concluding point, it must be emphasized that this entire research is based on numerous design decisions, which have steered the results in a specific direction, making them highly reliant on these decisions. While these choices were well-reasoned and aligned with the research's ultimate goal, their subjectivity cannot be overlooked. This underscores the importance of increasing the amounts of tests and research on the model, to obtain a complete indications of its capabilities.

Despite these limitations, the research outputs still provided valuable insights. The findings shed light on certain aspects and capabilities of WRF that enhance existing research and reveal new potential applications of the model. The final conclusion on the main research question, along with suggestions for further research and potential operational use of the model, are presented in the next and final chapter.

7

Conclusion and Recommendations

7.0.1. Conclusion

Coming to the end, this research aimed to answer the main research question: *"What is the performance of the Weather Research and Forecasting (WRF) model in forecasting the local onset of West Africa's annual rainy season in Ghana?"* To address this question, three experimental setups were designed, each focusing on one of the three sub-questions formulated to guide the research. All three experiments centered on precipitation amounts, the most commonly used indicator for local onset definitions, and on a time period around the timing of the associated regional onset, which is crucial for predicting the transition from dry to rainy conditions.

As the research progressed, it became evident that there is no straightforward answer to this question as it depends on a lot of factors. Previous literature suggested that the predictability of the rainy season onset depends not only on the model's performance but also on the onset definition used [41]. Different definitions require different model performance qualities, which can vary significantly. The choice of definition depends on the final purpose or which definition works best in a specific context.

However, the results revealed promising insights into WRF's precipitation forecasting performance. Four different model configurations based on four Planetary Boundary Layer (PBL) schemes—ACM2, YSU, MYJ, and MYNN2—were tested, showing varying results across different zones of Ghana. This highlighted the importance of the spatial context in selecting a PBL scheme. On average, ACM2 showed the lowest quantitative errors, followed by YSU, making it the most recommended and 'safe' PBL schemes for Ghana in general. However, the experiment also showed that for the Forest zone, where a lot of agricultural grounds and weather stations are present, making it one of the most important zones, there was no PBL scheme clearly outperforming others. This suggested further research for this zone. Detailed analysis focusing on smaller domains within the Forest zone, provided additional insights. It was found that MYNN2 was the best and most balanced in predicting classified rainfall within this zone, even though its quantitative errors were again higher than those of ACM2.

Additionally, the added value and advantages of WRF as a downscaling model over global models like ECMWF were clearly demonstrated in the final chapter. Generally, ECMWF was unable to capture local variations in rainfall, showing aggregated values that failed to capture heavy rainfall amounts. In contrast, WRF could capture detailed temporal and spatial variations, although the exact placement could be off. WRF consistently performed well in predicting true positives of 6-day average rainfall over multiple threshold values, whereas ECMWF only managed to predict low rainfall amounts.

Based on these findings, it can be concluded WRF shows significant potential for general weather prediction modeling and, more specifically, for predicting the onset of rainfall in Ghana. Its adaptability to different locations within the country, leading to more accurate forecasts, is a valuable feature that can be further refined. Furthermore, the research indicated that the model offers more detailed and localized weather information, capturing a wide range of rainfall patterns that the widely used ECMWF global model failed to capture.

7.0.2. Recommendations

For further research the following ideas are suggested:

- As mentioned in the Discussion, increasing the number of simulations would yield more statistically robust and meaningful outcomes. Extending the simulation period to cover three months of weather data, from the onset dates in the south of Ghana in February/March to the north in May, would add significant value. This extension addresses the limitation of this study and is highly recommended for future research.
- Direct comparison of resolution effects: Another potential research objective which is more in parallel to this study and not so much an expansion, is to study the direct effects of implementing a convection-solving resolution (e.g., 3 km or smaller) compared to a non-resolving resolution (e.g., 10 km). This study assumed that the higher resolution, where the cumulus (CU) scheme could shut off, would yield better results. However, whether this was indeed the case by implementing a direct comparison to verify this assumption was not conducted.
- Verification based on local onset definitions: Choosing specific local onset definitions at the start of the research and basing all verification metrics on these definitions could provide a clearer indication of the model's performance for each definition. This approach would allow for recommendations on which configuration works best for each definition.
- Comparing other PBL schemes with ECMWF: In the last experimental setup, only the model output of the MYNN2 scheme was compared with ECMWF's forecast. However, it would be valuable to also compare ACM2 and YSU, two PBL schemes that showed promising results, especially in experimental setup 1, with the global model.

Regarding the potential operational usage of the WRF model, several key points have emerged. Results indicated that WRF's primary advantage over the global model ECMWF is its ability to achieve very high resolution and to capture heavier rainfall amounts. Although the research focused on forecasting rain at the onset of the rainy season, it has demonstrated WRF's high potential for predicting extreme weather events on a local scale—a notable deficiency in ECMWF's operational forecast.

The implementation of WRF to predict the onset of the rainy season depends on the context and the preferred onset definition. For definitions based on relatively low rainfall amounts, ECMWF's data may be sufficient, as it has shown comparable, if not superior, performance compared to WRF. Additionally, higher resolutions than 9 km is most of the time not necessary for onset prediction. However, if definitions based on various types and amounts of rainfall are required, WRF is recommended due to its robust capabilities in predicting multiple rainfall types.

One important consideration is the computational time required by WRF. This depends heavily on the desired simulation time, resolution, and spatial domain. If longer predictions are needed for a relatively large domain at high resolution, the computation time can increase significantly, potentially posing challenges for operational model runs. For instance, in the last experimental setup of this research, which involved 5 simulations on a 100 x 100 gridpoint domain at a 3km resolution for 10 forecast days, the computation time was approximately 9 days (using 4 parallel processes). In the initial experimental setup, a single simulation took an average of 17 hours to complete. This emphasizes the importance of designing the model with careful considerations.

In conclusion, it should be mentioned that the research and implementation possibilities of WRF are endless. This study has only scratched the surface of WRF's capabilities and serves as one of many informative reports on the model's performance and qualities. As such, this contribution may have shed light on new aspects and hopefully inspired further research endeavors.

References

[1] Thomas Allen, Phillip Heinrigs, and Inhoi Heo. “AGRICULTURE, FOOD AND JOBS IN WEST AFRICA Club SAHEL AND WEST AFRICA Secretariat”. In: (Apr. 2018). DOI: 10.1787/dc152bc0-en. URL: https://doi.org/10.1787/dc152bc0-en.

[2] Ghana Statistical Service. “GCA_National Report_2020”. In: (June 2019).

[3] FAO. *Land use indicators*. URL: https://www.fao.org/faostat/en/#data/.

[4] Nick van de Giesen, Rolf Hut, and John Selker. “ The Trans□African Hydro□Meteorological Observatory (TAHMO) ”. In: *WIREs Water* 1.4 (July 2014), pp. 341–348. ISSN: 2049-1948. DOI: 10.1002/wat2.1034.

[5] Arlindo Meque et al. “Numerical weather prediction and climate modelling: Challenges and opportunities for improving climate services delivery in Southern Africa”. In: *Climate Services* 23 (Aug. 2021). ISSN: 24058807. DOI: 10.1016/j.cliser.2021.100243.

[6] Yan Wang et al. “Verification of operational numerical weather prediction model forecasts of precipitation using satellite rainfall estimates over Africa”. In: *Meteorological Applications* 30.1 (Jan. 2023). ISSN: 14698080. DOI: 10.1002/met.2112.

[7] Anke Kniffka et al. “An evaluation of operational and research weather forecasts for southern West Africa using observations from the DACCIWA field campaign in June–July 2016”. In: *Quarterly Journal of the Royal Meteorological Society* 146.728 (Apr. 2020), pp. 1121–1148. ISSN: 1477870X. DOI: 10.1002/qj.3729. URL: https://www.researchgate.net/publication/368452457_Verification_of_operational_numerical_weather_prediction_model_forecasts_of_precipitation_using_satellite_rainfall_estimates_over_Africa.

[8] Mekonnen Gebremichael, Haowen Yue, and Vahid Nourani. “The Accuracy of Precipitation Forecasts at Timescales of 1–15 Days in the Volta River Basin”. In: *Remote Sensing* 14.4 (Feb. 2022). ISSN: 20724292. DOI: 10.3390/rs14040937. URL: https://www.researchgate.net/publication/338140314_An_Evaluation_of_Operational_and_Research_Weather_Forecasts_for_Southern_West_Africa_using_Observations_from_the_DACCIWA_Field_Campaign_in.

[9] Imoleayo E. Gbode et al. “Sensitivity of different physics schemes in the WRF model during a West African monsoon regime”. In: *Theoretical and Applied Climatology* 136.1-2 (Apr. 2019), pp. 733–751. ISSN: 14344483. DOI: 10.1007/s00704-018-2538-x.

[10] Imoleayo E. Gbode et al. “Simulation of wet and dry West African monsoon rainfall seasons using the Weather Research and Forecasting model”. In: *Theoretical and Applied Climatology* 138.3-4 (Nov. 2019), pp. 1679–1694. ISSN: 14344483. DOI: 10.1007/s00704-019-02912-x. URL: https://www.researchgate.net/publication/334097669_Simulation_of_wet_and_dry_West_African_monsoon_rainfall_seasons_using_the_Weather_Research_and_Forecasting_model.

[11] P. Moudi Igri et al. “Assessing the performance of WRF model in predicting high-impact weather conditions over Central and Western Africa: an ensemble-based approach”. In: *Natural Hazards* 93.3 (Sept. 2018), pp. 1565–1587. ISSN: 15730840. DOI: 10.1007/s11069-018-3368-y.

[12] N. M.J. Hall and P. Peyrillé. “Dynamics of the West African monsoon”. In: *Journal De Physique. IV : JP*. Vol. 139. Dec. 2006, pp. 81–99. ISBN: 2868839436. DOI: 10.1051/jp4:2006139007.

[13] Rory G.J. Fitzpatrick et al. “The West African monsoon onset: A concise comparison of definitions”. In: *Journal of Climate* 28.22 (2015), pp. 8673–8694. ISSN: 08948755. DOI: 10.1175/JCLI-D-15-0265.1.

[14] Leonard K. Amekudzi et al. “Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana”. In: *Climate* 3.2 (June 2015), pp. 416–434. ISSN: 22251154. DOI: 10.3390/cli3020416.

[15] WRF. *WRF Users Guide Documentation*. URL: `https://www2.mmm.ucar.edu/wrf/users/wrf_users_guide/build/html/physics.html`.

[16] Richard Yao Kuma Agyeman et al. “Optimal Physics Parameterization Scheme Combination of the Weather Research and Forecasting Model for Seasonal Precipitation Simulation over Ghana”. In: *Advances in Meteorology* 2017 (2017). ISSN: 16879317. DOI: `10.1155/2017/7505321`.

[17] Emmanouil Flaounas, Sophie Bastin, and Serge Janicot. “Regional climate modelling of the 2006 West African monsoon: Sensitivity to convection and planetary boundary layer parameterisation using WRF”. In: *Climate Dynamics* 36.5-6 (Mar. 2011), pp. 1083–1105. ISSN: 09307575. DOI: `10.1007/s00382-010-0785-3`.

[18] Erik Noble, Leonard M. Druyan, and Matthew Fulakeza. “The sensitivity of WRF daily summertime simulations over West Africa to alternative parameterizations. Part II: Precipitation”. In: *Monthly Weather Review* 145.1 (2017), pp. 215–233. ISSN: 15200493. DOI: `10.1175/MWR-D-15-0294.1`.

[19] Agostino N. Meroni et al. “Sensitivity of some African heavy rainfall events to microphysics and planetary boundary layer schemes: Impacts on localised storms”. In: *Quarterly Journal of the Royal Meteorological Society* 147.737 (Apr. 2021), pp. 2448–2468. ISSN: 1477870X. DOI: `10.1002/qj.4033`.

[20] Camille Le Coz and Nick Van De Giesen. “Comparison of rainfall products over sub-saharan africa”. In: *Journal of Hydrometeorology* 21.4 (Apr. 2020), pp. 553–596. ISSN: 15257541. DOI: `10.1175/JHM-D-18-0256.1`.

[21] Jiyang Tian et al. “Numerical rainfall simulation with different spatial and temporal evenness by using a WRF multiphysics ensemble”. In: *Natural Hazards and Earth System Sciences* 17.4 (Apr. 2017), pp. 563–579. ISSN: 16849981. DOI: `10.5194/nhess-17-563-2017`.

[22] William C Skamarock et al. *A Description of the Advanced Research WRF Model Version 4*. Tech. rep. URL: `http://library.ucar.edu/research/publish-technote`.

[23] Jonathan E. Pleim. “A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing”. In: *Journal of Applied Meteorology and Climatology* 46.9 (Sept. 2007), pp. 1383–1395. ISSN: 15588424. DOI: `10.1175/JAM2539.1`.

[24] Mikio Nakanishi and Hiroshi Niino. “An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog”. In: *Boundary-Layer Meteorology* 119.2 (May 2006), pp. 397–407. ISSN: 00068314. DOI: `10.1007/s10546-005-9030-8`.

[25] Song-You Hong, Yign Noh, and Jimy Dudhia. *A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes*. Tech. rep. 2006.

[26] Zaviša I Janjić. *Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model*. Tech. rep. 2001.

[27] *OpenTopography DEM*. URL: `https://opentopography.org/`.

[28] Lorenzo Occelli. *Detection and Prediction of the Rainy Season Onset in West Africa*. Tech. rep. Utrecht Univeristy, July 2023.

[29] Aldo S. Moya-Álvarez et al. “Influence of PBL parameterization schemes in WRF_ARW model on short - range precipitation’s forecasts in the complex orography of Peruvian Central Andes”. In: *Atmospheric Research* 233 (Mar. 2020). ISSN: 01698095. DOI: `10.1016/j.atmosres.2019.104708`.

[30] Ghana Forestry Commission’s Resource Management Support Centre (RMSC). *Ghana National Land Use*. 2021. URL: `https://ghana-national-landuse.knust.ourecosystem.com/interface/`.

[31] Tobias Becker, Peter Bechtold, and Irina Sandu. “Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations”. In: *Quarterly Journal of the Royal Meteorological Society* 147.741 (Oct. 2021), pp. 4388–4407. ISSN: 1477870X. DOI: `10.1002/qj.4185`.

[32] John H. Marsham et al. “The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations”. In: *Geophysical Research Letters* 40.9 (May 2013), pp. 1843–1849. ISSN: 00948276. DOI: `10.1002/grl.50347`.

[33] Ramchandra Karki et al. "Quantifying the added value of convection-permitting climate simulations in complex terrain: A systematic evaluation of WRF over the Himalayas". In: *Earth System Dynamics* 8.3 (July 2017), pp. 507–528. ISSN: 21904987. DOI: `10.5194/esd-8-507-2017`.

[34] Kouakou Kouadio et al. "Does convection-permitting simulate better rainfall distribution and extreme over Guinean coast and surroundings?" In: *Climate Dynamics* 55.1-2 (July 2020), pp. 153–174. ISSN: 14320894. DOI: `10.1007/s00382-018-4308-y`.

[35] Rachel C. North, Marion P. Mittermaier, and Sean F. Milton. "Using SEEPS with a TRMM-Derived Climatology to Assess Global NWP Precipitation Forecast Skill". In: *Monthly Weather Review* 150.1 (Jan. 2022), pp. 135–155. ISSN: 15200493. DOI: `10.1175/MWR-D-20-0347.1`.

[36] T Haiden et al. *Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score*. Tech. rep. 2012. URL: `http://www.ecmwf.int/publications/`.

[37] Sinan Jasim Hadi and Mustafa Tombul. "Comparison of Spatial Interpolation Methods of Precipitation and Temperature Using Multiple Integration Periods". In: *Journal of the Indian Society of Remote Sensing* 46.7 (July 2018), pp. 1187–1199. ISSN: 09743006. DOI: `10.1007/s12524-018-0783-1`.

[38] Tomohito J. Yamada et al. "Seasonal variation of land-atmosphere coupling strength over the West African monsoon region in an atmospheric general circulation model". In: *Hydrological Sciences Journal* 58.6 (2013), pp. 1276–1286. ISSN: 02626667. DOI: `10.1080/02626667.2013.814914`.

[39] Romain Marteau, Vincent Moron, and Nathalie Philippon. "Spatial coherence of Monsoon onset over Western and Central Sahel (1950-2000)". In: *Journal of Climate* 22.5 (Mar. 2009), pp. 1313–1324. ISSN: 08948755. DOI: `10.1175/2008JCLI2383.1`.

[40] J. Bayo Omotosho. "Onset of thunderstorms and precipitation over Northern Nigeria". In: *International Journal of Climatology* 10.8 (Dec. 1990), pp. 849–860. ISSN: 0899-8418. DOI: `10.1002/joc.3370100807`.

[41] Talardia Gbangou et al. "Seasonal variability and predictability of agro-meteorological indices: Tailoring onset of rainy season estimation to meet farmers' needs in Ghana". In: *Climate Services* 14 (Apr. 2019), pp. 19–30. ISSN: 24058807. DOI: `10.1016/j.cliser.2019.04.002`.

A

Source Code

A.1. Point Validation

The following code includes verification functions for point comparison, focusing on temporal and spatial biases.

Temporal Analysis

Listing A.1: Temporal Analysis Script

```python

# TEMPORAL ANALYSIS
def temp_stats_points(WRF, stations_data):
    # Import WRF and stations dataframes
    error_df = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)
    error_squared = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)
    error_absolute = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)

    for time in WRF.index:
        cal_error = WRF.loc[time] - stations_data.loc[time]
        error_df.loc[time] = cal_error
        error_squared.loc[time] = cal_error**2
        error_absolute.loc[time] = abs(cal_error)

    # Max, Min, Mean WRF
    maximum = WRF.max()
    minimum = WRF.min()
    mean = WRF.mean()

    # Mean Bias Error
    MBE = error_df.mean(axis=0)

    # Mean Squared Error
    MSE = error_squared.mean(axis=0)

    # Mean Absolute Error
    MAE = error_absolute.mean(axis=0)

    # Normalized MAE
    relative = stations_data.mean()
    MAE_norm = MAE / relative
    inf_index = np.isinf(MAE_norm)
    MAE_norm[inf_index] = np.NaN

    return MBE, MSE, MAE, MAE_norm, error_df, mean, maximum, minimum
```

Temporal Standard Deviation (SD) and Correlation

Listing A.2: Temporal SD and Correlation Script

```
def SD_temp(WRF, stations_data):
    SD_model = WRF.std(0)
    SD_obs = stations_data.std(0)

    covariances = pd.DataFrame(data=None, index=['cov', 'r'], columns=WRF.columns)
    for column in WRF.columns:
        cov = WRF[column].cov(stations_data[column])
        covariances.loc['cov', column] = cov
        covariances.loc['r', column] = cov / (SD_model.loc[column] * SD_obs.loc[column])

    return SD_model, SD_obs, covariances.iloc[1,:]
```

Temporal Bias Summary

Listing A.3: Temporal Bias Summary Script

```
def temporal_biases(zones, names, stats):
    temporal_bias_df = pd.DataFrame(data=None, index=names, columns=['Mean␣[mm/d]', 'MBE␣[mm/d]', 'MSE␣[mm/d]²', 'MAE␣[mm/d]', 'r_temporal'])
    for i in range(len(zones)):
        temporal_bias_df.iloc[i, 0] = stats[0][zones[i].index].mean()
        temporal_bias_df.iloc[i, 1] = stats[1][zones[i].index].mean()
        temporal_bias_df.iloc[i, 2] = stats[2][zones[i].index].mean()
        temporal_bias_df.iloc[i, 3] = stats[3][zones[i].index].mean()
        temporal_bias_df.iloc[i, 4] = stats[4][zones[i].index].mean()

    return temporal_bias_df
```

Spatial Analysis

Listing A.4: Spatial Analysis Script

```
# SPATIAL ANALYSIS
def stats_points(WRF, stations_data):
    error_df = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)
    error_squared = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)
    error_absolute = pd.DataFrame(data=None, index=WRF.index, columns=stations_data.columns)

    for time in WRF.index:
        cal_error = WRF.loc[time] - stations_data.loc[time]
        error_df.loc[time] = cal_error
        error_squared.loc[time] = cal_error**2
        error_absolute.loc[time] = abs(cal_error)

    MBE = error_df.mean(axis=1)
    MSE = error_squared.mean(axis=1)
    MAE = error_absolute.mean(axis=1)

    return MBE, MSE, MAE
```

Spatial Standard Deviation (SD) and Correlation

Listing A.5: Spatial SD Script

```
def SD(WRF, stations_data):
    SD_model = WRF.std(1)
    SD_obs = stations_data.std(1)

    covariances = pd.DataFrame(data=None, index=WRF.index, columns=['cov', 'r'])
    for index in WRF.index:
        cov = WRF.loc[index].cov(stations_data.loc[index])
        covariances.loc[index, 'cov'] = cov
        covariances.loc[index, 'r'] = cov / (SD_model.loc[index] * SD_obs.loc[index])

    return SD_model, SD_obs, covariances.iloc[:, 1]
```

Spatial Bias Summary

The averaged spatial MBE, MSE, and MAE are the same values as the averaged temporal MBE, MSE, and MAE since they are aggregated over both dimensions. However, the spatial and temporal correlations differ due to the different covariances.

Listing A.6: Spatial Bias Summary Script

```python
def spatial_biases(zones, names, stats):
    spatial_bias_df = pd.DataFrame(data=None, index=names, columns=['MBE␣[mm/d]', 'MSE␣[mm/d]²', 'MAE␣[mm/d]', 'r_spatial'])

    spatial_bias_df.iloc[:, 0] = stats[0].mean(0)
    spatial_bias_df.iloc[:, 1] = stats[1].mean(0)
    spatial_bias_df.iloc[:, 2] = stats[2].mean(0)
    spatial_bias_df.iloc[:, 3] = stats[3].mean(0)

    return spatial_bias_df
```

A.2. IDW Interpolation Function

The following code defines the Inverse Distance Weighting (IDW) interpolation function. This function includes a maximum distance threshold to limit the influence of distant points.

Listing A.7: IDW Interpolation Function

```python
from scipy.spatial import cKDTree
import numpy as np

def IDW(points, data, grid, power, max_dist):
    tree = cKDTree(points)
    dist, idx = tree.query(grid, k=len(points))

    # Apply maximum distance threshold
    mask = dist > max_dist
    dist[mask] = np.inf

    #Calculates the weight
    weight = 1 / (dist**power)

    # Apply weights and calculate interpolated values
    weighted_vals = weight * data[idx]
    interpolated = np.sum(weighted_vals, axis=1) / np.sum(weight, axis=1)

    # Handle cases where sum of weights is 0 (no nearby points)
    interpolated[np.sum(weight, axis=1) == 0] = np.nan

    return interpolated
```

A.3. Automating WRF

A.3.1. Bash script

This bash script is specifically designed for this research, but can easily be adapted for other research experiments.

Listing A.8: WRF Model Automation Script

```bash
# Set the base directory for WRF and WPS
WRF_DIR=/home/weatherimpact/WRFmodel/WRF/run/
WPS_DIR=/home/weatherimpact/WRFmodel/WPS/
DATA_DIR=/home/weatherimpact/WRFmodel/Build_WRF/ECMWFdata/2018_02_forecast_lastexp/
BASE_RUN_DIR=/home/weatherimpact/WRFmodel/runs
PBL_SCHEME='MYNN2' #Change for each experiment!

# Set the maximum number of executions
MAX_RUNS=10  #subject to change
echo "Maximum␣number␣of␣executions␣set␣to:␣$MAX_RUNS"

```

```
# Get the last run date from the file and calculate the new start date
if [ ! -f last_run_date.txt ]; then
    echo "Please␣create␣last_run_date.txt␣with␣the␣initial␣start␣date␣in␣the␣format␣YYYY-MM-DD"
    exit 1
fi

LAST_RUN_DATE=$(cat last_run_date.txt)

# Add one day to the last run date
START_DATE=$(date -d"${LAST_RUN_DATE}␣+1␣day" +"%Y-%m-%d_00:00:00")
START_DATE_FORMATTED=$(date -d"${LAST_RUN_DATE}␣+1␣day" +"%Y-%m-%d")
END_DATE=$(date -d"${LAST_RUN_DATE}␣+10␣day" +"%Y-%m-%d_00:00:00")
END_DATE_FORMATTED=$(date -d"${LAST_RUN_DATE}␣+10␣day" +"%Y-%m-%d")

# Read the current run count from a file, default to 0 if the file doesn't exist
if [ ! -f run_count.txt ]; then
    echo "0" > run_count.txt
fi
RUN_COUNT=$(cat run_count.txt)

# Check if the run count has reached the maximum number of executions
if [ "$RUN_COUNT" -ge "$MAX_RUNS" ]; then
    echo "Maximum␣number␣of␣executions␣reached.␣Exiting."
    exit 0
fi

# Increment the run count
RUN_COUNT=$((RUN_COUNT + 1))
echo "Incremented␣run␣count:␣$RUN_COUNT"

# Extract the year, month, day, and hour for downloading ECMWF data
YEAR=$(date -d "$START_DATE_FORMATTED" +"%Y")
MONTH=$(date -d "$START_DATE_FORMATTED" +"%m")
DAY=$(date -d "$START_DATE_FORMATTED" +"%d")
HOUR=$(date -d "$START_DATE_FORMATTED" +"%H")

YEAR_END=$(date -d "$END_DATE_FORMATTED" +"%Y")
MONTH_END=$(date -d "$END_DATE_FORMATTED" +"%m")
DAY_END=$(date -d "$END_DATE_FORMATTED" +"%d")
HOUR_END=$(date -d "$END_DATE_FORMATTED" +"%H")

# Create directory for new run
RUN_DIR=${BASE_RUN_DIR}/RUN_IFSforecast_${START_DATE_FORMATTED}_$PBL_SCHEME
mkdir -p $RUN_DIR

# Link the new ECMWF data to WPS
cd $WPS_DIR
ln -sf ${DATA_DIR}ecmwf_sfc_data_${YEAR}${MONTH}${DAY}.grib .
ln -sf ${DATA_DIR}ecmwf_pl_data_${YEAR}${MONTH}${DAY}.grib .

# WPS
cd $WPS_DIR
# Set new date in namelist.wps
sed -i "s/start_date.*/start_date␣=␣'$START_DATE',/" $WPS_DIR/namelist.wps
sed -i "s/end_date.*/end_date␣=␣'$END_DATE',/" $WPS_DIR/namelist.wps
sed -i "s|opt_output_from_metgrid_path.*|opt_output_from_metgrid_path␣=␣'${RUN_DIR}',|" $WPS_DIR/namelist.wps
# Run geogrid
./geogrid.exe
# Ungrib surface data
./link_grib.csh ecmwf_sfc_data_${YEAR}${MONTH}${DAY}.grib
sed -i "s/prefix.*/prefix␣=␣'SFC',/" $WPS_DIR/namelist.wps
./ungrib.exe
# Ungrib pressure data
./link_grib.csh ecmwf_pl_data_${YEAR}${MONTH}${DAY}.grib
sed -i "s/prefix.*/prefix␣=␣'PL',/" $WPS_DIR/namelist.wps
./ungrib.exe
# Run metgrid
./metgrid.exe

```

```
# WRF
cd $WRF_DIR
ln -sf $RUN_DIR/met_em* $WRF_DIR
# Set namelist.input dates
sed -i "s/run_days.*/run_days␣=␣5,/" $WRF_DIR/namelist.input  # BE CAREFUL WITH THIS! Run hours are unchanged
sed -i "s/start_year.*/start_year␣=␣$YEAR,/" $WRF_DIR/namelist.input
sed -i "s/start_month.*/start_month␣=␣$MONTH,/" $WRF_DIR/namelist.input
sed -i "s/start_day.*/start_day␣=␣$DAY,/" $WRF_DIR/namelist.input
sed -i "s/start_hour.*/start_hour␣=␣$HOUR,/" $WRF_DIR/namelist.input

sed -i "s/end_year.*/end_year␣=␣$YEAR_END,/" $WRF_DIR/namelist.input
sed -i "s/end_month.*/end_month␣=␣$MONTH_END,/" $WRF_DIR/namelist.input
sed -i "s/end_day.*/end_day␣=␣$DAY_END,/" $WRF_DIR/namelist.input
sed -i "s/end_hour.*/end_hour␣=␣$HOUR_END,/" $WRF_DIR/namelist.input

# Run executables
./real.exe
mpirun -np 4 ./wrf.exe
mv wrfout* $RUN_DIR
mv wrfrst* $RUN_DIR

# Save the current run date and run count for the next iteration
cd /home/weatherimpact/WRFmodel/
echo "$START_DATE_FORMATTED" > last_run_date.txt
echo "$RUN_COUNT" > run_count.txt

at now <<ENDMARKER
/home/weatherimpact/WRFmodel/automating.sh
ENDMARKER
```

A.4. SEEPS

The code below provides the calculation of the SEEPS skill score at each grid-point, together with the associated error fractions.

Listing A.9: SEEPS Calculation Script

```
"TA00047␣seems␣to␣be␣the␣best␣from␣2018-2023"
plt.plot(stations4_data['TA00047'])

selected_station = stations4_data['TA00047']
mask = (selected_station.index <= '2023-01-01␣00:00')
selected_station = selected_station[mask]

# Make it daily
h24_prec = selected_station.rolling(window=24).sum()
h24_prec = h24_prec.dropna()
# Sort values and calculate CDF
sorted_rain = h24_prec.sort_values()
cdf = sorted_rain.rank(method='max') / len(sorted_rain)

# Dry threshold
t1 = 1
p1 = cdf[sorted_rain >= t1].min()

# Calculate p2 and p3 and find thresholds
remain = 1 - p1
p2 = remain * (2 / 3)
p2_added = p1 + p2

t2 = np.interp(p2_added, cdf, sorted_rain)
t3 = sorted_rain.max()

"Make␣the␣scoring␣matrix"
Smatrix = [
    [0, 1 / (1 - p1), 4 / (1 - p1)],
    [1 / p1, 0, 3 / (1 - p1)],
    [1 / p1 + 3 / (2 + p1), 3 / (2 + p1), 0]
]
```

```
Smatrix = np.array(Smatrix)

"Import␣and␣process␣data"
# Importing interpolated TAHMO datasets
path_tahmo = "C:\\Users\\Sophie.Verheugd\\Documents\\export_2506\\TAHMO_modified\\"
file_daily = "interpolated_daily.nc"
tahmo = xr.open_dataset(path_tahmo + file_daily)

# Import WRF daily
base_path = "C:\\Users\\Sophie.Verheugd\\Documents\\export_2506\\3km_MYNN2\\daily\\"
file = "5daily_prec_{date}.nc"
start_date = "2018-02-12"
end_date = "2018-02-21"
PBL = 'MYNN2'
date_range = pd.date_range(start_date, end_date)

def import_datasets(date_str):
    file_path = base_path + file.format(date=date_str)
    data = xr.open_dataset(file_path)
    return data

wrf = []
for date in date_range:
    date_str = date.strftime("%Y-%m-%d")
    data = import_datasets(date_str)
    wrf.append(data)

tahmo.sel(Time=wrf[0].Time)

counts = pd.DataFrame(data=None, index=['Total', 'errors'], columns=['dry', 'light', 'heavy'])
"Calculate␣SEEPS"
def categorize(value, t1, t2):
    if value < t1:
        cat = 0
    elif t1 <= value < t2:
        cat = 1
    elif value >= t2:
        cat = 2
    return cat

def SEEPS(tahmo, wrf, t1, t2, matrix):
    categorized_grid = xr.DataArray(np.zeros_like(wrf, dtype=int),
                                    dims=['Time', 'latitude', 'longitude'],
                                    coords={'Time': wrf.Time,
                                            'latitude': wrf.latitude,
                                            'longitude': wrf.longitude})
    categorized_grid_obs = xr.DataArray(np.zeros_like(wrf, dtype=int),
                                        dims=['Time', 'latitude', 'longitude'],
                                        coords={'Time': wrf.Time,
                                                'latitude': wrf.latitude,
                                                'longitude': wrf.longitude})
    counts = pd.DataFrame(data=0, index=['Total', 'errors'], columns=['dry', 'light', 'heavy'])
    tahmo = tahmo.sel(Time=wrf.Time)
    for t_idx, time in enumerate(wrf.Time):
        for lat_idx, lat in enumerate(wrf.latitude):
            for lon_idx, lon in enumerate(wrf.longitude):
                value_tahmo = tahmo.sel(Time=time, latitude=lat, longitude=lon).item()
                value_wrf = wrf.sel(Time=time, latitude=lat, longitude=lon).item()

                # Determine categories
                position_tahmo = categorize(value_tahmo, t1, t2)
                position_wrf = categorize(value_wrf, t1, t2)

                # Calculate SEEPS
                S = (1 / 2) * matrix[int(position_wrf), int(position_tahmo)]

                # Count errors
                if position_tahmo == 0:
                    counts.iloc[0, 0] += 1
```

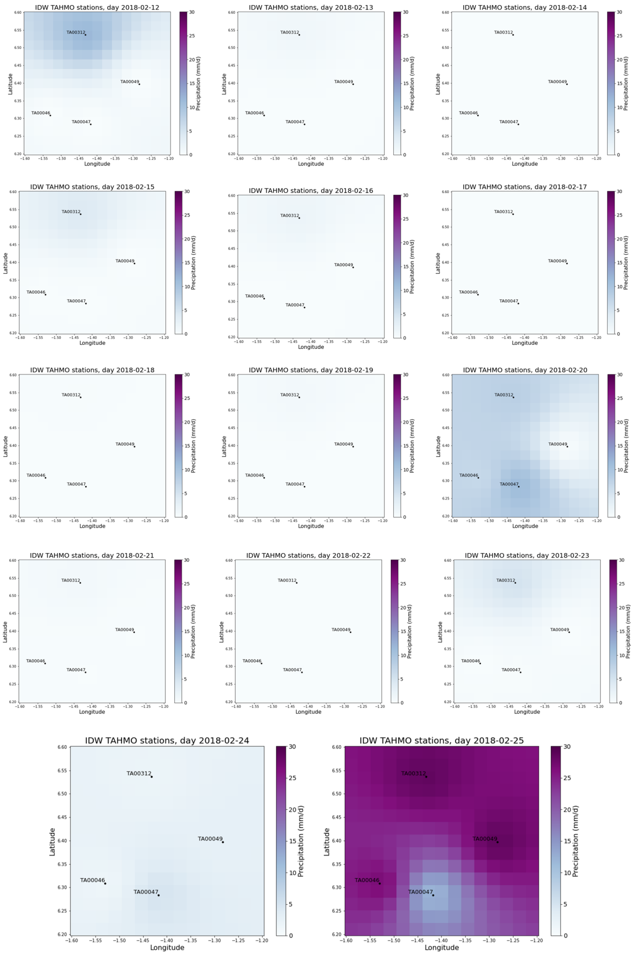
```
                    if S != 0:
                        counts.iloc[1, 0] += 1
                if position_tahmo == 1:
                    counts.iloc[0, 1] += 1
                    if S != 0:
                        counts.iloc[1, 1] += 1
                if position_tahmo == 2:
                    counts.iloc[0, 2] += 1
                    if S != 0:
                        counts.iloc[1, 2] += 1

                # Store in grid
                categorized_grid[t_idx, lat_idx, lon_idx] = S
                categorized_grid_obs[t_idx, lat_idx, lon_idx] = position_tahmo

    # Take mean over time
    categorized_grid_mean = categorized_grid.mean(dim='Time')
    categorized_grid_obs_sum = categorized_grid_obs.sum(dim='Time')
    # Rewrite to skill score
    categorized_skillscore = 1 - categorized_grid_mean
    categorized_skillscore = categorized_skillscore.where(categorized_skillscore >= 0, other=0)
    return categorized_skillscore, counts, categorized_grid_obs_sum

seeps = []
counts = []
categories = []
for count in range(len(wrf)):
    seeps_element, counts_element, categorie = SEEPS(tahmo.Daily, wrf[count].__xarray_dataarray_variable__, t1, t2, Smatrix)
    seeps.append(seeps_element)
    counts.append(counts_element)
    categories.append(categorie)
seeps_concat = xr.concat(seeps, dim='stack')
seeps_average = seeps_concat.mean(dim='stack')
counts_sum = sum(counts[1:], counts[0])
counts_sum.loc['error␣fraction'] = counts_sum.iloc[1, :] / counts_sum.iloc[0, :]
path_modified = 'C:\\Users\\Sophie.Verheugd\\Documents\\export_2506\\Experiments_part2\\Modidified_data\\'
# counts_sum.to_csv(path_modified + f'{PBL}_count_errors_t1is{t1}.csv')
categories_concat = xr.concat(categories, dim='stack')
categories_sum = categories_concat.sum(dim='stack')
```

B

Figures

B.1. IDW Interpolation Experimental Setup 2

Beneath you can find the IDW-interpolation output of the daily precipitation data of TAHMO, associated to experimental setup 2.

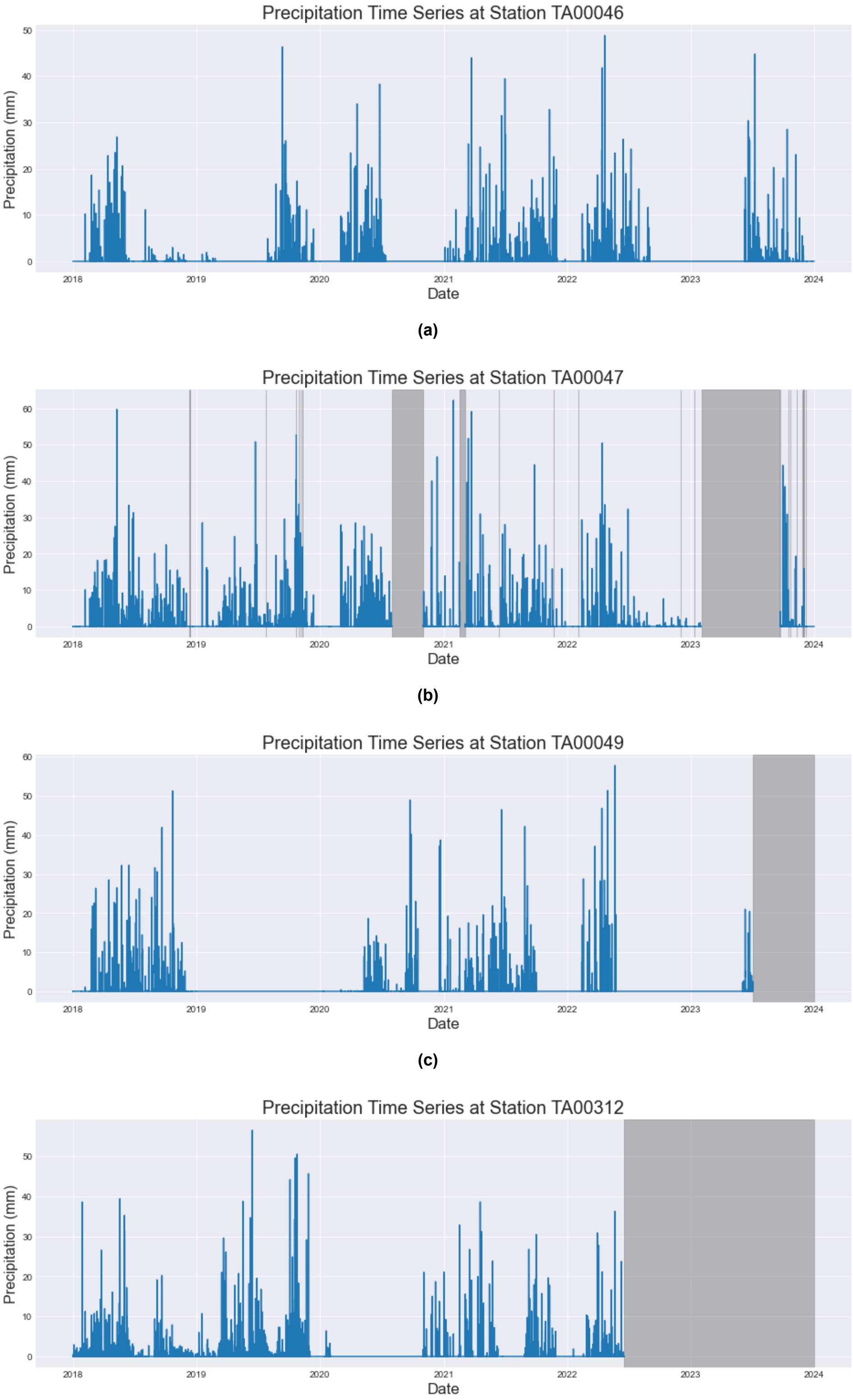


IDW TAHMO stations, day 2018-02-12
IDW TAHMO stations, day 2018-02-13
IDW TAHMO stations, day 2018-02-14
IDW TAHMO stations, day 2018-02-15
IDW TAHMO stations, day 2018-02-16
IDW TAHMO stations, day 2018-02-17
IDW TAHMO stations, day 2018-02-18
IDW TAHMO stations, day 2018-02-19
IDW TAHMO stations, day 2018-02-20
IDW TAHMO stations, day 2018-02-21
IDW TAHMO stations, day 2018-02-22
IDW TAHMO stations, day 2018-02-23
IDW TAHMO stations, day 2018-02-24
IDW TAHMO stations, day 2018-02-25

Figure B.1: IDW-interpolation output of daily precipitation (mm/d) of TAHMO for experimental setup 2

B.2. SEEPS Experimental Setup 2

Extra figures concerning the calculation of SEEPS of experimental setup 2 are depicted in this section.



(a)

(b)

(c)

(d)

Figure B.2: Precipitation time series at the four selected TAHMO stations.The gray band shows periods with no data values.

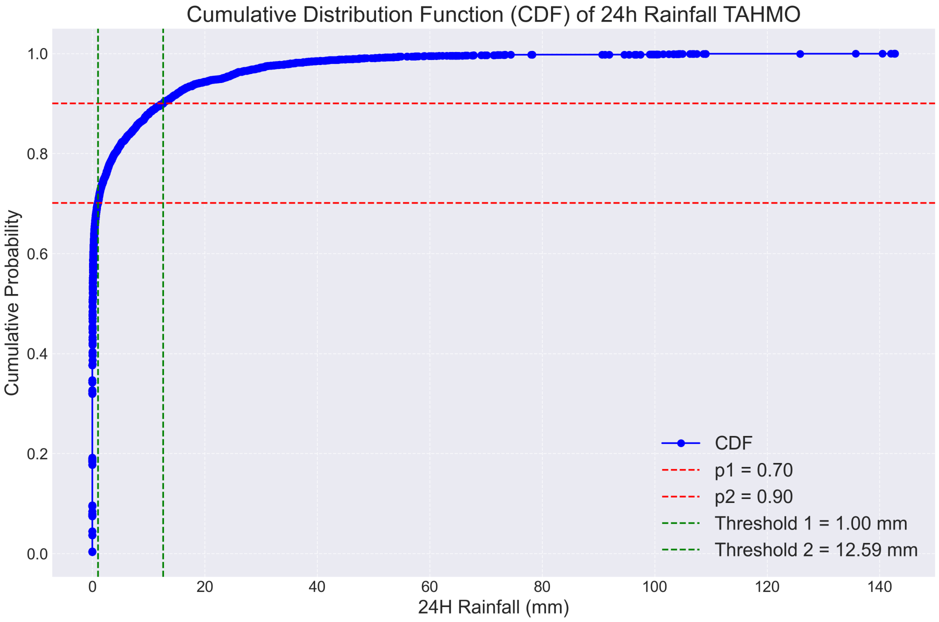


Figure B.3: Cumulative distribution for 24-h precipitation at station TA00047 for the period of 2018-2023 with t1=1mm

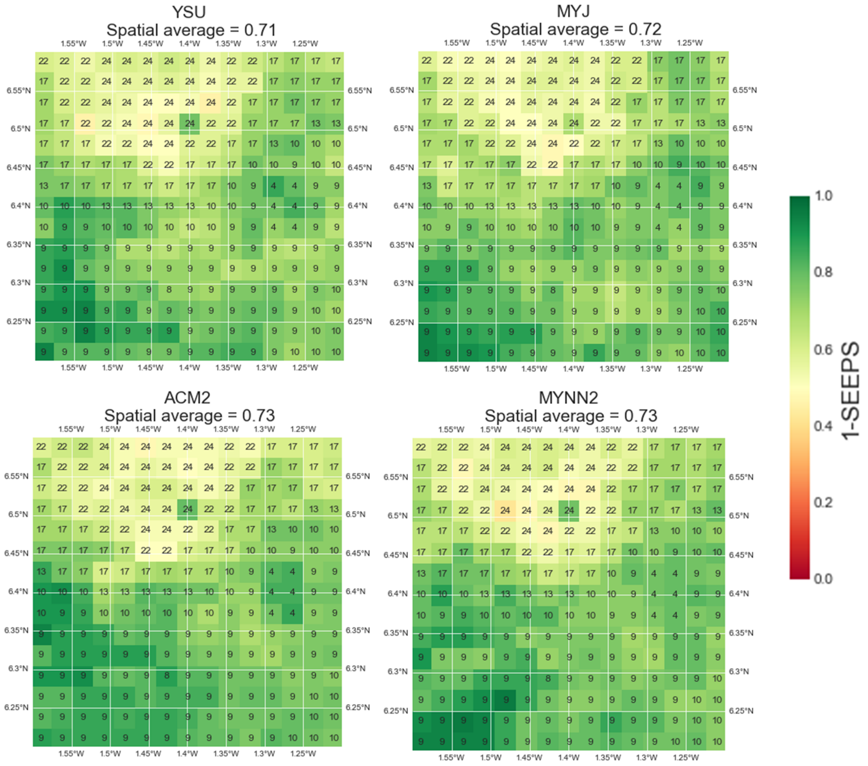


Figure B.4: SEEPS, with a t1 value of 1 mm. The numbers inside each grid cell depict the rainfall amount, ranging from 0 (no rainfall on all simulation days) to 100 (heavy rainfall on all simulation days). The spatial average is the mean SEEPS skill score for the entire validation domain.

B.3. Extra Results Experimental Setup 2

Evolution of the MBE and MAE

Table B.1: Evolution of the MBE and MAE against threshold values, with all daily rainfall above the threshold masked.

Threshold	**YSU**		**MYJ**		**ACM2**		**MYNN2**	
	MAE	**MBE**	**MAE**	**MBE**	**MAE**	**MBE**	**MAE**	**MBE**
1	0.148	−0.117	0.180	−0.099	0.136	−0.119	0.144	−0.114
3	0.382	−0.266	0.433	−0.216	0.361	−0.279	0.379	−0.261
5	0.593	−0.311	0.637	−0.259	0.566	−0.367	0.588	−0.312
7	0.742	−0.257	0.790	−0.199	0.700	−0.381	0.744	−0.235
9	1.083	−0.110	1.057	−0.036	0.939	−0.349	1.077	−0.096
11	1.316	0.078	1.315	0.168	1.088	−0.306	1.325	0.077
13	1.515	0.273	1.529	0.381	1.202	−0.209	1.527	0.271
15	1.740	0.513	1.725	0.585	1.340	−0.079	1.721	0.455
20	2.304	1.051	2.407	1.268	1.737	0.230	2.275	1.022
25	2.967	1.562	3.099	1.798	2.141	0.411	2.775	1.412
30	3.706	1.825	3.846	2.070	2.689	0.476	3.437	1.654
35	4.134	2.245	4.299	2.533	3.000	0.810	3.789	2.005
NaN	5.841	3.988	6.215	4.430	3.993	1.775	5.748	3.977