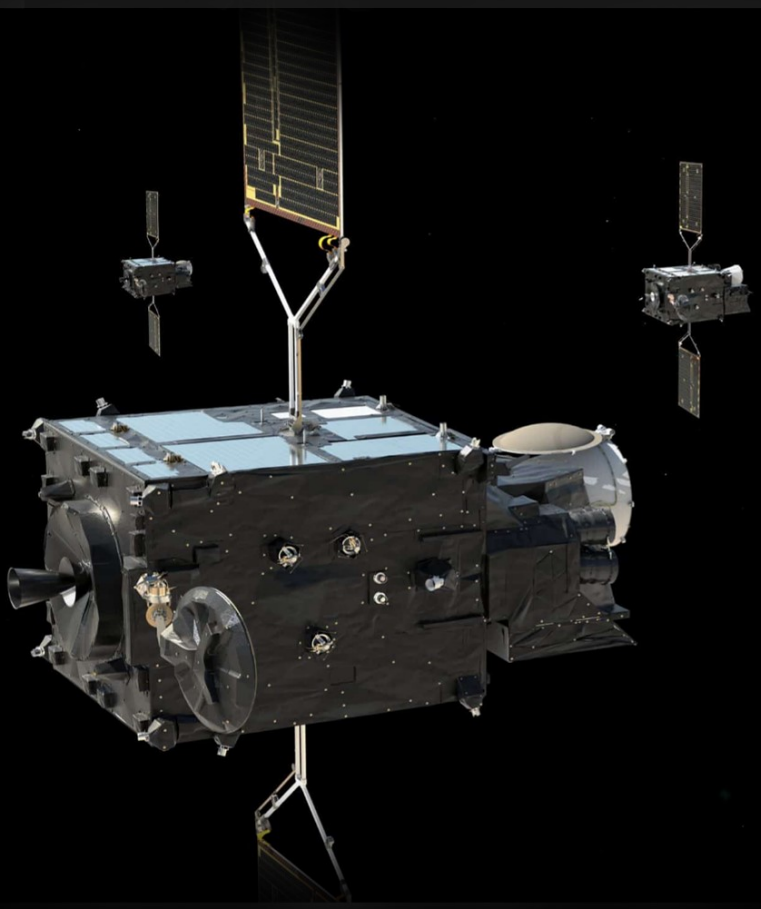
Implementation of a space-time transformer retrieval algorithm for high rainfall rates from the Meteosat satellite applied in Africa

Cecile Marcella Kwa



Delft University of Technology

TU Delft

HKV LIJN IN WATER

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Implementation of a space-time transformer retrieval algorithm for high rainfall rates from the Meteosat satellite applied in Africa

by

Cecile Marcella Kwa

Student Name	Student Number
Kwa, Cecile	4830652

Chair: Prof. dr. Ir. van de Giesen (Technical University Delft)
Other committee members: dr. M. A. Schleiss (Technical University Delft), dr. Y. Shapovalova (Radboud University)
Additional Supervisors: Ir. D. Lugt (HKV), J. Aerts (HKV), dr. J. F. Meirink (KNMI)
Project Duration: September, 2024 - March, 2025
Faculty: Faculty of Civil Engineering, Technical University Delft

TU Delft

HKV LIJN IN WATER

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Preface

This report marks the finalization of my six-year-long learning journey as a student at TU Delft, during which I encountered some downs but, mostly, many ups. I want to thank all my friends and family who made these last six years such an amazing time, with whom I laughed and cried, and who helped me through my thesis project by distracting me and patiently listening to all my problems.

I want to thank all the employees at HKV for making my final thesis project such an enjoyable end to my studies, especially Dorien Lugt and Jerom Aerts, my daily supervisors during this project. You always pointed me in the right direction and encouraged me, even when I doubted myself. You supported me at a professional level, always saying the right things, but you also genuinely cared about how I was personally feeling throughout this project. Thank you for checking in on me and for your enthusiasm, which gave me the courage to continue.

I also want to thank my committee members, Nick van de Giesen, Marc Schleiss, and Yuliya Shapovalova. Nick, thank you for grounding me throughout this project and ensuring I remained connected to the practical implications of my research. Marc, your thorough feedback and insightful questions challenged me to consider different perspectives. Yuliya, I am especially grateful for your guidance in the sometimes overwhelming world of Machine Learning.

I also would like to thank Jan Fokke Meirink for sharing his valuable expertise on satellites and remote sensing techniques.

And lastly, I want to thank my housemates, who supported me no matter what, my parents for always believing in me and Jelle for giving structure to my life and my project.

Cecile Marcella Kwa
Delft, March 2025

Abstract

In Ghana, flash floods are often triggered by severe storms. Flood Early Warning Systems (FEWS) can mitigate flood impacts but require accurate, near real-time rainfall data. Implementing FEWS in Ghana is challenging due to sparse ground-based data and a lack of accurate, rainfall data with short latency. While satellite-based rainfall products offer a promising alternative, they often show significant discrepancies compared to ground observations, limiting their use for effective FEWS.

The Meteosat satellite images provide a valuable source of data for near-real-time applications, due to its short latency (within 15 minutes) and relatively high temporal (15-minute) and spatial resolution of (3 x 3 km at sub-satellite point). This study explores the use of Earthformer, a space-time transformer model, to improve rainfall intensity estimates with minimal latency, using data from the Meteosat satellite. By evaluating the potential of Earthformer to create rainfall estimates, this research aims to contribute to more reliable FEWS, ultimately strengthening disaster risk management in flood-prone regions of Ghana.

The Earthformer model is trained on IMERG-Final, a satellite-based rainfall product known for its relative high accuracy but with delayed availability of several months. With the application of FEWS in mind, it was investigated whether the model could be adapted to improve the estimation of higher rainfall intensities. Two model setups were tested: one using a mean squared error (MSE) loss function during the training of the model and another with a balanced weighting loss function to emphasize higher intensities.

The model's accuracy was first evaluated by comparing its outputs with IMERG-Final on a test dataset, and secondly with ground station observations from the Trans-African Hydro-Meteorological Observatory (TAHMO), and Ghana Meteorological Services (GMET) for the year 2022. IMERG-Early was used as a benchmark for near real-time performance. The comparison with IMERG-Final revealed that both Earthformer models outperformed IMERG-Early for lower rainfall intensities in terms of probability of detection (POD), success rate (SUCR) and the Critical Succes Index (CSI) and that the balanced loss model also outperformed IMERG-Early for higher intensities. However, as rainfall intensity increased, the performance of both Earthformer models and IMERG-Early decreased.

Further comparisons with ground station data highlighted weak correlations at 30-minute intervals between all satellite rainfall estimates and ground observations (including IMERG-Final, IMERG-Early and both Earthformer models). However, when the data was aggregated to daily intervals, correlations improved significantly, suggesting that timing errors could play a role, however, further investigation is needed to quantify their impact.

Additional analysis revealed that peak rainfall was often underestimated, while lower intensities were overestimated. This discrepancy could be explained by several factors: the coarser spatial resolution of satellite estimates compared to gauge stations, the displacement of rainfall from observed clouds, and difficulty in capturing warm rain processes. Additionally, it was concluded that capturing spatial variability within Mesoscale Convective Systems (MCSs) is challenging. This is potentially due to anvil cloud tops obstructing the satellite's view, similar cloud-top temperatures for different rainfall intensities, and strong wind shear increasing the risk of rainfall misallocation.

This research demonstrates the potential of a space-time transformer model for near real-time rainfall estimation as it shows improved performance for most intensities when compared to IMERG-Early, with IMERG-Final set as the reference truth. However, the reduced performance at higher intensities and discrepancies with ground observations of all satellite based products underscore the need for further model development to improve extreme rainfall detection and better align satellite estimates with ground truth data. These improvements are essential for the model's utility in FEWS and contributing to more effective disaster risk management in flood-prone regions.

Contents

1

Introduction

The prediction of extreme weather events in Africa is becoming increasingly important due to climate change. It is expected that climate change will lead to more frequent and intense extreme weather events in the sub-Saharan region, outpacing the population's capacity to cope with the resulting disasters (Codjoe and Atiglo, 2020). These events are anticipated to have a severe impact on various socio-economic factors, including health and well-being, education, access to clean water and sanitation, and the sustainability of communities and cities (Codjoe and Atiglo, 2020). Atiah et al. (2023) states that flood-related fatalities and associated economic losses have already increased in Southern West Africa, with Ghana as one of their main examples. Ansah et al. (2020) show an increasing trend of annual daily maximum rainfall in Accra.

Flood Early Warning Systems (FEWS) have proven effective in reducing flood-related damages and casualties (Perera et al., 2019). Since floods in African urban environments are often triggered by extreme rainfall (Ansah et al., 2020), gaining insights into when and where these events will occur can help mitigate damage and save lives. Nowcasting, an algorithm integrated into FEWS, is a relatively new meteorological method that uses current weather data to forecast weather conditions over the next 0 to 6 hours. However, implementing such systems in Africa is challenging due to the limited spatial coverage and high temporal resolution of accurate rainfall observations, which is essential for the quality of real-time observations (Estébanez-Camarena, Curzi, et al., 2023). Remote sensing techniques from satellites offer a potential solution to the sparsely available ground-based data, but the latency, defined as the time between observation and the availability of data, often makes satellite-derived products unsuitable for real-time nowcasting. The Meteosat satellite offers the potential to overcome this issue. It is a geostationary satellite that observes visible light reflectances and infrared brightness temperatures using the Spinning Enhanced Visible and Infrared Imager (SEVIRI) over both Europe and Africa. With its relatively short latency of 15 minutes, high temporal resolution of 15 minutes, and spatial resolution of 3 x 3 km at sub-satellite point, it holds great potential for nowcasting purposes.

A study by Glas (2024) focused on developing a nowcasting model for flash floods in urban areas of Ghana. However, after completing the research, Glas (2024) identified the satellite rainfall estimates as one of the factors contributing to the low accuracy of the nowcasting model. This conclusion was drawn from a comparison between multiple satellite-based rainfall products and observations from the Trans-African Hydro-Meteorological Observatory (TAHMO) stations, which showed a significant discrepancy. The discrepancy between gauge stations and multiple satellite rainfall products shows a potential to improve spatial rainfall estimates for near real-time purposes from space.

To leverage the potential of the Meteosat satellites, **the goal of this research** is to develop a low-latency rainfall retrieval model that can support an effective FEWS using SEVIRI channel data from the Meteosat satellites, while agreeing with the ground observations. Instead of prioritizing precise rainfall rate estimates, an effective FEWS requires rainfall estimates that can enhance accurate predictions on whether extreme rainfall might lead to flooding, thereby improving preparedness and enabling timely warnings.

Machine learning techniques provide a computationally efficient approach for rainfall retrieval from satellite images. Many studies use deep learning techniques to estimate rain rates from satellite images, some showing promising results (Estébanez-Camarena, Curzi, et al., 2023; Estébanez-Camarena, Taormina, et al., 2023; Gao et al., 2022; Hong and Hsu, 2004; Küçük, Atencia, and Dabernig, 2024; Wang et al., 2021). From these models the Earthformer model architecture adapted by Küçük, Atencia, and Dabernig (2024) is selected as the main method of this study, as it outperforms the state of the art model architecture U-Net (Gao et al., 2022; Küçük, Atencia, and Dabernig, 2024), which is frequently used in other studies to estimate rainfall (D'Adderio et al., 2023; Han et al., 2022; Rahimi et al., 2023; Wang et al., 2021; Yang et al., 2021) (Section 3.1).

To train the model, IMERG-Final (Huffman et al., 2023) is chosen as the target dataset because it is recognized for its relatively high accuracy compared to other satellite-based rainfall products (Dezfuli et al., 2017; Pradhan et al., 2022). IMERG-Final incorporates gauge-based corrections using Global Precipitation Climatology Centre (GPCC) observations (Schneider et al., 2013), ensuring better alignment with ground observations. IMERG is part of the Global Precipitation Measurement (GPM) mission, which integrates observations using passive microwave (PMW) data from Low Earth Orbiting (LEO) satellites with the high temporal resolution data of geostationary (GEO) satellites like Meteosat. It provides gridded rainfall estimates at 30-minute intervals, making it one of the most widely used precipitation datasets for research (Pradhan et al., 2022). However, IMERG-Final has a latency of approximately 3.5 months (Huffman et al., 2023), prohibiting its use in near-real-time applications. This delay creates a gap for operational forecasting, where timely rainfall estimates are essential.

To achieve the goal of this research, this study investigates whether the Earthformer transformer model (Gao et al., 2022) adapted by Küçük, Atencia, and Dabernig (2024) can be used to approximate IMERG-Final's accuracy using only Meteosat satellite images as input. Using the high temporal resolution and low latency of the Meteosat satellite data, the study aims to reduce the latency from 3.5 months to less than 15 minutes while trying to maintain similar accuracy and ultimately support FEWS purposes. The study further evaluates how well these satellite-derived rainfall estimates correspond to local gauge station observations, using data from the Ghana Meteorological Agency (GMET) and TAHMO. By comparing model outputs with ground observations, this research aims to guide the future development of high-accuracy, low-latency rainfall retrieval models for West Africa, improving Meteosat-based rainfall products.

1.1. Research questions

Main Research Question:

How can the Earthformer transformer model be applied to improve the accuracy of rainfall rate estimates across different intensities for near real-time applications using SEVIRI channel data from the Meteosat Second Generation satellite in Ghana?

Subquestions:

- *How accurately does the Earthformer transformer model retrieve rainfall rate estimates for different rainfall rates from the SEVIRI channels of the Meteosat Second Generation satellite in Ghana?*
- *What modifications can be made to the Earthformer transformer model to improve the accuracy of high-intensity rainfall estimates?*
- *How do satellite based rainfall rate estimates relate to ground station observations?*

2

Theoretical background

This chapter provides the necessary background knowledge for understanding the research. Section 2.1 discusses various space-based observation strategies for rainfall and their key advantages. Section 2.2 explains the mechanisms responsible for rainfall formation. Finally, Section 2.3 highlights the important climatic factors affecting Africa, with a specific focus on Ghana.

2.1. Rainfall observation satellites

Satellites play an important role in monitoring global weather patterns, including the retrieval of rainfall data, since they provide spatial coverage of weather patterns (Prigent, 2010). The provided data can be valuable for forecasting, water management and climate research. The used satellites for estimating rainfall can be divided in two categories; Low Earth Orbit (LEO) satellites and Geostationary Earth Orbit (GEO) satellites. Each of these satellites have different spatial coverage and contain different measurement equipment to observe the rainfall from space.

Low Earth Orbit (LEO) satellites orbit around the earth at an relative low altitude. Because they orbit around the entire earth the satellites can cover a wide area, however they revisit the exact same location only about once every few hours to once in a few days. In contrary, since they orbit relatively close to the earth's surface they can observe the earth in great detail. A LEO satellite used to observe rainfall is typically equipped with either or both a Passive Microwave Radiometers and/or an Active Microwave radar. The Passive Microwave Radiometers detects the natural emission of microwaves by the atmosphere and the earth's surface, while active radar actively sends out radar pulses and measures the return signal. Both measurement techniques are sensitive to the presence of rainfall and cloud structure, but the Active Microwave provides the most accurate and direct measurement of rainfall, compared to the passive microwaves. However, since LEO satellites cannot provide data at a high temporal resolution in a single location, it makes them less useful for short-term weather monitoring. Therefore it is not suitable for tracking fast-changing weather patterns, such as convective rainfall (see section 2.3). They are, however, relatively accurate estimations from rainfall on a global scale (Fischer and Winterrath, 2021; Sun et al., 2018).

GEO satellites can in its turn provide data at a high temporal resolution, because they have the same rotational speed as the earth and observe the earth from a fixed position when looking from the earth's surface. Because of its high temporal resolution they are suitable for fast changing weather systems, such as convective rainfall. However, because the GEO satellites are located much farther from the earth compared to LEO satellites, their spatial resolution is also lower and they can not directly measure rainfall. On board of the satellites are Visible and Infrared Imagers that can monitor rainfall related parameters, such as cloud cover, humidity and temperature. From these parameters the actual rainfall can be estimated. This is however an indirect method and can therefore result in substantial variability between different retrieval products from VIS/IR sensors (Fischer and Winterrath, 2021; Sun et al., 2018). Some GEO satellites also have a Lightning detector on board. This can be useful for observing heavy rainfall and severe storms associated with convective storms. The Meteosat Third Generation

is equiped with such a system.

So where LEO satellites can provide more accurate and detailed information about rainfall, GEO satellites provide less accurate rainfall estimates, but on a higher temporal resolution. To overcome the limitations of both satellites some retrieval products integrate the two techniques, combining the higher temporal resolution of VIS/IR with the more direct precipitation estimates from microwave sensors (EUMETSAT, 2019; Fischer and Winterrath, 2021; Sun et al., 2018).

Table 2.1: Comparison of LEO and GEO Satellites for rainfall monitoring.

Feature	LEO Satellites	GEO Satellites
Altitude with respect to the Earth's surface	Relatively low	High
Spatial coverage	Global coverage, revisits the same location every few hours to days	Fixed location, taking observations from the same location
Temporal resolution	Low (less frequent revisits of the same location)	High (relatively frequent observations of the same location)
Spatial resolution	Lower	Higher
Measurement equipment	Passive Microwave Radiometers, Active Microwave Radar	Visible/Infrared Imagers, Geostationary Lightning Mapper
Rainfall measurement technique	Direct (microwave sensors)	Indirect (inferred from cloud cover, humidity, temperature)
Strengths	Accurate and detailed rainfall data	High temporal resolution, ideal for fast-changing weather
Weaknesses	Low temporal resolution	Less accurate rainfall estimates

2.2. Cloud dynamics

Precipitation and cloud formation occur when air becomes fully saturated and can no longer hold water, a condition known as reaching the dew point (National Weather Service La Crosse, 2014). When air is cooled beyond this point, it must release its moisture in the form of clouds or fog. As warm air rises and typically cools with altitude, clouds form.

There are three main mechanisms that cause cloud formation: The first one is heating. The sun warms the Earth's surface, which in turn heats the air above it. As the air warms, it rises, and when it reaches its dew point, clouds begin to form. If this mechanism results in rainfall it is called convective rain. This normally covers smaller areas of rain and can result in heavy rainfall. They are quite unpredictable, because of ground surface variations. The air close to the earth is heated at different rates, resulting in rainfall that can pop up anywhere and blown around by the wind directions. The second mechanism that forms rain, is frontal movement and occurs when cold air meets warm air. The movement of the air masses pushes the warm air over the cold air fronts, which forces the warm air to rise and form clouds. If this results in rainfall it is called frontal rain. A similar process happens when air is forced to rise over geographical features, like mountains, As it ascends and cools to its dew point, clouds form. This is called Orographic Lifting. When resulting in rainfall it is called orographic rain.

Clouds can be categorized into three main types: cumulus, stratus, and cirrus (Manishsiq, 2024), as shown in Figure 2.1.Cumulus clouds are formed when warm, moist air is forced upward. If the atmospheric conditions are unstable, the cumulus clouds can develop into thunderstorms. The Stratus clouds are formed by the lifting of large air masses. Typically they are precipitation-free, though they might produce light drizzle. Cirrus clouds are formed at high-altitudes and are made out of ice crystals. They usually indicate the direction of the wind at their altitude.

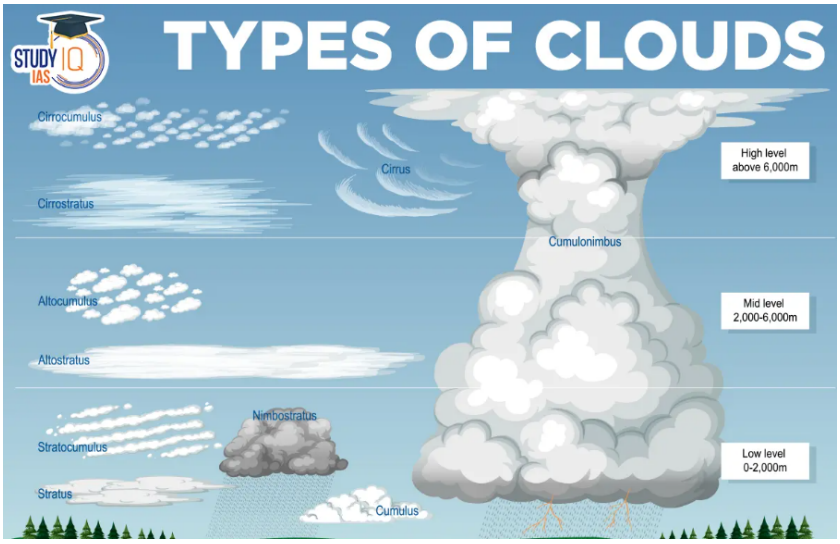


Figure 2.1: The atmosphere can be divided in three different levels; low-level, mid-level and high-levels. Each level has its own characteristics, such as temperature and pressure, resulting in different cloud structures. Large cumulus clouds can form heavy precipitation, while low stratiformal clouds can result in light drizzle (Manishsiq, 2024).

Precipitation formation is a combination of large scale processes, such as lifting air masses and small scale interactions within the cloud itself, which is called the micro-physical cloud dynamics. Precipitation within the clouds is formed by two different main micro-physical processes (Kikuchi et al., 2023). In 'warm clouds' no ice particles are present. With collision and coalescence cloud droplets of different sizes collide and merge into larger droplets. As these droplets grow larger, they become heavier and fall faster. Once they reach a size that the cloud can no longer hold, they fall as rain. The warm rain process is therefore strongly related to the droplet size. The other process is the ice-crystal process, which typically takes place within 'cold-clouds'. When both ice crystals and water droplets are present in a cloud, the water vapor from the droplets is attracted to the ice crystals and deposits directly onto the them. This causes the ice-crystals to grow. Once they become large enough, the ice crystals fall. On their downfall they collide with other crystals and droplets, increasing their particles seize. If the surface temperature is cold enough, this precipitation falls in the form of snow. However, when temperatures are high, the particles melt and fall as rain.

2.3. African climate

As described in the previous section the type of cloud has a significant effect on the precipitation formation. There exist a great difference between cloud formation around the equator in arid areas compared to cloud formation in cooler and moist areas.

Africa consists of eight distinct climatic regions: hot desert, semiarid, tropical wet and dry, equatorial (tropical wet), Mediterranean, humid subtropical, marine warm temperate upland, and mountain regions. The differences between climates are caused by the movement of different air masses over the continent. The location of the Intertropical Convergence Zone (ITCZ) is a key driver in the timing of the rainy seasons (Middleton et al., 2024).

The ITCZ is formed by the sun heating up the Earth's surface. This effect is intensified near the equator as this location receives the most direct sunlight. This causes the air to heat up more quickly and rise. The rising air creates low-pressure areas, enhancing the formation of convective rainfall. Thus, the ITCZ is defined as a low-pressure zone across the equator of the earth, which is often associated with convective rainfall. North and south of the ITCZ, the surface is heated more slowly, as the sunlight it receives is more indirect. In contrast to warm air the cooler air sinks, causing little to no rainfall.

Due to the Earth's tilt, the region that receives the most direct sunlight shifts throughout the year. For half of the year, regions just north of the equator receive the most direct sunlight, while for the other half, the area south of the equator experiences maximum heating. This movement of the ITCZ causes

the alternating rainy and dry seasons (as shown in Figure 2.2). Additionally, because the land heats more quickly than the ocean, the ITCZ is not a straight line and moves faster upwards or downwards overland compared to oversea.

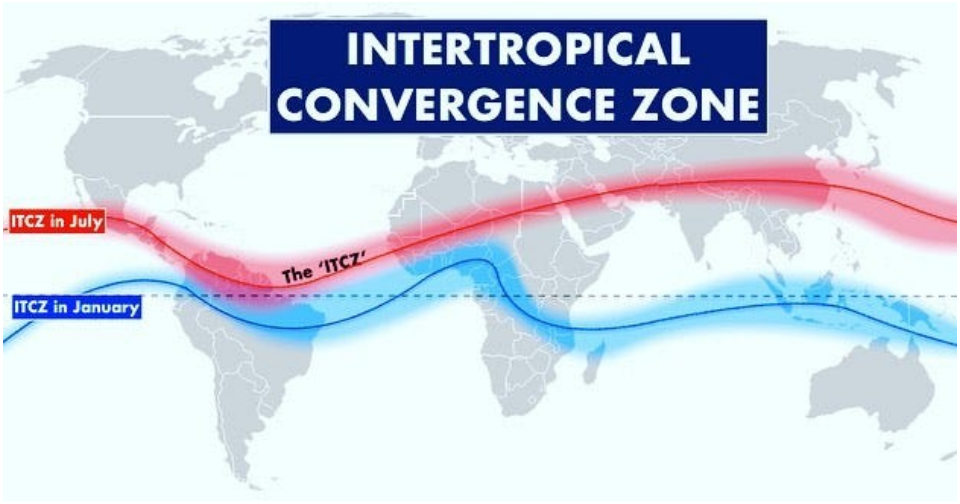


Figure 2.2: The Inter Tropical Convergence Zone (ITCZ) is a zone close to the equator, receiving he most direct sunlight, which causes extreme precipitation. Due to the tilt of the earth the location of the ITCZ changes with the time of the year, resulting in wet and dry seasons in areas around the equator. (Nikolaj Vinicoff, 2021)

As seen in Figure 2.2, in West Africa, the ITCZ is located where the continent meets the ocean in January. Maritime air from the Gulf of Guinea moves toward the ITCZ, as this is a low-pressure area. The air is heated by the Earth's surface and begins to rise. Another air mass, originating from the Sahara Desert, is dry and also rises. However, as the air rises, the moisture in the air mass coming from the Gulf of Guinea begins to condensate, often causing heavy precipitation. If the ITCZ moves further north, the air mass from the sea is forced to travel over land. This brings moisture to the dry areas in the north, but as it moves farther from its moisture source, it becomes less humid. This results in less intense rainfall in the northern regions of Ghana, caused by the ITCZ alone.

In addition to this pattern, there are also effects from El Niño and La Niña. El Niño and La Niña years are caused by changes in trade winds over the Pacific Ocean. Normally, these winds move from east to west. The winds push the warm surface water towards the west, near East Asia and Australia. Here the warm water piles up. In contrast, in the east, near Central and South America, the warm surface water is pushed away from its location, which allows cold water to well up. On one side, the air is heated by the relatively warm ocean, while on the other side, it is cooled by the ocean. In warmer parts of the ocean, the additional heating of the air by the warm ocean causes it to rise, and because it contains moisture, this results in precipitation. In contrast, at the cooler part of the ocean the air is cooled and sinks. Together, this reinforces the Easterly winds.

During an El Niño event, the intensity of the trade winds decreases or even reverses due to slight changes in Pacific weather systems or slow shifts in ocean currents around the equator. The decreased intensity of the trade winds result in less warm water being pushed westward, which in turn reduces the upwelling of cold water in the east. As a result, the eastern part of the ocean warms up, making the normal temperature gradient less pronounced. A larger portion of the ocean experiences warm temperatures and therefore releases more heat into the atmosphere, contributing to an increase in global temperatures. Additionally, the warmest part of the ocean shifts eastward, carrying moisture and warmth with it. This changes precipitation patterns and wind directions on a global scale.

In contrast to an El Niño event, the La Niña event is caused by the intensification of the normal trade winds. This pushes the warm water even more westward and increases the upwelling of cold water in the east. The cooled ocean reduces global temperatures and causes opposite effects of the El Niño. The El Niño and La Niña effects are also present in Africa, causing extreme precipitation and flooding or severe droughts (Amediegwu, 2024).

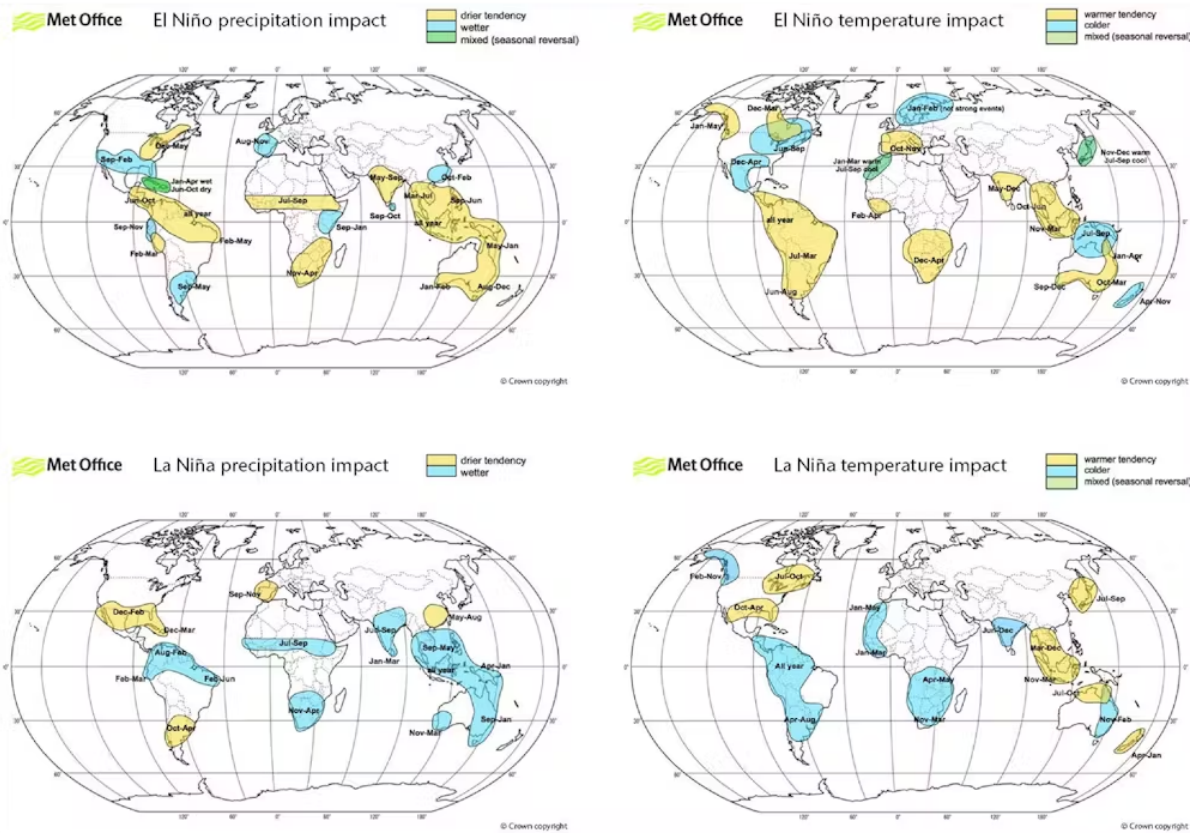


Figure 2.3: During El Niño years additional energy is available in the atmosphere, generally resulting an increased temperature and a drier climate in the largest part of Africa. During La Niña years, the effects are opposite as the El Niño years and result in colder temperatures and a wetter climate in Africa. (Amediegwu, 2024).

.

2.3.1. Climate of Ghana

Ghana is located in West Africa along the Guinea Coast, between latitudes 4° and 12° N. Due to it location near the equator, Ghana experiences a tropical climate with one or two rainy seasons each year, depending on the region. These rainy seasons are driven by the movement of the Intertropical Convergence Zone (ITCZ) over Ghana, as illustrated in Figure 2.2. As the ITCZ oscillates between the northern and southern tropics, it crosses Ghana's northern region once a year, resulting in a single rainy season there. In contrast, the ITCZ passes over the southern region twice a year, creating two rainy seasons.

In general, the primary rainy season in the southern region occurs from March to July, followed by a shorter wet season from September to November. In the northern region, the rainy season typically begins in April and ends in October, as shown in Figure 2.4. The highest rainfall levels are recorded between June and August (Lizcano et al., 2008).

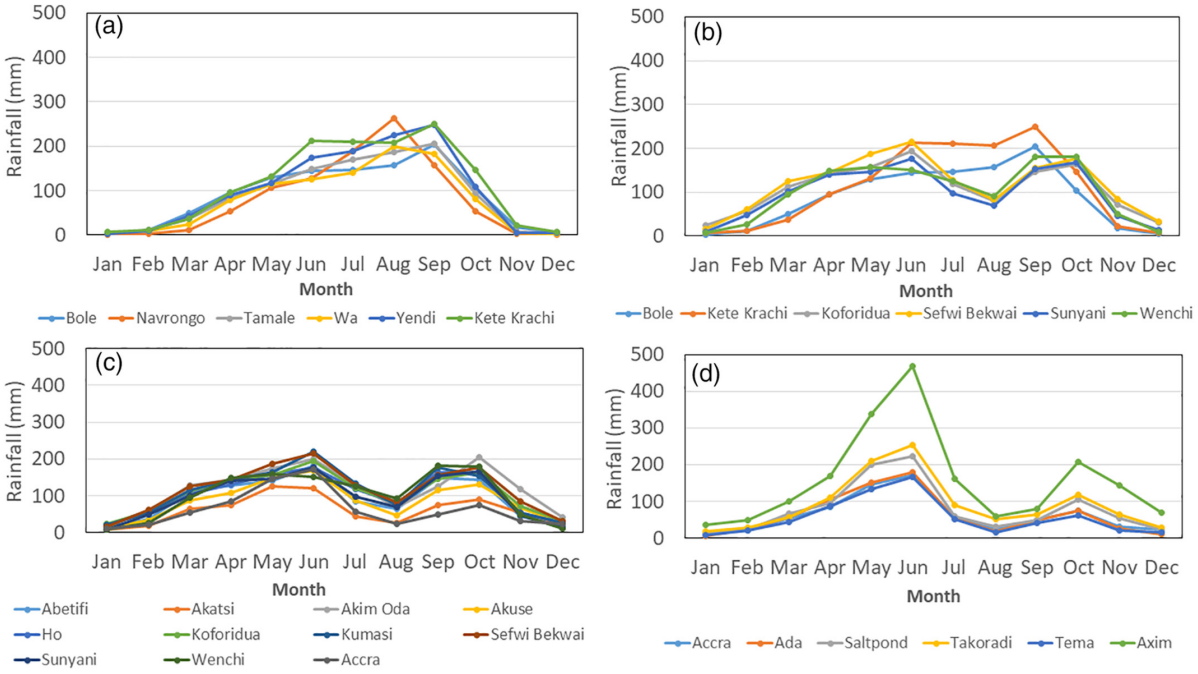


Figure 2.4: Long-term montly mean of rainfall for the period 1976–2018 for (a) cluster 1 (northern region) (b) cluster 2 (transition cluster between northern and central region) (c) cluster 3 (central region) and (d) cluster 4 (coastal region) (Bessah et al., 2022).

However, the timing, movement, and intensity of the ITCZ are also influenced by the El Niño–Southern Oscillation (ENSO). During El Niño years, West Africa tends to experience drier conditions (Figure 2.3), while La Niña years bring wetter conditions (Lizcano et al., 2008).

Based on climatological characteristics such as rainfall, minimal temperature, maximum temperature and relative humidity Bessah et al. (2022) identifies three climatological zones in Ghana that share similar characteristics: the northern, central, and coastal regions (Figure 2.5). The rainfall patterns are largely influenced by latitude (Bessah et al., 2022).

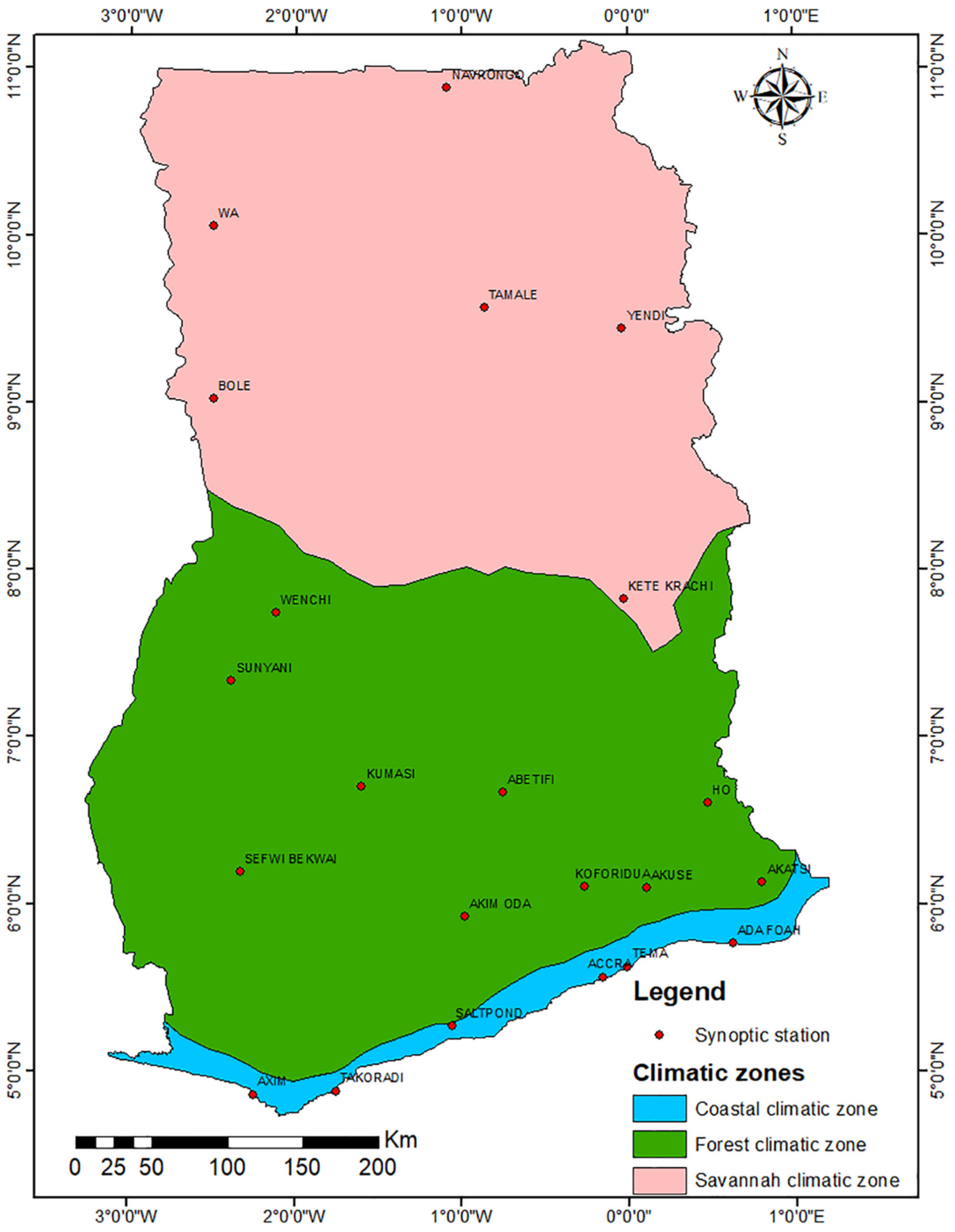


Figure 2.5: Climatic zones of Ghana-based on rainfall, relative humidity and temperature records from 1976 to 2018 (Bessah et al., 2022)

2.3.2. Squall lines and Mesoscale Convective systems over Ghana

In Ghana, complex interactions between wind patterns and the Earth's heating influence weather systems. Peters and Tetzlaff (1988) investigated the formation of squall lines, which are organized bands of thunderstorms, over West Africa. As previously explained, the ITCZ is a zone near the equator where the northeasterly trade winds from the Northern Hemisphere meet the southeasterly trade winds from the Southern Hemisphere. The winds move to the low pressure area of the ITCZ, which result in the convergence of these winds and when the ITCZ moves land inwards, brings moisture to the region.

At the same time, the strong temperature gradient between the hot, dry Sahara and the cooler, humid

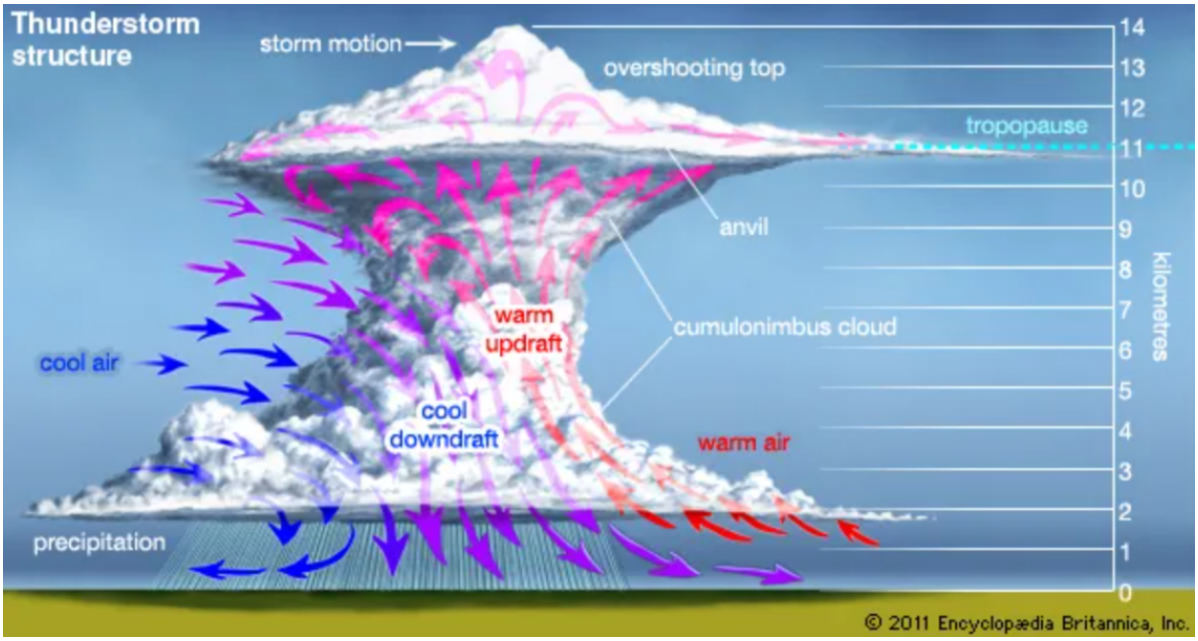


Figure 2.6: The formation of a thunderstorm structure. Due to the vertical wind shear the warm air is pushed over the cold air and rises, causing the formation of thunderstorms (Kreider, 2024).

Gulf of Guinea leads to the formation of the African Easterly Jet (AEJ) at mid-level altitudes and flows from east to west (Peters and Tetzlaff, 1988). When the AEJ is disturbed, it creates wave like motions of the jet, which are called the African Easterly Waves (AEWs) (Estébanez-Camarena, Curzi, et al., 2023; Peters and Tetzlaff, 1988). These waves result in alternating low- and high-pressure areas that move westward along with the AEJ. The low-pressure regions (thoughts) of these waves also cause convergence and helps thunderstorms to develop. The AEJ itself creates vertical wind shear (changing wind speed with height), which sets favourable conditions for thunderstorms to develop.

The time of day also influences convective storm formation. During the day, the surface is heated, which allows air to rise more rapidly. Many of the strongest storms in West Africa are initiated in the late afternoon or early evening when surface temperatures peak and the lifting of the air is at its maximum (Klein et al., 2018). Subsequently, during the evening hours, the wind is intensified and the air is cooled down. The intensified wind increases the lifting of the air, while cooled air releases more moisture into the atmosphere. This helps long-lived convective systems, such as squall lines, to intensify during the nighttime.

When the moist air from the Gulf of Guinea condensates, it releases heat to the atmosphere and helps the air to rise even further. Air masses that rise upward are called convective cells (Kreider, 2024). The AEJ helps them to merge into larger cells, eventually growing into large thunderstorms. Once reaching the tropopause (the air layer on top of the troposphere where vertical motion is less active), the air spreads out horizontally, instead of increasing more upward (Figure 2.6). Frequently, the part with the strongest convective energy can escape this equilibrium causing an overshooting top.

Mesoscale Convective Systems (MCSs) are large storm systems that include multiple convective cells. When an MCS organizes into a long, narrow band, it is classified as a squall line (M. Zhang, 2022). These systems are common in Ghana, particularly in the northern regions, where stronger AEW activity, drier air intrusions from the Sahara, and greater wind shear create an environment favorable for their formation (Peters and Tetzlaff, 1988). The effect of the AEJ is more pronounced in the north, making squall lines more frequent compared to the south.

2.3.3. Challenges in capturing rainfall patterns in Africa

Although the rain rate and cloud top temperature are strongly related, estimating exact intensities still introduces some challenges. Klein et al. (2018) explores the relationship in West Africa between cold

cloud tops and extreme rain rates and shows a strong relationship of cloud tops colder than -80°C with extreme rainfall (≥30 mm/h).

Geerts and Dejene (2005) pointed out that the convective storms over Africa are typically deeper compared to the ones in the Amazon. The deeper the system, the higher the cloud rises and the more rainfall potentially falls from it. This relationship can be partially captured with the cloud top temperature as observed with IR imagery, since higher altitudes correspond to colder temperatures. However, when systems rise beyond the troposphere and reach the tropopause (the air layer above the troposphere where vertical motions of air are weaker), the cloud-top temperatures do not decrease as significantly with height as they do in the troposphere. Therefore for very deep convective systems the cloud top temperature is a less relevant indicator of the depth of the system and thus, for the exact intensity of extremer rainfall as well.

Additionally, instead of vertically rising the air spreads out horizontally, making it more difficult to identify the core of the convective system from the cloud top temperature only. As explained in Section 2.3.2, an overshooting top often occurs, indicating the region of strongest vertical motion within the cloud. However, due to wind shear, the convective core of the cloud underneath it may be displaced horizontally relative to the overshooting top, making it difficult to identify the exact location of extreme rainfall when using only information on the cloud top temperatures. Besides, the anvil top (horizontally spread out top of the cloud) makes it difficult to see what happens underneath it. If multiple convective cores are active, which is the case with MCS often occurring in West Africa (Atiah et al., 2023; Estébanez-Camarena, Curzi, et al., 2023; Klein et al., 2018; Peters and Tetzlaff, 1988), some of these may be covered up underneath the anvil cloud top. However, with the water vapour channels the Meteosat satellites are able to look a bit deeper down into the clouds and observe convective motions within the cloud (Estébanez-Camarena, Curzi, et al., 2023). The complex interactions of the AEJ, the movement of ITCZ and converging trade winds at different levels of the atmosphere also contribute to the complex relation between the satellite observations and rainfall rates on the ground.

Aditionally, Geerts and Dejene (2005) points out how climate top temperatures and rain rate relations vary with climatological factors. The IMERG product was mostly developed and validated over Continental Unites States (CONUS) (Huffman et al., 2023), with different climatological factors compared to Africa and especially Ghana. The rainfall rate in Africa may correspond to different cloud top temperatures than over the CONUS. These differences in climatological conditions and satellite calibration might explain why higher rainfall rates are underestimated in satellite observations. However, Li et al. (2024) observed the underestimation of peak events over the CONUS as well.

Analysis also shows that light rainfall is overestimated at the temporal boundaries of rainfall events. Mccollum et al. (2000) names two potential reasons for the overestimation of rainfall in Africa. the winds from the Sahara bring more dust particles in the air, leading to a higher concentration of Cloud Condensation Nuclei (CNN), resulting in smaller droplets that struggle to grow and fall as rainfall and the fact that convective clouds form at higher altitudes in dry conditions and thus allow for more rainfall to evaporate before reaching the ground.

Research by Peinó et al. (2025) found that accurate rainfall estimates are typically associated with high Cloud Optical Thickness (COT) and Liquid Water Path (LWP) values, whereas lower values correspond to higher false alarm ratios, indicating poorer performance in estimating rainfall from warm clouds. These clouds, common in humid regions such as coastal Ghana, are a challenge for satellite-based rainfall detection. Prigent (2010) attributes the underestimation of warm rainfall to relatively warm and spatially homogeneous cloud-top temperatures, which show little distinction between raining and non-raining areas. However, the multispectral bands of the SEVIRI installation may help mitigate this issue. Still, as discussed in Section 2.2, warm rain processes primarily depend on droplet size. However, only the 1.6 µm near infrared channel can distinguish between ice and water clouds (Schmetz et al., 2002). Consequently, observing warm rain processes solely from Meteosat imagery remains difficult, particularly at night.

3

Literature review

This chapter justifies the selection of two key aspects of this research through a literature review. Section 3.1 examines various Machine Learning model architectures. As most machine learning models require target data to train to, Section 3.2 explores different potential sources to train the model on.

3.1. Machine Learning model architectures

Machine learning is a common method for retrieving rainfall from satellite data. The U-Net architecture, a type of Convolutional Neural Network (CNN), is widely used for rainfall retrieval, as shown in the literature (D'Adderio et al., 2023; Han et al., 2022; Rahimi et al., 2023; Wang et al., 2021; D. Zhang et al., 2023). In the domain of ML-based models for IR/VIS satellite data, the PERSIANN-CCS product is considered one of the most reliable for rainfall retrieval (Estébanez-Camarena, Taormina, et al., 2023; Wang et al., 2021). This product follows four key steps: first, it identifies cloud patches; second, it extracts features from these patches, such as area and minimum temperature; third, it uses an unsupervised learning technique using a Self-Organizing Feature Map (SOFM) to classify the cloud patches into 400 different classes; finally, these classifications are used as input to estimate rainfall rates (Hong and Hsu, 2004).

Temporal consistency is important for nowcasting purposes. Nowcasting algorithms identify cloud groups and track their movement over time to derive motion vectors. If the changes in values and shape between time steps are too large, the algorithm may fail to identify and track these groups accurately, leading to incorrect motion vectors and incorrect forecasting. To address this challenge, a convolutional Long Short-Term Memory (ConvLSTM) network is considered, as it is designed to capture both spatial and temporal relationships. However, prior research indicates that including temporal relations does not improve rainfall estimates significantly, although it may reduce the number of model parameters required (Estébanez-Camarena, Taormina, et al., 2023).

The perceptron architecture is also considered for this research. It is one of the most straightforward deep-learning architectures. Unlike the LSTM or CNN, a perceptron does not capture the spatial and temporal relationships between pixels. This can be a drawback since cloud and rainfall formations are spatiotemporal phenomena. However, the simplicity of the perceptron architecture makes training faster and more efficient. Research by Taravat et al. (2015) demonstrated that a Multi-Layer Perceptron (MLP) could identify clouds with 88.96% accuracy during daylight conditions using the 0.6, 0.8, 1.6, 3.9, 6.2, and 10.8 µm channels. Since MLPs are often used as classifiers, and the PERSIANN-CCS product showed significant improvements through the introduction of a classification step compared to the earlier PERSIANN model (Hong and Hsu, 2004), including an MLP as a classifier could be an interesting way to improve the model in the future.

Recent developments in Transformer models show considerable potential (Gao et al., 2022; Küçük, Atencia, and Dabernig, 2024; Vaswani et al., 2017). Earthformer, developed by Amazon, is a space-time transformer model specifically designed to handle high-dimensional spatiotemporal data (Gao et al., 2022). It was created for Earth forecasting systems and is based on the Transformer architecture.

The Transformer, a relatively recent innovation in deep learning, allows for significantly more parallelization during training, enhancing efficiency (Vaswani et al., 2017). Although originally designed for Natural Language Processing, the Transformer architecture has been adapted for other applications, such as video prediction (Gao et al., 2022). Applying Transformer models to Earth systems, which are often chaotic and high-dimensional, remains computationally intensive. Therefore, Earthformer has been further developed to address these challenges (Gao et al., 2022).

The Earthformer model has outperformed state-of-the-art architectures such as U-Net and ConvLSTM (Gao et al., 2022). Küçük, Atencia, and Dabernig (2024) developed the Earthformer for INCA model by adapting the Earthformer model for rainfall predictions from satellite images, radar images and station data, using radar images as reference data, and reported promising results.

Model architecture selection

From the literature review the Earthformer model (Gao et al., 2022) adapted by Küçük, Atencia, and Dabernig (2024) is viewed as the most promising and thus selected for this research. Appendix A provides an elaborate explanation of the model.

3.2. Potential target data source

The target source must be available in the region of Ghana and should be a gridded rainfall product, which is often the case with satellite-based products (Section 3.2.1). However, ground observations also have their advantages, as described in Section 3.2.2.

3.2.1. Satellite based products

Integrated Multi-Satellite Retrievals for GPM (IMERG), provides global precipitation data between 60°N and 60°S (Huffman et al., 2023). Its performance depends on regional factors such as season, climate, and topography. While few studies have examined IMERG's performance in Africa (Pradhan et al., 2022), existing research indicates variability in accuracy influenced by these factors (Dezfuli et al., 2017).

IMERG combines precipitation estimates from various satellite instruments into a global dataset (Huffman et al., 2023). It is part of the Global Precipitation Measurement (GPM) mission, an international collaboration led by NASA and JAXA to improve global precipitation estimates. IMERG integrates data from low Earth orbit (LEO) satellites, which measure precipitation using passive microwave (PMW) sensors (Figure 3.1, and geostationary (GEO) satellites, which provide visible and infrared (VIS/IR) imagery. The GPM Core Observatory, equipped with the Dual-frequency Precipitation Radar (DPR), plays an important role by providing highly accurate PMW and radar data for intercalibration.

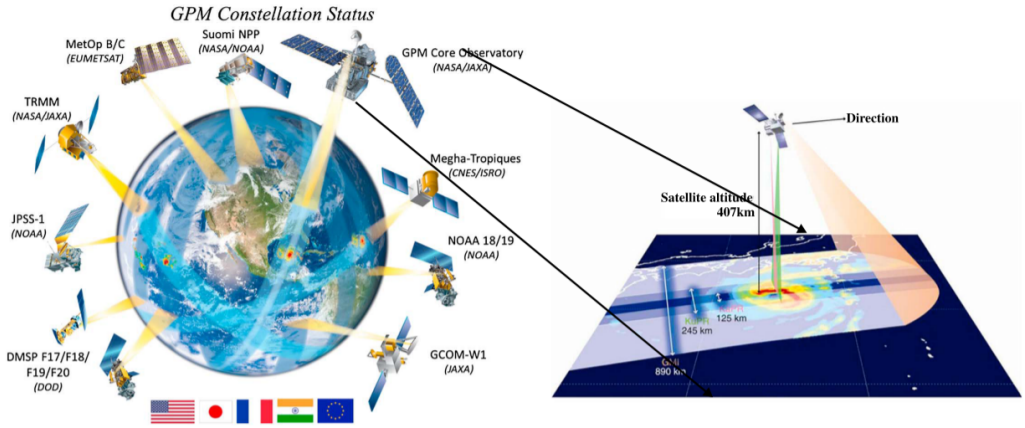


Figure 3.1: The GPM constellation with its LEO satellites on the left and the GPM core observatory on the right (Sun et al., 2018)

IMERG has a 30-minute temporal resolution and 0.1° × 0.1° spatial resolution, and merges data from multiple satellite sources. The core processing steps involve gridding, inter-calibration with the GPM

Combined Radar Radiometer Analysis product, and aggregation into half-hourly fields (Huffman et al., 2023). Since PMW data can be sparse in certain regions, IMERG supplements it with estimates derived from the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)-like algorithm, which utilizes GEO satellite VIS/IR data. The Climate Prediction Center Morphing-Kalman Filter (CMORPH-KF) then refines precipitation estimates by tracking cloud and rain system movement, improving temporal alignment.

The IMERG dataset contains three main versions:

- IMERG-Early (4-hour latency) is primarily used for near-real-time applications, using only satellite based estimates.
- IMERG-Late (14-hour latency), uses only satellite based estimates as well, but includes a backward pass in the morphing algorithm, to correct timing errors.
- IMERG-Final (3.5-month latency), is designed for research and incorporates Global Precipitation Climatology Centre (GPCC) monthly gauge analysis for bias correction. For the years before 2020, the GPCC full reanalysis was used, while for the other years, the the monitoring product of GPCC, as the full analysis product is not availabel yet for these years (Huffman et al., 2023).

The GPM core observatory provides the most accurate estimates of precipitation as it also has the DPR on board (Huffman et al., 2023). Using this as a target is however challenging as it normally revisits the same location only within the order of magnitude of days (Brasjen, 2014). D'Adderio et al. (2023) used the DPR product (GPM 2BCMB) as a reference for their machine-learning model to retrieve rainfall estimates over the full disk of the Meteosat satellite. However, for this study focusing on Ghana, the usability of GPM 2BCMB depends on LEO satellite overpass frequency, which can result in low temporal resolution and may provide not enough data for this research.

D. Zhang et al. (2023) used IMERG Early MWprecipitation (before called HQprecipitation) as reference data to train their model. The MWprecipitation field in IMERG provides a global, gridded, half-hourly dataset (0.1° × 0.1° resolution) that merges precipitation estimates from multiple LEO satellites. While this approach ensures spatial and temporal coverage, the data quality may suffer in areas with low satellite coverage, leading to gaps in the data as well (Huffman et al., 2023).

Table 3.1: Overview of IMERG and intermediate products.

IMERG Product	Temporal Resolution	Data Sources	Corrections Applied
IMERG Early	30 min	PMW (LEO) + IR (GEO) + DPR	No correction on gauge stations
IMERG Late	30 min	PMW (LEO) + IR (GEO) + DPR	Inclusion of a backward pass in the morphing algorithm correcting timing errors, no correction on gauge stations
IMERG Final	30 min	PMW (LEO) + IR (GEO) + DPR + GPCC Gauge Data	Inclusion of a backward pass in the morphing algorithm correcting timing errors, monthly bias correction using gauge data
GPM 2BCMB (DPR)	one in a few days	GPM core observatory	No correction on gauge stations or integration with other products
IMERG MWprecipitation (HQprecipitation)	30 min, global, but contains gaps	PMW (LEO) from multiple sensors	Merged multi-satellite precipitation field of MW satellites. No correction on gauge stations.

3.2.2. In-situ observations

All satellite measurements are essentially indirect estimates of surface rainfall (Fischer and Winterrath, 2021). Therefore, using direct observations from gauge station data as a target can also be beneficial. Moraux et al. (2019) trained their model directly on gauge station data both as a reference and as input. They however had access to a dense network of gauge stations. Converting data from gauge stations into a spatial dataset in less dense areas poses a challenge. It can be done through interpolation between the stations, but when too few stations are available, the interpolation may provide an inaccurate representation of reality (Estébanez-Camarena, Taormina, et al., 2023). Training a model on such data is thus risky. Estébanez-Camarena, Taormina, et al. (2023) addressed this issue by using point-based instead of grid-based estimations of rainfall for each TAHMO station cell. However, this method does not result in a spatial database that can be used for nowcasting purposes.

Target selection

Considering the challenges of using the ground-based observations directly as a target, IMERG-Final is selected for this research as it includes the stations through a monthly bias correction method and it has consistent temporal resolution, unlike the PMW or DPR product only.

4

Data description

This chapter provides a description of the data used in this research. Section 4.1 discusses the target data used for model training and the ground observations for validation. Section 4.2 outlines the input data provided to the model. Finally, Section 4.3 examines the benchmark products.

4.1. Target data

4.1.1. IMERG-Final

In 3.2 IMERG-Final (V07B) was selected as the target of the model. The section already described the used satellites and the algorithm behind it. As explained before, the satellite estimates are corrected on a monthly scale using the GPCC data. The bias correction is done by inverse variance weighting of the satellite estimates and GPCC gauge stations (Huffman et al., 2023). The exact number of used stations varies per month, but an example map with the number of stations included over Ghana is provided in Figure 4.1.

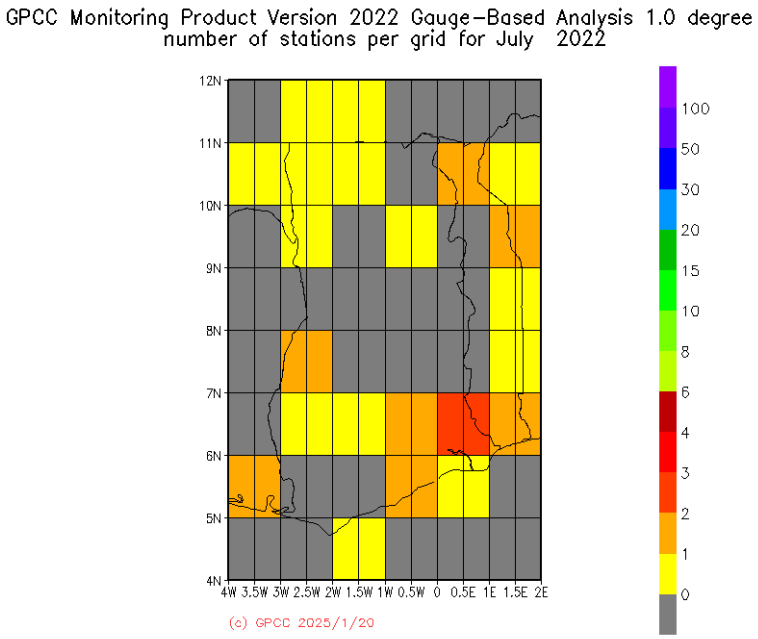
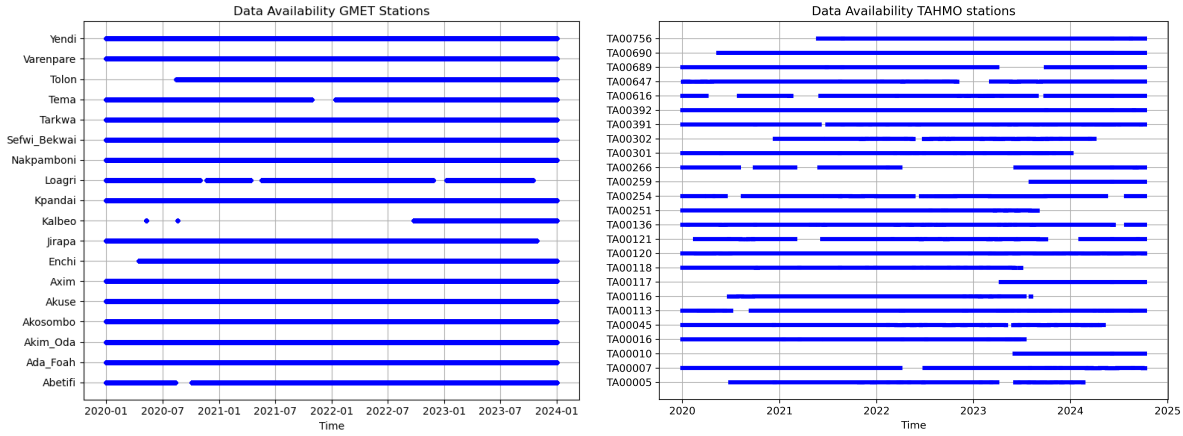


Figure 4.1: Number of stations per grid cell of 1° x 1° used in the GPCC Monitoring Product Version 2022 for July 2022 (Deutscher Wetterdienst, n.d.).

4.1.2. Ground observations

A relative large portion of the ground stations contains data gaps (Figure 4.2). The cumulative rainfall per month for the year 2022 (see Appendix B) was compared to Figure 2.4, to check whether the station aligned with local rainfall patterns. A considerable amount of data did not align with local rainfall patterns, often reporting only zeros, indicating the stations were malfunctioning or clogged. These stations were removed from the validation set to ensure a more reliable analysis. The used stations and their locations are displayed in Figure 4.3.



(a) GMET stations

(b) TAHMO stations.

Figure 4.2: Data availability of the ground stations. The blue lines indicate when data is available.

Ghana Meteorological Service (GMET) stations

Part of the ground station data was received from the Ghana Meteorological Service (GMET). The instruments used to measure the rainfall per station by GMET are unknown to the author of this research.

Trans-African Hydro-Meteorological Observatory (TAHMO) stations

The Trans-African Hydro-Meteorological Observatory (TAHMO) is an initiative that aims to cover the sub-saharian region in Africa with a dense network of hydro-meteorological monitoring stations (van de Giesen et al., 2014). TAHMO uses Atmos 41 gauge stations to measure different meteorological variables, one of them being rainfall. The rainfall is collected with a funnel and drips the water onto an electrode, which counts the amount of droplets. The funnel makes sure that every droplet has the same size for easy conversion of droplet count towards mm rainfall. However, when the station is not placed perpendicular to the ground, droplets could miss the electrode. Currently, this effect is not corrected in the data. The available data used in this research could be seen in Figure C.2.

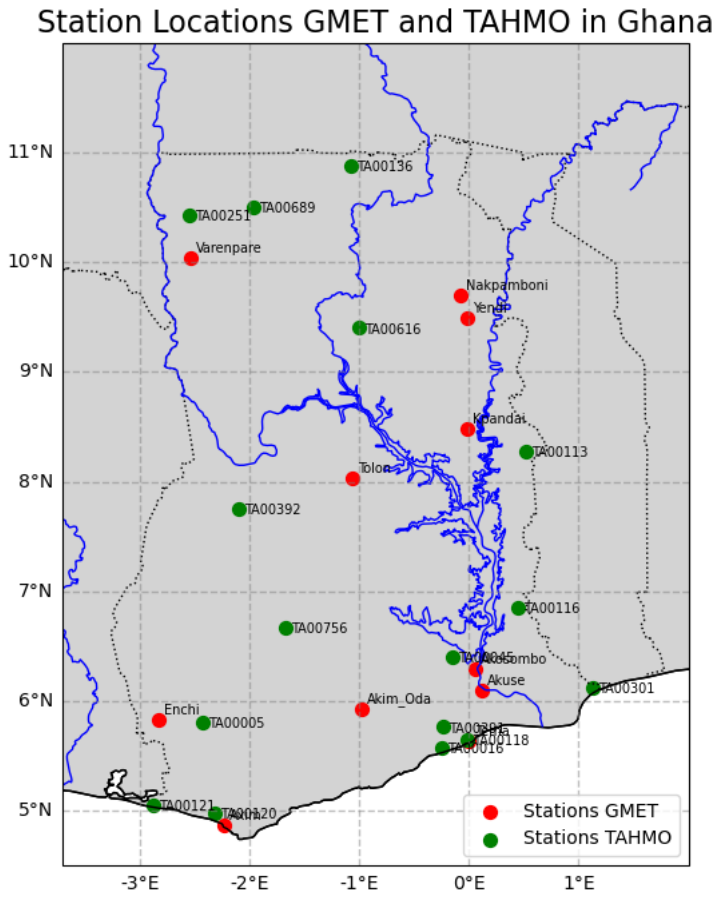


Figure 4.3: The used ground stations for this research plotted on the map in Ghana.

4.2. Input data

4.2.1. Meteosat Satellites equipped with Spinning Enhanced Visible and Infrared Imager (SEVIRI)

The Meteosat satellites are GEO satellites. This research focuses on the second generation of Meteosat satellites, which have a latency time of approximately 15 minutes, a spatial resolution of approximately 3 x 3 km at sub-satellite point, a temporal resolution of 15 minutes and primarily covers Europe and Africa (Schmetz et al., 2002).

The Meteosat Second Generation (MSG) satellite is equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI) system. SEVIRI takes observations of the Earth every 15 minutes using 11 spectral channels, along with a high-resolution visible (HRV) channel. While the 11 spectral channels cover the full disk of the Earth, the HRV channel only covers half of the full disk in the east-west direction but the entire disk in the north-south direction. The scanning process occurs from east to west and south to north. SEVIRI includes eight thermal infrared channels and three solar channels.

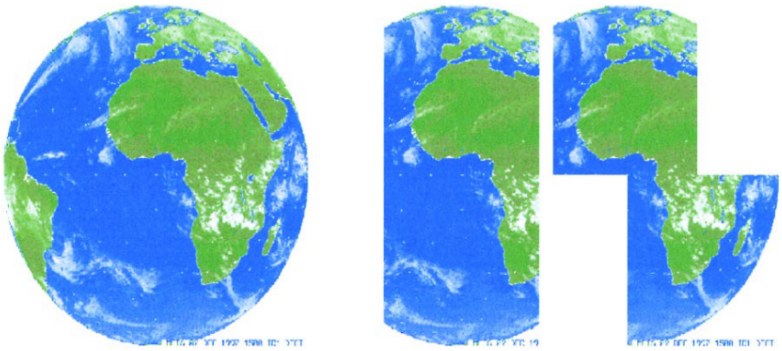


Figure 4.4: . Coverage with SEVIRI on MSG, in its nominal position at 0° lon, for a repeat cycle of 15 min for channels 1–11 (see Table 1). The full disk image has 3712 × 3712 pixels. The HRV (right-hand side of figure) covers only half of the earth in the E–W direction with 11 136 × 5568 pixels; however, the area of imaging can be selected. Scanning of SEVIRI is from east to west and south to north. (Schmetz et al., 2002)

A complete image of the Earth's full disk is 3712 x 3712 pixels for channels 1-11 (Schmetz et al., 2002). The SEVIRI channels observe radiance in mW·m□²·sr□¹·(cm□¹)□¹. For the visual channels (1 t/m 3 and HRV) this is converted to the Bidirectional Reflectance Factor (Equation 4.1) (Eumetsat, 2012) and for the infrared channels (4 t/m 11) to Brightness Temperature (K) (Equation 4.2) (Eumetsat, 2005). This conversion is already applied, when downloading the Native formatted files from the EUMETSAT database. Therefore, to provide the model with richer physical data, the Native files are used.

$$r_{\lambda_i} = \frac{\pi R_{\lambda_i} d^2(t)}{I_{\lambda_i} cos(\theta(t,x))} \tag{4.1}$$

In which i is the channel number, R_{λ_i} is the measured radiance (mW·m□²·sr□¹·(cm□¹)□¹), $d(t)$ is the Sun-Earth distance in AU at time t, I_{λ_i} the band solar irradiance for the channel λ_i at 1 AU in mW·m□²·sr□¹·(cm□¹)□¹ and $\theta(t,x)$ is the Solar Zenith Angle in Radiances at time t and location x.

$$T_b = \frac{[\frac{C_2 v_c}{log(\frac{C_1 v_c^3}{R}+1)} - B]}{A} \tag{4.2}$$

In which T_b is the Brightness Temperature (K). The constants $C_1 = 2hc^2$ and $C_2 = \frac{hc}{k}$ in which c is the speed of light (299792458 ms^{-1}), h the Planck constant (6.62606876 $10^{-34}Js$) and k the Boltzmann constant 1.380650324 $10^{-}23JK^{-1}$. v_c is the central wavenumber of the channel and A and B are parameters found from a non-linear regression pre-calculated lookup table that can be found in Eumetsat (2005).

The channels are selected to provide detailed information on clouds, the Earth's surface, water vapor, and ozone (Table 5.1). Combining several channels can offer additional insights, such as analyzing atmospheric instability (Schmetz et al., 2002). Below an overview is provided of the SEVIRI channels.

- VIS0.6 and VIS0.8: mainly used for cloud tracking, scene identification, aerosol and land surface and vegetation monitoring (Schmetz et al., 2002).
- NIR1.6: used to distinguish between snow and cloud, ice and water clouds and also provides aerosol information (Schmetz et al., 2002). The channel is chosen centered around a wavelength for which the absorption for ice particles is much larger that the absorption for water particles (Brasjen, 2014).
- IR3.9: mainly used for cloud and fog detection. It also helps with land and sea surface temperature measurements at night. Additionally it increases the low level wind coverage from cloud tracking (Schmetz et al., 2002).

- WV6.2 and WV7.3: Are used for observing the water vapor and winds. Additionally it also supports height allocation of semitransparent clouds (Benas et al., 2017). In the research of Estébanez-Mendez et al. (2023) the channel 6 (WV 7.3 μm) was used in combination with channel 9 (IR10.8) to improve a model for binary rain/no rain classification, since this channel could look deeper into the cloud compared to channel 9, and in Ghana mostly deep convective clouds form (Estébanez-Camarena, Curzi, et al., 2023). Earlier research showed that additional information about water vapour can improve the performance of a ML model (Estébanez-Camarena, Taormina, et al., 2023; Wang et al., 2021). Other research uses the difference between the WV 6.2 - WV7.3 μm, because higher differences are associated with a warmer higher upper troposhere and thus contains more moisture and therefore could be associated with higher changes of rainfall rates (Brasjen, 2014).
- IR8.7: Provides information for thin cirrus clouds and supports distinguishing between ice and water clouds (Schmetz et al., 2002).
- IR9.7: Observes absoprtion by the ozone, which can be used for numerical weather prediction (NWP), wind motion in the lower stratosphere and monitoring the evolution of the total ozone field (Schmetz et al., 2002).
- IR10.8 and IR12.0 are important channels to retrieve sea, land surface and cloud-top temperatures. They can be used for the detection of cirrus clouds and volcanic ash clouds as well (Schmetz et al., 2002). IR10.8 corresponds to the top of the cloud and is one of the most important channels for convective rainfall retrieval, due to a strong relationship between cloud top temperatures and rainfall and little affection of adsorption (Brasjen, 2014; Estébanez-Camarena, Taormina, et al., 2023; Schmetz et al., 2002). As a rule of thumb, the colder the infrared cloud top temperature, the higher the rainfall rate (Brasjen, 2014). This is however quite a simplification, since the formation of rainfall is not only dependent on the temperature of the cloud. Although channel 9 provides important information for cloud rainfall, it doesn't capture the internal cloud structure, since it can only look at the top of the cloud (D'Adderio et al., 2023; Estébanez-Camarena, Curzi, et al., 2023; Estébanez-Camarena, Taormina, et al., 2023; Hong and Hsu, 2004; Wang et al., 2021). Therefore it is useful to also account for other factors. Such as the fact that clouds in humid environments produce more rainfall than clouds in dry atmospheric conditions (Brasjen, 2014; Mccollum et al., 2000).
- IR13.4: This channel can be used for improving height allocation of tenuous cirrus clouds and in cloud-free areas to investigate atmospheric stability (Schmetz et al., 2002).

Table 4.1: Channel Characteristics of Main Gaseous Absorbers or Windows (Schmetz et al., 2002).

No.	Spectral Band	λ_{cen} (µm)	λ_{min} (µm)	λ_{max} (µm)	Main Absorber/Window
1	VIS0.6	0.635	0.56	0.71	Window
2	VIS0.8	0.81	0.74	0.88	Window
3	NIR1.6	1.64	1.50	1.78	Window
4	IR3.9	3.90	3.48	4.36	Window
5	WV6.2	6.25	5.35	7.15	Water vapor
6	WV7.3	7.35	6.85	7.85	Water vapor
7	IR8.7	8.70	8.30	9.10	Window
8	IR9.7	9.66	9.38	9.94	Ozone
9	IR10.8	10.80	9.80	11.80	Window
10	IR12.0	12.00	11.00	13.00	Window
11	IR13.4	13.40	12.40	14.40	Carbon dioxide
12	HRV	Broadband	(about 0.4 – 1.1)		Window/Water vapor

The channels 1 - 11 serve as input for the model, assuming that the model will learn meaningful patterns by only attending to the visual channels during daytime.

4.2.2. Auxiliary data

Besides from the SEVIRI channels the model is also provided with additional information about the region. This includes the Digital Elevation Model (DEM) obtained from the NASA dastabase, the time of the day and hour of the day as well as the latitude longitude grid in WGS84 projection of the earth.

4.3. Benchmark products

IMERG-Early

As explained in Section 3.2, IMERG-Early is a precipitation product part of the Global Precipitation Mission, specifically suitable for operational purposes, such as flood forecasting, due to its short latency ($\tilde{4}$h). Unlike IMERG-Final, it doesn't incorporate intensive corrections on gauge stations and only includes a forward pass in the morphing algorithm to define the motion vectors from the VIS/IR channels (Huffman et al., 2023). IMERG-Late also includes a backward pass, but has a latency of approximately 14h, thus is not as real-time as IMERG-Early. Therefore IMERG-Early is selected as a suitable benchmark. The latest version (V06) available through the GPM-API python library is used.

5

Methodology

This Chapter explains the methods used for this research, aiming to address the limitations of current near real-time rainfall retrieval algorithms by adapting a deep learning approach for estimating rainfall from the Meteosat satellite. Additionally, it seeks to validate the model output on local gauge station data, to guide future research.

The study follows a systematic approach, involving a series of steps that combine various input and reference data sources. The first step ensures the spatial and temporal alignment of the reference and input data, as described in Section 5.1. The model's architecture builds on the Earthformer for INCA model, originally developed by Amazon (Gao et al., 2022) and later adapted by Küçük, Atencia, and Dabernig (2024), as outlined in Section 5.2.1. How the model is trained is explained in Section 5.2.2.

To determine the accuracy of the model (Section 5.3), the model is validated both on IMERG-Final and the ground observations. The test dataset is compared to the target IMERG-Final and benchmarked on IMERG-Early and explained in Section 5.3.1. The validation on the ground observations is outlined in Section 5.3.2.

An overview of the entire research setup can be found in Figure 5.1.

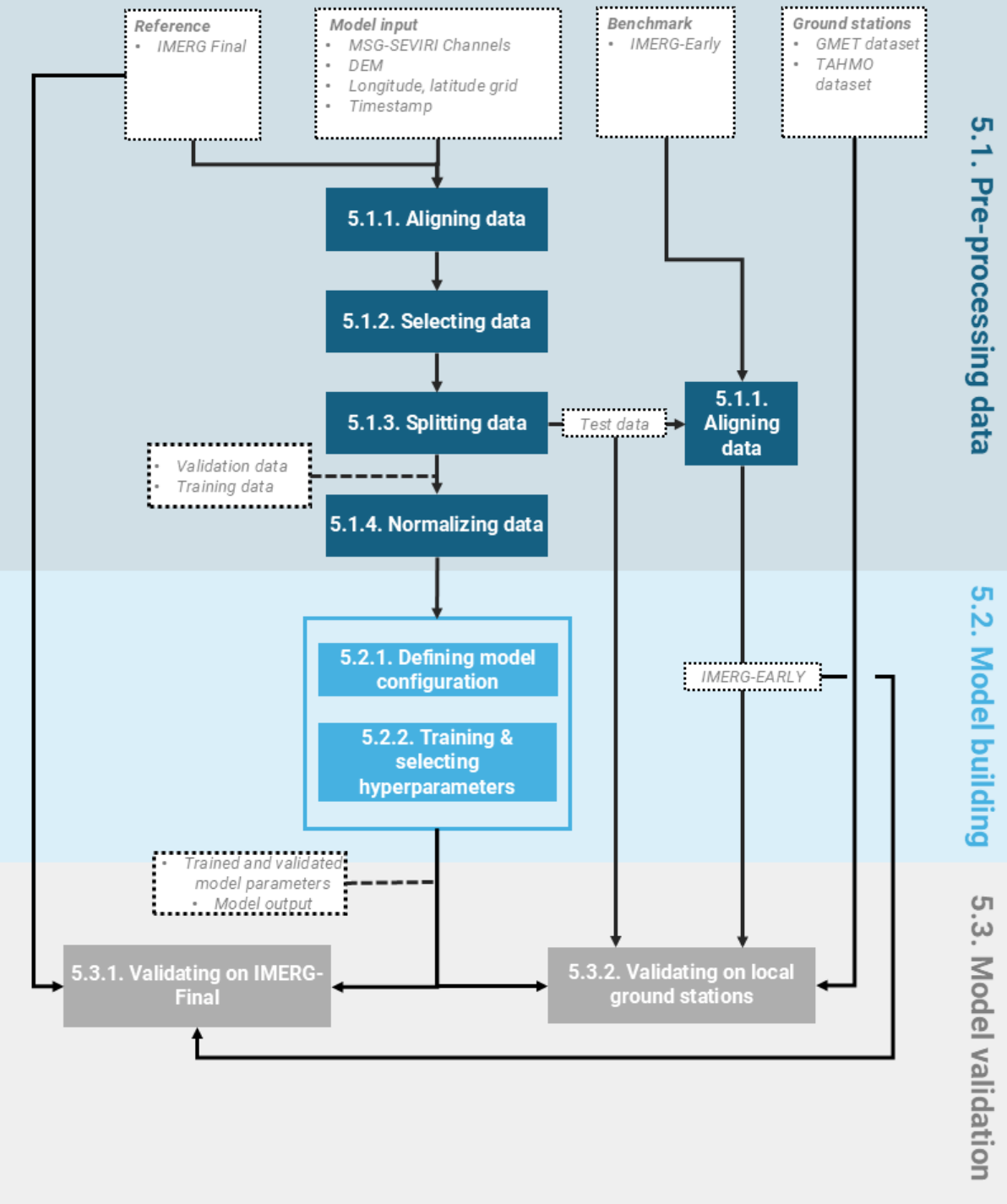


Figure 5.1: Overview of the methodology of the research. The research consists out of three main steps; pre-processing of the data, building the model and validating and optimizing the model output.

5.1. Data preprocessing

5.1.1. Aligning data

The SEVIRI data is retrieved from the EUMETSAT database, using the customization process from the Eumdac Python library. The data is then reprojected from the geostationary projection to WGS84 projection, using the Satpy Python library. The IMERG-Final data is retrieved from the ges-disc using the GPM-API Python library.

The Meteosat data and the IMERG-Final data are aligned by regridding and resampling the IMERG-Final data to the Meteosat grid using nearest neighbor resampling. The coverage of both the Meteosat and IMERG-Final data is longitude (3.658607 W, 1.308607 E) and latitude (4.541295547 S, 11.28623482 N) in WGS84 projection and has 248 number of grid cells in latitude direction and 184 grid cells in longitude direction, resulting in a resolution of approximately 0.027° x 0.027°/ 3 x 3 km.

The data is aligned in time by grouping the data in event files of a timespan of 2 hours. The IMERG-Final data in the given timespan, is the target data and corresponds to the SEVIRI data available every 15 minutes two hours prior to the event.

The Digital Elevation model is downloaded from the NASA database and resampled to the same grid using bi-linear interpolation.

The time of the day and hour of the day are transformed into sine and cosine wave signals (Equations 5.1, 5.2, 5.3, 5.4)) a strategy taken over from Küçük, Atencia, and Dabernig (2024).

$$DoY_{sintrans} = sin(\frac{DoY * 2\pi}{365.2425}) \tag{5.1}$$

$$DoY_{costrans} = cos(\frac{DoY * 2\pi}{365.2425}) \tag{5.2}$$

$$ToD_{sintrans} = sin(\frac{ToD * 2\pi}{365.2425}) \tag{5.3}$$

$$ToD_{costrans} = cos(\frac{ToD * 2\pi}{365.2425}) \tag{5.4}$$

In which the ToD is the time of the day and DoY, the day of the year.

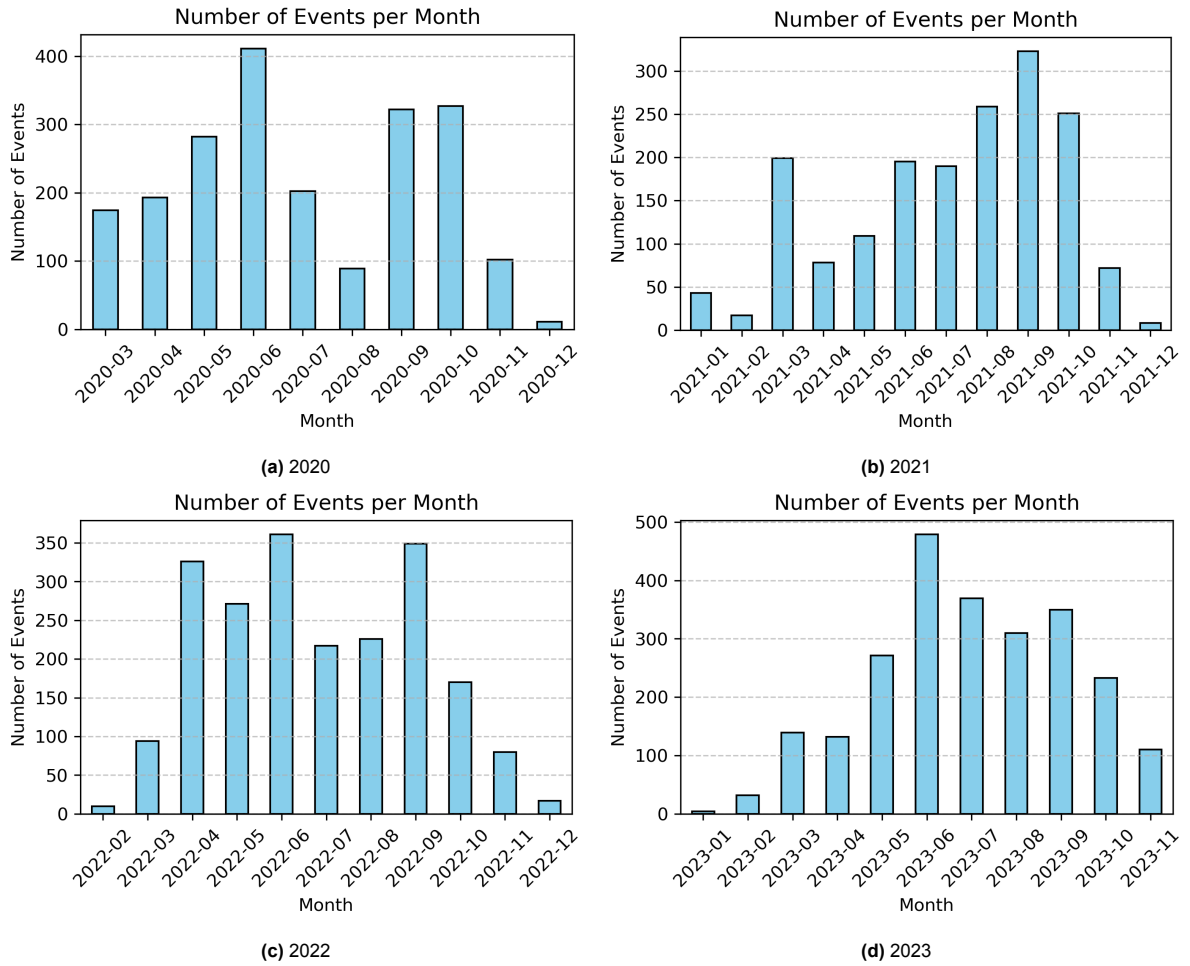
For benchmarking on IMERG-Early, the same spatial boundaries and temporal timestamps as the test dataset were selected to compare it with IMERG-Final.

5.1.2. Selecting data

Machine learning models often struggle with datasets dominated by zeros, as they may achieve good loss scores while failing to represent reality, leading to underestimation of extreme rainfall and low detection probabilities.

Therefore, to estimate extreme rainfall for flood early warning systems, the dataset of the years 2020 - 2023 is conditioned on a minimum spatial coverage of 2% (8,214 km^2) with at least 1 mm/h rainfall and a minimum maximum rainfall rate of 20 mm/h, retrieved from IMERG-Final. Samples with missing SEVIRI data are excluded, resulting in 7,676 samples. Due to seasonal variability in Ghana, differences in file distributions per month occurred (Figure 5.2).

A less strict selection criterion (1% spatial coverage, 5 mm/h minimum maximum) produced 25,940 samples but resulted in worse model performance, probably due to the high representation of zeros in the data set. Therefore, it was not explored further.



(a) 2020 **(b)** 2021 **(c)** 2022 **(d)** 2023

Figure 5.2: The number of samples selected for each month for each year with a condition of a spatial coverage of rainfall rate > 1 mm/h > 2% and a maximum rainfall rate > 20 mm/h.

5.1.3. Splitting data

A stratified method is used to split the data into training, test and validation datasets. The entire dataset is split into 70% training 21% validation and 9% testing. The spatial averaged (IMERG-Final) rainfall of each sample in the conditioned dataset is calculated and divided into 20 equally spaced bins between the minimum and maximum value of the mean rainfall. The distribution of the samples across each bin is persevered in a training, test and validation dataset.

Since the most extreme rainfall sample was rare in the dataset and resulted in only one count in the last bin, the dataset couldn't be split equally. Therefore the sample in the last bin was merged with the nearest bin with a lower mean rainfall eventually resulting into 19 bins (Figure: 5.3).

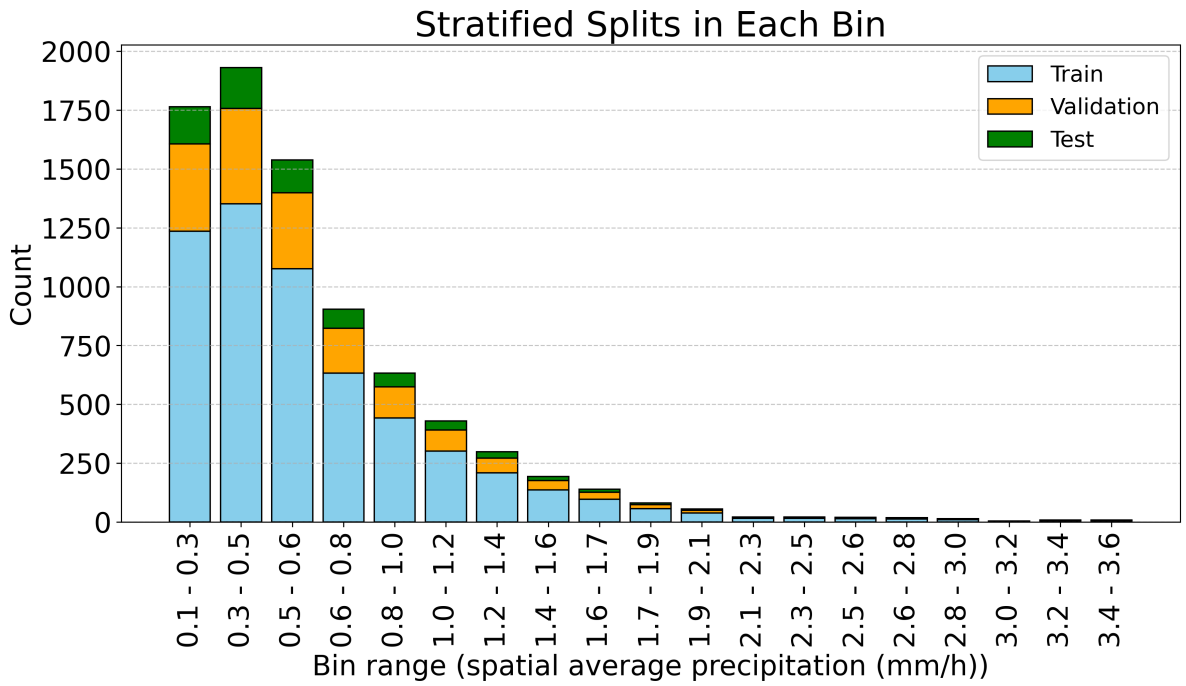


Figure 5.3: The binned distribution of the spatial averaged rainfall (mm/h) calculated from the target data (IMERG-Final) for all the values in the conditioned (spatial coverage of rainfall > 1 mm/h > 2% and maximum rainfall > 20 mm/h) dataset.

5.1.4. Normalizing data

To address the skewed distribution of the IMERG-Final rainfall data the data is logtransformed, using the strategy of Küçük, Atencia, and Dabernig (2024) and Leinonen et al. (2023) (Equation 5.5). The discontinuity of the transformed rainfall gives the model a clear distiction of rain and no rain values.

$$f(RR) = \begin{cases} \log_{10} RR & \text{if } RR \geq 0.1 \text{ mm h}^{-1}, \\ \log_{10} 0.02 & \text{if } RR < 0.1 \text{ mm h}^{-1}. \end{cases} \tag{5.5}$$

In which $f(RR)$ is the transformed rain rate and RR is the rain rate.

The other data sources are normalized (Equation 5.1.4) over all the grid cells using the global mean of all the grid cells for calculating the standard deviation (Equation 5.6) of all the grid cells in the entire dataset.

$$\sigma_{global} = \sqrt{\frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{i=j}^{M} (x_{i,j} - \mu_{global})^2} \tag{5.6}$$

In which N is the total number of frames for the given data source in the selected data set. M is the total number of grid cells in a specific frame of the given data source and $x_{i,j}$ the value of each individual grid cell.

$$z_i = \frac{x_i - \mu_{global}}{\sigma_{global}} \tag{5.7}$$

In which z_i is the normalized value of each individual grid cell in the selected dataset.

Normalizing the log transformed rainfall values was tested, but gave similar results as the log transformed values were already quite close to normalized values.

5.2. Model building

5.2.1. Defining model configuration

An extensive description of the model architecture can be found in Appendix A. To enable the division of the input tensor for patch and merging in the downsampling layers, the borders are padded on the right with zero values when necessary. The Earthformer for INCA model from Küçük, Atencia, and Dabernig (2024) has been adapted to allow unequal padding across multiple dimensions.

The configuration of the model has been left mostly the same and is described more specifically in Appendix A. More gradual downsampling of the convolutional downsampling layers was tested by increasing the number of downsampling layers from two to three and increasing the channel dimensions more gradually each layer to 16, 64 and 128 respectively. However, this did not improve model performance.

The number of heads was increased from 16 to 32, so the model could focus on an increased number of structures in the data simultaniously (Appendix A.1). The model was adapted so it would only estimate one rainfall frame and in such a way it could handle the provided input data of this research.

5.2.2. Training and selecting hyperparameters

During the training phase, the hyperparameters are specified. To prevent the model from converging too rapidly when updating weights, a learning rate schedule with a warm-up phase is applied, ensuring gradual adjustments at the start.

A one cycle learning rate scheduler (Appendix A.2.6) with warm-up percentage of 22% is introduced to ensure the model learns meaningful attention patterns at the beginning of training (Figure 5.4). Over 22% of the total training steps the learning rate is increased from an initial value 25 times smaller than the maximum learning rate (8.0e-5), until the maximum learning rate is reached. This gradual increase prevents oscillations or divergence during the early training phase. To further refine the model after meaningful patterns have been learned the learning rate is gradually decreased over the remaining steps using Cosine decay with a final learning rate 10 times smaller than the initial learning rate.

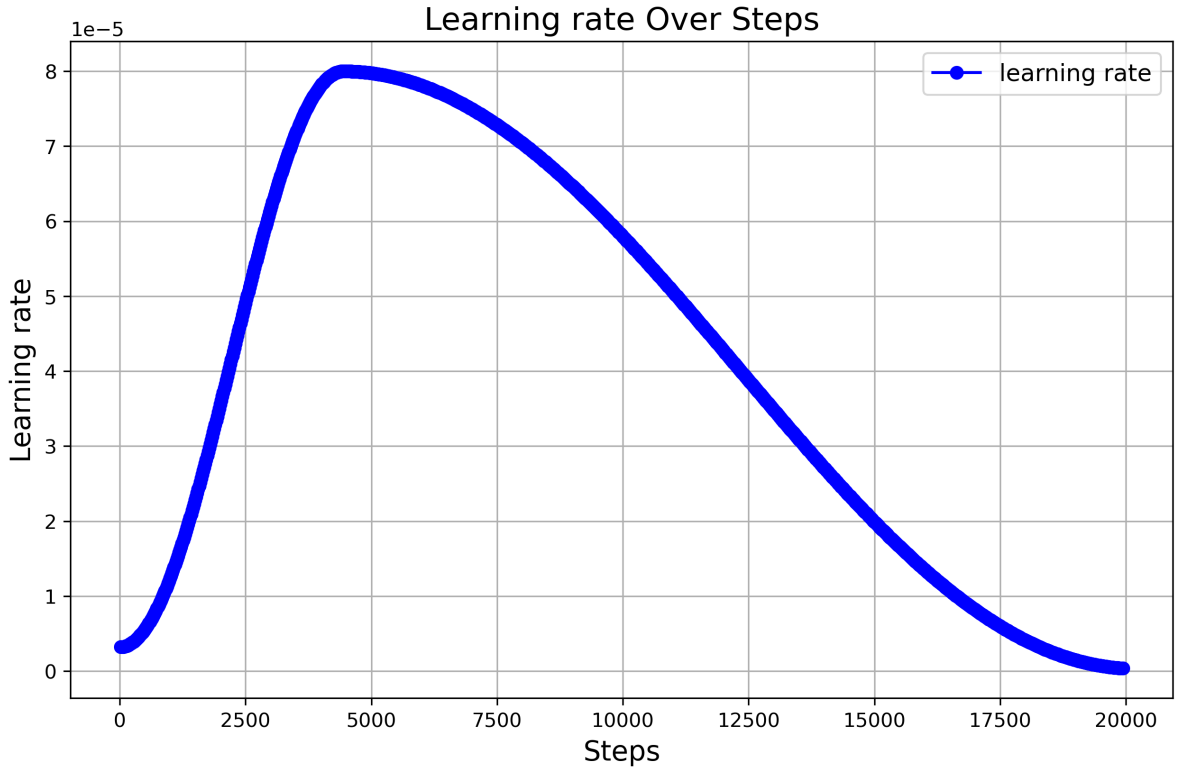


Figure 5.4: The learning rate against the trainingsstep during training of the Earthformer model with MSE loss.

Additionally, to prevent exploding gradients, gradient clipping using an L_2-norm is applied (Appendix A.2.3). When a gradient exceeds the L_2-norm threshold, it is clipped to the defined limit, which in this model is set to 1.0.

Early stopping (Appendix A.2.5) and an AdamW optimizer with weight decay (Appendix A.2.4) are used to prevent overfitting. The training is stopped if the validation loss doesn't improve for 20 epochs. The best performing weights for the validation set were selected as the optimal parameters after training. The weight decay was set to 1.0e-5.

The choice of the loss function could potentially influence the model's performance. The loss function should be selected based on the specific objectives of the model. For this research two different loss functions are explored; the Mean Squared Error (MSE) loss function is employed, as defined in Equation 5.8 and the balanced loss function with weighting similar to Shi et al. (2017) as defined in Equation 5.10. The last loss function assigns greater importance to extreme values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(RR_i - \hat{RR}_i)^2 \tag{5.8}$$

In which n is the total number of grid cells, RR_i, the estimated output of the model in grid cell i and $\hat{RR}_i$ the target value in grid cell i.

$$Loss = BMSE + BMAE \tag{5.9}$$

$$BMSE = \frac{1}{n}\sum_{i=1}^{n} w_i(RR_i - \hat{RR}_i)^2 \tag{5.10}$$

$$BMAE = \frac{1}{n}\sum_{i=1}^{n} w_i|RR_i - \hat{RR}_i| \tag{5.11}$$

$$w_i = \begin{cases} 1 & \text{if } RR_i \geq 1 \text{ mm h}^{-1}, \\ R_i & \text{if } < RR_i \geq 30 \text{ mm h}^{-1} \\ 30 & \text{if } RR_i > 1 \text{ mm h}^{-1}, \end{cases} \tag{5.12}$$

The model is trained on an NVIDIA-SMI 535.183.01 GPU with 9.77 GiB of available memory. Using a GPU significantly accelerates training due to the parallelizable nature of deep learning computations. However, GPU memory now limits the model performance. To address issues such as out of memory, several strategies are implemented in the model: data is loaded in batches of four using the PyTorch Lightning library to conserve memory, while gradient accumulation (Appendix A.2.2) allows weight updates after processing 32 samples, simulating a batch size of 32. Additionally, mixed-precision training (Appendix A.2.1), which uses both 32-bit and 16-bit precision for weights, further reduces memory usage. The training of the model on a GPU with 9.77 GiB took about 2 days and was stopped after 114 epochs for the MSE loss function with an early stopping criterium of 20 epochs and the model with the balanced loss function fully used the 120 maximum epochs.

5.3. Model validation

5.3.1. Validating on IMERG-Final

After the training phase, the model is tested on the test dataset. The model generates rainfall estimates based on the given input data and is compared to the reference data (IMERG-Final). The model's outputs are gridded estimations of rainfall in mm/h averaged over 30 minutes, which are compared to the gridded rainfall rates of IMERG-Final. Based on this comparison, the test metrics are calculated for specific thresholds.

Keeping the final integration into a FEWS in mind, the performance of the model is assessed. A FEWS requires accurate spatial estimates of rainfall rates, focusing more on correctly identifying extreme rainfall locations than on precise intensity values. The Earthformer model, which translates satellite images into rainfall rate estimates, is assessed, while this aspect is kept in mind. Important indicators that quantify the trade-off between the false alarms and misses are the Probability of Detection (POD) (Equation 5.13), Critical Success Index (CSI) (Equation 5.14), Bias (Equation 5.15) and the Success Ratio (SUCR) (Equation 5.16).

The performance indicators are assessed for the thresholds 1.2, 3.4, 5.4 10, 16.8 and 30 mm/h averaged over 30 minutes, with 30 mm/h being the 99th percentile in West Africa (Klein et al., 2018). Defining a clear threshold for 'extreme' rainfall causing flooding, helps assessing the performance of the model. However, whether extreme rainfall results in flooding is not only dependent on the intensity, but also on local factors such as the topography, land use, and drainage infrastructure, making it difficult to establish a threshold for the whole of Ghana.

$$POD = \frac{hits}{hits + misses} \tag{5.13}$$

$$CSI = \frac{hits}{hits + misses + falsealarm} \tag{5.14}$$

$$Bias = \frac{hits + falsealarms}{hits + misses} \tag{5.15}$$

$$SUCR = \frac{hits}{hits + falsealarm} \tag{5.16}$$

The number of hits corresponds to the number of cells where both the model and the target detect rainfall exceeding the threshold. Misses represent the number of cells where the target detects rainfall exceeding the threshold, but the model does not. False alarms refer to the number of cells where the model detects rainfall exceeding the threshold, but the target does not. These values are calculated separately for each threshold.

The POD indicates the probability of detecting rainfall. A POD value of 1 represents a perfect score. The CSI is an indicator of the overall performance of the model; the closer the CSI is to 1, the better the performance. Bias measures the tendency of the model to overestimate or underestimate rainfall detection by evaluating the number of false alarms relative to hits and misses. A score higher than 1 indicates overdetection and a score below 1 underdetection. The perfect score is 1, but this doesn't necessarily indicate good performance of the model. The Success Ratio evaluates how much of the model's predicted rainfall corresponds to rainfall in the target dataset and tells something about the precision of the model. Again 1 reflects a perfect score.

To asses the performance against other near real-time satellite products the computed metrics from comparison to the target IMERG-Final are compared to the the metrics computed from IMERG-Early to IMERG-Final. IMERG-Early is compared to IMERG-Final for the same timestamps as the test dataset and for the same spatial boundaries. However, the comparison of IMERG-Early to IMERG-Final is in the original resolution of IMERG (0.1 x 0.1 degrees), while the output of the model is compared to the resampled resolution of the Meteosat satellite images (0.027 x 0.027 degrees).

To visually assess the performance of the model, the root mean square error (RMSE) (mm/h) averaged over 30 minutes is calculated for each pixel individually with Equation 5.17.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(RR - \hat{RR})^2} \tag{5.17}$$

In which n is the number of frames/ timesteps within the test dataset RR, the value of IMERG-Final in the pixel and $\hat{RR}$ the value of either the output of the model or IMERG-Early in the pixel.

5.3.2. Validating on local ground stations

After the model's output has been spatially compared to the target data, the model's output is compared to the individual gauge stations of both GMET and TAHMO. The GMET and TAHMO data is aggregated over 30min intervals to compare to the rain rate (mm/h) averaged over 30 min of the satellite products. To convert the rain rate into an accumulative rainfall value over 30 min (mm), the rain rate is divided by two. The comparison is drawn between the nearest grid cell to the station location and the station data itself. The Pearson correlation coefficient (Equation 5.18) is used to calculate the correlation between the stations and the nearest grid cell. For the analysis zero data is excluded. Other correlation coefficients such as Spearman and Kendall rank correlation gave similar results. The Pearson correlation coefficient seemed most suitable, as this is the quantification of a linear correlation.

$$r = \frac{[(\sum_{i=1}^{n} x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i]}{\sqrt{[(\sum_{i=1}^{n} x^2) - (\sum_{i=1}^{n} x_i)^2][(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2]}} \tag{5.18}$$

In which x_i are the independent variables, y_i the dependent variables and n the sample size.

The same performance calculations used for the spatial comparison, as described in Section 5.3.1, are applied here. Given that the primary objective of this research is to predict extreme rainfall, the model's performance is also assessed across various rainfall intensities.

In this research rainfall intensities are classified as follows:

Table 5.1: Rainfall intensities classification.

Class	Intensity (mm/h)	Cumulative rainfall (mm) fallen within 30 minutes
Light rain	0.1 - 2.7	0.1 - 1.3
Moderate rain	2.8 - 8.6	1.4 - 4.3
Heavy rain	8.6 - 50	4.4 - 25
Extreme rain	≥ 50	≥ 25

Ansah et al. (2020) investigated the meteorological causes of extreme rainfall and analyzed two flood events in Accra and Kumasi. Their study reported daily rainfall totals of 124.8 mm, 90.1 mm, and 109.0 mm at different stations in Accra during an 8-hour flood event. This suggests that rainfall rates of approximately 15 mm/h can contribute to flooding, although the duration and peak intensities of rainfall also play an important role.

To establish appropriate thresholds for evaluating model performance, the 99th percentile of observed rainfall of the ground stations was analyzed for different time intervals. For 30-minute intervals, the 99th percentile of rainfall (≥ 0.1 mm) across all available station data ranged from 16.57 mm to 31.96 mm, with one outlier at Tarkwa station (52.36 mm). For daily intervals, the 99th percentile varied between 37.6 mm and 93.5 mm, except for TA00120 and Tarkwa stations, which recorded 107.9 mm and 165.5 mm, respectively.

Based on these findings and to ensure alignment with the validation on IMERG-Final, the following threshold values were chosen:

30-minute intervals: 0.6, 1.4, 4.3, 10, 15 and 25 mm
Daily intervals: 5, 10, 15, 30, 40, and 60 mm

As the comparison of IMERG-Final was in mm/h averaged over 30 minutes and the comparison of the ground stations is in mm rainfall fallen within 30 minutes.

This comparison is also conducted for both the IMERG-Early and IMERG-Final product to evaluate and compare the performance of these products to the model performance, when validated on gauge stations.

To further investigate the relationship between satellite-based rainfall estimates and ground stations, a time series analysis is conducted. Additionally, the conditional probability of the brightness temperature of channel 9, given rainfall ≥ 1 mm, is visualized to identify which rainfall processes are in place.

6 Results

In this chapter, the results of the research are presented. Section 6.1 provides a visual representation of the model results for different events. In Section 6.2, the model outputs are compared to IMERG-Final, which serves as the reference truth. Lastly, in Section 6.3, the satellite estimates are validated against ground station observations.

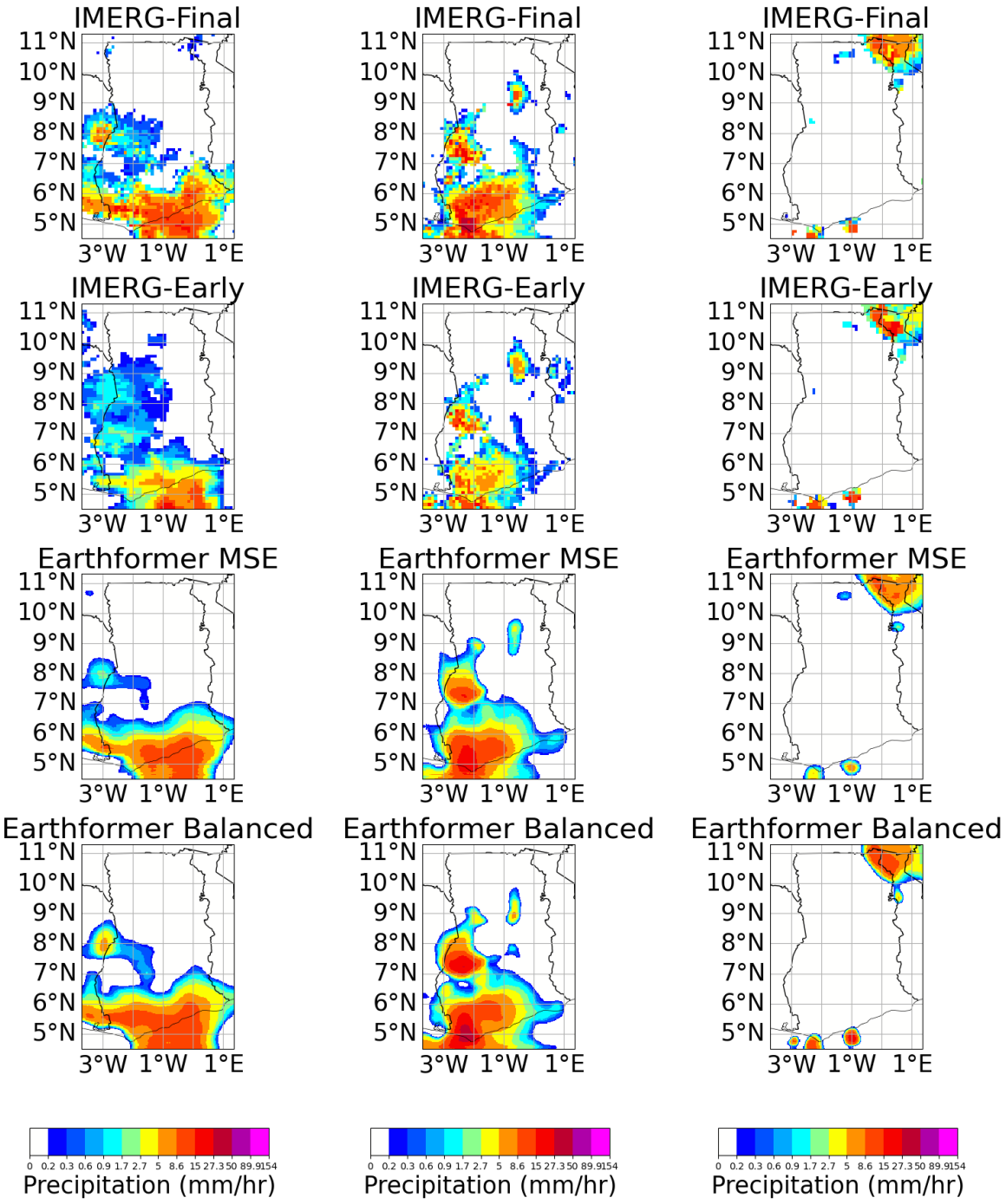
6.1. Visual representation output results

Figure 6.1 shows three seperate example rainfall events from the test dataset as estimated from satellites. The first row shows the rainfall estimates of IMERG-Final, the target to which the model is trained. In the second row the estimates of IMERG-Early are displayed, which is the rainfall product on which the output of the product is benchmarked. The third row shows the output of the Earthformer model with MSE loss and the last row shows the model with balanced loss. A brief visual inspection already reveals that both the Earthformer model with MSE loss and the model with balanced loss tend to smooth out rainfall patterns compared to IMERG-Final.

The first event (Figure 6.1a) shows how IMERG-Early misplaces the areas of higher rainfall intensity, while both Earthformer models provide a more accurate estimation. However, both Earthformer models tend to under detect light rainfall, whereas IMERG-Early over detects it.

In the second event (Figure 6.1b), the Earthformer model with MSE loss fails to capture the most intense rainfall, whereas the model with balanced loss successfully detects it. Despite the smoothing effect, the Earthformer model's output closely resembles IMERG-Final. While IMERG-Early also shows similarities to IMERG-Final in terms of the shape, its estimates underestimate the intensities.

In the third event (Figure 6.1c), the Earthformer model with MSE loss underestimates the intensity of a small cluster of rainfall cells in the lower part of the figure but successfully detects light rainfall in the upper right corner. In contrast, the Earthformer model with balanced loss misses this light rainfall entirely, but observes the higher intensities of rainfall in the lower part of the Figure. IMERG-Early observes the small rainfall cells correctly, but in the large cell both overestimates and underestimates the rainfall intensity, although getting the overall shape correctly.



(a) Event 1 **(b) Event 2** **(c) Event 3**

Figure 6.1: Different examples of rainfall events in Ghana of the test dataset; event 1 (2020-04-26 13.00), event 2 (2022-04-24 17.30) event 3 (2020-05-17 10.00). From top to bottom: IMERG-Final, IMERG-Early, Earthformer with MSE loss, and Earthformer with balanced weighting loss. The colours indicate the different rainfall intensities.

6.2. Comparison to IMERG-Final (target)

6.2.1. Performance indicators comparison

Figure 6.2 compares the performance of the Earthformer model, trained with different loss functions, against IMERG-Early using various rainfall rate thresholds, with IMERG-Final serving as the reference ('truth'). The metric scores are derived by comparing the test dataset's output frames to IMERG-Final, while IMERG-Early is evaluated over the same time steps and spatial boundaries (see Section 5.3.1 for details on metric computation).

The x-axis represents increasing rainfall thresholds, while the y-axis shows the corresponding performance scores. CSI, POD, and SUCR range from 0 to 1, with higher values indicating better performance. Bias ranges from 0 to $+\infty$, where 1 is ideal; values below 1 indicate underestimation (more missed detections), while values above 1 suggest overestimation (more false alarms).

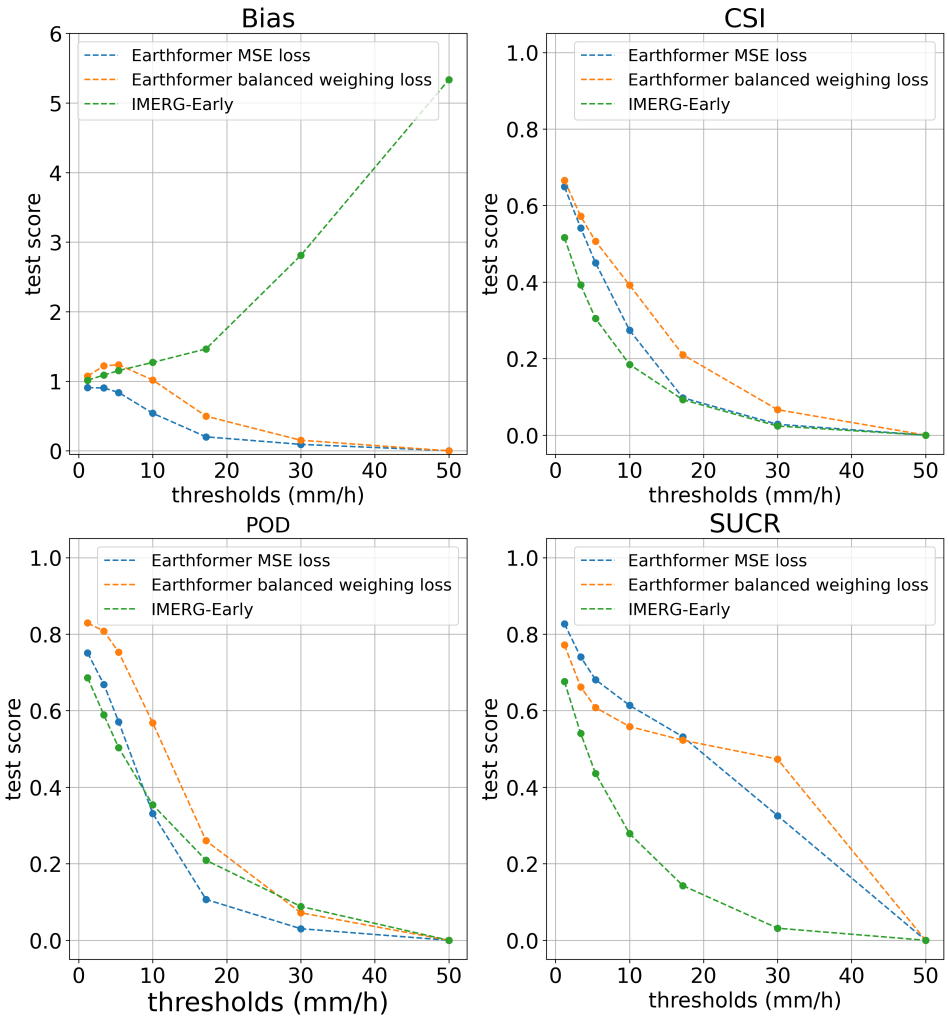


Figure 6.2: The figure shows the metric scores for varying intensities of the Earthformer models and IMERG-Early, compared to IMERG-Final. The top left panel displays the Bias (perfect score = 1), the top right shows the Critical Success Index (perfect score = 1), the bottom left presents the Probability of Detection (perfect score = 1), and the bottom right illustrates the Success Rate (perfect score = 1).

Looking at the Critical Success Index (CSI), which is the performance indicator to show the overall performance of the model, the Earthformer model with balanced loss obtains a higher score compared to the model with MSE loss, especially at higher rainfall rate thresholds. However, when examining the biases, the balanced loss function's emphasis on higher rainfall rates leads to slightly more frequent rainfall detections, at the cost of overestimating the presence of rainfall for lower thresholds. In contrast,

the MSE loss model shows a slight under detection for smaller thresholds. This results in a higher POD for the balanced loss model, but a lower SUCR at smaller thresholds.

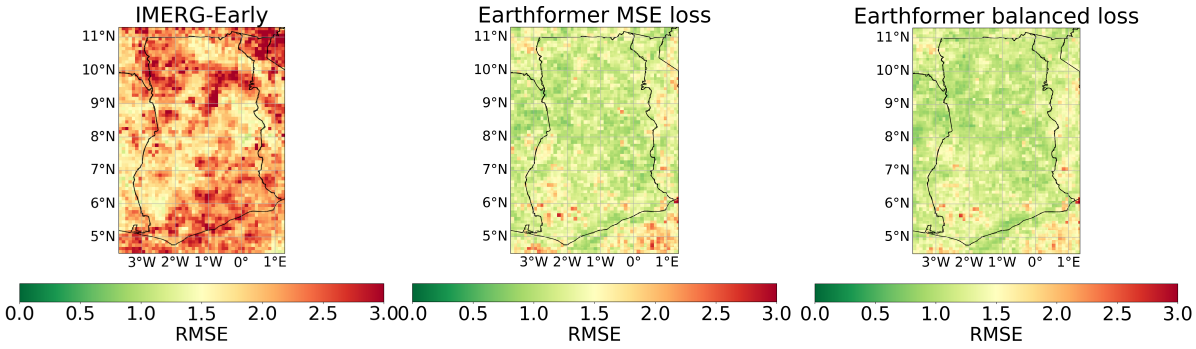
The Earthformer model with balanced loss generally outperforms IMERG-Early across all metrics. On the other hand, the Earthformer model with MSE loss performs better in terms of SUCR for lower and mid thresholds. However, looking at the CSI, as the thresholds increase, its performance becomes more similar to IMERG-Early, since it struggles to detect higher rainfall rates and strong decrease in POD.

Both Earthformer models tend to miss rainfall more frequently at higher thresholds (Bias below 1 and decreasing POD), while the Bias of IMERG-Early increases as the threshold rises, indicating that for higher thresholds a larger proportion of its detected rainfall is a false detection. The Earthformer models on the other hand tend to miss rainfall for higher thresholds more frequently. It is important to note that all the satellite estimates show a decreasing performance trend as the rainfall threshold increases.

6.2.2. Spatial RMSE comparison

Figure 6.3 presents a spatial comparison of RMSE (mm/h) within the study area, comparing the Earthformer model outputs for the test dataset against IMERG-Final. IMERG-Early is similarly evaluated against IMERG-Final for the same time steps and spatial boundaries (a detailed explanation is found in Section 5.3.1).

IMERG-Early exhibits an RMSE ranging from approximately 1.5 to 3, while both Earthformer models show lower errors, generally between 0.5 and 1.5 across most of the domain. The Earthformer models with MSE and balanced loss produce similar results, though the MSE loss model shows slightly higher errors, particularly over coastal regions. This increased error over coastal areas is likely due to higher observed rainfall rates in these regions.



(a) RMSE IMERG-Early **(b)** RMSE Earthformer MSE loss function **(c)** RMSE Earthformer balanced loss function

Figure 6.3: The RMSE (mm/h) averaged over 30 minutes calculated for each grid cell individually.

6.3. Comparison to ground stations

The comparison between the satellite outputs and the ground stations is done for the year 2022 by comparing the nearest grid cells to the gauge station observations (both GMET and TAHMO). However, it is important to note that approximately 8% of this dataset are estimates from the training dataset.

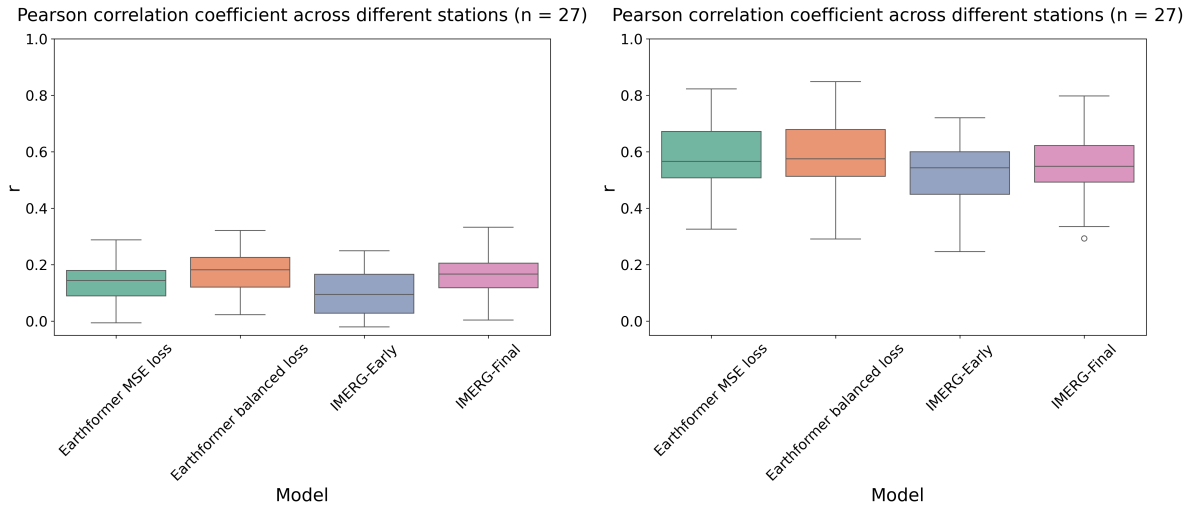
6.3.1. Correlation coefficient

The Pearson correlation is calculated by comparing the nearest grid cells to gauge station observations (both GMET and TAHMO) for the year 2022, with zero-data points excluded (see Section 6.3 for details). An analysis including zero-data, as well as separate evaluations for GMET and TAHMO stations, and comparison only on the test dataset can be found in Appendix D.

Figure 6.4 presents a boxplot of Pearson correlations of each satellite product with all the used stations (Figure 4.3). The line within each box represents the median, while the box boundaries indicate the first

and third quartiles. Whiskers extend to the minimum and maximum values, excluding outliers, defined as points beyond 1.5 times the interquartile range (IQR). Correlations at a 30-minute interval are shown on the left (Figure 6.4a), while daily interval correlations appear on the right (Figure 6.4b).

For short 30-minute intervals, the correlation across all rainfall products is relatively low, with quite some variation between stations. When aggregated to daily intervals, the correlation increases, but the spread between stations remains relatively high and even increases. This suggests that either the accuracy of the satellite rainfall products vary by region or that the accuracy of station measurements differs across stations. The Figure also shows that the Earthformer models achieve a similar correlation with gauge stations as IMERG-Final.



(a) 30 min time interval

(b) daily time interval

Figure 6.4: Pearson correlation coefficient by comparing the nearest grid cell to the ground station observations of both GMET and TAHMO. Each boxplot represents the Pearson correlation among different stations for different models.

To assess which stations show higher correlations relative to others, their locations are visually mapped in Figure 6.5 and Figure 6.6. The same analysis, but with inclusion of zero data can be found in Appendix C. From visual inspection, stations situated in the Savannah climate region (Figure 2.5) generally show higher correlations compared to those in the coastal or forest climatic zones. Additionally, stations that show weaker correlations at the 30-minute interval also tend to show weaker correlations at the daily interval, and vice versa.

Notably, some stations show particularly low correlations across all satellite-based estimates. These include Tolon, Enchi, TA00113, and TA00391, indicating potential challenges in capturing rainfall variability at these locations.

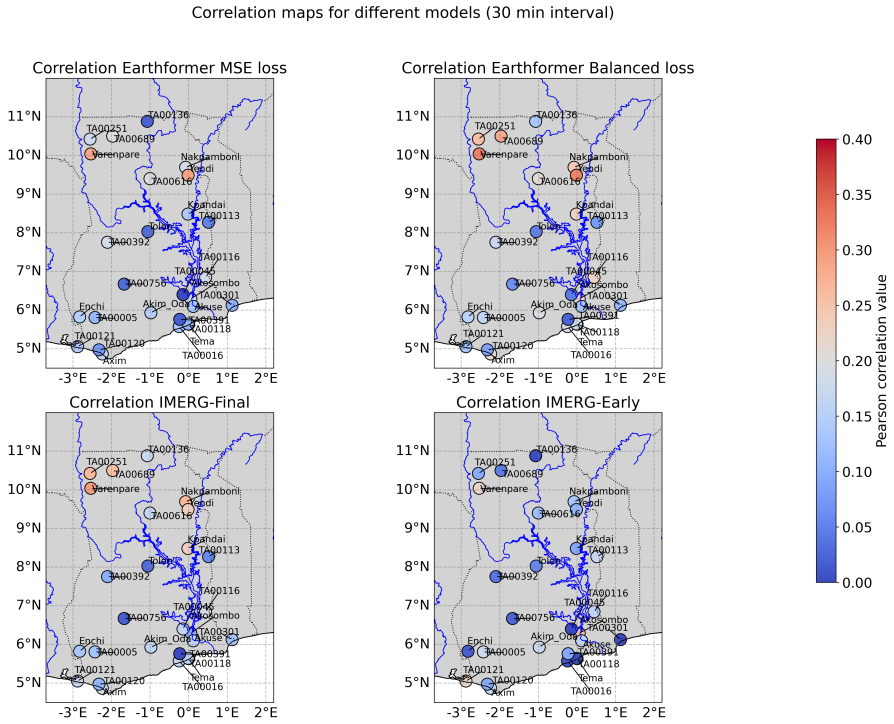


Figure 6.5: The Pearson correlation for each used ground station in Ghana, plotted on the map for a 30 minute interval. The top left shows the correlation of the Earthformer model with MSE loss, the top right the correlaiton of the Earthformer with balanced loss, the bottom left the correlation with IMERG-Final and the bottom right the correlation with IMERG-Early.

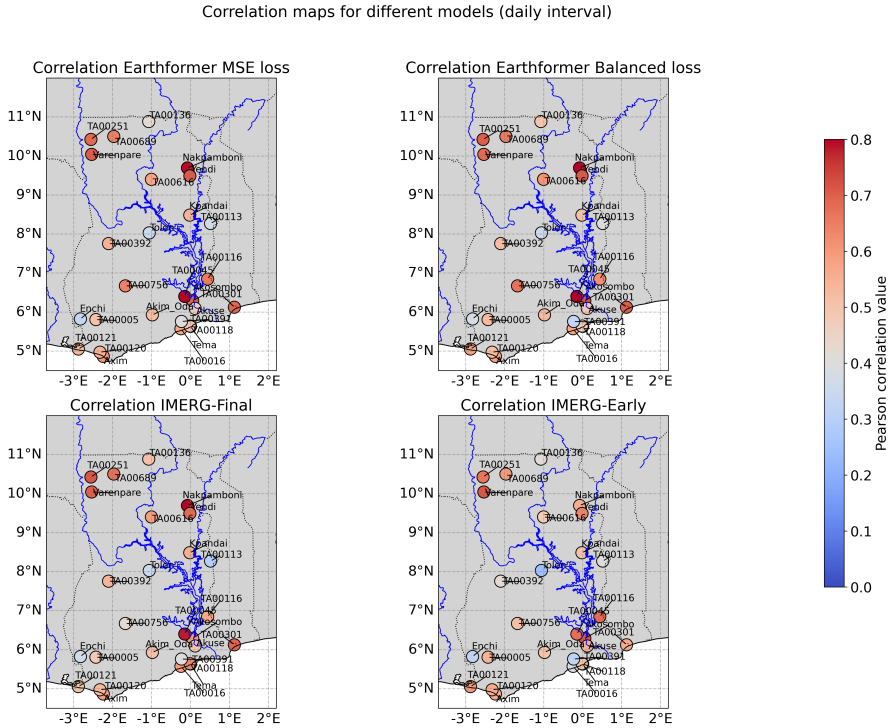


Figure 6.6: The Pearson correlation for each used ground station in Ghana, plotted on the map for a daily interval. The top left shows the correlation of the Earthformer model with MSE loss, the top right the correlaiton of the Earthformer with balanced loss, the bottom left the correlation with IMERG-Final and the bottom right the correlation with IMERG-Early.

6.3.2. Performance indicators compared to ground stations

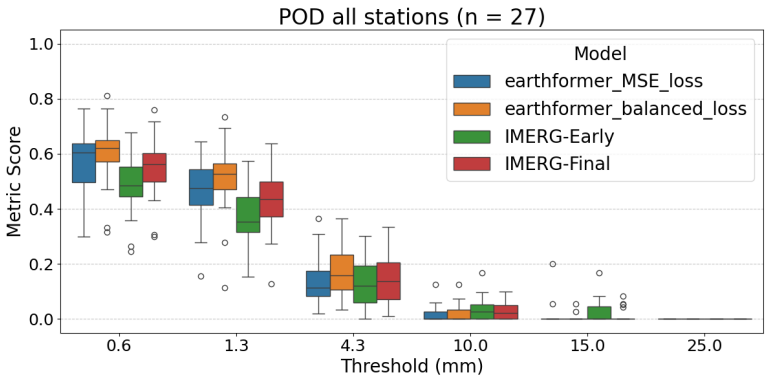
Figure 6.7 presents the performance indicators calculated from comparing the nearest satellite pixel to all available ground stations over 30-minute intervals in 2022, for multiple satellite rainfall products and various thresholds, again using a boxplot to show the performance score among different stations. Figure 6.8 shows the performance indicators at a daily timestep. A separate comparison of satellite rainfall products with GMET and TAHMO stations is available in Appendix E.

From Figures 6.7 and 6.8, it is clear that all satellite rainfall products perform worse at higher rainfall intensities when compared to ground observations.

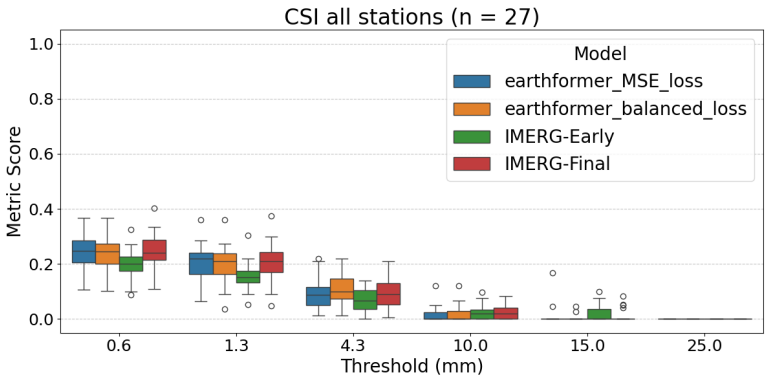
When comparing the Earthformer model with balanced loss to the Earthformer model with MSE loss, the balanced-loss model shows a slightly higher POD across most thresholds. However, it also demonstrates a slightly lower SUCR (Figures 6.7c and 6.8c). Both Earthformer models perform similarly to IMERG-Final, the dataset on which they were trained, and seem to outperform IMERG-Early at the 30-minute interval. However, for a cumulative rainfall threshold of ≥15 mm over 30 minutes, only IMERG-Early shows notable performance, even when compared to IMERG-Final. This suggests that IMERG-Final rarely estimates higher rainfall rates, a bias that the Earthformer models seem to take over.

While the Earthformer model with balanced loss consistently underperforms compared to the MSE-loss version in terms of SUCR when validated against ground stations (Figures 6.7 and 6.8), its overall performance is slightly better for higher thresholds (Figures 6.8b and 6.8b). This is due to its higher POD (Figures 6.7a and 6.8a).

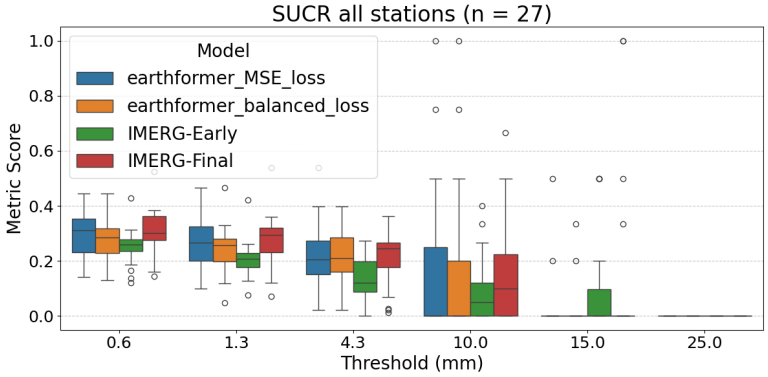
When examining the bias (Figures 6.7d and 6.8d), all satellite products tend to generate more false alarms for lower rainfall thresholds relative to missing rainfall events. However, for higher thresholds, they are more likely to miss rainfall rather than falsely detect it. IMERG-Early overdetects rainfall for lower thresholds, just like IMERG-Final. However, the Earthformer model with balanced loss shows the highest over-detection for lower thresholds of rainfall, suggesting it is overly sensitive to light and moderate rain. In contrast, the MSE model shows the least over-detection at higher thresholds but tends to under-detect heavy rainfall more often than the other satellite products.



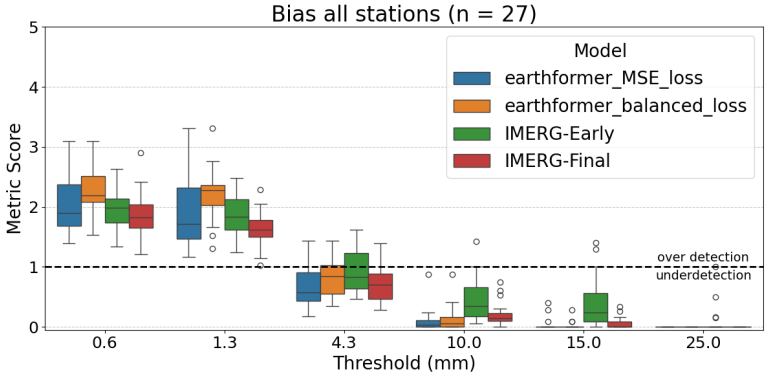
(a) POD 30min interval (perfect score = 1)



(b) CSI 30 min interval (perfect score = 1)

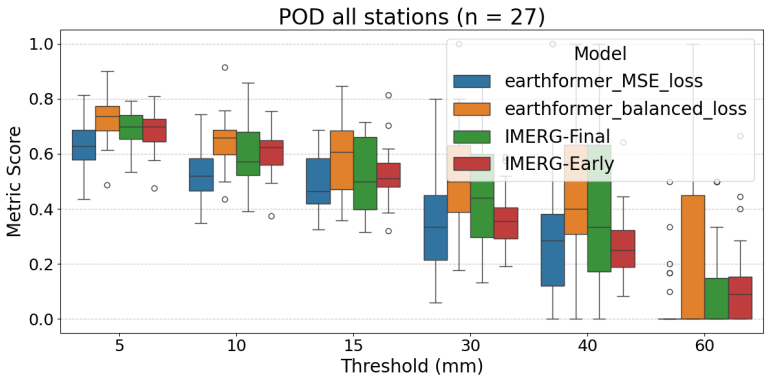


(c) SUCR 30 min interval (perfect score = 1)

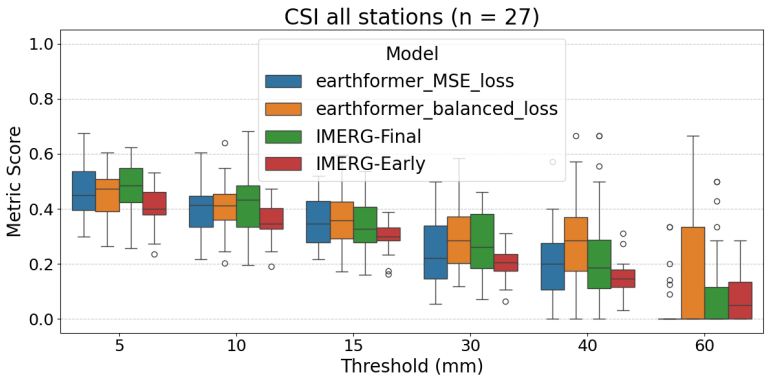


(d) Bias 30 min interval (perfect score = 1)

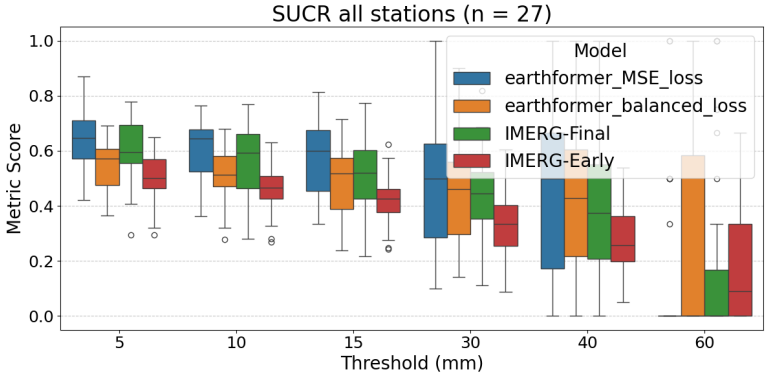
Figure 6.7: Metrics of mm rainfall calculated of the year 2022 over a 30min time interval by comparing gauge station measurements with the nearest pixel estimate of satellite products.



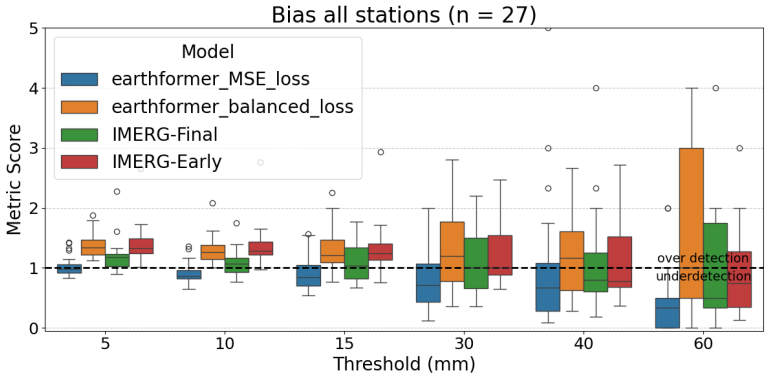
(a) POD daily interval (perfect score = 1)



(b) CSI daily interval (perfect score = 1)



(c) SUCR daily interval (perfect score = 1)

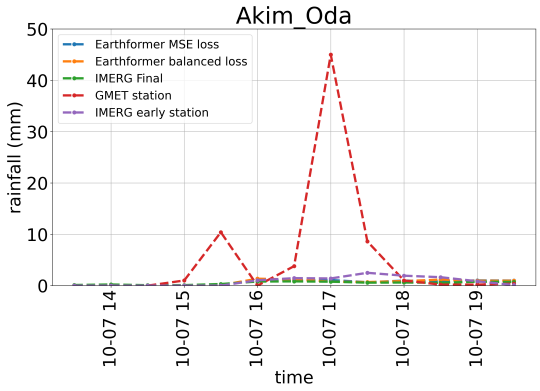


(d) Bias daily interval (perfect score = 1)

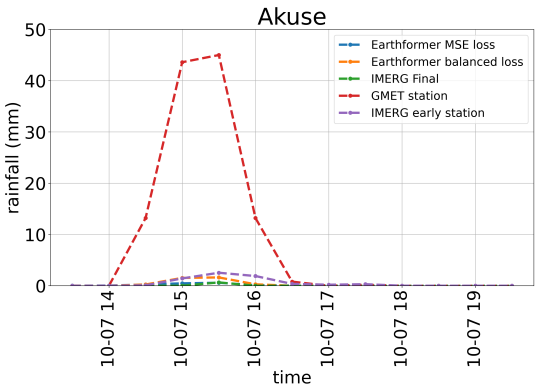
Figure 6.8: Metrics of mm rainfall calculated of the year 2022 over a daily time interval by comparing gauge station measurements with the nearest pixel estimate of satellite products.

6.3.3. Time series analysis of peak events

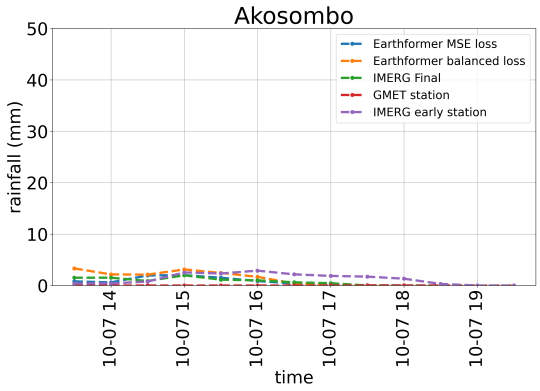
To further investigate the low correlation between station observations and satellite rainfall estimates, several extreme rainfall events are examined over time. Figure 6.9 illustrates a peak event where the Akim Oda and Akuse stations of GMET report more than 40 mm of rainfall within 30 minutes, occurring just 2 hours apart, despite being 123 km apart. Interestingly, none of the satellite rainfall products capture this event at either station. To explore this further, nearby stations are checked to see if they also experienced similar extreme rainfall. However, none of the other stations, even those much closer to one of the affected locations, report such extreme rainfall. This suggests either very localized convective activity in the region or potential errors in the station data, although the last one is less likely the case, when we further investigate the events observed from space.



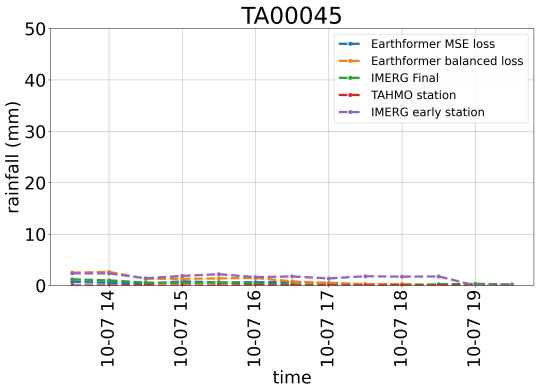
(a) Akim Oda (distance to Akuse +- 122 km)



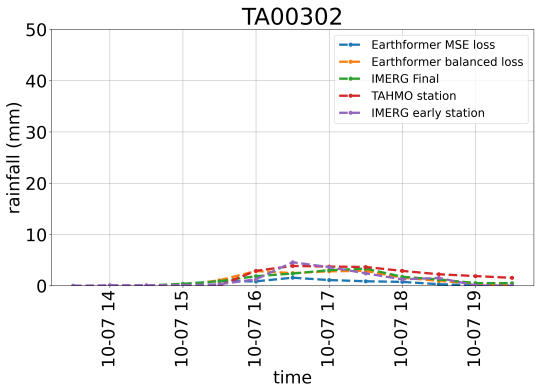
(b) Akuse



(c) Akosombo (distance to Akuse +- 23 km)



(d) TA00045 (distance to Akuse +- 45 km)



(e) TA00302 (distance to Akuse +- 42 km)

Figure 6.9: Time series of the day 2022-10-07 for stations located relatively nearby. Akuse and Akim Oda both show a peak rainfall event of more than 40 mm of rainfall fallen in 30 minutes. However, when compared to other neighboring stations, which are even closer located, these peaks are completely invisible. Satellite rainfall products don't observe the peak as well.

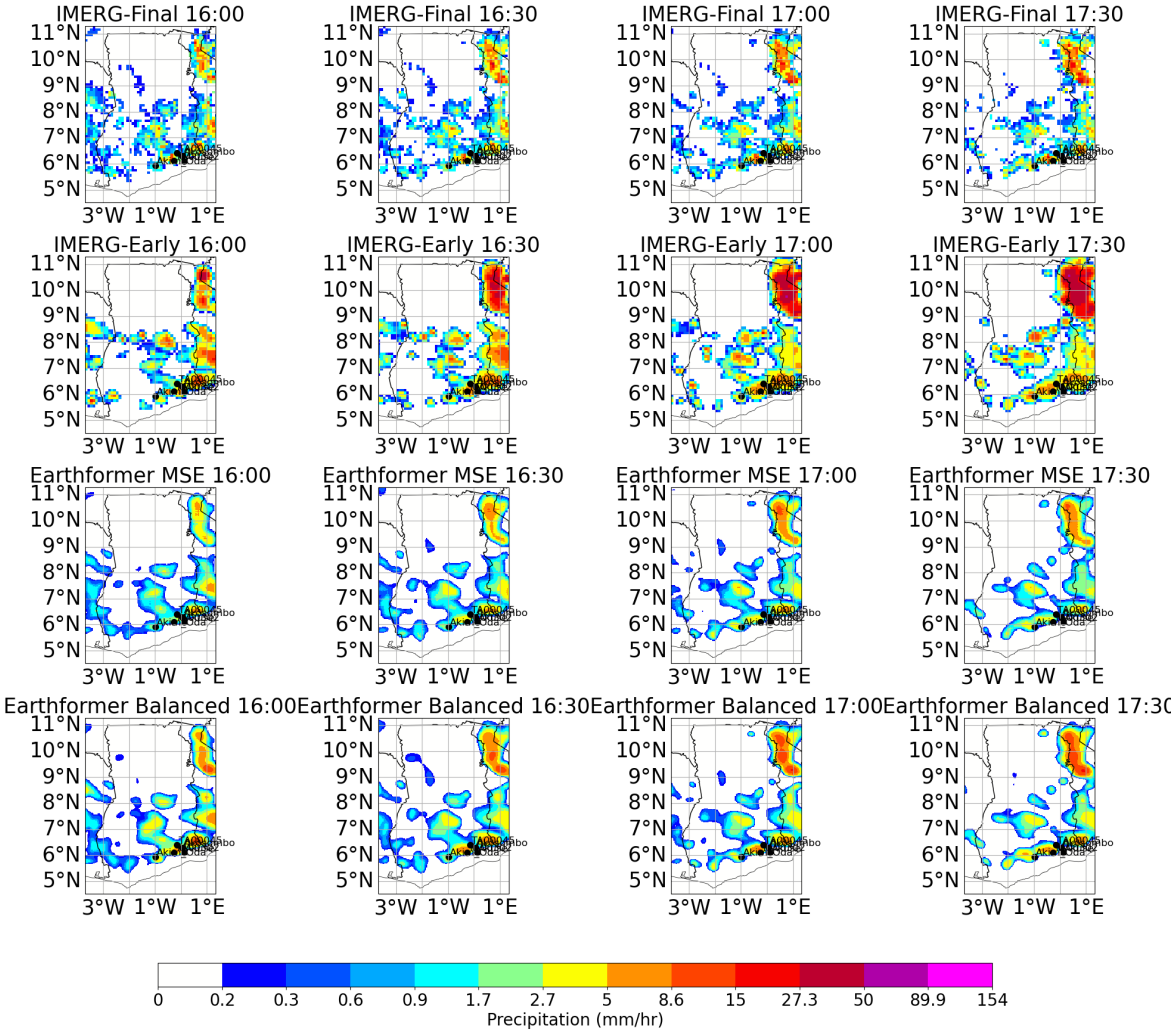


Figure 6.10: 2022-10-07 16.00 - 17.30. Spatial distribution of rainfall observed from space from Figure 6.9.

This is further analyzed by observing the rainfall event from space and examining its spatial distribution (Figure 6.10). Visually, it can be concluded that an organized line of convective cells (a squall line) is moving westward over the group of stations. Figure 6.11 provides a closer view of these stations, revealing that Akim Oda and Akuse are located near different convective cores within the squall line.

IMERG-Final exhibits greater spatial variability within these convective cores, whereas IMERG-Early and the Earthformer models display a more uniform rainfall distribution. The peak intensity of the rainfall is located at slightly different locations by the different satellite estimate products. Neighboring grid cells of the stations show higher rainfall intensities, but none capture the peak intensities observed at the Akuse and Akim Oda stations. Furthermore, none of the satellite products detect high intensity rainfall moving over the Akuse station, suggesting both spatial misplacement of the rainfall distribution when observed from space and an underestimation of peak intensities.

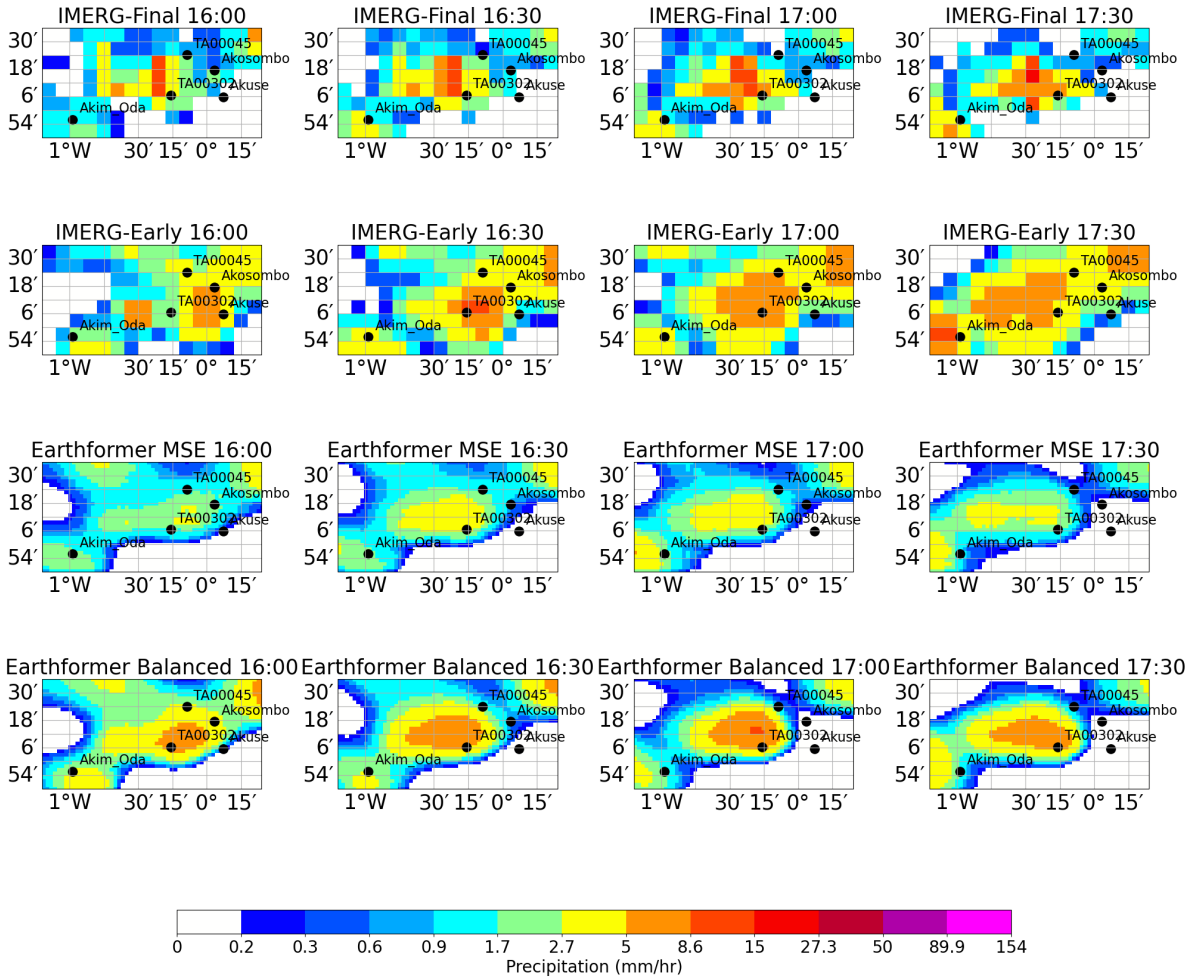
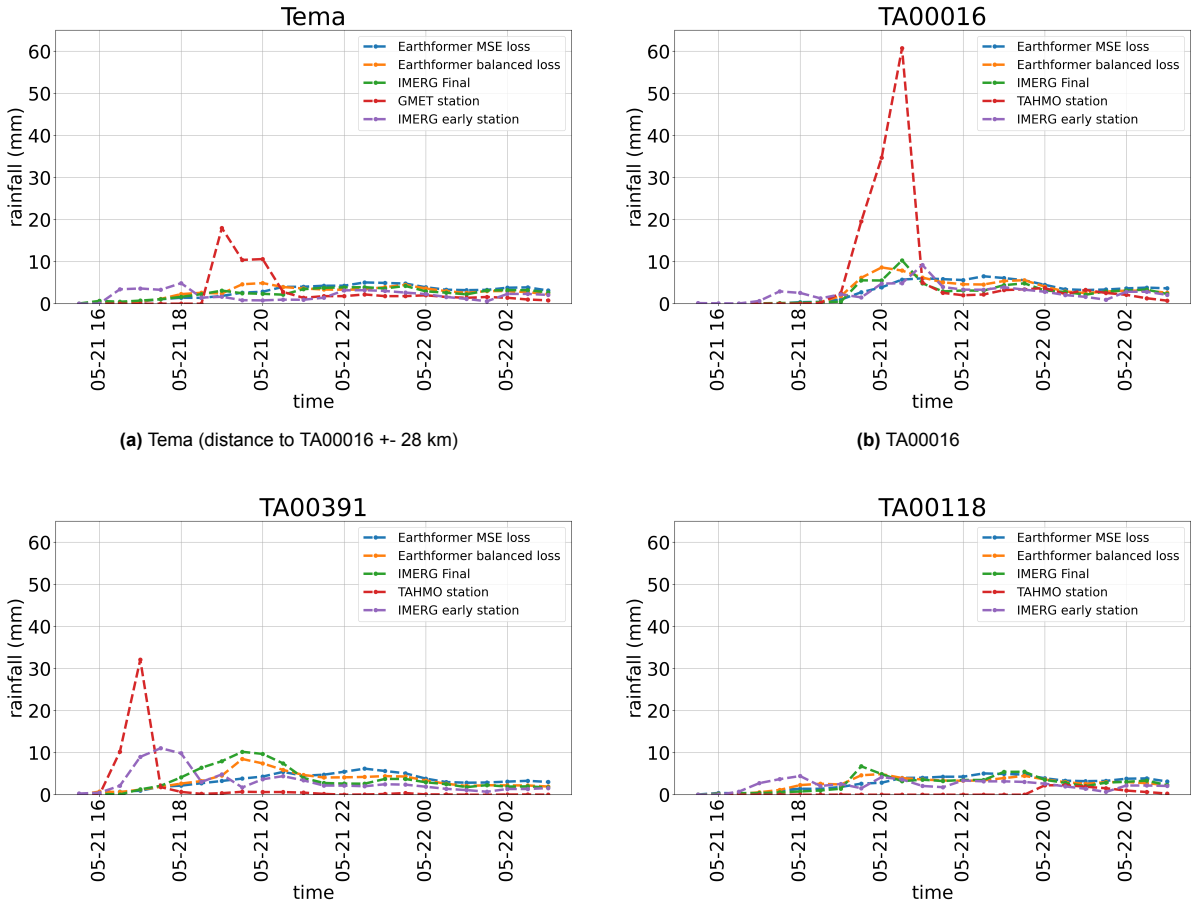


Figure 6.11: 2022-10-07 16.00 - 17.30. Spatial distribution of rainfall observed from space zoomed in on the group of stations from Figure 6.9.

Figure 6.12 presents a group of stations in the Accra region. Three of the four stations report high rainfall accumulation within 30 minutes, exceeding 15 mm, while the TA00016 station records an extreme peak of over 60 mm in just 30 minutes. In contrast, the TA00118 station does not report any peak event. All satellite estimates underestimate the peak intensities, whereas lower intensities are slightly overestimated.



(a) Tema (distance to TA00016 +- 28 km)

(b) TA00016

(c) TA00391 (distance to TA00016 +- 21 km)

(d) TA00118 (distance to TA00016 +- 27 km)

Figure 6.12: Time series of the day 2022-05-21 for stations located relatively nearby in the region of Accra. TA00016 shows a very extreme peak of more than 60 mm rainfall fallen in 30 min at 21.00. Both the Tema station and TA00391 show high intensities of more than 15 and more than 30 mm respectively, however at different timing (19.00 and 17.00). The TA00118 station, although located only 1.8 km from the Tema station shows no such peak.

The rainfall event from Figure 6.12 and its spatial distribution are also analyzed (Figures 6.13 and 6.14). The rainfall is observed moving southwestward over Ghana, forming a MCS (Figure 6.13).

In Figure 6.14, at the peak of the TA00016 station at 20:30, the highest rainfall rates within the area are attributed to this location. However, none of the satellite products capture rainfall of such high intensity as observed by the station itself in this region. Earlier, at 19:00, the peak rainfall recorded at the Tema station suggests that, from a satellite perspective, the station is near the boundary of a rainfall cell. This may explain why TA00118 does not record any rainfall at that time. It could be just outside the active precipitation area. At the same time (19:00), the most extreme rainfall is located near the TA000391 station, yet no rainfall was recorded by the station itself.

These observations suggest multiple possibilities as the cause of the underestimation of rainfall from satellite images: (1) an overall underestimation of rainfall intensities within this convective cell, (2) very localized rainfall, as TA00018 does not record a peak, while the Tema station does and (3) a misplacement of the system's core, when looking from space, although at 20.30 the highest rainfall intensity within the area seems to be attributed to the correct location.

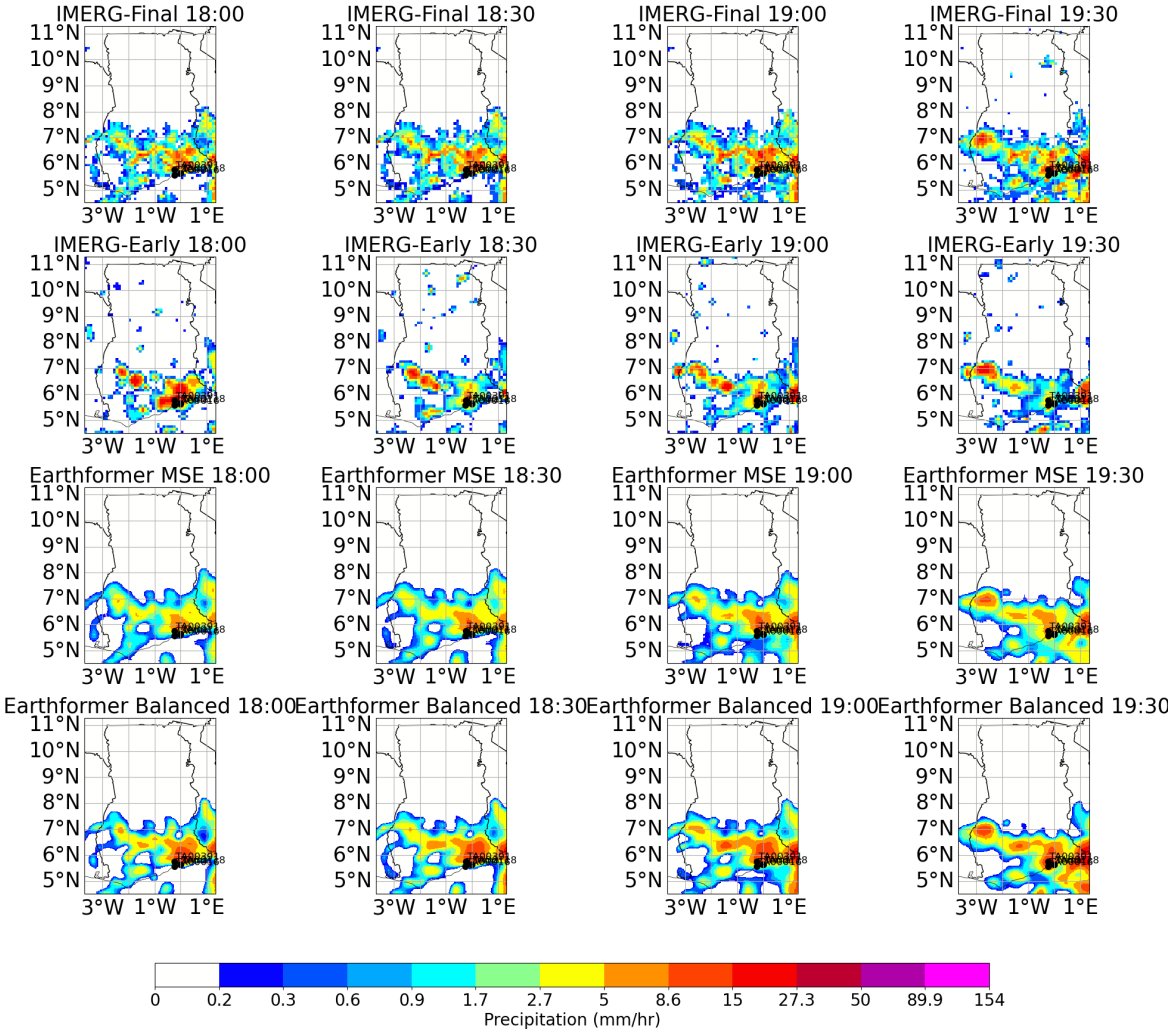


Figure 6.13: 2022-05-21 18.00 - 19.30. Spatial distribution of rainfall observed from space on the group of stations from Figure 6.12.

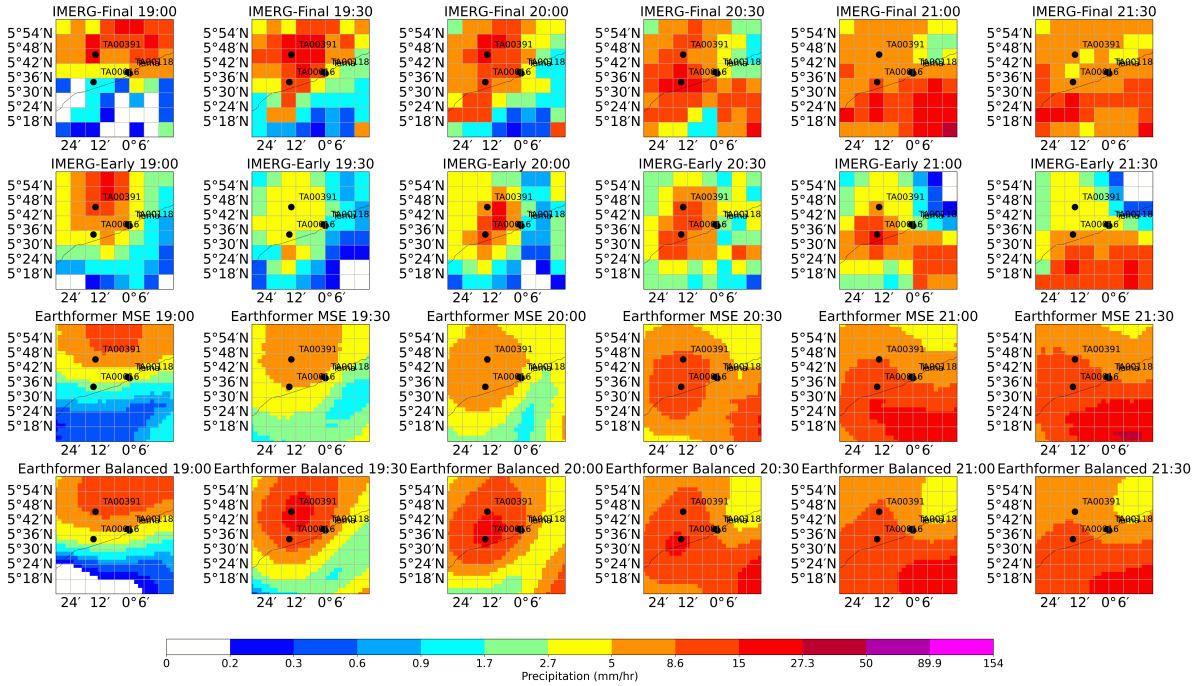
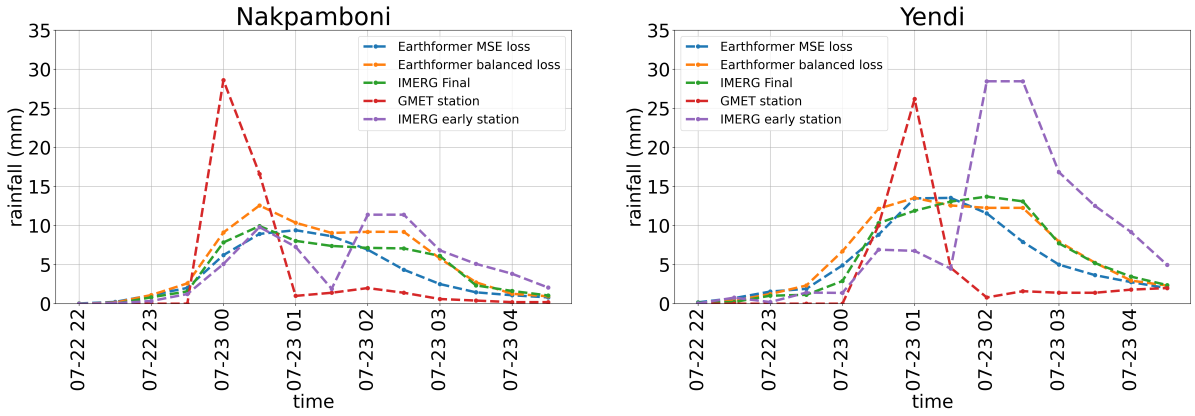


Figure 6.14: 2022-05-21 18.00 - 19.30. Spatial distribution of rainfall observed from space zoomed in on the group of stations from Figure 6.12.

Figure 6.15 displays rainfall data from the Nakpamboni and Yendi stations, located in the Central region of Ghana, at a distance of 23.6 km from each other. Both stations record a peak of approximately 25 mm one hour apart from each other, which is underestimated by most satellite products. IMERG-Early is the only product that correctly captures the peak intensity at Yendi, though it misplaces the timing. However, for Nakpamboni, also IMERG-Early fails to detect the peak intensity.



(a) Nakpamboni(distance to Yendi +- 23 km)

(b) Yendi

Figure 6.15: Time series for 2022-07-22 showing rainfall at two stations in central Ghana, located within 23 km of each other. Both stations recorded a peak rainfall event exceeding 25 mm one hour apart from each other. All satellite estimates fail to capture the peak, smoothing it out over time, except for IMERG-Early, which detects it but with a timing offset.

Figures 6.16 and 6.17 show the spatial distribution of rainfall during the event recorded by the Nakpamboni and Yendi stations in Figure 6.15. Once again, a large mesoscale convective system (MCS) is observed moving southwestward over the stations.

The first peak is recorded at 00:00 by the Nakpamboni station, which is located northeast of the Yendi station. During this peak, the satellite products do not estimate such high rainfall intensities. However,

at 01:00, when the Yendi station records its peak rainfall, the satellite products show higher intensities, but still not as high as the station's observed peak.

One hour later, IMERG-Early detects very extreme rainfall matching the peak recorded earlier by the Yendi station. However, at this time, the Yendi station itself does not observe any extreme rainfall. The satellite-based estimates indicate a widespread area of intense rainfall across different time steps. However, ground stations suggest a much more localized rainfall pattern.

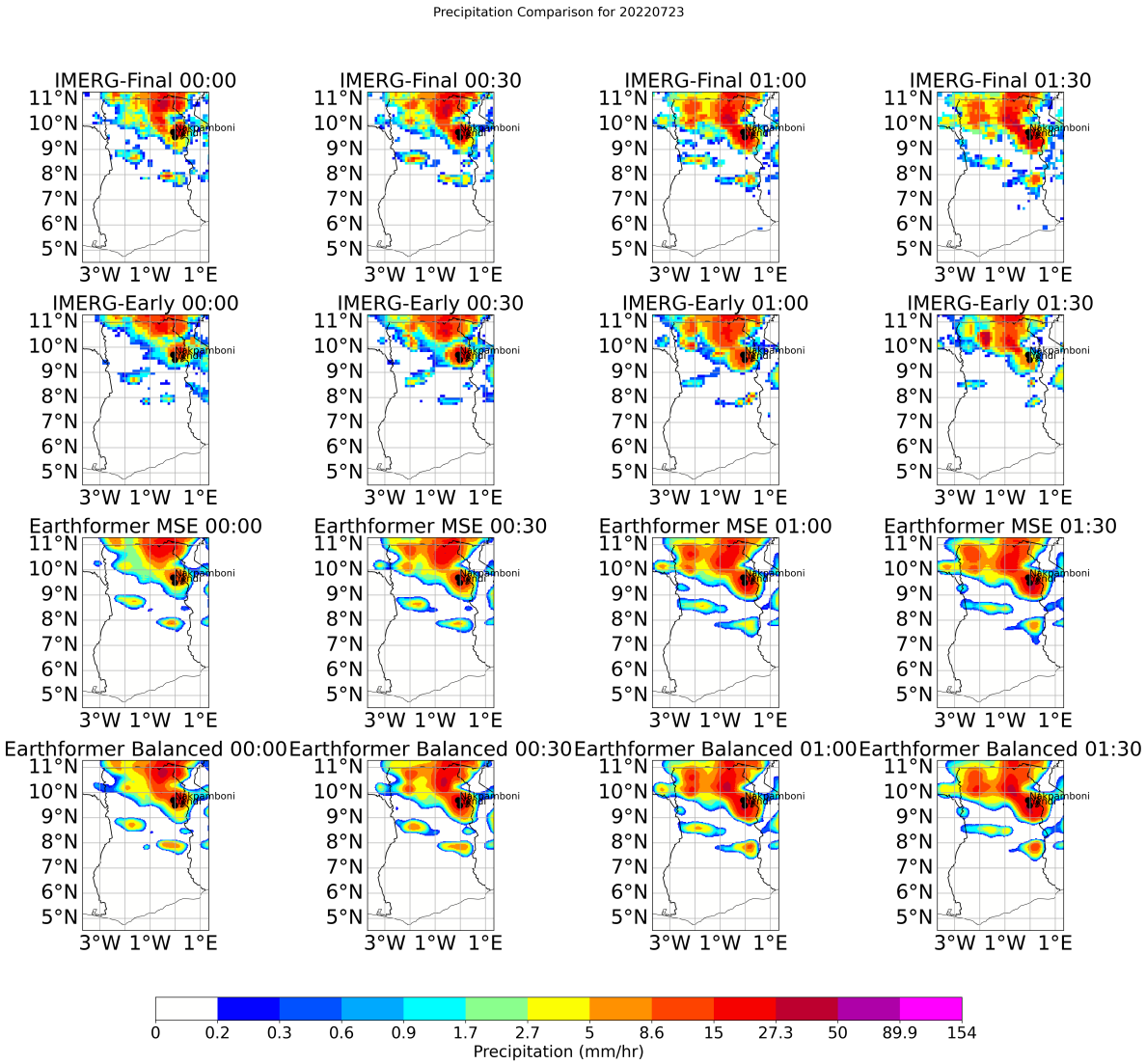


Figure 6.16: 2022-07-23 00.00 - 01.30. Spatial distribution of rainfall observed from space on the group of stations from Figure 6.15.

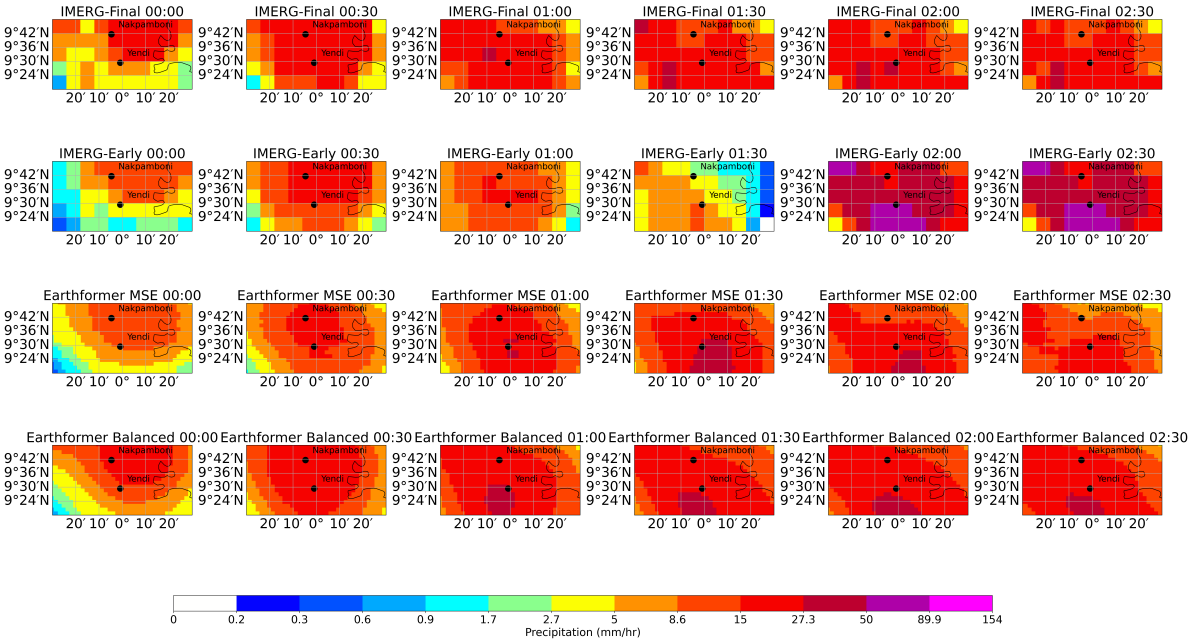
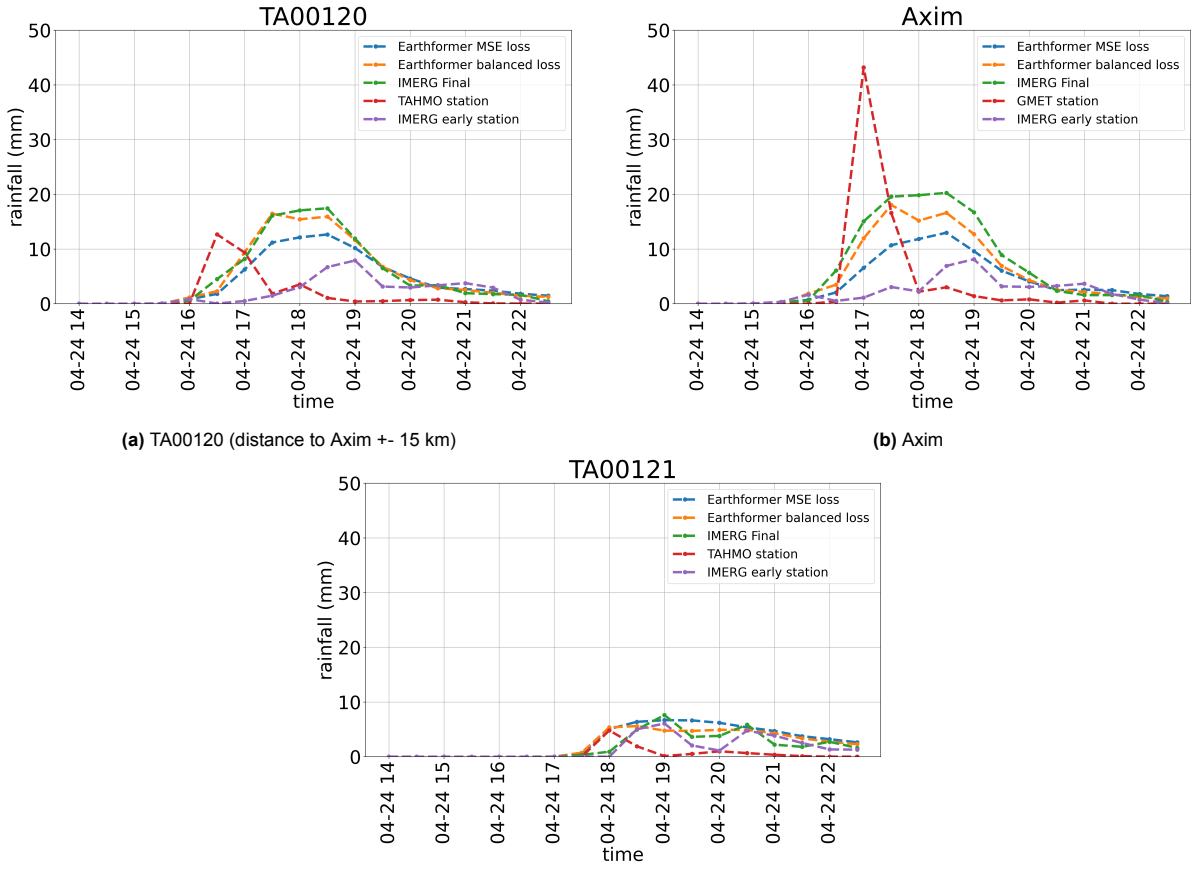


Figure 6.17: 2022-07-23 00.00 - 01.30. Spatial distribution of rainfall observed from space zoomed in on the group of stations from Figure 6.15.

The Axim station, located in Ghana's coastal region, records an extreme peak of over 40 mm of rainfall in 30 minutes (Figure 6.18). Interestingly, neighboring stations do not observe this peak event. While satellite-based products detect significant rainfall in the area, they miss the extreme peak, instead they spread it out over multiple time steps. However, for the other stations, the satellite estimates also report lower rainfall amounts, aligning with the station observations.



(a) TA00120 (distance to Axim +- 15 km)

(b) Axim

(c) TA00121 (distance to Axim +- 74 km)

Figure 6.18: Time series of 2022-04-24 located in the coastal region of Ghana. The Axim station observed a peak event of more than 40 mm, while the station relatively nearby only observed a peak less than 20 mm. IMERG-Final and the Earthformer with balanced loss observe rainfall of approximately 20 mm, but smoothens the peak out in time.

When observed from space, once again a MCS is visible moving over the station group in Figure 6.18 (Figure 6.19) at the corresponding time steps.

At 17:00, the peak rainfall recorded by the Axim station (Figure 6.18b) corresponds to a peak observed in one of the nearby grid cells of IMERG-Final in Figure 6.20, though not in the nearest grid cell. All other satellite-based estimation products underestimate this peak rainfall.

As time progresses, the satellite rainfall products continue to detect intense rainfall, whereas the Axim station does not. At 17:00, the satellite-based rainfall estimates align with the intensity recorded at the TA00120 station, but at 16:30, all products fail to capture its peak. While all satellite products correctly detect the absence of rainfall at TA00120 initially, they overestimate rainfall intensities as time progresses. This suggests that satellite-based estimates depict a larger raining area than what is observed on the ground. Additionally, satellite images show a relatively large region of high rainfall intensities across different time steps, whereas station data indicates a more localized pattern.

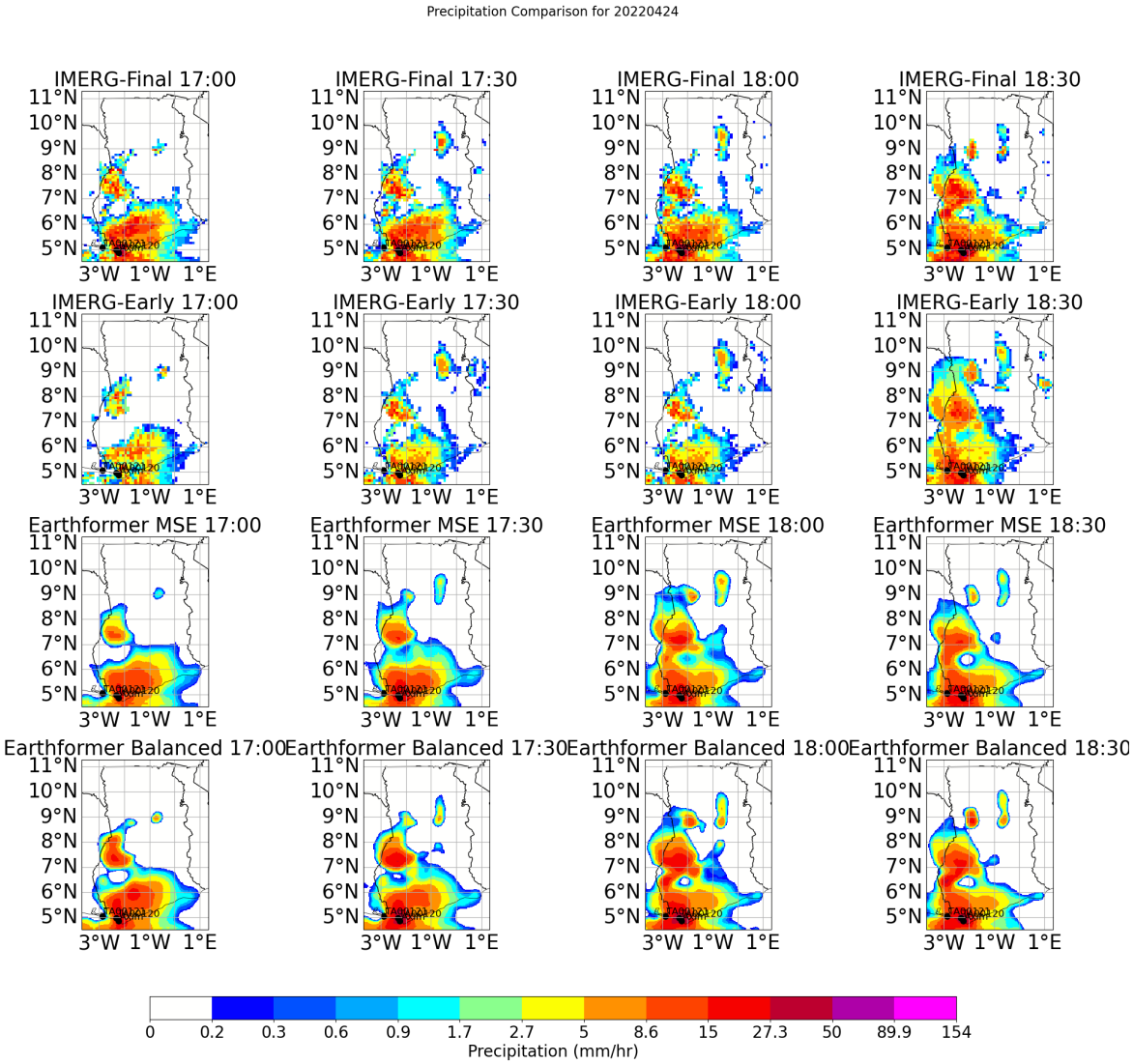


Figure 6.19: 2022-04-24 17.00 - 18.30. Spatial distribution of rainfall observed from space on the group of stations from Figure 6.18.

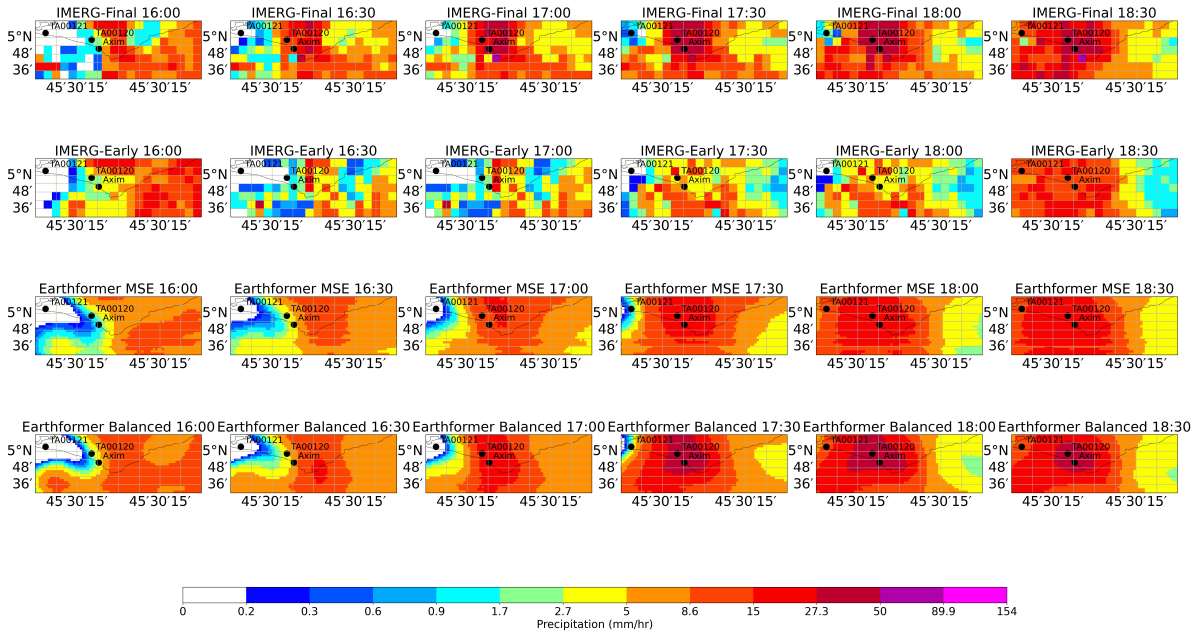


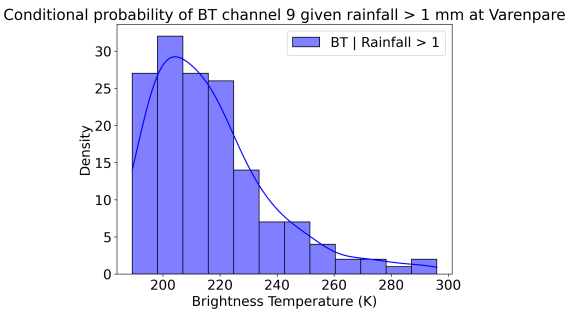
Figure 6.20: 2022-05-21 16.00 - 18.30. Spatial distribution of rainfall observed from space zoomed in on the group of stations from Figure 6.18.

6.3.4. Rainfall process

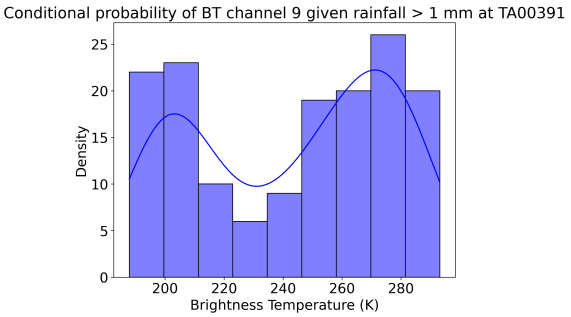
As discussed in Section 2.3, different rainfall processes exist across regions. To understand which rainfall mechanisms play an important role in each region, the conditional probability distribution of the brightness temperature of channel 9, given rainfall > 1 mm within 30 minutes was analyzed for the year 2022.

Figure 2.5 provides an example of stations within the Savannah climatic zone, demonstrating that rainfall formation in this region is mostly dominated by cold rain processes. This is shown in Figure 6.21a, where rainfall mostly occurs when cloud tops are colder than 220 K, a pattern consistent with other stations in the Savannah zone (see Appendix F).

In contrast, stations in the forest and coastal climatic zones show a mix of cold and warm rain processes. Figure 6.21b illustrates an example from these regions. Similar to most other stations (see Appendix F) in the forest and coastal zones, the peak of the brightness temperature around 280 K (Figure 6.21b) suggests the presence of warm rainfall formation processes.



(a) The Varenpare station, located in the Savannah climatic zone.



(b) The TA00391 station, located in the coastal climatic zone.

Figure 6.21: Conditional probability distribution of the brightness temperature of channel 9 given rainfall > 1 mm fallen within 30 minutes

7

Discussion and recommendations

7.1. Discussion

7.1.1. Limitations ground station comparison

This study compares ground station data to satellite products using only the year 2022. Ground station performance varied over time, due to either clogged or malfunctioning stations reporting mostly zeros. This made long-term comparisons challenging and a full quality check across multiple years very time-consuming.

Additionally, approximately 8% of the 2022 comparison dataset contained frames from the training dataset when compared to ground stations. This was done to assess model outputs over time and to aggregate results into daily intervals. However, this overlap may lead to a misinterpretation of results and a potential overestimation of model performance. However, Appendix D shows that the median Pearson correlation of the test dataset for the 30-minute time interval shows similar results to that of the entire test dataset comparison but with a larger spread. This may be attributed to either the pre-selection of data, which emphasises more extreme rainfall events, or the fact that station data was only quality-controlled for 2022, whereas the test dataset includes multiple years. Additionally, the smaller number of events used to calculate the correlation could also contribute to the increased spread.

7.1.2. Quality of ground observations

A visual inspection was conducted by comparing cumulative monthly rainfall to the annual rainfall pattern in Figure 2.4. Stations that showed clear misalignment with regional seasonal rainfall patterns were removed from the dataset to prevent errors from malfunctioning or obstructed stations that predominantly recorded zero rainfall. While this raises some concerns about the quality of the remaining data, it is reasonable to assume that if a station records rainfall, it is functioning correctly, as few plausible mechanisms would cause a station to report rainfall in the absence of actual rainfall. Consequently, potential inconsistencies in the station data can largely be attributed to rainfall underestimation. Thus, the observed underestimation of peak rainfall intensities in satellite-based products is unlikely to result from errors in the ground station data.

The qualitative comparisons between nearby stations indicate highly variable and localized weather patterns across Ghana. For example in Figure 6.12 it can be seen how the Tema station observes rainfall of almost 20 mm within 30 min, while the TA00118 station does record no such peak, although the stations are located within 1.8 km distance from each other. This may suggest that the TA00118 station was clogged or obstructed during this event. However, earlier that month, TA00118 recorded a peak intensity of approximately 20 mm within 30 minutes similar to the Tema station, suggesting the station was not obstructed or clogged at all, and thus highly variable rainfall patterns play a role. The visual inspection of the satellite-based estimates revealed mainly MCS were the cause of peak rainfall. Potentially both the misallocation of peak rainfall and underestimation of rainfall intensity play an important role of the discrepancies between satellite based estimates and ground station observations within such convective systems.

7.1.3. Comparison of satellite pixels and gauge station observations

Comparing pixels to points has some difficulties for multiple reasons. Le Coz et al. (2021) explains how part of these discrepancies can be explained with positional and timing issues. First of all, the Meteosat satellites are looking from space towards the top of the clouds above the Earth. The distance of the cloud to the Earth itself is usually unknown and, since this can be quite high, especially in Ghana, a lot could happen before rainfall formed within the clouds reaches the ground. For instance, wind could take the precipitation a few kilometers elsewhere, resulting in a mismatch of precipitation located from space and the actual precipitation location on the ground . When comparing the nearest pixel to a gauge station it is possible the gauge station measures precipitation, while looking from space at that specific location there is no precipitation (Figure 7.1).

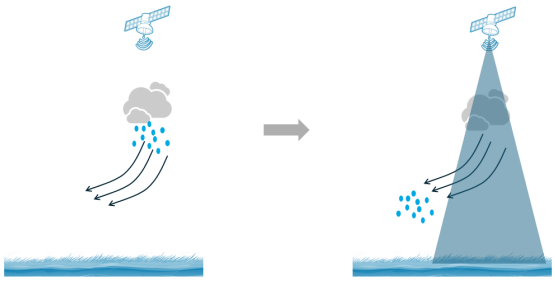


Figure 7.1: The Figure shows an example of causing discrepancy between observations from space an ground observations. The rainfall observed from space measures rainfall at the location of the cloud. Due to other factors such as wind the correctly observed rainfall from space ends up at a different location on the ground, thus causing a discrepancy between the ground observations and observations from space.

Besides, with convective rainfall events the variability of the rainfall can be very high and localized. For IMERG-Final, when comparing (10 x 10 km) grid cells that take an average rainfall over a specific area over 30 min, to an actual point is difficult. The point measurement might capture a very localized intense rainfall event, while averaged out over the grid cell this effect is dampened (Figure 7.2). Additionally, the rainfall estimations of IMERG-Final are not only averaged over space, but also over time, contributing to the smoothing effect of extreme rainfall.

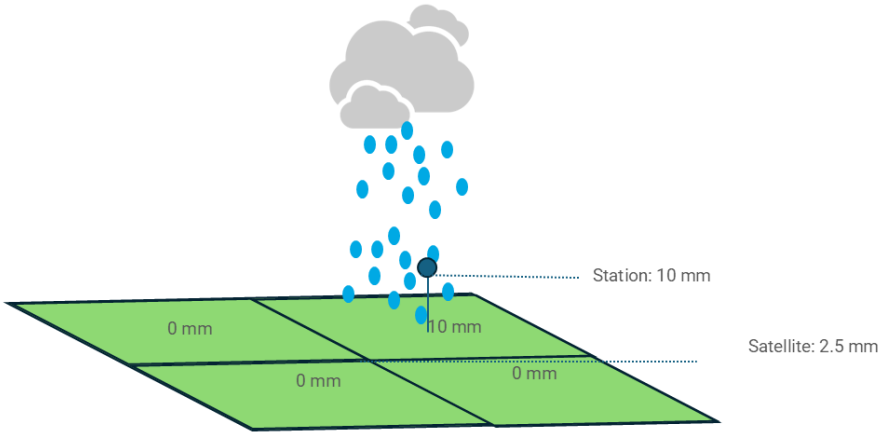


Figure 7.2: This figure illustrates how both a satellite-based rainfall product and a ground-based station can accurately capture rainfall observations, but still show discrepancies between them. The satellite product provides a correct average rainfall estimate for a given pixel, while the station accurately records a localized peak event. Although both observations are correct, the discrepancy is caused by the difference in the spatial resolution of the satellite data against the localized observation of the station measurement.

7.1.4. Influence of reference and input data

For this research, IMERG-Final was selected as the reference dataset. Although resampled to the resolution of SEVIRI images, the model wasn't able to produce model outputs at higher resolution compared to IMERG-Final, thus effectively reducing its resolution to 0.1° x 0.1°. Climate Prediction Center Morphing dataset (CMORPH) was not considered; however, in retrospect, it would have been a good alternative, as it also includes bias adjustment based on the GPCC dataset (NOAA, 2018). Additionally, it has the same temporal resolution of 30 minutes but a higher spatial resolution of 8 × 8 km (Joyce et al., 2004). Besides, according to Pradhan et al. (2022), CMORPH has similar performance to IMERG over Africa.

From comparison to the ground stations it is known that IMERG-Final struggles to capture the extreme rainfall rates and overestimates lower rainfall rates. This is in line with the research of Li et al. (2024) and Bogerd et al. (2024). This issue is common in rainfall products derived from infrared imagery, as estimates are based on observations viewed from space explained earlier.

Interestingly, IMERG-Early appears more capable of capturing higher rainfall rates but often struggles with accurate timing. As explained in 3.2.1, IMERG-Final incorporates a backward pass in its data assimilation process to enhance temporal consistency and alignment with climatological products (Huffman et al., 2023). However, a potential side effect of this approach is the smoothing of peak rainfall events over time. This effect is evident in the analysis and explains why IMERG-Final rarely estimates higher rainfall rates. In contrast, IMERG-Early is less affected by this smoothing but more frequently exhibits timing errors.

Training a model on IMERG-Final risks propagating its biases and errors. Li et al. (2024) highlights how these biases vary throughout precipitation events and suggests incorporating event-stage information and environmental variables in future algorithms.

Le Coz et al. (2021) discusses positional and timing errors and how to correct them. However, the proposed method is computationally expensive and impractical for operational nowcasting. But, when providing the Earthformer model with ground-based data, instead of a satellite base product, it might

be able to learn those patterns autonomously, leading to more computationally efficient precipitation estimation.

7.1.5. Influence of data splitting method

A potential limitation of this study is the assumption that individual frames were independent during data splitting. For long-lived systems such as squall lines and MCS, some temporal correlations may exist between frames, which could introduce dependencies in the training, validation and test set. However, the model was also evaluated on the entire year of 2022 by comparing it to the ground stations, where most of the data had not been used in training, validation, or testing. The fact that the model demonstrated similar performance to IMERG-Final on this dataset suggests that it captures meaningful rainfall patterns and still holds potential for real-world applications.

7.1.6. Correcting biases with ground stations

A closer examination of the discrepancies between ground observations and satellite rainfall estimates reveals that false alarms in satellite data mainly come from timing errors and a smoothing effect over time. This smoothing not only spreads out rainfall events but also leads to an underestimation of peak intensities. To address these biases, two correction approaches were explored.

The first approach involved adjusting satellite estimates using an accumulation factor. This factor was calculated by dividing the cumulative rainfall over a specific time interval at a station by the cumulative rainfall of the nearest satellite grid cell, using a rolling window. The factor was then linearly interpolated between stations and applied to scale the satellite rainfall estimates at each timestep. However, this method had one disadvantage; when the model missed a peak event, subsequent timesteps were significantly overcorrected, even though light precipitation was already known to be overestimated. While applying thresholds could help mitigate this issue, this was not further explored in this study.

The second approach investigated the use of quantile mapping, focusing on scaling extreme values. A rolling window was used to obtain the cumulative distribution function (CDF) for each grid cell over time, and each present grid cell value was mapped to the corresponding quantile of the nearest station. However, this method also presented challenges, particularly due to the seasonal variability in Ghana's rainfall patterns. Differences between rainy and dry months made it difficult to ensure that past CDFs accurately represented current conditions. A possible solution would be to correct each month using historical data from previous years, though this approach is complicated by variations between wetter and drier years as explained in Section 2.3.1.

Literature also showed other more advanced options. Moazami et al. (2022) employed an advanced spatiotemporal bias adjustment for IMERG over Canada by using quantile mapping derived from gauged locations. They interpolated the bias adjustment parameter on ungauged grids based on climatic similarities and the spatial distribution of rain gauges, increasing the correlation with gauge stations from 0.3 to 0.9. Another approach found in the literature involves using classifiers to categorize rainfall into intensity classes and then applying a regression model (a random forest) to predict the rainfall amount for each intensity class (Ziarh et al., 2021). This method is particularly interesting because it scales rainfall of varying intensities differently. However, a bias correction is mostly useful for systematic errors and if timing and location errors play an important role, this might not be as useful.

7.1.7. Optimizing hyperparameters and model architecture

Due to the relatively long training time of the model (approximately two days), limited time was spent optimizing its hyper parameters. The model was trained, using a base unit of 128 values for the attention mechanism. Increasing this value might help the model to capture more complex relations, which is especially important when using additional input data. Additionally, increasing the depth of the model by increasing the number of hierarchies might helps the model to look for multiple relations at different spatial or temporal levels, which is relevant for precipitation estimations. This was not tested in this research due to the limiting available memory of the GPU, but is advised to check whether it could improve performance. However, both with increasing depth or base unit values the model becomes more computationally expensive.

8

Conclusion and recommendations

8.1. Conclusion

This research explored how the Earthformer transformer model can improve rainfall rate estimation using SEVIRI channel data from the Meteosat Second Generation satellite for near real-time applications in Ghana. The main research question is addressed as follows:

"How can the Earthformer transformer model be applied to improve the accuracy of rainfall rate estimates across different intensities for near real-time applications using SEVIRI channel data from the Meteosat Second Generation satellite in Ghana?"

This study demonstrated how the Earthformer transformer model can be applied by training the model with the Meteosat SEVIRI channel data as input and IMERG-Final as the target data. Its findings suggest that the Earthformer model, using Meteosat SEVIRI satellite images as input, is a promising alternative to IMERG-Early due to its faster near real-time availability (within 15 minutes) and its improved performance across most rainfall intensities when evaluated against IMERG-Final as the reference truth. Although the Earthformer model performs relatively well for the lower intensities, it still struggles to capture heavy and extreme rainfall events accurately. This limitation reduces the models' suitability for FEWS, where accurately localizing extreme events is essential. While precise intensity estimates are not a strict requirement for FEWS, the models must still effectively identify the occurrence of extreme events. This is something the Earthformer models currently fail to achieve.

8.1.1. Accuracy of rainfall estimation for varying intensities

The first subquestion is adressed in this section.

"How accurately does the Earthformer transformer model retrieve rainfall rate estimates for different rainfall rates from the SEVIRI channels of the Meteosat Second Generation satellite in Ghana?"

The accuracy of the model is evaluated by comparing it both to IMERG-Final and ground station observations.

- When compared to IMERG-Final on the test dataset, the Earthformer models perform well at lower rainfall intensities but show decreasing performance at higher intensities, similar to IMERG-Early. However, both Earthformer models outperform IMERG-Early at lower intensities, and the model with balanced loss also performs better at higher intensities. The balanced loss function improves the detection of heavy rainfall rates but also increases the overestimation of light and moderate rainfall.
- Comparing satellite-based estimates with ground station observations for the year 2022 reveals weak correlations when zeros are excluded from the analysis (median ≈ 0.2 and values range from 0 to 0.4) at 30-minute intervals for all satellite estimates. The weak correlation may be caused by multiple factors: (1) timing errors in satellite observations (2) the spatial resolution of satellite estimates not aligning with localized rainfall variations or (3) inconsistencies in ground

station measurment, although the thrid one seems unlikely as explained in the discussion. Daily aggregation improves the correlations (median ≈ 0.6 and values range from 0.3 to 0.9), suggesting that timing discrepancies play a role, but further investigation is needed to quantify their impact. Performance indicators show that all satellite-based estimates, including Earthformer, perform poorly when compared to gauge station data at 30-minute intervals, particularly for extreme rainfall events. This poor performance could partially be explained by the limitations of satellite resolution in both time and space, as well as by the very localized and high variability of rainfall. An analysis of several extreme peak events supports this conclusion, showing that satellite estimates consistently underestimate peak rainfall while overestimating rainfall at the temporal boundaries of events. Especially for IMERG-Final and the Earthformer models, which inherit its biases. The time series analysis further shows that IMERG-Early occasionally captures extreme rainfall intensities accurately but misaligned the timing of this event. Additionally, IMERG-Early is the only satellite product demonstrating any (although limited) performance for higher rainfall intensities (≥ 15 mm within 30 minutes). This finding suggests that the backward pass used in the IMERG-Final algorithm, intended to improve temporal consistency with other data sources (e.g., climatological records), may reduce extreme values, while improving timing accuracy. By training on IMERG-Final, the Earthformer models inevitably adopt these biases.

Overall, while both Earthformer models demonstrate improved performance at lower rainfall intensities compared to IMERG-Early, their accuracy declines for higher intensities, partly due to biases inherited from IMERG-Final. However, the Earthfomer with balanced loss also outperforms IMERG-Early for the higher rainfall intensities when compared to IMERG-Final. The weak correlation with ground station observations of all satellite based rainfall products at short timescales highlights challenges in capturing highly variable and localized rainfall events, though daily aggregation improves correlations. These findings underscore the need for further refinement in satellite-based rainfall retrieval methods, particularly for extreme events.

8.1.2. Adapting the model for extreme rainfall estimation

The second sub-question is addressed in this section.

"What modifications can be made to the Earthformer transformer model to improve the accuracy of high-intensity rainfall estimates?"

This study explored whether modifying the model's loss function to emphasize extreme values (balanced loss) would improve performance for high-intensity rainfall. The model with balanced loss was compared to the model with MSE loss. While the balanced loss function led to improvements at higher intensities, the Earthformer model with balanced loss still underestimated peak intensities and also resulted in decreased accuracy for the lower intensities. Additionally, as discussed earlier, both Earthformer models inherit biases from IMERG-Final, which affects its peak intensity estimates.

In conclusion, using a balanced loss function in the Earthformer model improves estimates of higher rainfall intensities but comes at the cost of light to moderate-intensity accuracy. Furthermore, training on IMERG-Final causes Earthformer to inherit its biases, leading to missed extreme peak intensities.

The recommendation section discusses other possible adaptations of the model to enhance its performance for extreme rainfall in future research.

8.1.3. Relation to ground stations

This section discusses the third subquestion.

"How do satellite based rainfall rate estimates relate to ground station observations?"

The accuracy assessment of the model already covered key relationships to ground stations, including relevant metrics and statistical correlations. A qualitative analysis of various peak events revealed the underestimation of extreme rainfall and a smoothing of the peak intensities in time. Notably, all investigated peak events in this study were associated with MCSs, suggesting that peak rainfall events in Ghana are often linked to MCSs. However, this observation is anecdotal and requires further investigation to establish statistical significance.

Visual inspection from the satellite estimates for the peak events revealed how the spatial variability of

rainfall intensities within MCS systems is difficult to capture from satellite observations and showed a more uniform spatial rainfall pattern than what the ground station observations indicate, leading to an underestimation of peak rainfall and an overestimation of rainfall at the edges of rainfall cells. A few potential reasons that might explain the difficulties in accurately capturing the rainfall intensities within MCS are; (1) The anvil cloud top obstructs the vision of the satellite (2) Deep convective systems reaching the tropopause their cloud top temperature is similar for various intensities (3) MCS, typically associated with strong wind shear increases the risk of misallocation of rainfall.

Mapping the spatial correlation of rainfall stations revealed that stations within the Savannah climatic zone in northern Ghana generally show higher correlations with satellite-based rainfall estimates compared to those in the coastal and forest climatic zones in central and southern Ghana. Further analysis of rainfall processes using conditional probabilities of brightness temperatures associated with rainfall highlighted a key distinction between these regions. Stations within the Savannah climatic zone are predominantly influenced by cold rain processes, indicative of deep convective rainfall. In contrast, the coastal and forest climatic zones show a stronger presence of warm rain processes, as shown by the conditional probability distributions. This difference might explain the greater discrepancies between satellite-based rainfall estimates and ground station observations in these zones. Since warm rain processes are not associated with cold cloud tops, they are more challenging to detect from space, leading to potential underestimations by satellite products.

Additionally, the discussion already outlined two key explanations that influence the discrepancy between satellite-based rainfall product and ground observations. The first one is the resolution of the satellite-based rainfall estimates, compared to a single-point measurement of a gauge station. Due to spatial averaging of the satellite-based estimates, both station observations and satellite-based estimates can correctly observe rainfall, but discrepancies remain. Secondly, the rainfall can be displaced from the clouds observed from space before it reaches the ground.

Overall, satellite-based rainfall estimates show varying degrees of agreement with ground station observations, with stronger correlations in the Savannah climatic zone and greater discrepancies in the coastal and forest zones. These differences are likely due to the presence of warm rain processes in the latter regions, which are harder to detect from space. Additionally, the inability to fully capture spatial rainfall variability within MCSs contributes to the underestimation of peak intensities and overestimation at event edges of peak events. While satellite estimates provide valuable insights into rainfall distribution, their limitations highlight the need for improved retrieval methods, particularly for extreme events and warm rain processes.

8.2. Recommendations

This section provides recommendations for future research to improve model performance, particularly for extreme rainfall events.

Firstly, the stratified splitting method used in this research might introduce temporal dependencies between the training, validation, and test sets, potentially affecting the generalization of results. To address this, future research could explore alternative data-splitting strategies, such as separating these sets by different years. This approach could enhance generalization across independent weather events.

Additionally, incorporating more environmental variables into the model could improve its performance, as suggested by Li et al. (2024). Potential variables include temperature, lightning imagery, and Convective Available Potential Energy (CAPE). Küçük, Giannakos, et al. (2024) demonstrated the importance of lightning images, particularly for the localization of extreme rainfall. Since MCSs contribute significantly to extreme rainfall in West Africa (Klein et al., 2018) and are often associated with high lightning activity, the availability of lightning data from the Meteosat Third Generation satellites and TAHMO station sensors presents a valuable opportunity.

Surface temperature data could help distinguish stratiform from convective rainfall, while CAPE provides insights into atmospheric instability and the potential for large convective systems to form. Including wind field data could also help identify how rainfall is displaced from clouds before reaching the ground.

To gain a deeper understanding of the model's interpretation of input data, future research should analyze the relative importance of each data channel and examine the correlation between cloud-top temperatures, rainfall rate estimates, and ground station observations more closely.

Further research is also recommended to identify which rainfall mechanisms most frequently lead to flooding, guiding future models to capture the most relevant physical interactions. As Peinó et al. (2025) highlighted, estimating warm rainfall using VIS/IR imagers is challenging because warm rain is not typically associated with cold cloud-top temperatures. Wind shear, often linked with MCSs, may cause misplacement of rainfall in satellite observations. Additionally, anvil cloud tops complicate the relationship between cloud-top temperatures and rainfall rates. Locating the convective core within an MCS remains a challenge and should be a priority for future studies, as qualitative analysis indicates that extreme rainfall is predominantly associated with these systems. For accurate FEWS, the exact intensities may not be as important. It could be more valuable for future research to focus on identifying the core of MCSs, rather than striving to precisely determine the exact rainfall intensities.

Future work should also consider training the Earthformer model on GPUs with higher memory capacity. This could reduce training times and enable more iterative hyperparameter tuning, as well as support models with increased depth and base units to capture more complex relationships within the data.

To mitigate discrepancies between ground observations and satellite-based estimates, especially the underestimation of peak events, future research should consider refining IMERG-Final or other satellite products (e.g., CMORPH) using direct measurements such as radar data or dense rain gauge networks. Alternatively, training models directly on high-quality regional rainfall data, when available, could improve performance. However, since such data is currently unavailable in Ghana, innovative approaches are needed to establish statistical relationships between satellite observations and the probability of peak rainfall intensities. Techniques such as Non-Parametric Bayesian Networks or Copula and Vine Networks could help address this challenge. Investigating the spatial variability within one grid cell by placing multiple ground stations within less than 10 x 10 km from each other can contribute to establishing the spatial rainfall variability, which is especially interesting for the convective cores of the MCS.

Lastly, future research should assess the practical implications of the model. Given the absence of a standardized definition for early warning system requirements, it is important to determine acceptable thresholds for false positives and missed extreme rainfall events to maintain the credibility of the system. Engaging with local authorities and communities is necessary to establish these thresholds and to ensure that the model is in line with practical needs.

8.3. Final remarks

Overall, the Earthformer model, particularly with balanced loss, provides a valuable near real-time alternative to IMERG-Early. Although its underestimation of extreme rainfall limits its use in early warning systems, the model still remains valuable for other applications, such as short-term weather monitoring in regions where ground-based observations are sparse. As the model performs relatively well for the lower and mid intensities it could help in monitoring dry spells and drought patterns in these regions. This might supports farmers in decision-making by monitoring the rainfall distribution and plan the irrigation accordingly.

References

Amediegwu, L. (2024, September). El Niño may have ended, but its legacy is greater hunger in sub-Saharan Africa.

Ansah, S. O., Ahiataku, M. A., Yorke, C. K., Otu-Larbi, F., Yahaya, B., Lamptey, P. N., & Tanu, M. (2020). Meteorological Analysis of Floods in Ghana. *Advances in Meteorology*, *2020*. https://doi.org/10.1155/2020/4230627

Atiah, W. A., Amekudzi, L. K., & Danuor, S. K. (2023, March). Mesoscale convective systems and contributions to flood cases in Southern West Africa (SWA): A systematic review. https://doi.org/10.1016/j.wace.2023.100551

Benas, N., Finkensieper, S., Stengel, M., Van Zadelhoff, G. J., Hanschmann, T., Hollmann, R., & Meirink, J. F. (2017). The MSG-SEVIRI-based cloud property data record CLAAS-2. *Earth System Science Data*, *9*(2), 415–434. https://doi.org/10.5194/essd-9-415-2017

Bessah, E., Amponsah, W., Ansah, S. O., Afrifa, A., Yahaya, B., Wemegah, C. S., Tanu, M., Amekudzi, L. K., & Agyare, W. A. (2022). Climatic zoning of Ghana using selected meteorological variables for the period 1976–2018. *Meteorological Applications*, *29*(1). https://doi.org/10.1002/met.2049

Bogerd, L., Pinto, R. B., Leijnse, H., Meirink, J. F., van Emmerik, T. H., & Uijlenhoet, R. (2024). Gauging the ungauged: estimating rainfall in a West African urbanized river basin using ground-based and spaceborne sensors. https://doi.org/10.1080/02626667.2023.2284871

Brasjen, A. M. (2014). *PRECIPITATION ESTIMATION FROM INFRARED SATELLITE IMAGERY* [Doctoral dissertation].

Codjoe, S. N., & Atiglo, D. Y. (2020, December). The Implications of Extreme Weather Events for Attaining the Sustainable Development Goals in Sub-Saharan Africa. https://doi.org/10.3389/fclim.2020.592658

D'Adderio, L. P., Casella, D., Dietrich, S., Panegrossi, G., & Sanò, P. (2023). A First Step towards Meteosat Third Generation Day-2 Precipitation Rate Product: Deep Learning for Precipitation Rate Retrieval from Geostationary Infrared Measurements. *Remote Sensing*, *15*(24). https://doi.org/10.3390/rs15245662

Deutscher Wetterdienst. (n.d.). GPCC VISUALIZER. https://kunden.dwd.de/GPCC/Visualizer

Dezfuli, A. K., Ichoku, C. M., Huffman, G. J., Mohr, K. I., Selker, J. S., De Giesen, N. V., Hochreutener, R., & Annor, F. O. (2017). Validation of IMERG Precipitation in Africa. https://doi.org/10.1175/JHM-D-17-0139.s1

Estébanez-Camarena, M., Curzi, F., Taormina, R., van de Giesen, N., & ten Veldhuis, M. C. (2023). The Role of Water Vapor Observations in Satellite Rainfall Detection Highlighted by a Deep Learning Approach. *Atmosphere*, *14*(6). https://doi.org/10.3390/atmos14060974

Estébanez-Camarena, M., Taormina, R., van de Giesen, N., & ten Veldhuis, M. C. (2023). The Potential of Deep Learning for Satellite Rainfall Detection over Data-Scarce Regions, the West African Savanna. *Remote Sensing*, *15*(7). https://doi.org/10.3390/rs15071922

EUMETSAT. (2019, December). *Algorithm Theoretical Baseline Document ATBD-60-63 (Product H60-P-IN-SEVIRI-PMW Product H63-P-IN-SEVIRI-PMW_E)* (tech. rep.). EUMETSAT.

Eumetsat. (2005, October). Radiances to brightness temperature conversion On the conversion from radiances to equivalent brightness temperature. http://physics.nist.gov/cuu/Constants/index.html

Eumetsat. (2012, October). Conversion from radiances to reflectances for SEVIRI warm channels. http://www.eumetsat.int

Fischer, F. K., & Winterrath, T. (2021, January). Precipitation measurement techniques, typical data sets, and their application in erosion research and extreme value statistics. In *Precipitation: Earth surface responses and processes* (pp. 147–172). Elsevier. https://doi.org/10.1016/B978-0-12-822699-5.00007-0

Gao, Z., Shi, X., Wang, H., Zhu, Y., Wang, Y., Li, M., & Yeung, D.-Y. (2022). Earthformer: Exploring Space-Time Transformers for Earth System Forecasting. http://arxiv.org/abs/2207.05833

Geerts, B., & Dejene, T. (2005). Regional and Diurnal Variability of the Vertical Structure of Precipitation Systems in Africa Based on Spaceborne Radar Data. *Journal of climate*, *18*(7), 893–916. https://doi.org/https://doi.org/10.1175/JCLI-3316.1

Glas. (2024, May). *Accuracy and suitability of satellite-based retrieval products for operational precipitation nowcasting in Ghana* [Doctoral dissertation]. https://essay.utwente.nl/98974/1/Glas_MA_ET.pdf

Han, L., Liang, H., Chen, H., Zhang, W., & Ge, Y. (2022). Convective Precipitation Nowcasting Using U-Net Model. *IEEE Transactions on Geoscience and Remote Sensing*, *60*. https://doi.org/10.1109/TGRS.2021.3100847

Hong, Y., & Hsu, K.-L. (2004). Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology and Climatology*, *43*(12), 1834–1853. https://doi.org/https://doi.org/10.1175/JAM2173.1

Huffman, G. J., Bolvin, D. T., Joyce, R., Kelley, O. A., Nelkin, E. J., Tan, J., Watters, D. C., & West, B. J. (2023, July). *Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation* (tech. rep.). Earth Sciences Division, NASA Goddard Space Flight Center.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). *CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution* (tech. rep.).

Kikuchi, M., Braun, S. A., Suzuki, K., Liu, G., & Battaglia, A. (2023, December). *Satellite Precipitation Measurements: What Have We Learnt About Cloud-Precipitation Processes From Space?* (S. C. Sullivan & C. Hoose, Eds.; 1st ed.). Geophysical Monograph Series. https://doi.org/https://doi.org/10.1002/9781119700357.ch15

Klein, C., Belušić, D., & Taylor, C. M. (2018). Wavelet Scale Analysis of Mesoscale Convective Systems for Detecting Deep Convection From Infrared Imagery. *Journal of Geophysical Research: Atmospheres*, *123*(6), 3035–3050. https://doi.org/10.1002/2017JD027432

Kreider, E. P. (2024, December). Thunderstorm. https://www.britannica.com/science/thunderstorm/Types-of-thunderstorms

Küçük, Ç., Atencia, A., & Dabernig, M. (2024). Integrated nowcasting of convective precipitation with Transformer-based models using multi-source data. http://arxiv.org/abs/2409.10367

Küçük, Ç., Giannakos, A., Schneider, S., & Jann, A. (2024). Transformer-Based Nowcasting of Radar Composites from Satellite Images for Severe Weather. https://doi.org/10.1175/AIES-D-23

Le Coz, C., Heemink, A., Verlaan, M., & van de Giesen, N. (2021). Spatial and time warping for gauge adjustment of rainfall estimates. *Atmosphere*, *12*(11). https://doi.org/10.3390/atmos12111510

Leinonen, J., Hamann, U., Nerini, D., Germann, U., & Franch, G. (2023). Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. http://arxiv.org/abs/2304.12891

Li, R., Guilloteau, C., Kirstetter, P.-E., & Foufoula-Georgiou, E. (2024). Understanding the error patterns of multi-satellite precipitation products during the lifecycle of precipitation events for diagnostics and algorithm improvement. *Journal of Hydrology*, 132610. https://doi.org/10.1016/j.jhydrol.2024.132610

Lizcano, G., McSweeney, C., & New, M. (2008, October). UNDP Climate Change Country Profile: Ghana. https://web.archive.org/web/20130921055503/http://ncsp.undp.org/document/undp-climate-change-country-profile-11

Manishsiq. (2024, September). Tyoes of Clouds, IMportance, Classificatios, Diagram. https://www.studyiq.com/articles/types-of-clouds/

Mccollum, J. R., Gruber, A., & Ba, M. B. (2000). Discrepancy between Gauges and Satellite Estimates of Rainfall in Equatorial Africa. *Journal of Applied Meteorology and Climatology*, *39*(5), 666–679. https://doi.org/https://doi.org/10.1175/1520-0450-39.5.666

Middleton, J. F., Smedley, A., Dickson, K. B., Clarke, J. I., Gardiner, R. K., Kröner, A., Mabogunje, A. L., McMaster, D. N., Nicol, D. S., & Steel, R. W. (2024). Climate of Africa. https://www.britannica.com/place/Africa/Climate

Moazami, S., Na, W., Najafi, M. R., & de Souza, C. (2022, October). Spatiotemporal bias adjustment of IMERG satellite precipitation data across Canada. https://doi.org/10.1016/j.advwatres.2022.104300

Moraux, A., Dewitte, S., Cornelis, B., & Munteanu, A. (2019). Deep learning for precipitation estimation from satellite and rain gauges measurements. *Remote Sensing*, *11*(21). https://doi.org/10.3390/rs11212463

National Weather Service La Crosse. (2014, May). Weather Basics: Clouds and Precipitation. https://www.youtube.com/watch?v=H2Pe4G6vfYY&t=1s

Nikolaj Vinicoff. (2021, September). The ITCZ. https://nikovinic.medium.com/the-itcz-b04c453af091

NOAA. (2018, July). *Climate Data Record (CDR) Program Bias-Corrected CMORPH High-Resolution Global Precipitation Estimates [Precipitation-CMORPH]* (tech. rep.). NOAA.

Peinó, E., Petracca, M., Polls, F., Udina, M., & Bech, J. (2025, January). *Intercomparison of H SAF and IMERG heavy rainfall retrievals over a Mediterranean coastal region* (tech. rep.). University of Barcelona. https://doi.org/http://dx.doi.org/10.2139/ssrn.5079222

Perera, D., Seidou, O., Agnihotri, J., Wahid, A., & Rasmy, M. (2019). Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects. *UNU-INWEH Report Series*, (Issue 08). https://doi.org/10.13140/RG.2.2.28339.78880

Peters, M., & Tetzlaff, G. (1988). *Meteorology and Atmospheric Physics The Structure of West African Squall Lines and Their Environmental Moisture Budget* (tech. rep.).

Pradhan, R. K., Markonis, Y., Vargas Godoy, M. R., Villalba-Pradas, A., Andreadis, K. M., Nikolopoulos, E. I., Papalexiou, S. M., Rahim, A., Tapiador, F. J., & Hanel, M. (2022, January). Review of GPM IMERG performance: A global perspective. https://doi.org/10.1016/j.rse.2021.112754

Prigent, C. (2010). Precipitation retrieval from space: An overview. *Comptes Rendus - Geoscience*, *342*(4-5), 380–389. https://doi.org/10.1016/j.crte.2010.01.004

Rahimi, R., Ravirathinam, P., Ebtehaj, A., Behrangi, A., Tan, J., & Kumar, V. (2023). Global Precipitation Nowcasting of Integrated Multi-satellitE Retrievals for GPM: A U-Net Convolutional LSTM Architecture. http://arxiv.org/abs/2307.10843

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002). Supplement to An Introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society*, *83*(7), 992–992. https://doi.org/10.1175/bams-83-7-schmetz-2

Schneider, D. P., Deser, C., Fasullo, J., & Trenberth, K. E. (2013). Climate Data Guide Spurs Discovery and Understanding. *Eos, Transactions American Geophysical Union*, *94*(13), 121–122. https://doi.org/10.1002/2013EO130001

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. http://arxiv.org/abs/1706.03458

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Taravat, A., Proud, S., Peronaci, S., Del Frate, F., & Oppelt, N. (2015). Multilayer perceptron neural networks model for meteosat second generation SEVIRI daytime cloud masking. *Remote Sensing*, *7*(2), 1529–1539. https://doi.org/10.3390/rs70201529

van de Giesen, N., Hut, R., & Selker, J. (2014). The Trans□African Hydro□Meteorological Observatory (TAHMO). *WIREs Water*, *1*(4), 341–348. https://doi.org/10.1002/wat2.1034

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. http://arxiv.org/abs/1706.03762

Wang, C., Tang, G., Xiong, W., Ma, Z., & Zhu, S. (2021). Infrared precipitation estimation using convolutional neural network for FengYun satellites. *Journal of Hydrology*, *603*. https://doi.org/10.1016/j.jhydrol.2021.127113

Yang, X., Sun, P., Zhang, F., Du, Z., & Liu, R. (2021). Multi-task collaboration deep learning framework for infrared precipitation estimation. *Remote Sensing*, *13*(12). https://doi.org/10.3390/rs13122310

Zhang, D., He, Y., Li, X., Zhang, L., & Xu, N. (2023). PrecipGradeNet: A New Paradigm and Model for Precipitation Retrieval with Grading of Precipitation Intensity. *Remote Sensing*, *15*(1). https://doi.org/10.3390/rs15010227

Zhang, M. (2022). An Analytical Model of Two-Dimensional Mesoscale Circulation and Associated Properties Across Squall Lines. *AGU Advances*, *3*(6). https://doi.org/10.1029/2022AV000726

Ziarh, G. F., Shahid, S., Ismail, T. B., Asaduzzaman, M., & Dewan, A. (2021). Correcting bias of satellite rainfall data using physical empirical model. *Atmospheric Research*, *251*. https://doi.org/10.1016/j.atmosres.2020.105430

A

Model explanation and setup

A.1. Transformer models explained

Transformer models are a recently developed deep learning approach, initially designed for large language models (Vaswani et al., 2017). The key mechanism that drives transformer models is the attention mechanism, which can capture contextual information. The attention mechanism captures the importance of tokens relative to other tokens. A token is a vector that describes a specific part of the data.

The attention mechanism consists of a Key Query and Value matrix. The attention score is calculated as described in Equation A.1.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k})V} \tag{A.1}$$

In which Q is the Query matrix, K the Key matrix, V the value matrix and d_k represents the dimensions of the K vector.

The Key, Query and value matrices are obtained by multiplying each token both by a weight matrix W_K and a weight matrix w_Q (Figure A.1). To obtain the relative importance of each token to other tokens the dot product between the Key (K) matrix and the Query ((Q) is obtained. The higher the value of the dot product the higher the relative importance. The Weight Key matrix and the weight Query matrix are the same for all the tokens within a sequence.

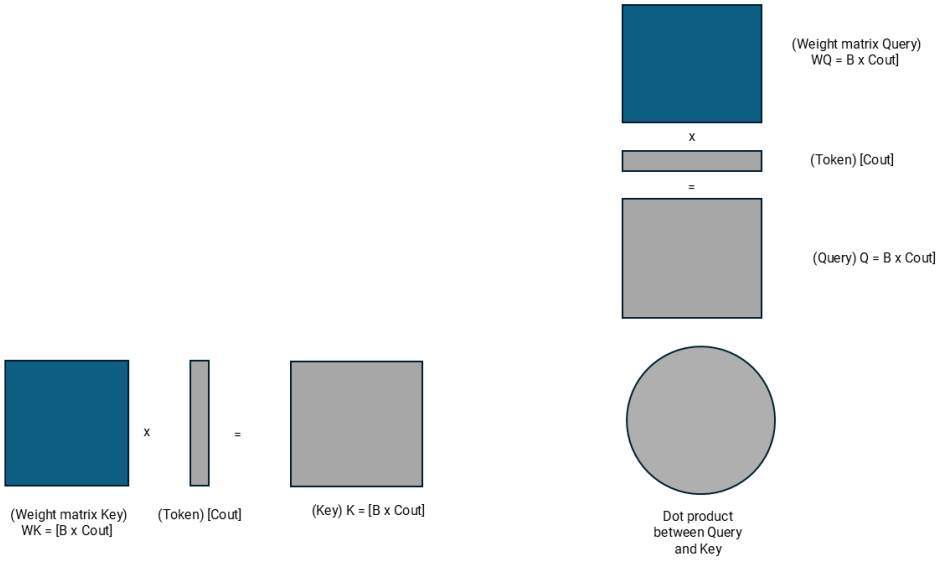


Figure A.1: The Query weight matrix and the Key weight matrix contain the trainable parameters of the model. Each token is multiplied by both the W_Q and the W_K matrix. To obtain the relative importance of each token to other tokens the dot product of the K matrix and Q is calculated.

The dot products of the Key and Query matrices for a sequence of input data can be gridded and show which other tokens are of relative importance related to the token (Figure A.2). By applying a softmax function this is translated into a probability distribution (the data is thus normalized and all the values are translated into values between zero and one). This grid is called the attention pattern. The probability distribution can be used to predict what comes next in the sequence of tokens, which is called self attention. Therefore only one input sequence can serve as many training samples as it is constantly predicting the next token. However, it is important with self attention that no information of 'future' tokens in the sequence is transferred to the previous tokens, as this already gives away the answer. However, when translating one dataset into another dataset, such as language translation tasks or translating satellite images into precipitation values, this becomes less important. The mechanism almost acts the same, but the Key and Query matrices are obtained from different datasets, instead of the same. This is called cross attention.

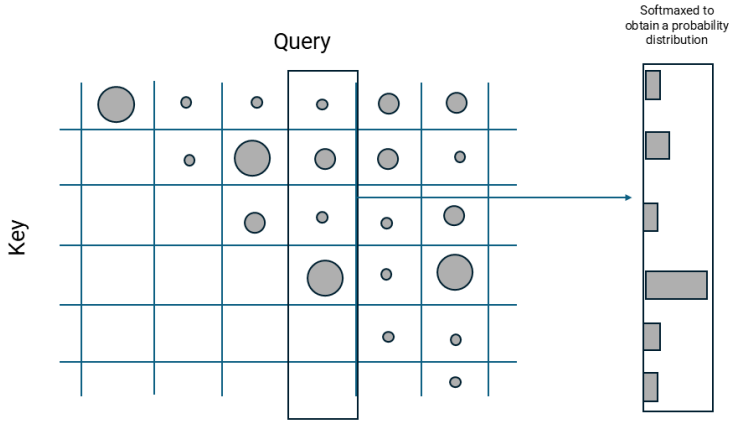


Figure A.2: The higher the dot product, the higher the relative importance of the token to the other tokens. By softmaxing the dot products a probability vector is created that can used to predict the next token.

To capture all the information of the different tokens into one vector the Value matrix comes in place. The Value weight matrix W_V is multiplied by each token again to obtain the value vector V. Subsequently, to preserve the information of the previous tokens in the sequence the value vector multiplied by the Softmaxed probability of the previous vector is added to the next token in the sequence, each token 'attending' to the other tokens. Eventually, the last token in the sequence contains all the information of the previous tokens as well (Figure A.3).

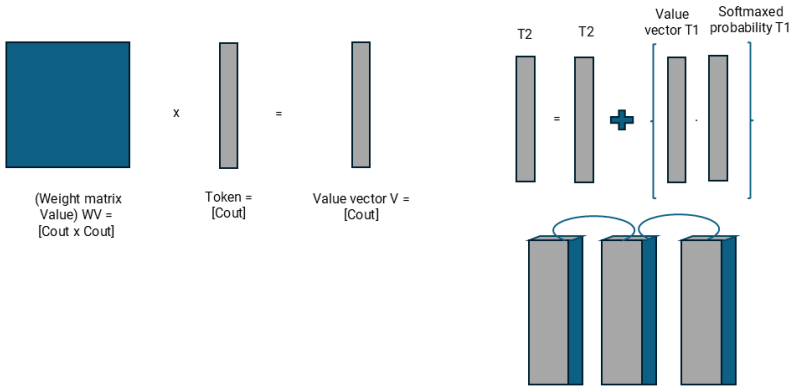


Figure A.3: The value weight matrix contains trainable parameters of the model and is multiplied by each token. The individual tokens attend to each other by multiplying the value matrix with the probability vector of the same token and adding it to the next token in the sequence.

All of this together is called one attention head. Typically, multiple attention heads are run simultaneously, so different heads can capture different structures in the data. Each head can focus on different parts of the input sequence or capture different types of relationships (e.g., local context, long-range

dependencies, etc.). The setup of the model of this research uses 32 attention heads. After each head computes its output, the results from all heads are concatenated and passed through a final linear projection layer.

A.2. Earthformer model setup

As a basis of the model for this research, the Eartformer for INCA model is used. This is a space-time transformer model especially designed for earth forecasting purposes. The Earthformer model was developed by Amazon (Gao et al., 2022) and further adapted to predict convective weather events in Austria using the Meteosat Second Generation and Integrated Nowcasting through Comprehensive Analysis (INCA) by Küçük, Atencia, and Dabernig, 2024. As described in section 3.1, the Earthformer model is a hierchical Transformer encoder-decoder model that uses Cuboid Attention. An overview of the layers of the model can be found in Figure A.4 and Table A.1).

The model requires input tensors with dimensions time, height, width, and the number of input layers $[T, H, W, C]$ in HDF5 format. In the Earthformer model the data is tokenized by downsampling the spatial dimensions and inceasing the channel dimensions by downsampling layers. The spatial resolution (H x W) of the input tensor are reduced first by a convolutional layer and secondly with a patch and merge algorithm. Patch and Merge takes a 2D input tensor that has C channels. It divides this tensor into small, non-overlapping patches, where each patch has a size of p×p (for example,2×2, 4×4, etc.). After creating these smaller patches, it then merges the spatial information from each patch into the channel dimension. This means that the spatial information (height and width) from the patches is condensed into the channel dimension, effectively reducing the spatial resolution while increasing the number of channels. In essence, patch and merge is a way of compressing the spatial information and expanding the feature space in the process. The used setup of the research uses two downsample layers. The first layer reduces the spatial dimension by a factor three and increases the channel dimension to 32. The second downsample layer reduces the spatial dimensions by a factor 2 and increases the channel dimension to 128. Note that the final channel dimension of the channel dimension should correspond to the base unit of cuboid attention mechanism as this is the embedding the model is expecting. As the model processes the data simultaneously, a positional encoder is used to save information of the position of the tokens in the sequence.

Block	**Layer**	**Resolution**	**Channels**
Input	-	248 $\times$ 184	11
2D CNN + Downsampler	Conv3 $\times$ 3	248 $\times$ 184	11 $\rightarrow$ 32
	Conv3 $\times$ 3	248 $\times$ 184	32
	GroupNorm16	248 $\times$ 184	32
	LeakyReLU	248 $\times$ 184	32
	PatchMerge	248 $\times$ 184 $\rightarrow$ 83 $\times$ 62	32 $\rightarrow$ 288
	LayerNorm	83 $\times$ 62	288
	Linear	83 $\times$ 62	288 $\rightarrow$ 32
2D CNN + Downsampler	Conv3 $\times$ 3	83 $\times$ 62	32 $\rightarrow$ 128
	Conv3 $\times$ 3	83 $\times$ 62	128
	GroupNorm16	83 $\times$ 62	128
	LeakyReLU	83 $\times$ 62	128
	PatchMerge	83 $\times$ 62 $\rightarrow$ 42 $\times$ 31	128 $\rightarrow$ 512
	LayerNorm	42 $\times$ 31	512
	Linear	42 $\times$ 31	512 $\rightarrow$ 128
Encoder Positional Embedding	PosEmbed	42 $\times$ 31	128

Block	Layer	Resolution	Channels
Cuboid Attention Block x 1	LayerNorm	42×31	128
	Cuboid $(T, 1, 1)$	42×31	128
	FFN	42×31	128
	LayerNorm	42×31	128
	Cuboid $(1, H, 1)$	42×31	128
	FFN	42×31	128
	LayerNorm	42×31	128
	Cuboid $(1, 1, W)$	42×31	128
	FFN	42×31	128
Downsampler	PatchMerge	$42 \times 31 \rightarrow 21 \times 16$	$128 \rightarrow 512$
	LayerNorm	21×16	512
	Linear	21×16	$512 \rightarrow 256$
Decoder Initial Positional Embedding	PosEmbed	21×16	256
Cuboid Attention Block x 1	LayerNorm	21×16	256
	Cuboid $(T, 1, 1)$	21×16	256
	FFN	21×16	256
	LayerNorm	21×16	256
	Cuboid $(1, H, 1)$	21×16	256
	FFN	21×16	256
	LayerNorm	21×16	256
	Cuboid $(1, 1, W)$	21×16	256
	FFN	21×16	256
Upsampler	Nearest NeighbourInterp	$21 \times 16 \rightarrow 42 \times 31$	128
	Conv3 $\times$ 3	42×31	128
Cuboid Attention Block x 1	LayerNorm	42×31	256
	Cuboid $(T, 1, 1)$	42×31	256
	FFN	42×31	256
	LayerNorm	42×31	256
	Cuboid $(1, H, 1)$	42×31	256
	FFN	42×31	256
	LayerNorm	42×31	256
	Cuboid $(1, 1, W)$	42×31	256
	FFN	42×31	256
2D CNN + Upsampler	Nearest NeighbourInterp	$21 \times 16 \rightarrow 42 \times 31$	$256 \rightarrow 64$
	Conv3 $\times$ 3	42×31	64
	GroupNorm16	42×31	64
	LeakyReLU	42×31	64

Table A.1: Represents the different layers of the model, showing the mechanism of each layer how the data flows through the model along with its dimensions.

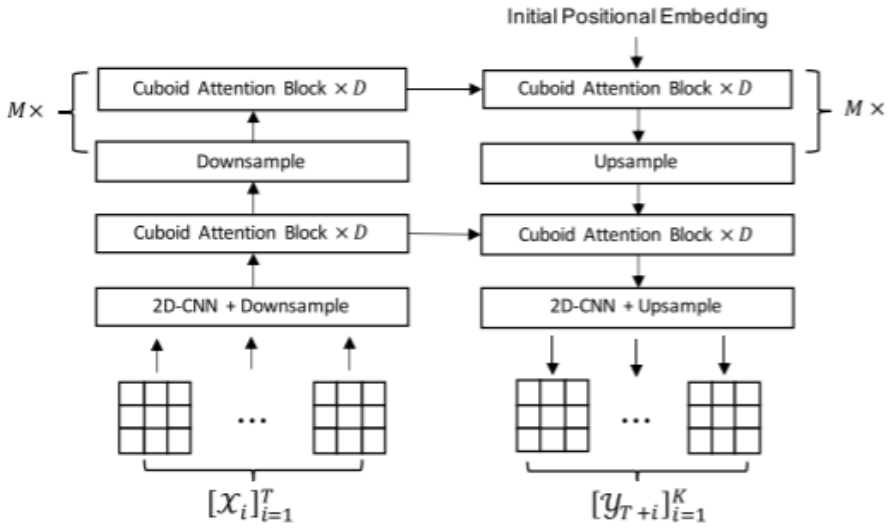


Figure A.4: A schematic overview of the Earthformer model, with the encoder on the left and the decoder on the right. The input tensor first passes through the 2D-CNN and Downsample layers, which reduce its spatial dimensions while increasing its channel dimensions through convolution, patch merging, tokenizing the data. The tokenized data is then processed by the cuboid attention block to capture contextual relationships. Here, M represents the number of hierarchical layers, while D denotes the number of cuboids within each layer. The model's upsampling is performed using NNI.

The convolved tensor passes through the cuboid attention mechanism. This mechanism involves decomposing the tensor, applying attention, and then merging the tensor back together (Figure A.5). The tensor can be decomposed using different decomposition strategies (Figure A.6). The attention pattern proven to be the most effective is the 'axial' pattern, which includes global vectors (Gao et al., 2022). In this approach, attention is first applied along the temporal direction, then along the height direction, and lastly along the width direction within the decomposed cuboids. Additionally, all decomposed cuboids attend to global vectors to capture the global dynamics of the system. For upsampeling Nearest Neighbour Interpolation (NNI) is used.

The model consists of two hierarchical layers (M), each containing a single cuboid attention block (D). The cuboid attention blocks are stacked and connected by a normalization layer and a feedforward network. Within each attention block, 32 attention heads are used, meaning that 32 parallel processes of cuboid attention mechanisms are run individually, each with its own Key (K), Query (Q), and Value (V) matrices. The outputs from these 32 parallel attention mechanisms are concatenated to form the final output of the cuboid attention block. The input and output of the model are hierarchically encoded and decoded to allow attention at different scales, enabling the model to capture both local patterns and system-wide dynamics.

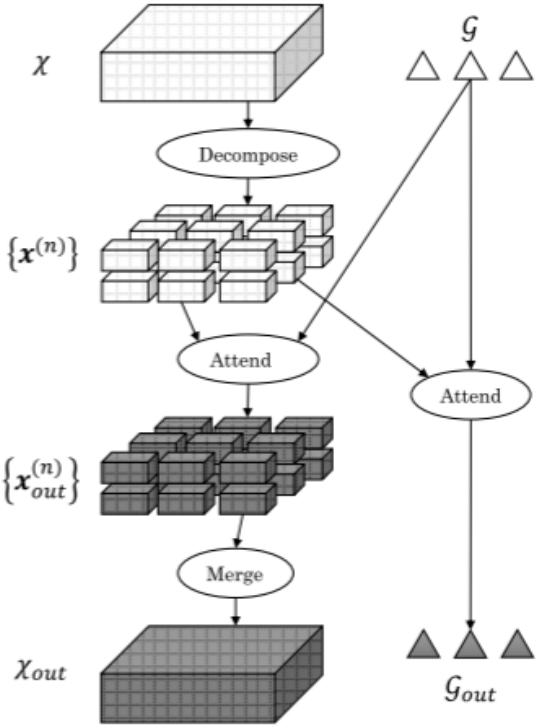


Figure A.5: The Figure shows how the attention mechanism within the cuboid is decomposed in smaller cuboids, for computational efficiency. Within each cuboid the attention mechanism is run, while all cuboids also attend to the global parameter to ensure system wide dynamics are captured as well. After the attention mechanism has run, the decompased blocks are merged back together.

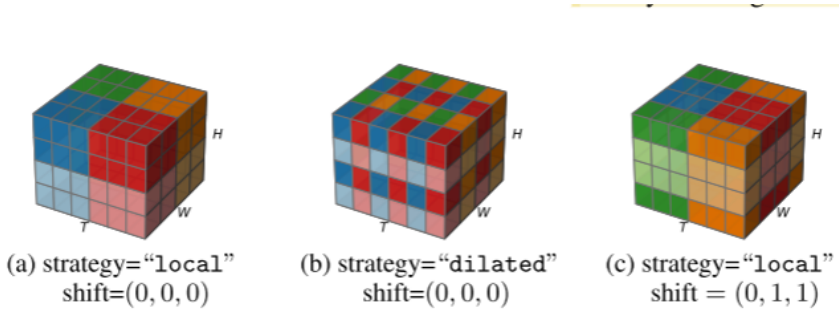


Figure A.6: The Figure shows different strategies to decompose the cuboid attention block.

The input and output of the model is hierarchically encoded and decoded so it can attend at different scales, thus finding both local patterns as well as system wide patterns.

A.2.1. Mixed precision

The model makes use of mixed precision, which means the model saves important parameters, such as the loss function in full precision (32-bits) and saves less sensitive parameters, such as the gradients in half precision (16-bits). Typically, gradients in Neural networks become smaller closer to the beginning of the network, when gradients are smaller than one and become bigger when gradients are bigger than one. To prevent underflow (gradients becoming too small to capture in mixed precision resulting in zero values) or overflow (gradients becoming too big to capture in mixed precision, resulting in NaN values) Pytorch Lighting dynamically scales the loss function up or down so the gradients are scaled

accordingly. After the backward pass and right before the optimization step the gradients are scaled back, so the 'true' gradients are used to update the model parameters. When the model accounts NaN values the dynamical scaler can drastically change its scaling, because it assumes overflow, resulting in vanishing gradients. Therefore no NaN values can be present in you traings dataset and your loss function should be defined in such a way it is always differentiable.

A.2.2. Gradient accumulation

To account for a limited amount GPU memory the model makes use of gradient accumulation, which mimics a larger batch size. The model takes the mean gradients of eight steps, with an actual batch size of four thus mimicking a batch size of 32, before updating its model parameters. The Pytorch Lighting library makes sure this is alligned with mixed precision and gradient clipping.

A.2.3. Gradient clipping

To prevent exploding gradients, which is a common issue in deep neural networks gradient clipping is applied. When the gradients of a vector exceed the gradient norm, which is essentially the length of the vector and can be calculated with the Formula A.2, the gradient norm is clipped to the clipping norm, which is 1.0 in the current model.

$$|x| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} \tag{A.2}$$

in which $|x|$ is the gradient norm and x the n gradients present in the vector.

A.2.4. AdamW optimizer

The optimizer used in this model is the Adam optimizer with weight decay, also called the Adaptive Moment Estimation optimizer. The adaptive optimizer ensures that even if a gradient is small or sparse, it can contribute to model performance. Additionally, it avoids that small gradients are completely ignored. This is done by adaptively changing the gradient for each parameter separately, by using the first and second order moments.

Weight decay helps the model to generalize and prevents overfitting, by preventing large parameter updates. The larger the weight decay the less likely it becomes for models to rely solely on a specific set of parameters, however when too big underfitting is also risked.

A.2.5. Early stopping

To prevent overfitting the model stopped the training if the validation loss doesn't improve over 20 epochs. For the final model parameters after training, the model weights that performed best on the validation dataset are selected as the optimal parameters of the model.

A.2.6. Learning rate scheduler

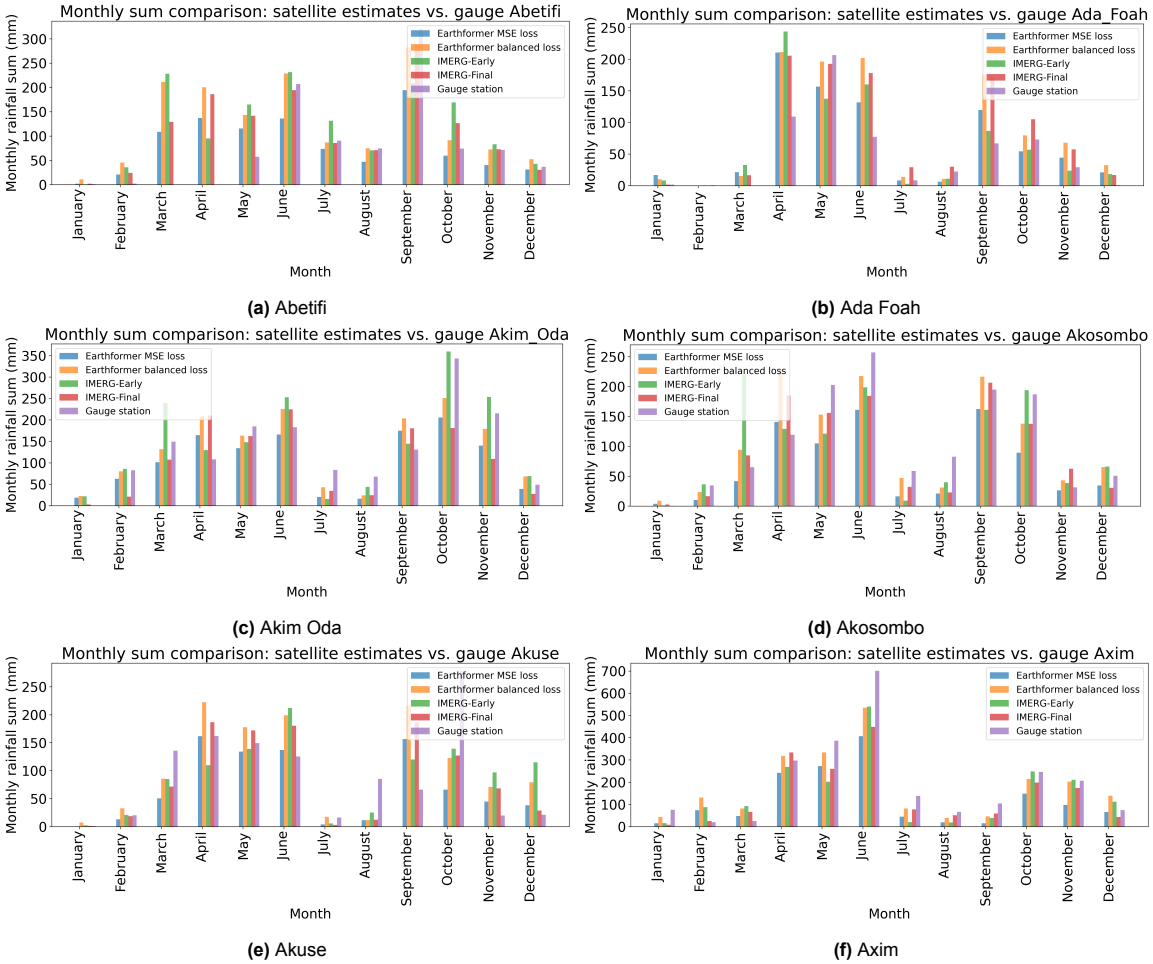
For Transformer models it is common to use an adaptive learning rate scheduler, as transformer models are very sensitive for its initial parameters. Therefore, the OneCycle learning rate scheduler is applied. In the beginning of training the model starts off with a relatively small learning rate (a division factor 25 of the maximum learning rate) and is gradually increased to the maximum learning rate, using Cosine annealing over 22% of the training phase. Afterwards the learning rate is gradually decreased again (with a division factor 10 of the initial learning rate), assuming that the model at the end of training approaches the optimum.

A.2.7. Activation functions

To prevent dead neurons and/or over saturated neurons, a GELU activation is used in the Feedforward network of the attention mechanism. In the convolutional layers leaky relu functions are used as activation functions. To ensure that the distribution of the activation layers in the model remain constant during training group normalization and layer normalization is applied. Group normalization normalizes the activation layers over specific groups, while layer normalization normalizes each activation neuron in a activation layer over the entire layer. Unlike batch normalization this is done for each data sample separately.

B

Cumulative rainfall per month

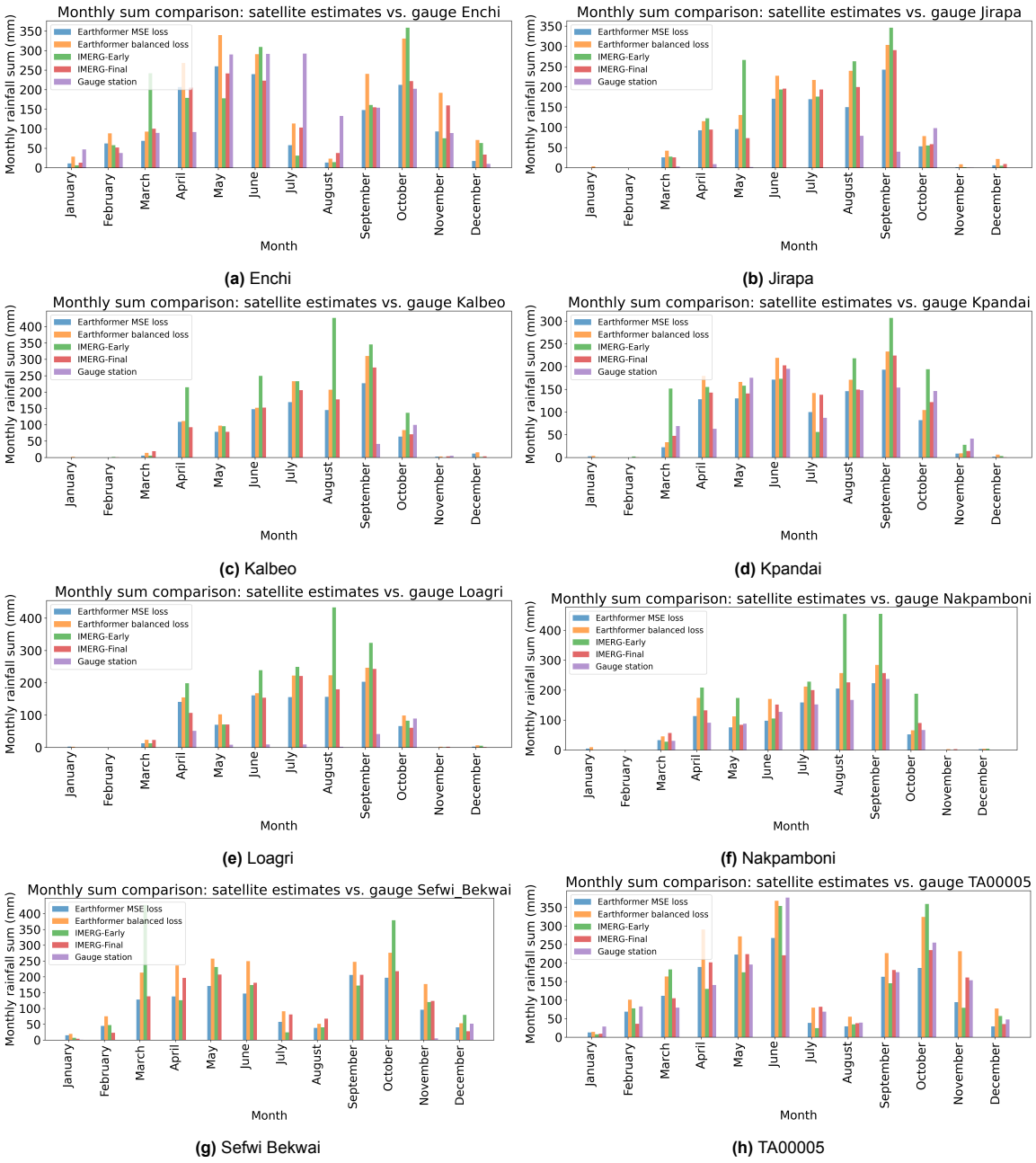


(a) Abetifi **(b)** Ada Foah

(c) Akim Oda **(d)** Akosombo

(e) Akuse **(f)** Axim

Figure B.1: Cumulative rainfall per month for various stations in 2022.



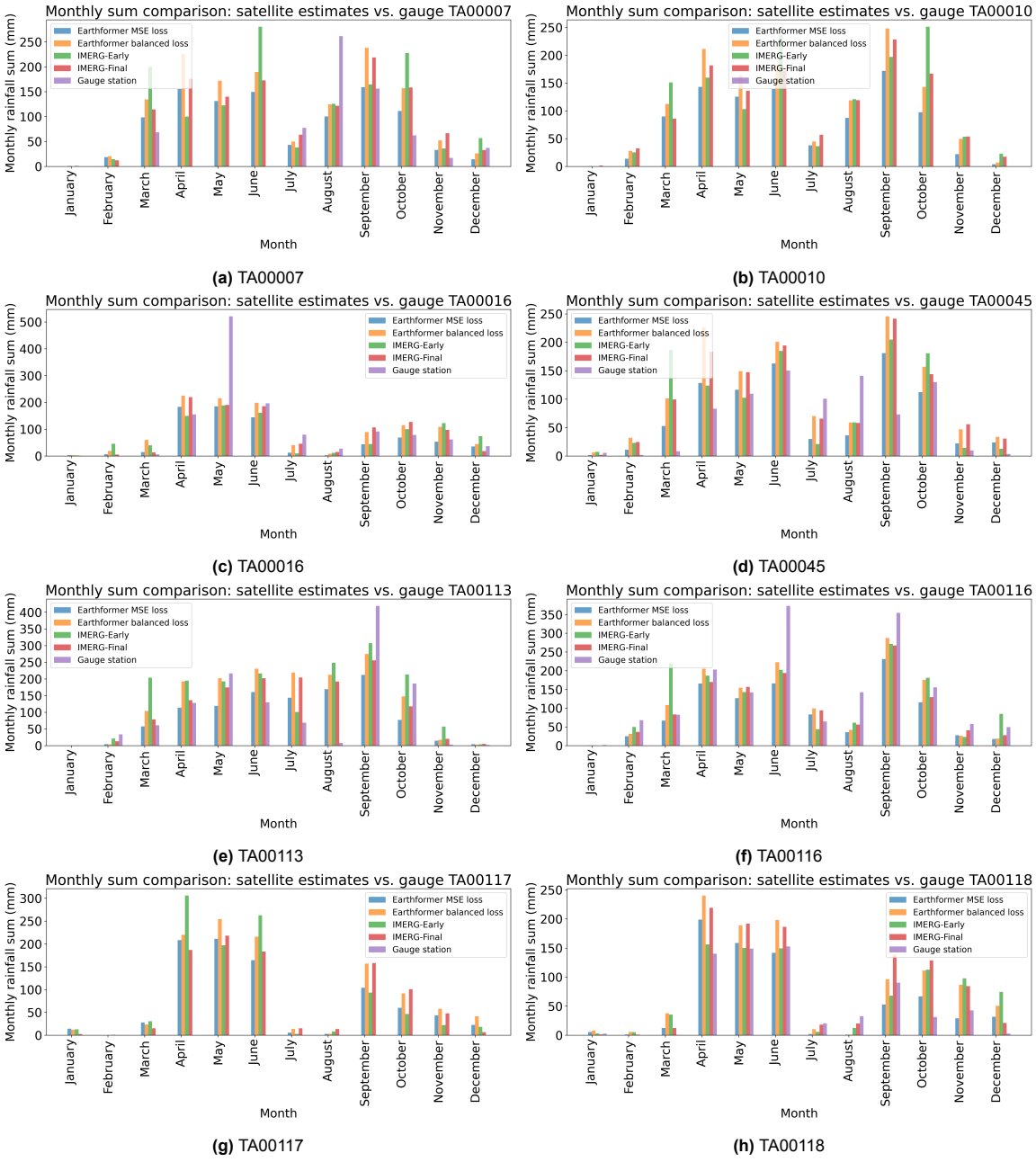
(a) Enchi **(b)** Jirapa

(c) Kalbeo **(d)** Kpandai

(e) Loagri **(f)** Nakpamboni

(g) Sefwi Bekwai **(h)** TA00005

Figure B.2: Cumulative rainfall per month for various stations in 2022.



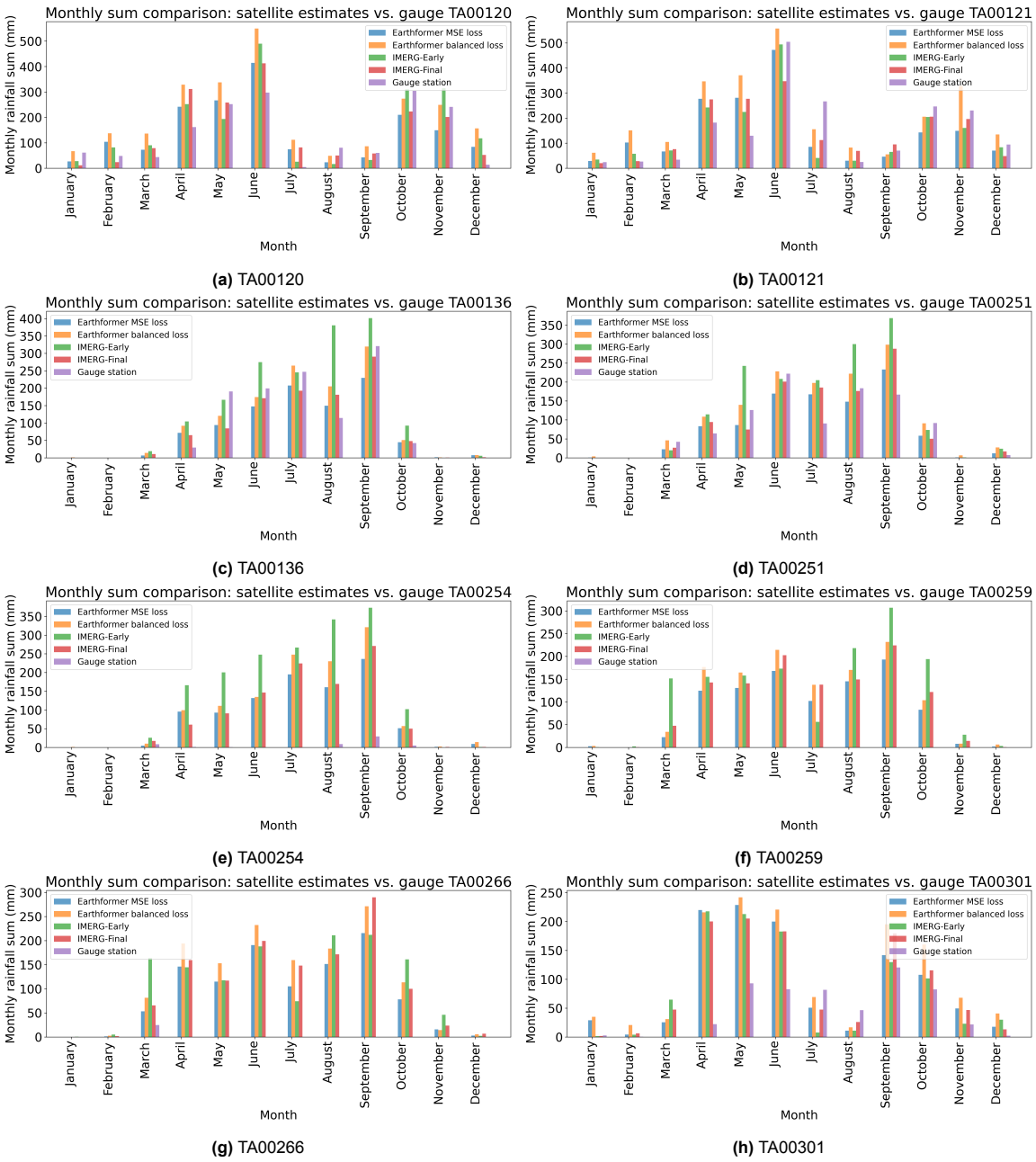
(a) TA00007 **(b)** TA00010

(c) TA00016 **(d)** TA00045

(e) TA00113 **(f)** TA00116

(g) TA00117 **(h)** TA00118

Figure B.3: Cumulative rainfall per month for various stations in 2022.



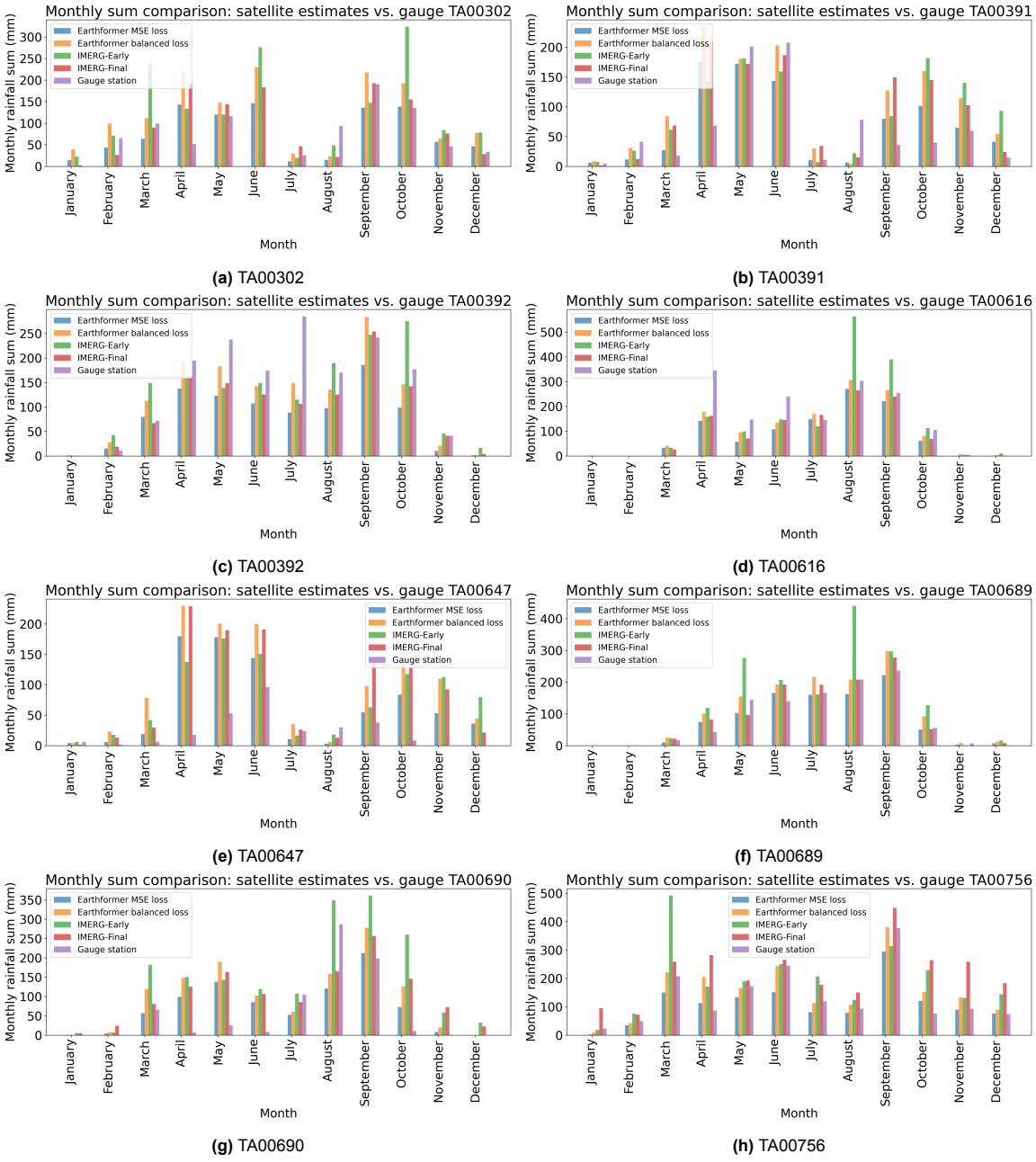
(a) TA00120 **(b)** TA00121

(c) TA00136 **(d)** TA00251

(e) TA00254 **(f)** TA00259

(g) TA00266 **(h)** TA00301

Figure B.4: Cumulative rainfall per month for various stations in 2022.



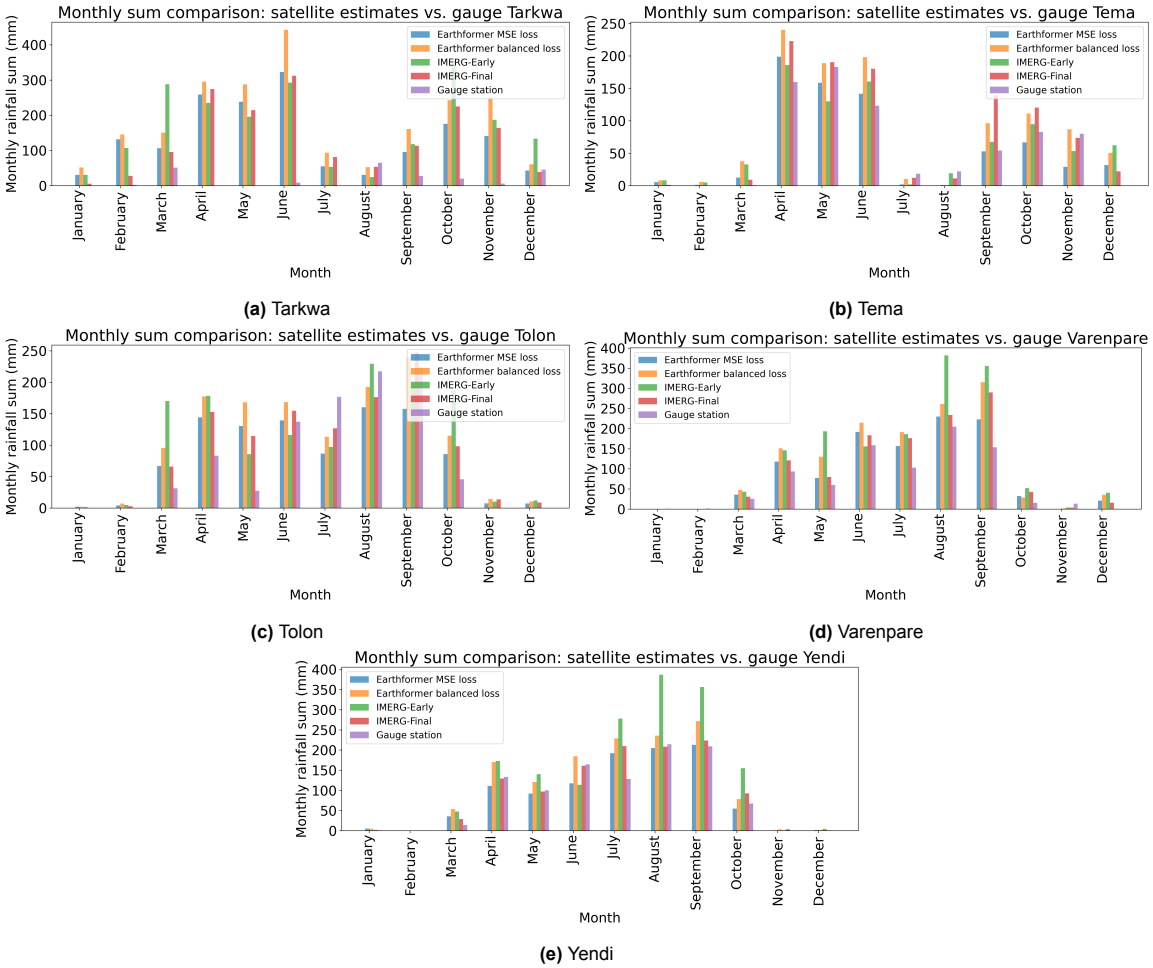
(a) TA00302 **(b)** TA00391

(c) TA00392 **(d)** TA00616

(e) TA00647 **(f)** TA00689

(g) TA00690 **(h)** TA00756

Figure B.5: Cumulative rainfall per month for various stations in 2022.



(a) Tarkwa

(b) Tema

(c) Tolon

(d) Varenpare

(e) Yendi

Figure B.6: Cumulative rainfall per month for various stations in 2022.

C

Spatial correlation plots with zeros

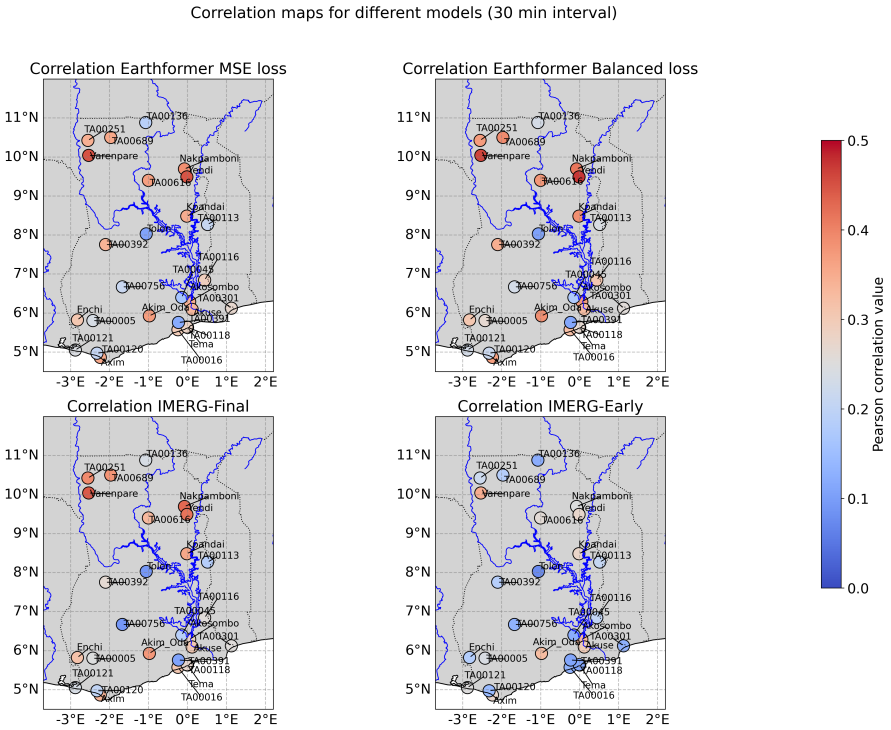


Figure C.1: The Pearson correlation including zero values for each used ground station in Ghana, plotted on the map for a 30 minute interval. The top left shows the correlation of the Earthformer model with MSE loss, the top right the correlaiton of the Earthformer with balanced loss, the bottom left the correlation with IMERG-Final and the bottom right the correlation with IMERG-Early.

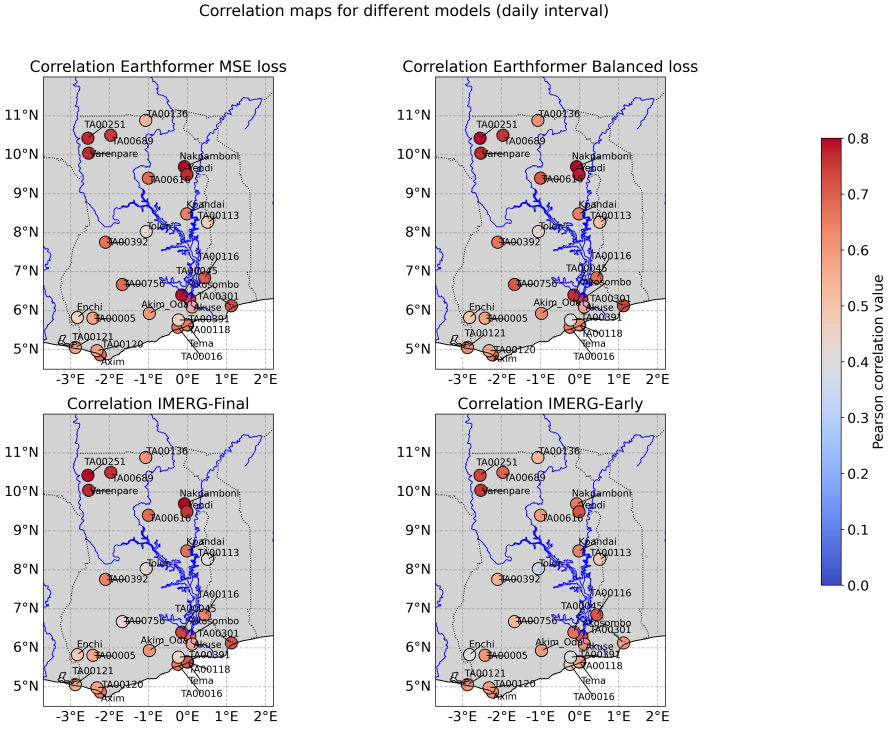
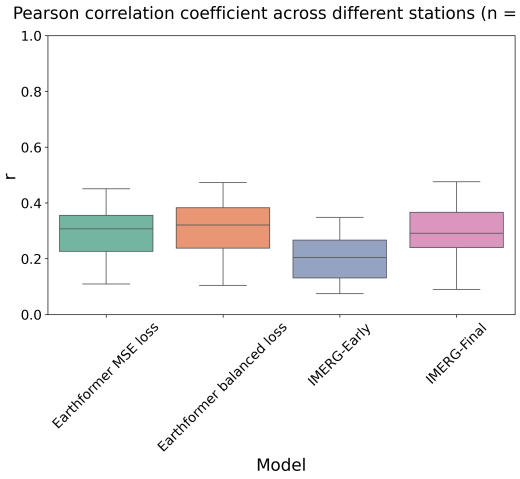


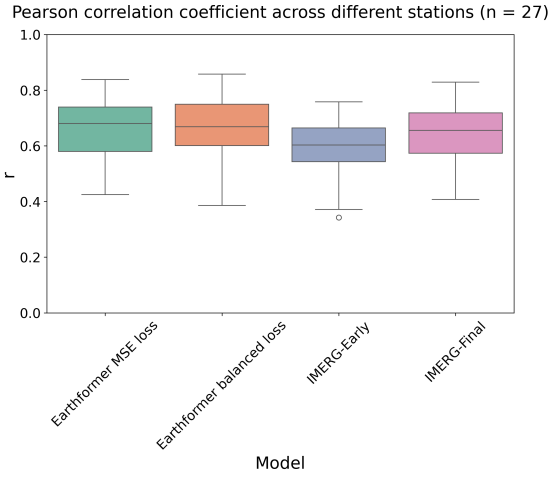
Figure C.2: The Pearson correlation including zero values for each used ground station in Ghana, plotted on the map for a daily interval. The top left shows the correlation of the Earthformer model with MSE loss, the top right the correlaiton of the Earthformer with balanced loss, the bottom left the correlation with IMERG-Final and the bottom right the correlation with IMERG-Early.

D

Correlation coefficients

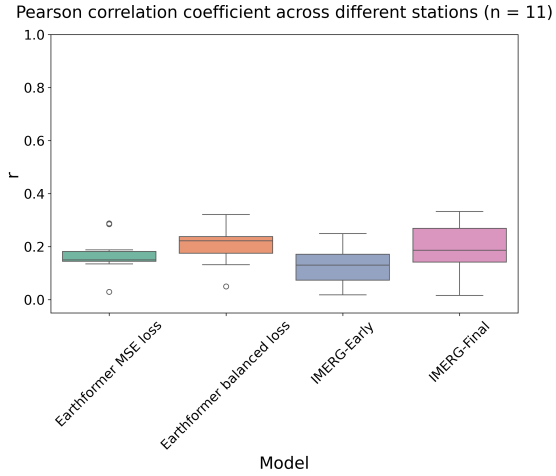


(a) Pearson correlation all stations zeros included 30 min time interval

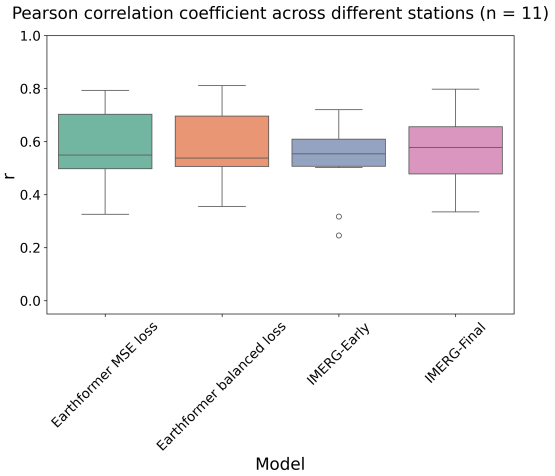


(b) Pearson correlation all stations zeros included daily time interval

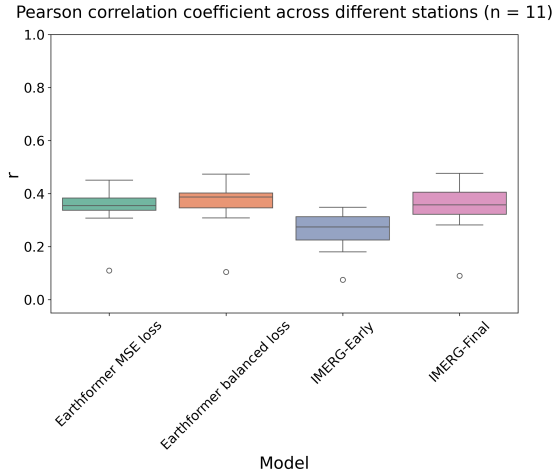
Figure D.1: Correlation coefficents with inclusion of zero data for different time intervals compared to all stations.



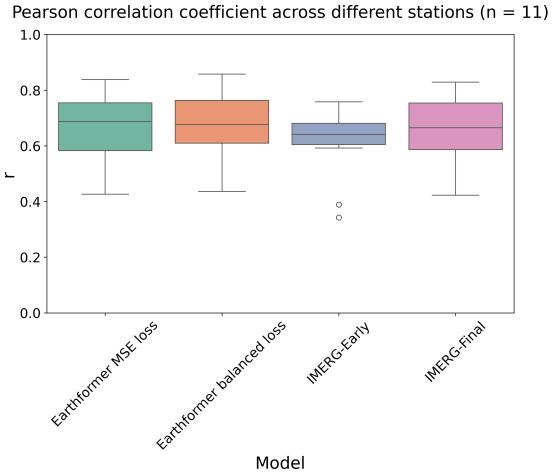
(a) Pearson correlation GMET zeros not included 30 min time interval



(b) Pearson correlation GMET zeros not included daily time interval

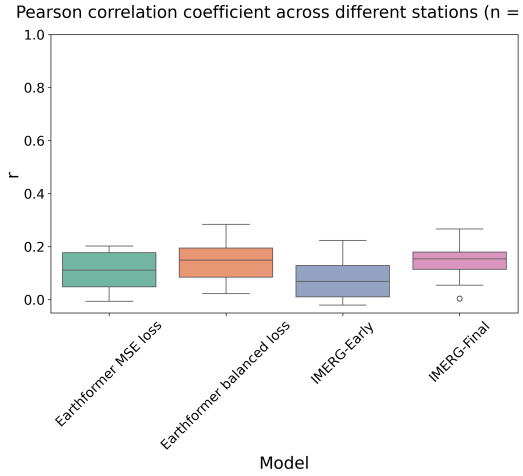


(c) Pearson correlation GMET zeros included 30 min time interval

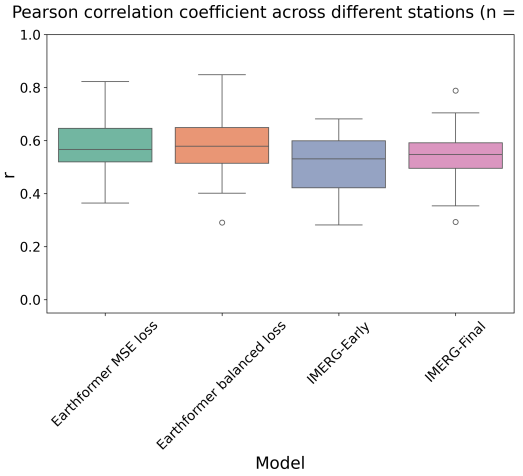


(d) Pearson correlation GMET zeros included daily time interval

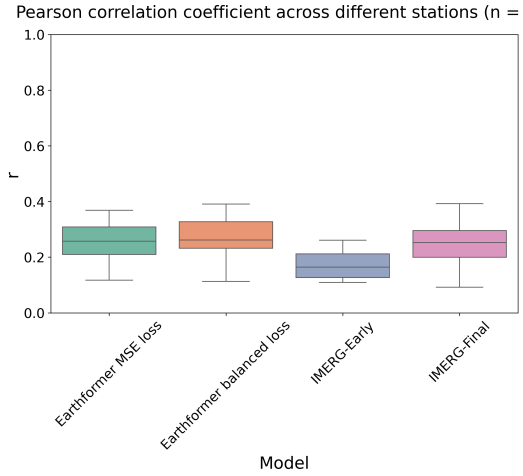
Figure D.2: Correlation coefficents with and without inclusion of zero data for different time intervals compared to GMET stations.



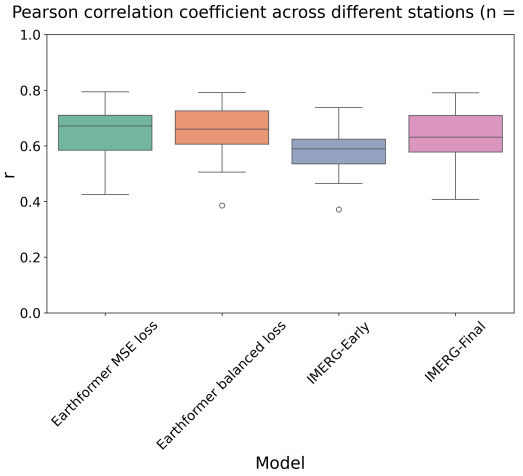
(a) Pearson correlation TAHMO zeros not included 30 min time interval



(b) Pearson correlation TAHMO zeros not included daily time interval

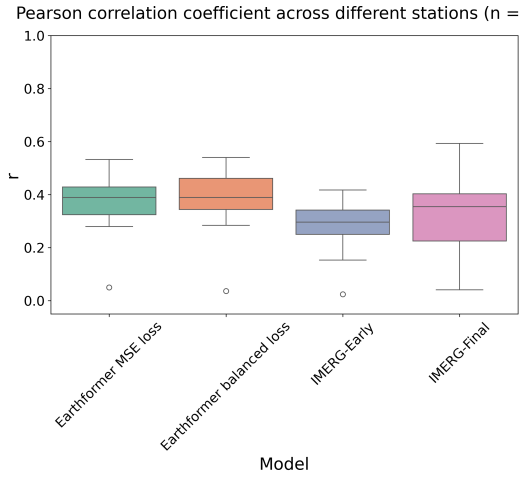


(c) Pearson correlation TAHMO zeros included 30 min time interval

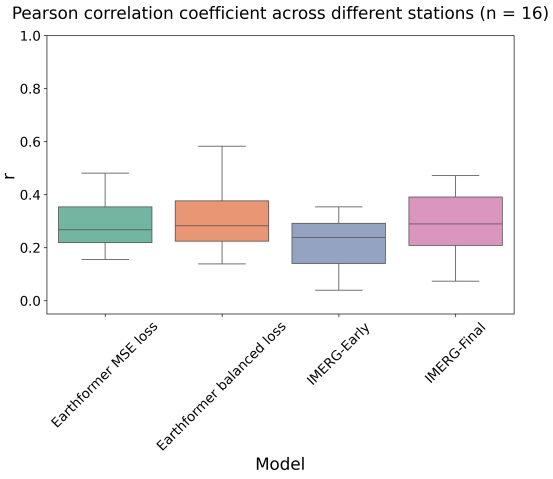


(d) Pearson correlation TAHMO zeros included daily time interval

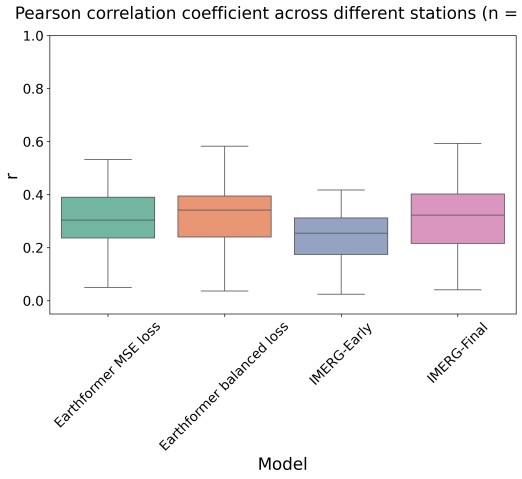
Figure D.3: Correlation coefficents with and without inclusion of zero data for different time intervals compared to TAHMO stations.



(a) Pearson correlation GMET stations zeros included 30 min time interval test dataset

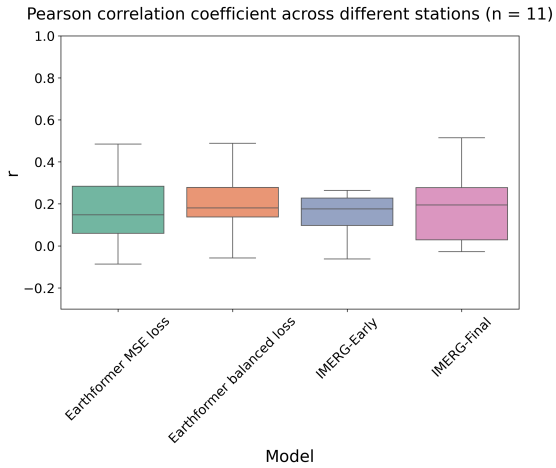


(b) Pearson correlation TAHMO stations zeros included 30 min time interval test dataset

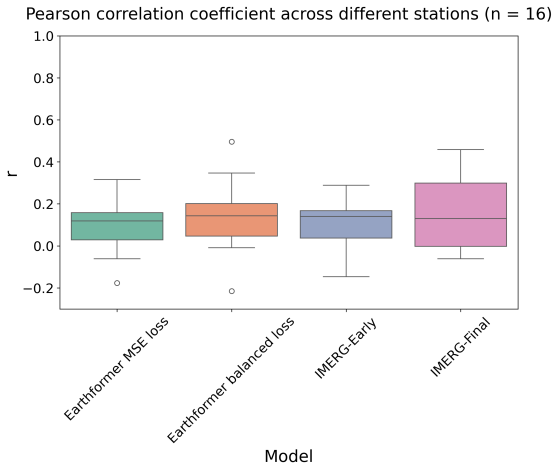


(c) Pearson correlation all stations zeros included 30 min time interval test dataset

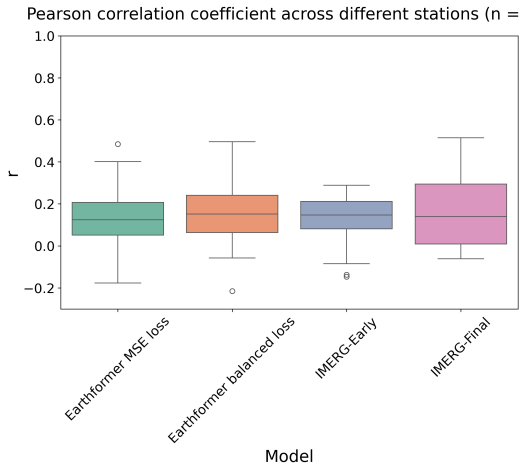
Figure D.4: Correlation coefficents with inclusion of zero data for 30 min time intervals compared to stations for the test dataset.



(a) Pearson correlation GMET stations zeros not included 30 min time interval test dataset



(b) Pearson correlation TAHMO stations zeros not included 30 min time interval test dataset

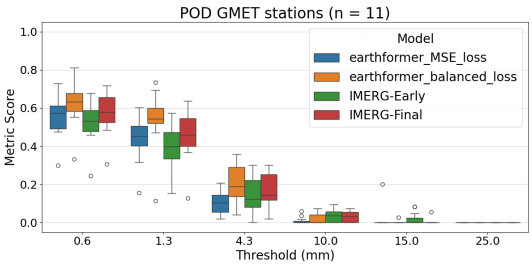


(c) Pearson correlation all stations zeros not included 30 min time interval test dataset

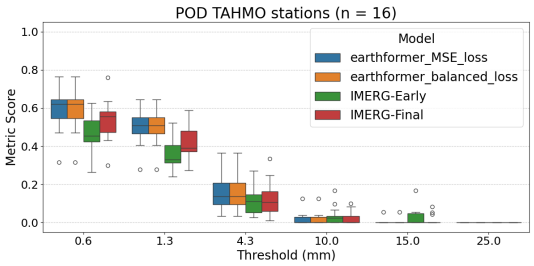
Figure D.5: Correlation coefficents without inclusion of zero data for 30 min time intervals compared to stations for the test dataset.

E

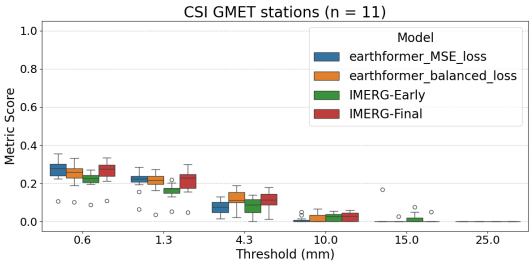
Metrics point pixel comparison



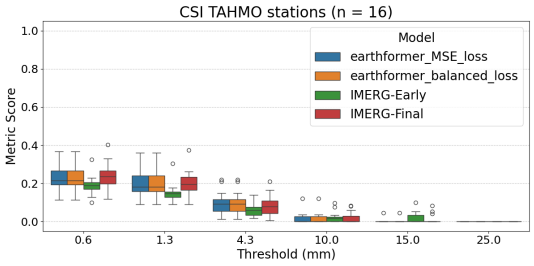
(a) POD GMET



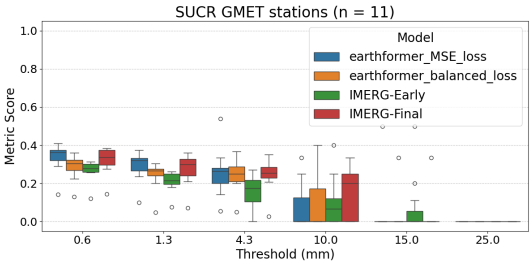
(b) POD TAHMO



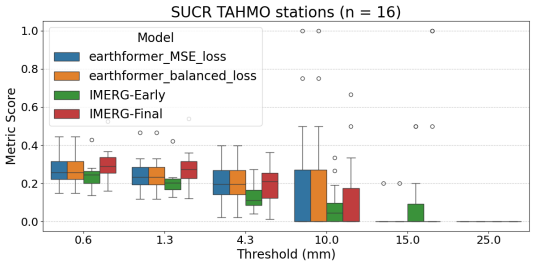
(c) CSI GMET



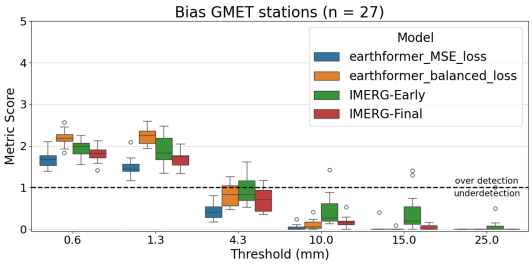
(d) CSI TAHMO



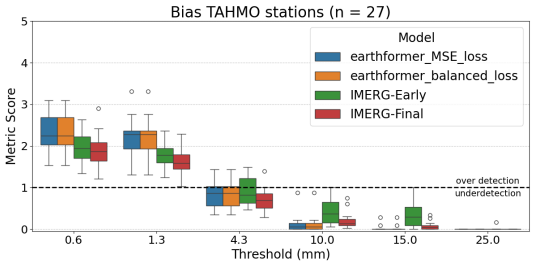
(e) SUCR GMET



(f) SUCR TAHMO

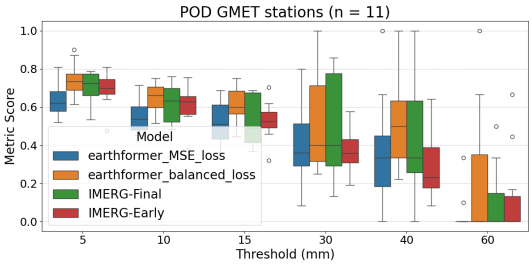


(g) Bias GMET.

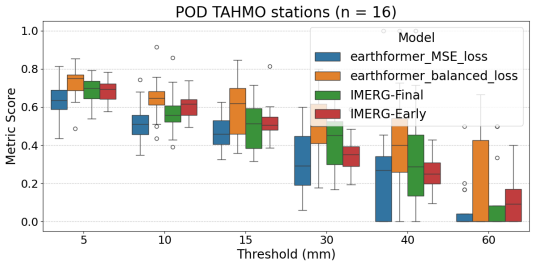


(h) Bias TAHMO

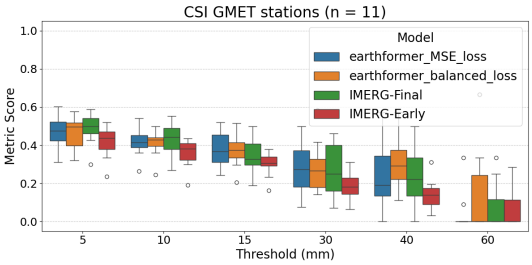
Figure E.1: Metrics calculated with mm precipitation over 30 min.



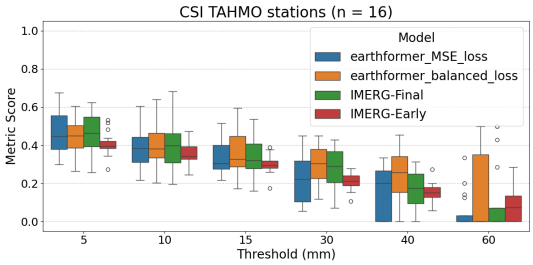
(a) POD GMET



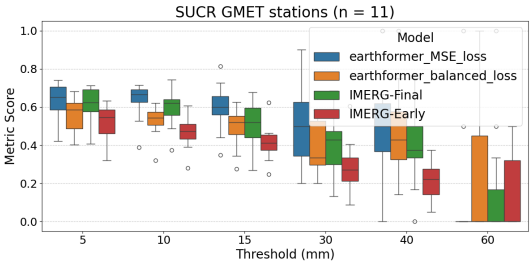
(b) POD TAHMO



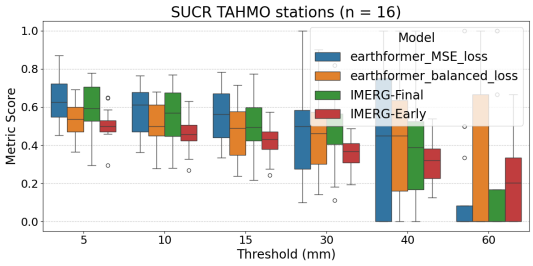
(c) CSI GMET



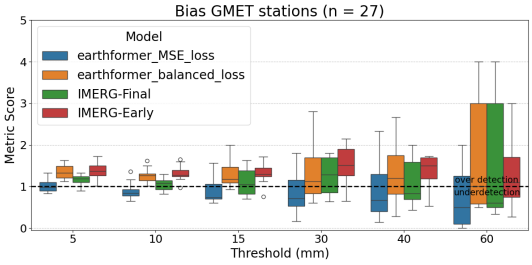
(d) CSI TAHMO



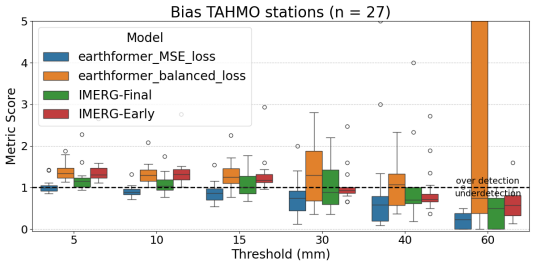
(e) SUCR GMET



(f) SUCR TAHMO



(g) Bias GMET

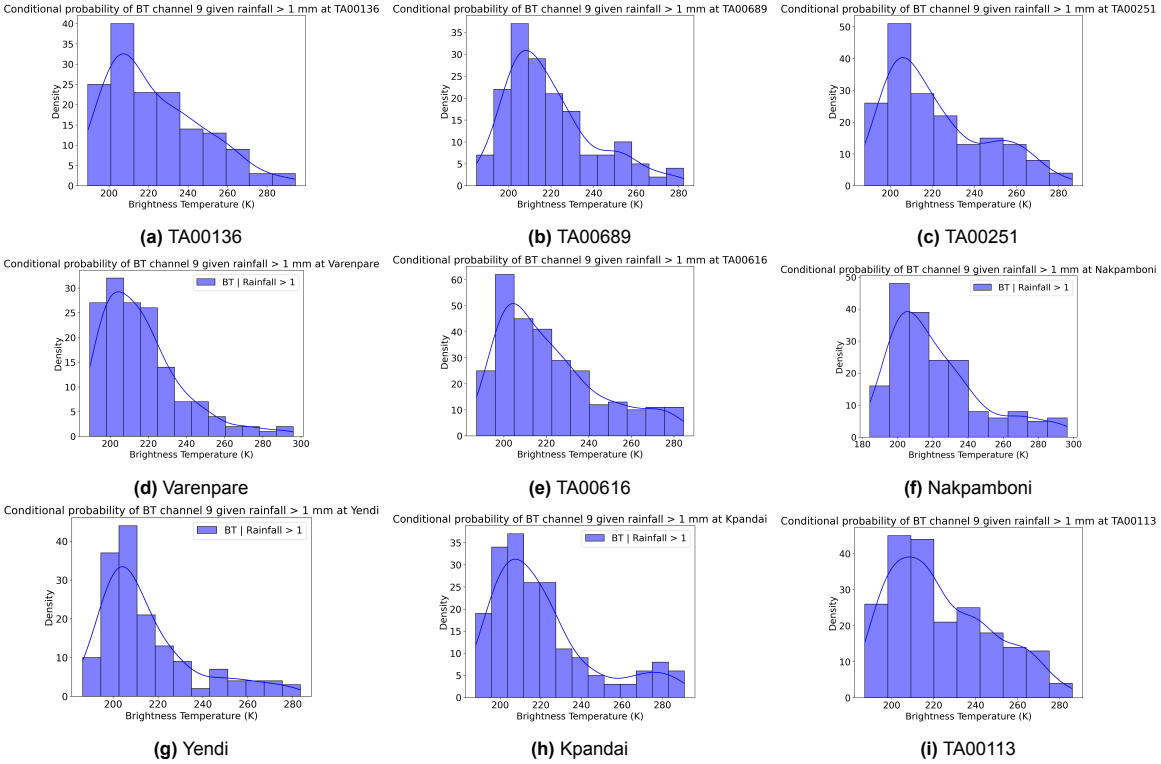


(h) Bias TAHMO. The extremely high bias for the Earthformer with balanced loss at the threshold of 60 mm is caused by one false alarm and zero hits and misses. The additional factor of 1e-6 to avoid divisions by zero causes the one false alarm to be divided by 1e-6.

Figure E.2: Metrics calculated for mm precipitation aggregated over a day.

Conditional probabilities BT channel 9

F.1. Savannah climatic zone



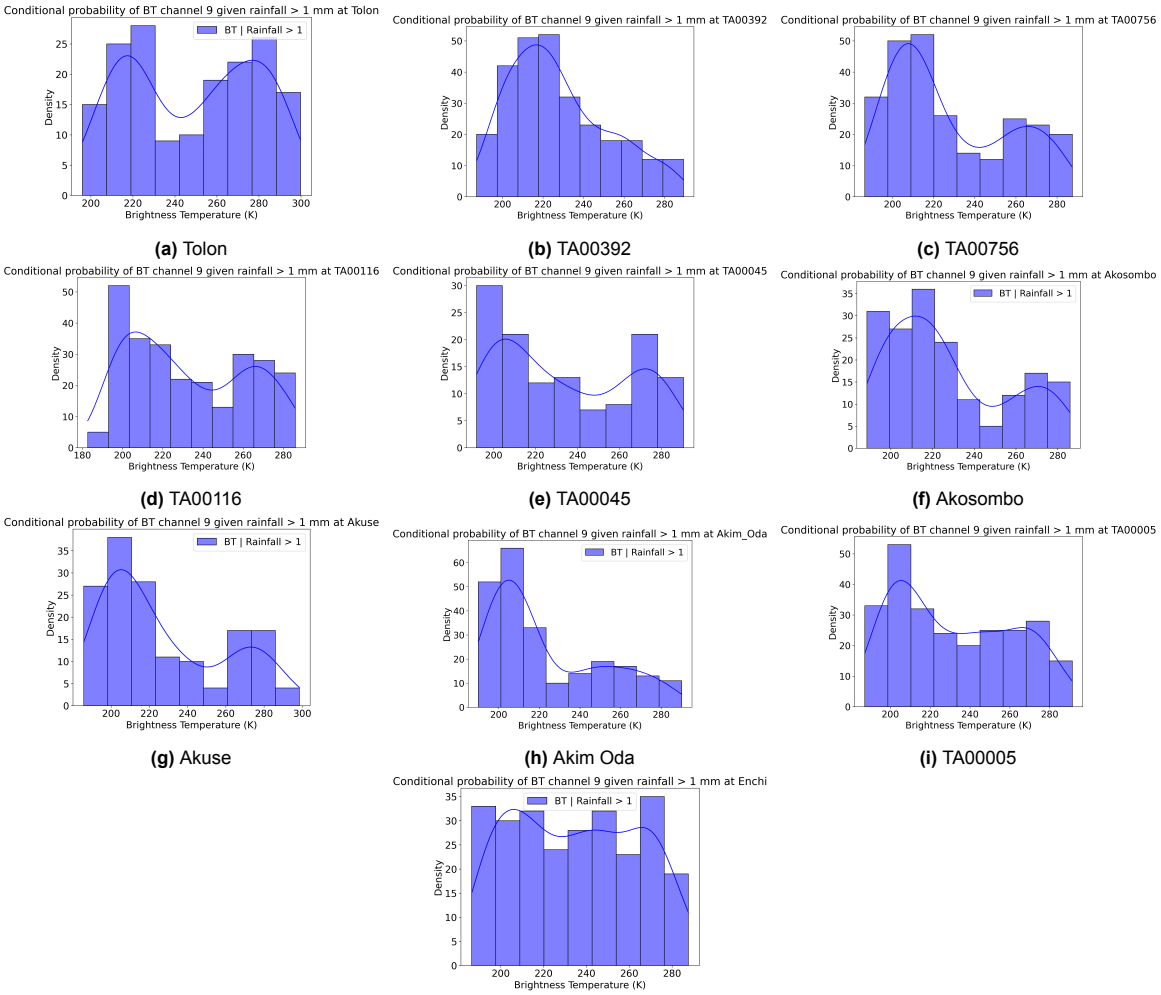
(a) TA00136 **(b)** TA00689 **(c)** TA00251

(d) Varenpare **(e)** TA00616 **(f)** Nakpamboni

(g) Yendi **(h)** Kpandai **(i)** TA00113

Figure F.1: Conditional probability of Brightness Temperatures (BT) of channel 9 of the nearest grid cell of the SEVIRI images, given that the rainfall is at least 1 mm fallen within 30 minutes within the Savannah climatic zone.

F.2. Forest climatic zone



(a) Tolon **(b)** TA00392 **(c)** TA00756

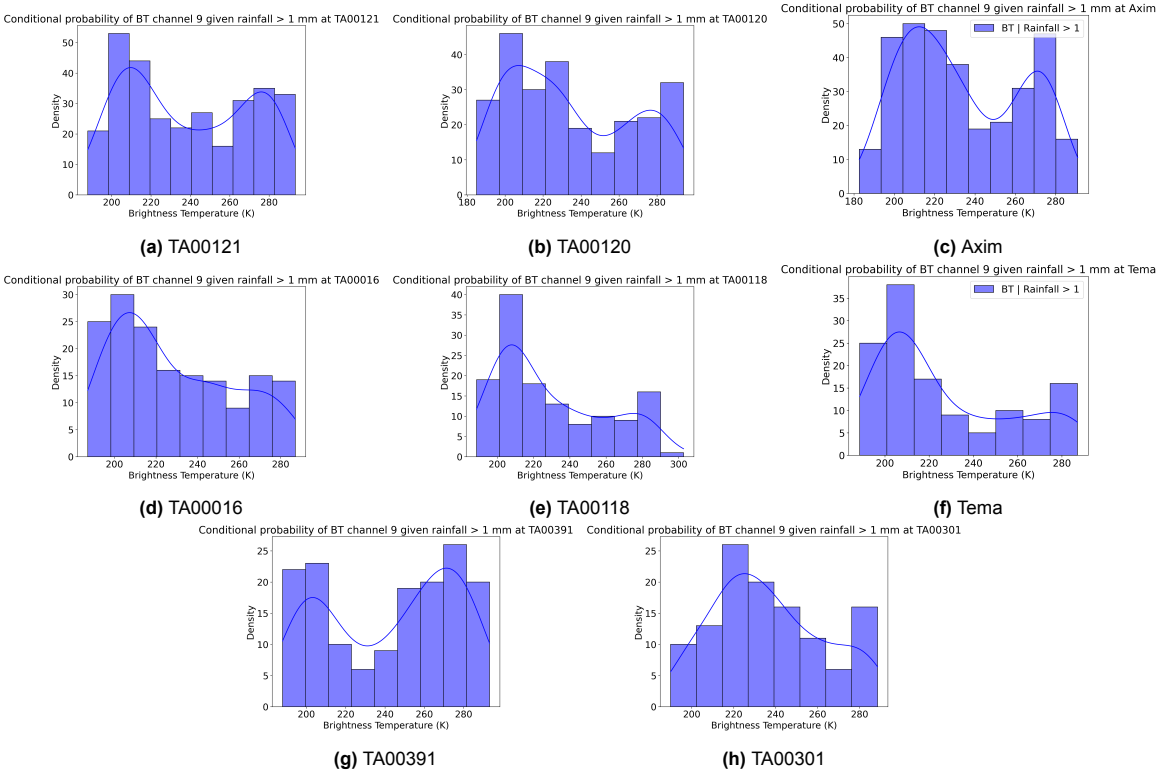
(d) TA00116 **(e)** TA00045 **(f)** Akosombo

(g) Akuse **(h)** Akim Oda **(i)** TA00005

(j) Enchi

Figure F.2: Conditional probability of Brightness Temperatures (BT) of channel 9 of the nearest grid cell of the SEVIRI images, given that the rainfall is at least 1 mm fallen within 30 minutes within the Forest climatic zone.

F.3. Coastal climatic zone



(a) TA00121 **(b)** TA00120 **(c)** Axim

(d) TA00016 **(e)** TA00118 **(f)** Tema

(g) TA00391 **(h)** TA00301

Figure F.3: Conditional probability of Brightness Temperatures (BT) of channel 9 of the nearest grid cell of the SEVIRI images, given that the rainfall is at least 1 mm fallen within 30 minutes within the Coastal climatic zone.