ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea Magistrale in Fisica del Sistema Terra

Enhancing Quantitative Precipitation Estimation over Vietnam through Stacked Random Forest Models using Satellite Multispectral Data.

Relatore:
Prof. Federico Porcù

Correlatore:
Dott. Giacomo Roversi

Presentata da:
Daniele Corradini

Anno Accademico 2022/2023

"Not all those who wander are lost. "

Abstract

La precipitazione riveste un ruolo cruciale nel determinare il ciclo idrologico, il clima e le condizioni atmosferiche. Alla luce dei cambiamenti climatici in corso, si stima un'accentuazione degli episodi di piogge intense, con conseguenti sfide in diversi ambiti socio-economici. Il nostro studio ambisce a migliorare l'accuratezza delle stime pluviometriche basate su dati satellitari attraverso metodi di apprendimento automatico, ponendo un'enfasi specifica sulla stagione monsonica in Vietnam.

Mediante l'uso dell'algoritmo Random Forest, abbiamo elaborato un modello che integra dati provenienti dal satellite geostazionario FengYun-4A con quelli dei radar terrestri. Questo approccio sfrutta sia i valori diretti dei canali che le differenze dei canali infrarossi, oltre a combinazioni spazio-temporali di pixel adiacenti, per stimare vari gradi di precipitazione. Abbiamo inoltre utilizzato un modello digitale di elevazione come variabile supplementare, considerando la relazione inversa che esiste tra altitudine e intensità della precipitazione.

L'algoritmo è stato ottimizzato per funzionare sia in condizioni diurne che notturne, tenendo conto della mancanza di dati dai canali visibili durante la notte. Il nostro studio è stato calibrato su una griglia con una risoluzione spaziale di 4 km e una risoluzione temporale di 1 ora, concentrandoci su ottobre 2020, un periodo in cui il Vietnam ha sperimentato piogge di intensità notevole. Per garantire l'affidabilità delle nostre analisi, abbiamo effettuato un'accurata pre-elaborazione e validazione dei dati satellitari.

La nostra metodologia inizia identificando i pixel con e senza pioggia, avanza per classificare varie classi d'intensità di precipitazione e infine effettua una regressione all'interno di ogni classe. Durante la valutazione delle prestazioni, gli output del nostro modello sono stati confrontati con le misurazioni dei pluviometri e altri benchmark regionali di precipitazione usando varie metriche.

Le analisi basate sull'algoritmo Random Forest hanno evidenziato una tendenza del modello a favorire dati spaziali piuttosto che valori di canali isolati. Oltre alle combinazioni di canali infrarossi tradizionalmente citate nella letteratura, il nostro modello sembra prediligere anche alcuni canali visibili durante le ore diurne. I risultati iniziali mostrano che le capacità del nostro modello sono comparabili, se non superiori, ai prodotti pluviometrici ottenuti dai geostazionari FengYun-4A e GeoKompsat-2A già disponibili sul territorio Vietnamita. Nonostante ciò, esistono delle complessità residue: ci sono difficoltà nella classificazione di eventi di pioggia estrema, data la loro bassa frequenza, e le stime di regressione tendono a normalizzare eccessivamente la reale distribuzione delle precipitazioni, avvicinandola ai suoi valori medi.

L'obiettivo principale è creare un modello di stima delle precipitazioni ad alta risoluzione in tempo quasi reale, specializzato per le regioni tropicali frequentemente colpite da tifoni. Integrando osservazioni da terra e dati satellitari, il nostro modello può riconoscere efficacemente le relazioni non lineari tra variabili atmosferiche e spettrali. Tuttavia, per garantire un funzionamento ottimale, il modello attuale necessita di ulteriori perfezionamenti e dati aggiuntivi.

Abstract

Precipitation is pivotal in shaping the water cycle, weather, and climate. With the advent of climate change, extreme precipitation events are anticipated to intensify, presenting challenges across socio-economic realms. This research seeks to bolster the precision of satellite-based precipitation assessments using machine learning techniques, spotlighting Vietnam's wet season.

Leveraging the robust, interpretable, and popular Random Forest algorithm, we have constructed a model that integrates geostationary satellite data from the Chinese FengYun-4A and ground-based radars. This model combines pure channel values, infrared channel differences, spatial patterns from pixel neighbourhoods, and temporal differences to predict precipitation. A digital elevation model was incorporated as an auxiliary variable, given its inverse relationship with rainfall intensity.

Notably, the algorithm adjusts for daytime and nighttime conditions, considering the absence of visible channels at night. We have calibrated our study using a 4 km spatial and 1-hour temporal resolution, with a focus on October 2020—a month marked by intense precipitation in Vietnam. A thorough preprocessing and validation of the satellite data were conducted to ensure the quality of subsequent steps.

Our methodology starts by identifying raining vs. non-raining pixels, advances to classify rain intensities, and finally conducts regression within each class. Performance assessment will compare our model's outputs against rain gauge measurements and other regional precipitation benchmarks using various metrics.

Insights from the Random Forest algorithm revealed a preference for spatial data over singular channel values. In addition to the commonly used infrared channel combinations found in the literature, our model also appears to favour several visible channels during the daytime. Preliminary results indicate our model's performance aligns closely with, if not surpasses, the rain products obtained by the geostationary satellites FengYun-4A and GeoKompsat-2A already available over Vietnam. However, challenges remain, particularly in the classification of extreme rain events due to their rarity, and in regression estimates that tend to diminish the true rainfall distribution around its average values.

The primary goal is to develop a high-resolution estimate model for rainfall in tropical regions, which are often impacted by typhoons. By integrating ground observations and satellite data, machine learning effectively captures nonlinear relationships between atmospheric and spectral variables. However, the existing model demands further enhancements and additional data for optimal performance.

Contents

Introduction

Precipitation plays a pivotal role in the Earth's hydrological cycle, influencing ecosystems, human endeavours, and a myriad of processes that sustain life on our planet [64]. Accurate measurements of precipitation are thus not only of academic interest but bear significant implications for fields ranging from agriculture and water management to public health and disaster mitigation and adaptation [58]. Historically, researchers have relied on traditional methods like rain gauges to measure rainfall. While effective in localized settings, they often falter in providing comprehensive spatial and temporal data [25]. Ground-based radar systems have bridged this gap to some extent, offering more extensive coverage. While radars have proven effective in capturing rainfall data over wider regions, they are not without limitations. Issues such as attenuation during intense precipitation episodes, the inability to capture data from challenging terrains, and occasional discrepancies during light rainfall episodes underscore the need for enhanced methodologies in quantitative precipitation estimation (QPE) [5]. Yet, as weather patterns evolve due to factors like climate change, there is a burgeoning need to refine and expand our precipitation measurement techniques [65].

In the realm of QPE, there exists a pressing need for methodologies that are both comprehensive and precise. Addressing this need, the present research proposes an innovative approach that integrates remote sensing data with machine learning (ML) techniques. This amalgamation offers a holistic, accurate, and adaptable solution for QPE, with potential implications for refined weather predictions, strategic water resource management, and nuanced climate modelling [38]. The central aim of this study is to augment QPE methodologies by synergizing multispectral satellite data within a machine learning framework. By doing so, it seeks to harness the combined strengths of radar and satellite data, mitigating their individual constraints, and producing an output that is superior in accuracy and utility.

The geographical context of this research is Vietnam, a region that bears the brunt of heavy rainfall during its monsoon season, frequently exacerbated by tropical cyclones. This unique climatic condition provides a fitting backdrop for the study's objectives. To lay the groundwork for our research, the initial phase involves the application of the proposed ML model to a dataset from October 2020, a month that was notably characterized by several episodes of heavy precipitation [33].

In this study, inspired by the study of Hirose et al. 2019 [19], we employed a sophisticated stacked model approach, wherein multiple random forest classifiers were sequentially implemented. Initially, these classifiers were tasked with segregating precipitation into distinct categories. Subsequent to this classification, a dedicated random forest regressor was deployed for each distinct class. For the calibration of our models, we utilized radar-based QPE as the ground truth. The feature set was primarily derived from multispectral satellite data, with an emphasis on their spatio-temporal derivations. Additionally, after thorough analysis, we integrated the digital elevation model into our feature set, recognizing its significance given the intricate relationship between precipitation patterns and Vietnam's orography. The

Random Forest framework allowed for a deeper investigation into feature importances, providing clarity on their individual contributions during model training. This understanding was pivotal for the subsequent feature selection. However, it must be noted that the model's performance on unseen data did not entirely meet expectations. We observed certain anomalies, particularly a tendency to converge rain rate distributions around class medians and challenges in effectively classifying extreme precipitation rates due to the rarity of such instances. Yet, upon juxtaposing the aggregated results of all regressions against other prevailing rain products for Vietnam (that were previously validated in the study of Roversi et al. 2023 [59]), our model showcased reasonable fidelity. Notably, it outperformed the geostationary-based products in certain metrics. Considering the nascent stage of our research, these results are promising. There remains ample scope for model refinement, including hyperparameter optimization and the incorporation of a more expansive dataset.

In the subsequent chapters, readers will embark on a comprehensive exploration of the challenges, methodologies, and potential solutions in the realm of quantitative precipitation estimation. Beginning with a short overview in Chapter 1, we shed light on the fundamental characteristics of rainfall, underscore the importance of accurate precipitation estimation, and elucidate traditional methods of QPE. This foundational understanding paves the way for an exploration of ML techniques, culminating in a synopsis of the state-of-the-art in QPE using ML. Transitioning to Chapter 2, we focus on our case study: Vietnam. Detailed descriptions of the region's climate, and datasets utilized—including satellite data, ground reference data, and other rain products—are provided. This chapter further outlines the methodological workflow, from preprocessing steps and feature engineering to the actual ML algorithm setup and performance assessment. Lastly, Chapter 3 presents the empirical results of our research. We initiate with a thorough preliminary analysis, discussing data quality, feature correlations, and more, before delving into the outcomes derived from the Random Forest models. We then wrap up with a conclusive chapter discussing the study's implications and avenues for future research.

Through this structured journey, we aim to offer a comprehensive perspective on ML-based QPE, specifically tailored to the climatic context of Vietnam. By fusing the strengths of satellite remote sensing and ML, this thesis strives to chart a forward-looking course in the continually evolving field of QPE.

Chapter 1

Quantitative Precipitation Estimation

1.1 Overview

1.1.1 Basic Characteristic of Rainfall

According to the World Meteorological Organization (WMO) Guide to Meteorological Instruments and Methods of Observation, precipitation encompasses both liquid and solid water forms that condense from water vapour and descend from clouds or deposit from air onto the ground. Precipitation intensity, typically measured in millimetres per hour (mm/h), denotes the precipitation amount during a set interval. Conversely, rainfall depth, in millimetres (mm), represents the cumulative precipitation over a designated period, reflecting the hypothetical vertical depth of water covering a flat Earth projection [54].

Rainfall is vital in the Earth's hydrological cycle, which circulates water between the atmosphere, land, and oceans. Formed from atmospheric water vapour condensation, warm, moist air rises, cools, and causes water vapour to condense into droplets or ice crystals. These particles amalgamate, growing heavy enough to fall as precipitation [11, 22].

The **microphysics** of rainfall encompasses intricate physical processes like nucleation, growth, and collision-coalescence. Nucleation can be homogeneous (spontaneous droplet formation) or heterogeneous (formation on existing surfaces like dust). Post-nucleation, droplets grow via mechanisms like diffusion or collision-coalescence. The forming rain's nature, whether from liquid droplets or ice crystals, is termed the warm or cold rain process, respectively [11].

Rainfall types include stratiform and convective. **Stratiform** rainfall emerges from extensive moist air layers that cool and result in widespread, consistent precipitation, often linked to large-scale weather systems. In contrast, **convective** rainfall arises from rapid air rises due to instability, leading to intense, localized precipitation, commonly accompanying thunderstorms. Other rain types include orographic (air rising over mountains), frontal (warm meets cold air mass), and monsoonal (seasonal wind reversal regions with distinct wet and dry seasons) [22, 12].

Precipitation displays a diverse range across the globe, influenced by factors such as temperature, atmospheric circulation, and topography. Its **global distribution** manifests distinctly in various regions like the equator, subtropics, mid-latitudes, and polar zones. The equatorial belt, extending approximately 10 degrees on either side of the equator, experiences abundant rainfall. This is attributed to the merging trade winds and the presence of the Intertropical Convergence Zone (ITCZ). Conversely, the subtropics, ranging between 20 to 35 degrees both north and south of the equator, are relatively arid due to dominant high-pressure systems that cause

air to descend. The mid-latitudes, encompassing 35 to 60 degrees from the equator, exhibit fluctuating precipitation patterns influenced by storm systems and weather fronts. Polar areas, from around 60 degrees to the poles, see sparse precipitation because of the prevailing cold, dry air masses. Within the tropics, which combine the equatorial and subtropical belts, rainfall patterns are majorly driven by the ITCZ's seasonal shifts and monsoonal circulations. However, annual variations in precipitation volume and timing can occur due to inherent climate variability and broader climate change trends [17].

1.1.2 The Importance of Precipitation

Precipitation, especially rainfall, is integral to the Earth's **climate system** [17]. It governs the planet's temperature, influences water distribution across its surface, and sustains diverse ecosystems like forests, grasslands, and wetlands, all vital habitats for myriad species [64].

Beyond natural ecosystems, precipitation is pivotal for human endeavours like agriculture, industry, and daily living. It's indispensable for hydroelectric power, fuelling turbines to produce electricity, and for replenishing essential freshwater resources for consumption and industrial activities [64].

Agriculture's reliance on rainfall is evident, as it ensures crop growth and livestock health. Extreme events like heavy rainfalls or floods can devastate agriculture by causing soil erosion, reducing fertility, and inducing waterlogged conditions harmful to crops. Floodwaters can also contaminate the soil, impacting both crops and water quality [58].

Shifts in precipitation patterns, be it long droughts or intense rainfalls, drastically affect ecosystems and human ventures. Droughts can reduce crop yields and intensify wildfires, while intense rainfalls may result in floods and landslides. As the **climate crisis** escalates, variations in these patterns, from heightened precipitation events to extended droughts, become crucial. It underscores the need to fathom precipitation's role in our climate system, urging refined estimations and predictions of its characteristics [65].

Existing rainfall measurement techniques often falter due to sparse monitoring infrastructure, potentially skewing precipitation estimates in regions with complex terrains or variable rainfall patterns. Traditional devices, such as rain gauges and radars, might miss out on capturing the full spectrum of precipitation types or the intensity of extreme events [25]. Additionally, these methods may inadequately track shifts in precipitation patterns stemming from climate change, as they tend to emphasize long-term averages over short-term variations [65]. The upcoming section will explore both conventional and advanced quantitative precipitation estimation (QPE) methods, highlighting their strengths and weaknesses. Recognizing the significance of precipitation in our climate system, nature and human activities underscores the importance of refining QPE techniques.

1.2 Traditional Methods for QPE

The traditional methods for QPE have been developed over many years and include rain gauges, radars, and satellites, that can be grouped into ground-based and spaceborne techniques. In recent years, there has been growing interest in combining multiple methods to improve the accuracy and reliability of QPE [21].

As will be discussed in the following sections, these methods have different principles, advantages, and limitations, and the choice of method for QPE depends on the specific application and the availability of data [62].

1.2.1 Ground-based Sources

Rain Gauges

Rain Gauges (RG) are one of the oldest and most widely used methods for precipitation estimation. These methods rely on the measurement of precipitation at ground-based weather stations that collect precipitation, which is then measured and recorded at regular intervals.

Several RG types are available, with a few significant ones described below (more details in [32]):

- **Storage gauges** are manual devices with a container that collects rainwater for volume measurement over time. Some have automatic emptying systems with siphon mechanisms, though they might miss some rainfall amounts during rapid events.
- **Tipping-bucket** gauges are popular due to their cost-effectiveness and reliability. They employ a balance and two equal buckets; rainfall causes one bucket to tilt upon filling, registering an electrical impulse for measurement. Regular maintenance to prevent funnel blockages is essential.
- **Weighing** gauges measure precipitation using a weight principle. While some use the tipping-bucket mechanism for high-resolution measurements, this can lead to underestimation during the tilting phase and demand more maintenance.
- **Catching-type** gauges use a calibrated nozzle and optical sensor to detect falling raindrop size and frequency. They work best for light rain but may struggle with irregular water fluxes and need regular maintenance.
- **Disdrometers** sense individual hydrometeors without needing a container. They offer data on particle size distribution and fall velocity, requiring minimal maintenance. However, they lag behind traditional gauges in performance due to non-standardized calibration methods.

The use of RG for precipitation estimation has several advantages, including their accuracy, low cost, and ease of maintenance. Additionally, they can be used to validate and calibrate other precipitation estimation methods. However, RGs also have some limitations. For example, they provide point measurements of precipitation that may not be representative of the surrounding area and they may not be available in remote or inaccessible locations. Additionally, they may be subject to measurement errors due to wind, evaporation, and other factors. Despite these limitations, RG remains a valuable tool for precipitation estimation, particularly in areas where other methods are not available or where high-precision measurements are required for specific applications [32].

Radars

Radar, specifically RAdio Detection And Ranging, is a pivotal tool in QPE. While traditional RG offers point measurements, they lack broad spatial representation. In

contrast, weather radars offer comprehensive spatial and temporal insights from a single site, enabling precise rainfall measurements across vast scales.

Polarimetric Doppler radar, a recent advancement, captures data in both horizontal and vertical polarizations, offering richer details on precipitation particles. These radars now oversee most populated regions globally, aiding in tasks from flood prediction to climatology. Moreover, they play a crucial role in refining satellite precipitation data and fostering algorithm innovations.

To interpret precipitation using radar, it's crucial to understand how electromagnetic signals from the radar interact with raindrops. This interaction is largely governed by the drop size distribution (DSD), which gives the average count of raindrops per size interval for every unit volume of air. The radar reflectivity factor, denoted by Z, is derived from the DSD and is defined by:

$$Z = \int N(D) D^6 dD \tag{1.1}$$

where N(D) represents the number of droplets with diameter D. The rain rate R, when accounting for DSD and a raindrop's terminal fall speed, can be connected to Z using empirical relationships like the Marshall and Palmer relation:

$$R = aZ^b \tag{1.2}$$

Here, a and b are coefficients which might change depending on the geographical location and season [41].

Weather radar systems, commonly mounted on towers to avoid local obstructions, employ a directional antenna encased in a radome. These systems emit electromagnetic pulses that scatter upon hitting meteorological targets. Some scattered energy reflects back as a radar echo. By incrementing the antenna's elevation angle with each rotation, the system captures volumetric atmospheric observations. The returned power provides precipitation insights within the resolution volume, influenced by precipitation particle types and their DSD. Radar resolution is determined by pulse length and antenna beam widths. Pulse volume positioning depends on antenna location, elevation angle, target range, and the radar beam's propagation path. As electromagnetic waves bend due to the atmosphere's vertical stratification, their trajectory can be predicted through temperature, moisture, and pressure profiles. Ground-based radars typically use S, C, or X frequency bands, each offering unique advantages. While S-band detects distant heavy rain, X-band is hydrometeor-sensitive but faces signal attenuation. C-band strikes a balance in reflectivity measurements, range, and cost.

Radar has several advantages, including its ability to provide high-resolution measurements of precipitation over large areas. Additionally, it can provide information on the spatial and temporal distribution of precipitation, as well as the intensity and type of precipitation. Radar can also be used in real-time for nowcasting and short-term forecasting of precipitation. However, there are some drawbacks to using this method. One major issue is the potential for large observation errors to occur, which can be caused by a variety of meteorological, topographic, and technical factors. These factors include the variability of the Z-R relationship, errors related to the range of the radar signal due to the vertical variability of reflectivity and incomplete or non-uniform beam filling, attenuation caused by intervening precipitation, ground clutter, beam blockage, and miscalibration of electronic components. Additionally, radar-based QPE may not accurately estimate the intensity of light precipitation, such as drizzle, and may underestimate or miss convective precipitation

events [5].

1.2.2 Spaceborne Sources

Satellites offer a formidable advantage in QPE with their global reach and frequent observations. They employ both **passive** and **active** sensors; the former captures Earth's natural radiation while the latter transmits signals to Earth, recording their reflection or scattering. Since the debut of meteorological satellites in 1960, many have observed Earth, yet not all target precipitation specifically. Despite this, the meteorological community has devised methods to extract precipitation data from varied satellite readings. Maximizing data availability involves using Geostationary (GEO) satellites with **visible** (VIS) and **infrared** (IR) sensors, and **microwave** (MW) sounders on Low Earth Orbit (LEO) satellites.

Satellite-based QPE plays a pivotal role in diverse **applications** like flood and weather forecasting, climate studies, and monitoring extreme weather events, particularly in regions underserved by ground measurements, such as the tropics and polar areas. Its **benefits** include global coverage, frequent rain monitoring, insights into precipitation patterns, types, intensities, and accessibility in remote areas.

However, it's not without **limitations**. Factors like cloud interference and variability in precipitation particles can introduce errors. Light drizzles might be underestimated, and convective events might be inaccurately portrayed. Enhancing the precision of satellite-based QPE, especially for subtle rainfall events and in complex terrains, remains a research priority. Validating satellite precipitation retrieval against surface data is crucial for its accuracy [24].

GEO Satellites

GEO satellites provide frequent and regular images over a full disc of the Earth from their location around the Equator. A constellation of about five GEO satellites is needed for quasi-global coverage. The GEO suite of sensors provides consistent observations every 30 minutes with a resolution of about 4 km around the Equator (60°N-60°S). Nowadays, better resolutions are possible in some sensors. Meteorological satellites first captured simple VIS/IR images of clouds to identify precipitation-bearing systems. VIS imagery has the highest spatial resolution among remote sensing techniques, but its potential products are constrained to daytime use only. As a result, retrieval schemes have primarily focused on IR imagery. IR-based techniques assume that clouds with lower cloud-top temperatures are taller and thicker, and thus more likely to produce precipitation. However, the relationship between cloud-top temperature and precipitation near the Earth's surface is indirect and influenced by regional variations and the cloud life cycle. Besides cloud top temperature, multi-spectral VIS/IR techniques can provide information on cloud top height, pressure, particle sizes, and phase [42, 43]. Although VIS/IR observations only provide information on cloud-top properties, they offer frequent observations and good spatial resolution. Multi-spectral channels may improve retrieval of cloud-top characteristics, but the indirectness of cloud-top properties to surface precipitation persists [24].

LEO Satellites

Direct observations of precipitation are possible using MW radiometers, which are currently only deployed on LEO satellites due to the large antenna required for good

spatial resolution. These measurements are available from about 10 LEO-based sensors, providing less frequent observations with coarser spatial resolutions with respect to GEO-based sensors. They provide typically two observations per day per sensor at the same location.

Passive MW (PMW) sensors employ imaging or sounding frequencies, with conically-scanning instruments favoured for precipitation retrievals. These sensors detect atmospheric liquid water and ice. Lower frequencies (up to 40 GHz) are more attuned to liquid hydrometeors, while higher frequencies (>40 GHz) detect ice particles. Over oceans, brightness temperatures (Tb) facilitate precipitation retrievals. Over land, the use of high-frequency channels is necessitated by high emissivity. While ice scattering reduces Tbs in intense events, not every cloud has ice, and the upper cloud ice may not reliably indicate surface rainfall. Physical retrieval schemes offer insights into surface precipitation and vertical profiles. However, PMW sensors have resolutions inferior to IR sensors, necessitating multiple satellites for comprehensive precipitating system analyses [24].

Active MW (AMW) sensors emit MW energy pulses that backscatter from precipitation, yielding a vertical precipitation profile. Despite their precision, their limited numbers and narrow swath width compromise temporal sampling. Examples include the Dual-frequency Precipitation Radar (DPR). While they deliver direct precipitation measurements, their periodicity and range restrict utility. Yet, they're instrumental in data generation for PMW retrieval and calibration/validation, offering valuable three-dimensional data for understanding precipitation systems [24].

1.2.3 Multi-Source Precipitation Products

While no single satellite precipitation product can comprehensively quantify surface precipitation, given varying user needs, integrating data from multiple sources can bolster accuracy. Merging disparate satellite observations primarily aims to refine both the temporal and spatial accuracy of precipitation readings.

Global initiatives have advanced precipitation satellite missions, leveraging space-based technology and integrating multiple data sources [68]. Notably, the Tropical Rainfall Measuring Mission (TRMM) (1997-2015) was a US-Japan venture employing cutting-edge sensors to scrutinize tropical rainfall patterns. More details can be found at `https://gpm.nasa.gov/missions/trmm`. Building on TRMM's legacy, the Global Precipitation Measurement (GPM) (initiated in 2014) enhances global precipitation tracking capabilities, playing an integral role in weather and climatic endeavours. Further information is available at `https://gpm.nasa.gov/missions/GPM`.

LEO and GEO **satellite** systems are complementary and often used together to gather precipitation data. In fact, no single satellite precipitation product can fully capture the true surface precipitation, so combining multiple observations is seen as a way to improve precipitation products. The PMW precipitation products tend to be infrequent but more direct, while the IR products are less direct but more frequent. Furthermore, the TRMM and GPM satellites used active and passive observations to reduce ambiguities by combining radar-radiometer observations in algorithms to improve precipitation measurements. Currently, there are few primary operational merged products that provide surface precipitation information at a scale of approximately 10 km per 30 minutes or better. A prominent example, which will also be utilized in this study, is NASA's Integrated Multi-satellite Retrievals for GPM (IMERG). These schemes have the potential to provide a continuous record of precipitation over time, but the generation of such products is complex and relies on the accuracy of the original PMW retrievals [11, 68].

Many satellite precipitation products also incorporate conventional **surface data** to correct for any systematic bias in the satellite estimates, and newer techniques have started to incorporate model information into their combined precipitation products. However, all combination schemes are only as good as their components, and large uncertainties can still exist in certain regions and situations. Satellite precipitation products often incorporate conventional surface data to correct bias and ensure consistency. For instance, the IMERG products combine precipitation estimates from multiple satellites and gauge analyses to produce global precipitation products. These products have been widely used for various applications, including weather forecasting, flood prediction, and water resource management [24].

Not only satellite retrievals are enhanced using multi-source products. The merging of rain gauge networks and radar measurements is done to improve the accuracy of rainfall estimates. Rain gauges measure precipitation at a single point, while radars provide volumetric measurements of precipitation over a large area. By combining the two, a more accurate representation of precipitation can be achieved. However, there are challenges in merging the two datasets due to differences in spatial and temporal scales and biases in the measurements themselves [5].

To integrate data from various sources like rain gauges, radar, and satellites, one approach involves using **statistical methods**. This includes techniques such as bias correction, Kriging-based approaches, linear regression, geographical difference analysis, Bayesian combination, Kalman filter calibration, and geographically weighted regression. For instance, methods like weighted averaging or Bayesian merging allocate weights to each QPE based on its reliability and accuracy, resulting in a unified precipitation estimate. However, considering the uncertainties in each QPE and the combination process is crucial. These uncertainties can stem from measurement errors, sampling discrepancies, or interpolation issues. Thus, it's essential to both quantify these uncertainties and factor them into the integration process to ensure dependable and precise QPE outcomes [11, 68].

1.3 Machine Learning Techniques for QPE

Conventional QPE methodologies, encompassing gauge, radar, and satellite-based techniques, often encounter challenges in accuracy and dependability, especially in intricate terrains or during severe weather patterns. Their inherent limitations arise from reliance on mathematical relationships and stringent assumptions.

With the capability to discern patterns among multiple estimates and formulate a unified product, **machine learning** (ML) offers a promising avenue to bypass some of the restrictions of traditional approaches. Recently, the application of ML for QPE has gained traction, consistently demonstrating enhancements in the precision and trustworthiness of precipitation evaluations. Distinctly advantageous, ML can adeptly handle complex, nonlinear relationships without the need for detailed statistical models. It's adaptive, able to learn from historical datasets and adjust to evolving scenarios. Coupled with its prowess in efficiently crunching vast data volumes, ML assures notable computational agility [38].

ML techniques, a subset of artificial intelligence (AI), possess the ability to learn from data autonomously, enabling them to predict or decide without predefined programming. Their applications span across diverse sectors, including meteorology, hydrology, and climate science, with considerable promise evident in the domain of QPE. This section intends to furnish a comprehensive look into the ML

techniques tailored for QPE. We will delve into the foundational principles, enumerate their strengths and shortcomings, and spotlight specific case studies where they have been instrumental for QPE.

1.3.1 Brief Introduction to ML

The methods based on ML involve teaching a computer to recognize patterns in data and make estimates or decisions based on those patterns. There are two main types of machine learning: supervised learning and unsupervised learning. In **supervised learning**, the algorithm is trained on a labelled dataset, where the correct output is provided for each input. The goal is to learn a mapping between the input and output so that the algorithm can make accurate estimates on new, unseen data. In **unsupervised learning**, the algorithm is trained on an unlabelled dataset, where the correct output is not provided. The goal is to learn the underlying structure of the data and identify patterns or groupings. The ML models can also be divided into parametric and non-parametric. **Parametric models** make assumptions about the distribution of the data and have a fixed number of parameters, while **non-parametric models** do not make any assumptions about the distribution of the data and have a flexible representation of the model. Parametric models are typically simpler and more computationally efficient, while non-parametric models are more flexible in capturing complex relationships in the data but can be computationally expensive and prone to overfitting [4].

ML has been increasingly applied in various fields, especially in QPE, since traditional methods are limited or unreliable. ML algorithms can be used to estimate precipitation based on a variety of input data sources, including satellite imagery, weather radar, and numerical weather prediction models. Moreover, it can also be used to improve the accuracy of precipitation estimates by identifying patterns and relationships between different variables, such as temperature, humidity, and topography. Commonly used supervised ML algorithms for QPE are [4]:

- **Artificial Neural Network** (ANN) is a commonly used ML technique for QPE. It's a supervised learning algorithm inspired by the human brain's structure and function. It's composed of interconnected neurons organized into layers that process input data. The activation function determines the neuron's output based on its inputs and bias term. During training, the network adjusts the weights and biases of each neuron to minimize the difference between its output and the desired output using backpropagation. ANNs are known for their ability to learn complex patterns from data and achieve high accuracy. However, they can be computationally expensive and challenging to interpret in some applications.

- **Support Vector Machine** (SVM) is a type of supervised machine learning algorithm used for classification and regression tasks. SVMs find the hyperplane that best separates the data into different classes or predicts the target variable in regressions. They can handle non-linearly separable data by using a kernel function. SVMs are known for their excellent generalization performance but can be computationally expensive for large datasets and sensitive to hyperparameter tuning.

- **Decision Trees** (DT) are a type of supervised machine learning algorithm used for classification and regression tasks. They use a tree-like model to represent decisions based on input data features. The tree is constructed by recursively

splitting the data into smaller subsets based on the value of a selected feature to maximize information gain or minimize impurity. DTs are easy to interpret and visualize but can be prone to overfitting. Ensemble methods like Random Forest (RF) and Gradient Boosting (GB) are often used to improve their performance. **Random Forest** combines multiple trees trained on different subsets of the data and features, while **Gradient Boosting** builds trees sequentially to correct errors. Random Forest is typically used for interpretability and speed, while Gradient Boosting is used for high accuracy.

- **Logistic and Linear Regression** are both types of supervised machine learning algorithms used for regression and classification tasks. Linear Regression is used for predicting a continuous output variable, while Logistic Regression is used for predicting a binary output variable. Both use a cost function and optimization algorithm to adjust their parameters during training.
- **K-Nearest-Neighbour** (KNN) is a non-parametric supervised ML algorithm used for classification and regression tasks. It classifies a new data point based on the class that appears most frequently among its K nearest neighbours in the training set. KNN is simple and easy to implement but can be sensitive to the choice of the distance metric and the number of neighbours. It can be useful in applications where a simple and interpretable model is desired and when the data lacks a clear linear separation.

There are many examples of QPE products that utilize ML algorithms to improve the accuracy and reliability of precipitation estimates. Each product has its unique strengths and weaknesses, and the choice of which product to use depends on the specific needs and applications of the user. For example, **IMERG** uses a combination of satellite data and gauge observations to estimate precipitation on a global scale. IMERG uses a combination of physical models and machine learning algorithms to improve the accuracy of its estimates. **PERSIANN-CCS** - Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System - is a QPE product developed by the University of California, Irvine. It uses a combination of satellite data and ML algorithms to estimate precipitation. PERSIANN-CCS is known for its high accuracy and ability to estimate precipitation in regions with limited gauge observations. **CHIRPS** - The Climate Hazards Group InfraRed Precipitation with Station data - is a QPE product developed by the Climate Hazards Group at the University of California, Santa Barbara. It uses a combination of satellite data and gauge observations to estimate precipitation on a global scale. CHIRPS uses ML algorithms to interpolate gauge data and fill in gaps in the satellite data, improving the accuracy of its estimates [68].

Advantages and Limitation of ML methods

Using ML in QPE can lead to many advantages [38]. ML algorithms can learn patterns and relationships between different variables in precipitation data that may be difficult or impossible for humans to identify. As a result, ML algorithms can often provide more **accurate** precipitation estimates than traditional methods. ML algorithms can process large amounts of data quickly and **efficiently**, making it possible to analyze data from multiple sources and generate precipitation estimates in near-real-time. They can be applied to a variety of input data sources, including satellite imagery, weather radar, and numerical weather prediction models. This **flexibility** allows for more comprehensive and accurate precipitation estimates, even in regions

where traditional measurement methods are limited or unreliable. ML algorithms can **adapt** to changing precipitation patterns and weather conditions, allowing for more accurate and timely precipitation estimates in both short-term and long-term forecasting.

However, ML models still have some challenges to be addressed [4]. The accuracy of ML algorithms depends heavily on the quality of **input data**. If the input data is incomplete, inaccurate, or biased, the ML algorithm may produce inaccurate or unreliable precipitation estimates. ML algorithms require large amounts of data to train and validate the model, which can be a challenge in regions where data is limited or difficult to obtain. Additionally, the complexity of their operations can hinder the identification and correction of model errors or biases. ML algorithms often lack **interpretability**, making it challenging to discern the reasoning behind a specific precipitation estimate. When the model is overly complex and fits the training data too closely, the model can be prone to **overfitting**, which can lead to poor performance on new, unseen data. Lack of **standardization** in ML can yield biased models, especially when trained on mismatched data sources. Ensuring data consistency often requires intensive preprocessing. For optimal QPE results, a unified guideline for data handling and model tuning is essential. **Reproducibility** is critical in ML because it allows others to verify the results and test the validity of the model. However, reproducing the results of an ML model can be challenging due to the complexity of the models, the randomness in the training process, and the use of different hardware and software environments. Reproducibility can be improved by documenting the code and using publicly available datasets in the training step [9].

Despite these limitations, ML has the potential to significantly improve precipitation estimation accuracy, particularly in regions where traditional measurement methods are limited or unreliable. As a result, researchers and practitioners are increasingly turning to ML algorithms to improve QPE accuracy and reliability [38].

1.3.2 State of the art

Data Sources

We begin by examining the diverse **data sources** referenced in prior research for QPE. The choice of data is often influenced by desired spatial-temporal resolutions and data availability.

- **Satellite-Based Data**: Studies have leveraged IR, Near-IR, and VIS channels from satellites such as MSG SEVIRI, Himawari-8 AHI, Fengyun-4A AGRI, and GOES-16. The appeal of these sources lies in their fine temporal and spatial granularity, enabling regional precipitation estimates [2, 16, 19, 48, 49, 60, 61, 70, 3, 44, 56, 8, 13].
- **MW Channels**: Deployed on LEO satellites like GMI on GPM, Aqua AMSR-E, and others, MW channels detect precipitation particle size and shape, facilitating assessments of precipitation intensity and type [8, 35, 36, 57, 10].
- **Radar-Based Data**: Ground-based radar and DPR have furnished insights on precipitation's intensity, verticality, and movement. Both ground-based and space-borne radar data have been harnessed as validation benchmarks for ML models [13, 63, 45, 19, 10].

- **Rain Gauges & Disdrometers**: Renowned for pinpoint accuracy, rain gauges offer localized precipitation measurements. Disdrometers, meanwhile, elucidate the size distribution of precipitation particles. Both have played a pivotal role as ground truths in ML model validation [13, 29, 48, 49, 61, 60, 70, 63, 56, 34, 45, 47].
- **Precipitation Products**: The literature references products like IMERG, CHIRPS and PERSIANN-CDR for QPE [29, 3, 67, 28, 47]. Derived from satellite and ground-based observations, these products offer varying spatial-temporal precipitation estimates. Notably, some have been utilized as benchmark data in ML model training and evaluation [70, 57, 35, 67, 36].
- **Soil Moisture Products**: Satellite-based measures of soil water content serve as indirect indicators of precipitation in some studies [29].
- **Forecast & Reanalysis Data**: Products like GFS, ERA, and MERRA, which present also non-rain parameters like temperature and humidity, aid in understanding precipitation-friendly atmospheric conditions [70, 10, 34].
- **Digital Elevation Models** (DEM): Providing insights on regional topography, DEMs are useful in pinpointing probable precipitation zones [47, 34, 67, 35, 3].

Spatial and Temporal Information

In QPE studies, **spatial resolution** significantly influences the accuracy and reliability of precipitation estimates. Different data sources often lead to varying spatial resolutions. Addressing this, some researchers have unified the data by reprojecting maps onto a standard grid using techniques like cubic spline interpolation [34, 28], bilinear resampling [44], and averaging [67, 35, 61, 8]. The resulting resolutions typically span 3 to 6 km [13, 8] or 0.05° to 0.25° [29, 19, 70]. An outlier in these studies is [35], opting for a coarser 15 and 45 km resolution.

Another key consideration is **parallax correction**, which can impact QPE's precision. Various methods have been employed to adjust for parallax, including referencing cloud top height [3] and aggregating radar pixels—5x5 for higher clouds and 3x3 for others [49, 61]. However, many studies overlook the mention of parallax correction.

Lastly, **study areas** in QPE research exhibit a broad range: from specific regions [13, 29, 28, 47] to whole countries [67, 8, 16, 48, 57, 35, 44, 31, 45], and even broader areas [70, 39, 3].

In QPE studies, the chosen time intervals critically shape the estimates' accuracy and reliability. Studies have explored varied **time intervals**, from specific seasons like the rainy season [13, 48, 60, 70, 3, 36, 39, 46], a single year [44, 8], multiple years [48, 10, 35, 28, 47, 16], to particular rainfall events or non-consecutive days [34, 57, 45].

Temporal resolution varies: some studies capture data in sub-hourly intervals, such as 10 or 30 minutes [2, 8, 13, 19], others opt for hourly [35, 34, 45] or daily resolutions [29, 28, 47]. Moreover, models often differentiate based on time, discerning between day and night [61, 60, 44, 31, 34, 56], or seasonally, like summer versus winter [8].

Samples and Classes

In QPE studies, **preprocessing** is pivotal as the quality of data samples directly affects outcomes. Some studies pruned the sample pool due to data quality, missing entries, or a pixel-per-scene threshold [8, 3, 63, 16, 35, 44]. Techniques like rain or cloud masking [44, 16, 70, 35, 67], random undersampling [10], and grayscale level reduction [44] have been employed for computational optimization. Data normalization [3, 36] and cross-validation during training or hyperparameter adjustment are also common procedures [61, 63, 34, 45, 44, 36].

For classification in QPE, varied **classes** are used. Some studies differentiate simply between rain and no rain [48, 67], others between stratiform and convective rain [61, 10]. Rain rate categories range from general classifications like light to heavy [29, 49, 70] to detailed divisions based on radar reflectivity intervals [2, 48]. To tackle class imbalances, approaches like SMOTE oversampling [70, 46], random undersampling [13, 3, 10], and bootstrap resampling [10] have been used.

Features

In QPE studies, **satellite-based features** predominantly inform precipitation estimates. Researchers commonly use **spectral** data like Tb, reflectances, and related derivatives from IR, Near-IR, water vapour, and VIS channels [2, 8, 16, 19, 13, 48, 49, 70, 3, 67, 44, 56, 34, 39, 31, 46], as well as MW channels [8, 57, 10, 35, 36]. From spectral data, various **cloud** attributes—like water path, type, mask, temperature, and texture—were extracted [16, 48, 61, 60, 3, 34, 39, 31, 46]. For texture analysis, features such as homogeneity and contrast were derived from the grey-level co-occurrence matrix, and entropy from the grey-level difference vector [61, 60, 44, 56, 39]. Soil moisture [29] and surface emissivity [10] also served as satellite-based predictors in some studies.

Furthermore, features span a wide range of sources to train ML models. From **Rain gauges**, metrics like rainfall intensity and the proximity of gauges to a given pixel are utilized [13, 47]. Data from **radar**, such as Surface Rain Intensity and Reflectivity at various elevations, offer additional insights [13, 45]. Combined **precipitation products** also contribute to the feature set [29, 28, 47].

Geographic information, which encompasses land-water differentiation, elevation, and various topographical indices, is integrated into the models [3, 35, 67, 34, 47]. **Spatiotemporal** features, such as the solar zenith angle and localized time and positioning data, play a role in these studies [8, 13, 70, 3, 10, 35, 34]. Additionally, **Weather** parameters, ranging from temperatures at various altitudes to metrics indicating atmospheric stability, enrich the ML models [70, 10, 34, 46].

ML Models

In QPE research, ML models cater to both classification and regression challenges. For **classification**, models such as logistic regression [3, 10], K nearest neighbors [2], multilayer perceptron [8, 48, 61, 3, 10, 44], and others including random forest and support vector machine [19, 70, 10, 60, 61, 67] have been deployed. For **regression**, tools such as K nearest neighbor [28, 45], linear regression [49, 3, 28], and several others, including multilayer perceptron and gradient boosting [48, 57, 3, 44, 36, 28, 45, 39, 36], have been utilized.

While many researchers have embraced **hyperparameter tuning** using methods like grid search [57, 36, 28], others have opted for default ML function parameters.

Moreover, **feature selection** plays a crucial role in enhancing model efficiency. Approaches include feature ranking [67, 35, 63, 46], selective feature utilization [44], inter-feature correlation assessments [36], and principal component analysis [39].

Performance Assessment

Performance in QPE models is assessed using a variety of **metrics** tailored for both classification and regression tasks. Classification metrics encompass a wide range, from the probability of detection and false alarm ratio to more specialized metrics like the equitable threat score and Area Under the Curve of the Receiving Operating Curve. Similarly, regression tasks might involve the use of simple metrics like the correlation coefficient, and more advanced ones, such as Kling-Gupta efficiency, among others.

The actual **performance** of these QPE models doesn't remain limited to a singular context. Various conditions are tested, such as: 1) Different altitudinal conditions [67]. 2) Surface variations like terrestrial, aquatic, or coastal zones [36, 8, 46]. 3) Different climatic zones [28]. 4) Locale variations, such as urban vs. rural settings [34].

The evaluation often includes analyzing accumulated rain rates over different timeframes, ranging from short durations like 3 hours to longer ones, like a month. This aids in comparing rain gauges and observing temporal trends [48, 61, 67, 56]. To test resilience, some studies have incorporated Gaussian noise into their datasets [36]. Unique approaches, like calibrating the probability density function of model outputs, have been implemented in some research [39]. Rain rate retrieval can also be categorized into various classes for a more detailed analysis [29]. Lastly, a benchmarking approach is common, where new models are pitted against previously established ones from the literature to ascertain their relative efficacy.

Chapter 2

Materials & Methods

2.1 Case Study Description

Vietnam is a long, narrow country in Southeast Asia, stretching over 1,600 kilometres from north to south. It has a diverse climate and geography, with many ecosystems and natural resources. The study specifically focuses on the wet season, as this is the time when the country is most vulnerable to heavy rainfall.

2.1.1 Vietnam's Climate

Vietnam has a unique geographical and climatic profile, situated between two prominent monsoon regions: the South Asian and East Asian monsoons. Stretching from low-lying plains to mountainous terrains, the country experiences a tropical monsoon climate characterized by seasonal shifts in weather patterns. The land is abundant with water resources, including nine major river systems and two significant gulfs, making hydrology a key issue for the region. During the summer months, the South Asian monsoon dominates the climate, bringing hot and wet conditions, especially to the southern parts of the country. This monsoon is influenced by various atmospheric pressure centres like the Asiatic Low and the Australian High. These conditions contribute to tropical cyclones, a common hazard during this season. On the other hand, the winter months, stretching from November to March, are affected by the cold and dry East Asian monsoon. Cold surges originating from the Siberian High significantly lower temperatures in Vietnam, particularly in the northern and central regions. The rainy season usually starts in May and ends in October, although there are variations depending on the region within Vietnam. Overall, the rainy season contributes to over 80% of the annual rainfall in the country [51].

In Figure 2.1, the right-hand plot distinctly illustrates the annual mean climatological patterns for temperature and precipitation. The data reveals that the wet season, spanning from May to October, coincides with the period of highest mean temperatures, ranging between 25 to 27.5°C. Conversely, the dry season from November to April experiences the lowest mean temperatures, falling within a range of 20 to 25°C.

The left-hand plot of Figure 2.1 provides a comprehensive visualization of Vietnam's diverse climate regions using the Köppen-Geiger climate classification. The northern areas predominantly feature a temperate climate characterized by dry winters and hot summers, though higher altitudes experience milder summers. The central lowlands primarily exhibit a tropical climate, notably monsoonal and, to some extent, rainforest conditions. In contrast, the southern regions are largely influenced by a tropical monsoon climate, with pockets of temperate conditions corresponding to the highlands.

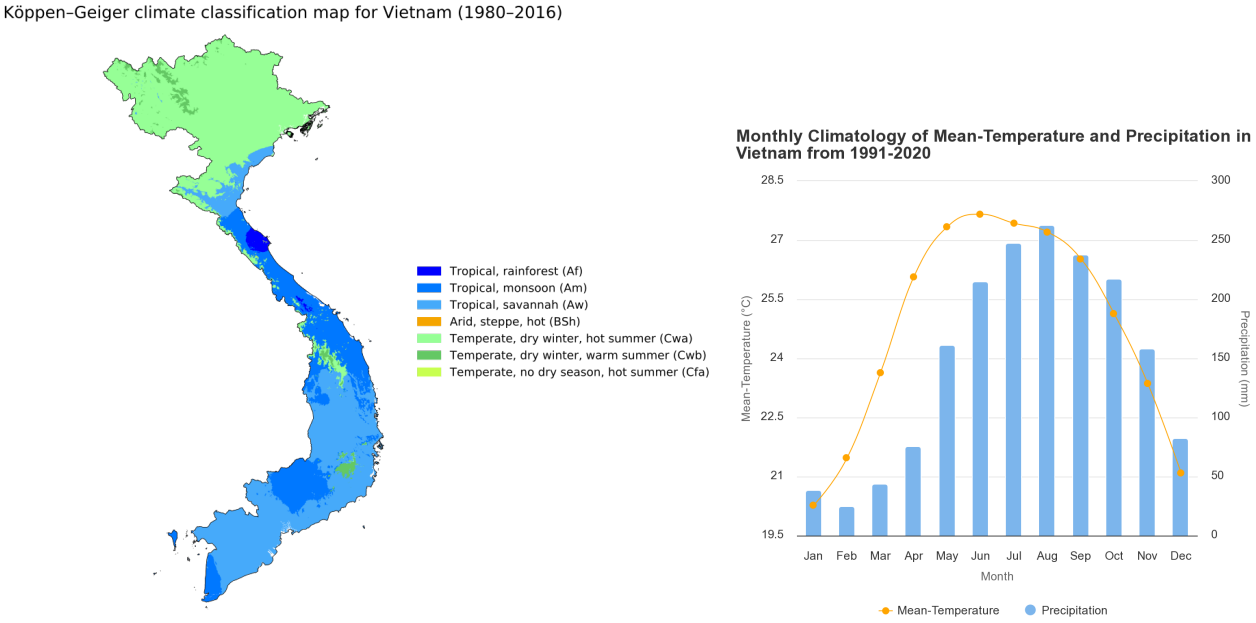

FIGURE 2.1: **Left**: Köppen-Geiger climate classification map for Vietnam (1980-2016) [1]. 1-km resolution climate maps are available here: `https://www.gloh2o.org/koppen/`. **Right**: Monthly Climatology of Mean-Temperature and Precipitation in Vietnam from 1991-2020. Image taken from `https://climateknowledgeportal.worldbank.org/country/vietnam`.

2.1.2 Study period

The year 2020 was marked by an unprecedented frequency of tropical cyclones affecting Vietnam. A total of ten major cyclonic events—Sinlaku, Noul, Linfa, Nangka, Saudel, Molave, Goni, Atsani, Etau, and Vamco—impacted the country (see Figure 2.2 for chronological order and Figure 2.3 for their respective tracks). Each of these storms was remarkable for its intensity, accompanied by exceptional levels of precipitation and elevated wind speeds. The month of October, in particular, witnessed a surge in severe meteorological incidents with devastating humanitarian and financial impacts [37].

In October 2020, the province of Thua Thien Hue and other regions in central Vietnam were severely affected by a sequence of tropical cyclones, including Linfa, Nangka, Ofel, Saudel, and Molave. This spate of storms caused unparalleled flooding and landslides. According to the Vietnamese Center for Disaster Prevention and Search and Rescue, tropical storm Linfa established a new rainfall record, with accumulations reaching up to 2,290 mm in Thua Thien Hue [52]. Notably, this extreme meteorological activity culminated in the loss of over 30 lives, the submersion of hundreds of thousands of residences, and financial damages estimated at $86.29 million [15]. Typhoon Molave was the fourth event in this series, with winds of approximately 161 km/h. This resulted in extensive structural damage, including the destruction of 90,000 homes, and also triggered several landslides. During this period, the rainfall levels in central Vietnam were without precedent; Danang recorded a total monthly rainfall of approximately 1.52 meters, while Hue reported an even more staggering 2.61 meters [33]. At the end of October, Super Typhoon Goni made landfall, albeit in a weakened state, becoming the fifth cyclone to impact Vietnam

within a single month. Despite its diminished intensity, Goni caused unusual flooding in coastal cities like Nha Trang and Quy Nhon, thereby worsening the existing hydro-meteorological crisis [33].

Given the unique meteorological characteristics exhibited during this period, it serves as an invaluable case study for exploring the challenges of QPE in tropical cyclones affecting Vietnam. Consequently, the training dataset for the random forest model was deliberately selected to span the period from October 1 to October 26, 2020. The testing dataset covers the interval from October 27 to October 31, 2020, as delineated by the grey boxes in Figure 2.2.

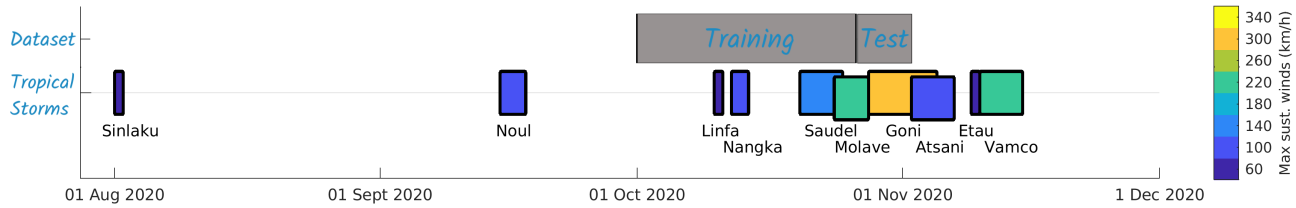

FIGURE 2.2: Coloured boxes show the chronological sequence of tropical storms impacting Vietnam in 2020, each annotated with its maximum sustained wind speed. The grey boxes within the figure denote the divisions between the training and testing datasets used for the random forest models.

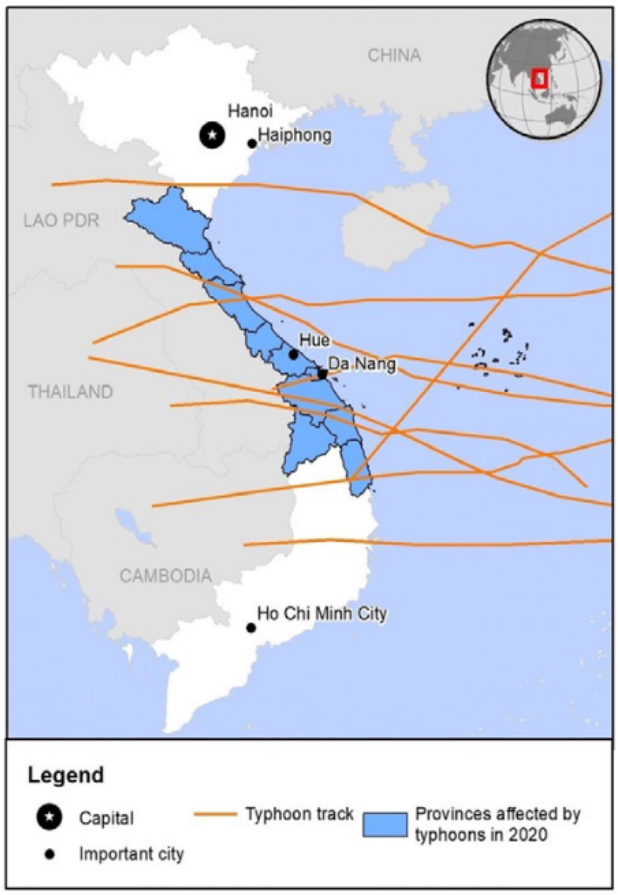

FIGURE 2.3: Typhoon tracks affecting Vietnam in 2020 [55].

2.2 Datasets

In this section, we present the datasets utilized for the analysis. Firstly, we illustrate the geostationary satellite data employed for QPE in the Vietnam region. The satellite data underwent calibration using radar data, which served as our ground truth. Lastly, we validated the ML model against pluviometer data, considered the gold standard, and compared its accuracy with other rainfall products.

2.2.1 Satellite Data

The Chinese Fengyun-4A (FY-4A) satellite was chosen for its alignment with Vietnam's longitude, notably minimizing parallax error compared to Hanoi's longitude of 106°E. Despite similar capabilities across regional geostationary satellites, FY-4A's position at 104.7°E negates the need for parallax correction, ensuring more accurate data for our analysis.

The sensor employed for precipitation estimation is the Advanced Geosynchronous Radiation Imager (AGRI). As detailed in Table 2.1, AGRI is capable of measuring radiances across 14 distinct channels, ranging from 0.47 μm to 13.5 μm. The spatial resolution varies from 0.5/1 km for the VIS and near-IR (NIR) channels to 2–4 km for the IR channels. Each channel is designed to retrieve specific types of information, ranging from cloud and vegetation coverage to water vapour and surface temperature.

The first six channels are categorized as part of the primary group of Visible channels, given that their units are specified in Reflectance (%). It's worth noting, however, that channels 4, 5, and 6 actually belong to the shortwave IR spectrum. Channels 7 to 14 are IR channels, expressed in Tb (K), encompassing mid-wave, longwave IR, and water vapor (WV) channels.

Channel	Type	Central wavelength	Bands	Spatial resolution	Main purpose
1	VIS/NIR	0.47 μm	0.45∼0.49μm	1 km	Aerosol
2	VIS/NIR	0.65 μm	0.55∼0.75μm	0.5-1 km	Fog, Cloud
3	VIS/NIR	0.825 μm	0.75∼0.90μm	1 km	Vegetation
4	Short-Wave IR	1.375 μm	1.36∼1.39μm	2 km	Cirrus
5	Short-Wave IR	1.61 μm	1.58∼1.64μm	2 km	Cloud, snow
6	Short-Wave IR	2.25 μm	2.1∼2.35μm	2-4 km	Cirrus, aerosol
7	Mid-Wave IR	3.75 μm	3.5∼4.0μm (high)	2 km	Fire
8	Mid-Wave IR	3.75 μm	3.5∼4.0μm (low)	4 km	Land surface
9	WV	6.25 μm	5.8∼6.7μm	4 km	High level water vapour
10	WV	7.1 μm	6.9∼7.3μm	4 km	Middle level water vapour
11	Long-Wave IR	8.5 μm	8.0∼9.0μm	4 km	Water vapour, cloud
12	Long-Wave IR	10.7 μm	10.3∼11.3μm	4 km	Surface temperature
13	Long-Wave IR	12.0 μm	11.5∼12.5μm	4 km	Surface temperature
14	Long-Wave IR	13.5 μm	13.2∼13.8μm	4 km	Cloud thickness

TABLE 2.1: Main characteristics of all the channels available in FY-4A AGRI sensor. Information is taken from the official website of the National Satellite Meteorological Center: `https://fy4.nsmc.org.cn/nsmc/en/instrument/AGRI.html`. VIS: Visible; NIR: Near Infrared; IR: Infrared; WV: Water Vapour.

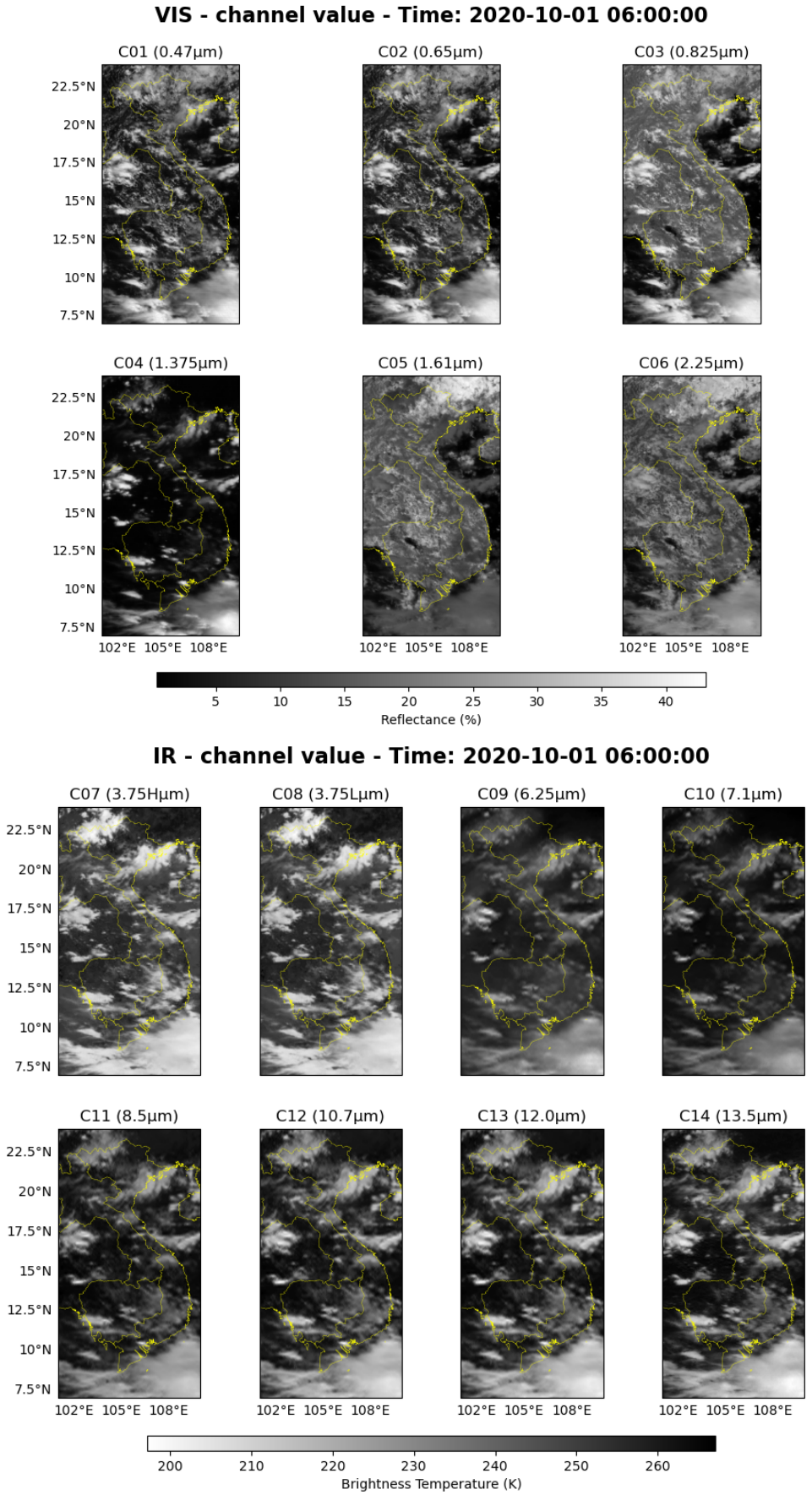

FIGURE 2.4: VIS (upper plot) and IR (lower plot) Channel maps for a specific time step.

In this study, the values from these channels serve as features for our ML models, as will be detailed in Section 3.1.1 of the Results & Discussion Chapter. For the scope of this introductory section, we present snapshots of channel data categorized into Infrared (IR) and Visible (VIS) spectra. These snapshots are depicted in Figure 2.4. This visual representation allows us to observe the distinct information each channel captures. While some channels accentuate cloud structures, others are more revealing of terrestrial characteristics.

2.2.2 Rain Products: Ground Reference Data

In this section, we provide an overview of the precipitation products employed in this study. Particular emphasis is placed on the radar data, which serves as the ground truth for calibrating the RF models. Additionally, we discuss the rain gauges used for subsequent validation, as they are often considered the gold standard in QPE. We also explore other rain products available in Vietnam to evaluate the effectiveness of the model developed in this study. A summary of the key characteristics of all these precipitation products can be found in Table 2.2 at the end of this Section.

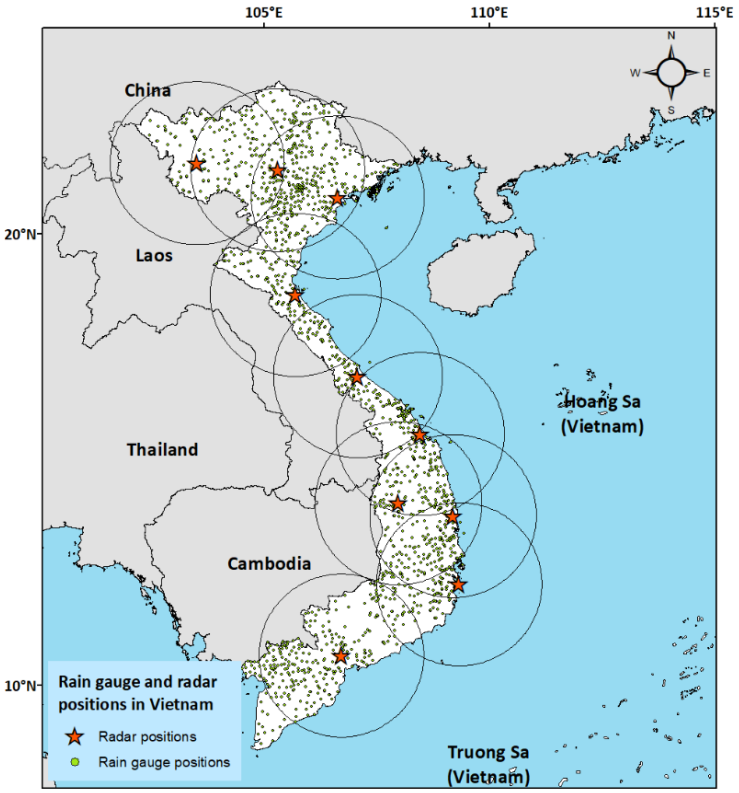

FIGURE 2.5: Location of the radars and rain gauges over Vietnam. Image taken from [59].

Radar

In 2020, Vietnam's weather radar infrastructure featured 5 dual-polarization units situated in cities such as Viet Tri, Nha Trang, Quy Nhon, Pleiku, and Pha Din Pass. Additionally, the network included 3 single-polarization C-band radars in Dong Ha, Tam Ky, and Nha Be, along with 2 single-polarization S-band radars in Phu Lien

and Vinh. The data for this study was sourced from these 10 radar locations and was partly drawn from the outcomes of the Vietnam National Research Project.

Prior to integrating this radar data into QPE calculations, a quality assurance process was undertaken to eliminate basic errors. The radar-based precipitation assessments were created using one-hour cumulative radar intensity, as per the original methodology developed by the Japanese Meteorological Agency [40]. This method was further tailored for the Vietnamese context by [26] and the Authority of Meteorology and Climate. These customizations included decoding the initial raw radar data and adjusting specific parameters.

The radar data offers hourly coverage across most of mainland Vietnam at a 1-kilometer resolution. Instrument density is at its highest in the Hanoi and Quy Nhon regions, whereas it is less comprehensive in central Vietnam and in the Ho Chi Minh area in the south (as illustrated in Figure 2.5).

Rain Gauges

The Vietnamese rain gauge network is a collaborative effort involving multiple entities, including the Viet Nam Meteorological and Hydrological Administration, as well as other partners. This expansive network features around 1300 automatic weather stations (AWS) scattered throughout the country, as shown in Figure 2.5. However, the network's density varies and is subject to occasional interruptions, resulting in an operational count of approximately 800-900 rain gauges at any given moment. These gauges take measurements every 10 minutes, which are then aggregated to calculate hourly average rainfall rates in millimetres per hour (mm/h).

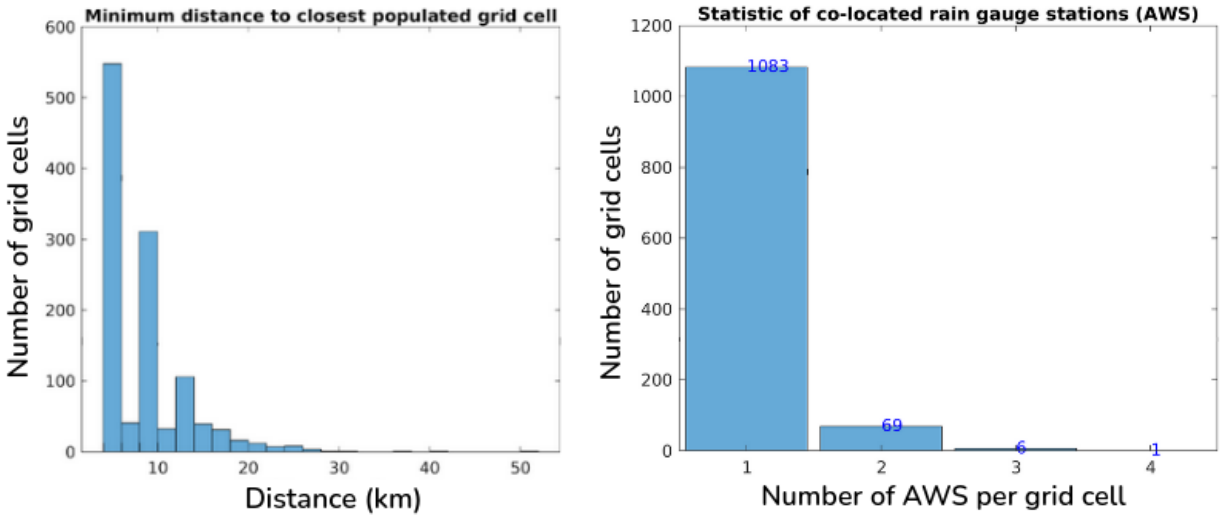

FIGURE 2.6: **Left**: Shortest Distance in Kilometers to the Nearest Populated Grid Cell with an AWS. **Right**: Frequency Distribution of Grid Cells Containing 1, 2, 3, or 4 AWS Units.

On average, the closest operational rain gauge is situated about 8.6 kilometers away, with the majority of the AWS spaced less than 30 kilometers apart. This distribution is depicted in the left-hand plot of Figure 2.6, which shows the minimum distance to the nearest populated grid cell. The right-hand side of Figure 2.6 illustrates the distribution of AWS within the FY-4A satellite grid used in this study. Specifically, the vast majority of grid cells (1083) contain just one AWS, while 69 cells contain two, six cells have three, and one cell contains four AWS.

We considered only those grid cells that had at least one rain gauge. A significant majority of AWS stations remained operational for over 80% of the research duration. However, there was a minor subset, comprising about 10%, that delivered valid data for less than half of the study period; these stations were subsequently omitted from the AWS dataset. If more than one AWS were present within a grid cell the average value was considered.

2.2.3 Other Rain Products

Geostationary satellite-based products

GeoKompsat-2A (GK-2A), operated by the Korea Meteorological Administration's National Meteorological Satellite Center, has been functional since December 2018 at longitude 128.2° E. The satellite's Advanced Meteorological Imager (AMI) has 16 spectral channels ranging from visible to thermal infrared (0.47 to 13.3 µm). It offers a 2 km spatial resolution at nadir for its infrared channels and scans the full disk area every 10 minutes. GK-2A's rain rate estimation employs Tb from five specific bands and utilizes probability density functions from recent GPM-DPR data to improve precipitation estimates [69]. The data for this study is sourced from the AMI Level-2 rainfall rate product, version 1.6.0.4, with technical documentation primarily available in Korean.

FY-4A AGRI characteristics are described in Section 2.2.1. While the specific algorithms used by FY-4A are not publicly disclosed, it's noted that the current version does not include a calibration step, and future updates may incorporate radar measurements [69]. The rainfall product from FY-4A is known as "FY4A AGRI L2 QPE" and is generated by version 1.0 of the respective software.

IMERG

The IMERG algorithm, version 3IMERGH_6.3, from the GPM Mission, offers surface-level precipitation estimates by integrating data from multiple satellites in both low-earth and geostationary orbits. These data are captured using the IR and MW parts of the spectrum. Version V06B of the IMERG products consists of three distinct runs: Early, Late, and Final. For a detailed explanation of the IMERG algorithm, refer to [20].

The complexity of the calibration procedures and quality controls escalates from the Early run to the Final run, accompanied by an increase in latency. The Early run, made available four hours post-data collection, is solely dependent on satellite data. The Late run, released 14 hours after data collection, incorporates additional satellite overpass data that were unavailable during the first four hours. Lastly, the Final run is published after a lag of 3.5 months and undergoes calibration with the Global Precipitation Climatology Centre's monthly monitoring analysis.

For the purposes of this study, we focused on the Early and Final runs, following the analysis adopted in [59]. The Early run offers a near-real-time (NRT) amalgamated product, while the Final run delivers the most accurate satellite-derived precipitation estimates for the region. We utilized the "precipitationCal" variable to extract rainfall data, which is available on a $0.1° \times 0.1°$ grid on an hourly basis.

ERA5-Land

ERA5 is a cutting-edge reanalysis product created by the European Centre for Medium-Range Weather Forecasts (ECMWF) [18]. One of its sub-components, ERA5-Land, is

specifically tailored to capture land-based variables. In this research, we utilized the "Total Precipitation" variable from ERA5-Land, which features a one-hour temporal resolution and a spatial resolution of 0.1°.

This Total Precipitation variable in ERA5-Land is calculated as the sum of two components: large-scale precipitation, computed by the model at scales beyond the individual grid box, and convective precipitation, which is computed within smaller spatial scales by the convective scheme of the ECMWF Integrated Forecasting System. To align with ERA5-Land's finer spatial resolution, the original precipitation data from ERA5, with an approximate resolution of 31 km, is rescaled using linear interpolation based on a triangular mesh.

It's crucial to recognize that the Total Precipitation figure from ERA5-Land represents the water depth distributed uniformly across each grid box. Consequently, this value may not align precisely with point-specific, instantaneous observations.

Overview of Key Rain Product Features

Dataset	Period	Grid resolution	Temporal sampling	Coverage	Data source	Latency
Rain Gauges	2008 -	point (avg. dist.: 8.6 km)	10 min/1 hour	country-wide	tipping bucket	∼10 min/ 30 min-1 hour
Radars	2019 -	1 km	1 hour	country-wide	C- and S-band	15-30 min
FY-4A	2017 -	4 km (nadir)	1h/3h/6h	full disk	VIS-IR	NRT
GK-2A	2018 -	2 km (nadir)	10 min	full disk	IR+DPR	NRT
IMERG Early run	2000 -	0.1°×0.1°	30 min	90°N-90°S	MW+DPR+IR	4 hours
IMERG Final run	2000 -	0.1°×0.1°	30 min	90°N-90°S	MW+DPR+IR +rain gauges	3.5 months
ERA5-Land	1950 -	∼9 km	1 hour	global	ECMWF model	2-3 months
FY-4A RF	October 2020	4 km (nadir)	1 hour	from 7°N-101°E to 24°N-111°E	FY-4A+radars	hours

TABLE 2.2: Overview of the key attributes of the precipitation data sets used for calibration, test and validation of the ML models. The information is obtained here [59]. The final row additionally provides details on the precipitation product obtained from the RF models in this study.

The key attributes of various rain products are summarized in Table 2.2. The table outlines essential characteristics such as the original period of data availability, grid resolution, temporal resolution, spatial coverage, and data source—specifying the instruments, spectral channels, or models used to derive the product. Additionally, the table indicates the latency required to convert raw data into the final rain product. It's noteworthy that rain gauges, radar, and certain satellite products exhibit low latency, often approaching near-real-time capabilities, due to minimal processing requirements. In contrast, products like IMERG and ERA necessitate extended processing time.

The final row of Table 2.2 contrasts these features with the product developed in this study, referred to as FY-4A_RF. This product leverages FY-4A satellite data via an RF model. Specifically, the study employs a spatial resolution of 4 km (native to the FY-4A) and a temporal resolution of 1 hour, focusing on the Vietnam region using a radar mask, since the radar product was used for calibration. While the initial model training takes several hours (accounting for data download, preparation, feature

extraction, and model preprocessing and training), subsequent product generations are faster, with latency mainly stemming from executing the ready-to-use model for estimates and the preparation of new unseen data.

2.2.4 Digital Elevation Model

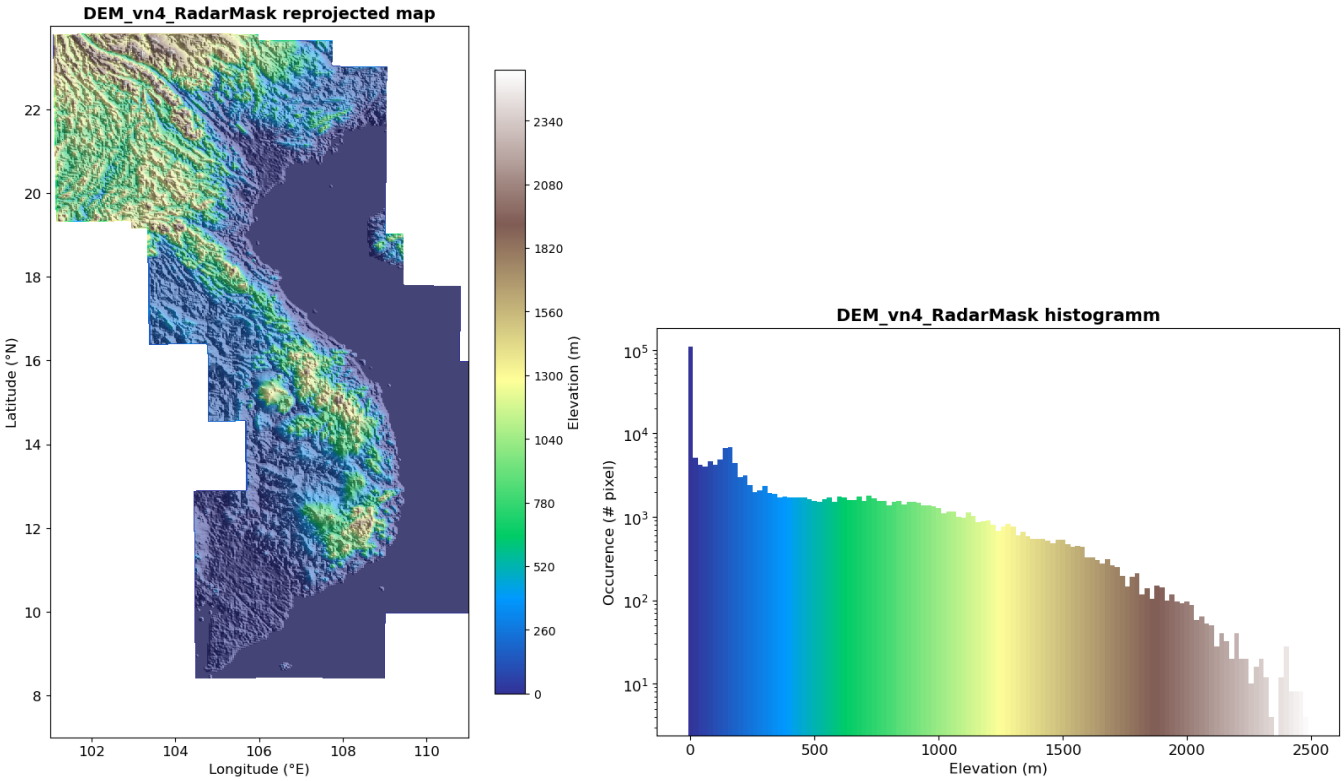

FIGURE 2.7: **Left**: DEM map with radar mask applied, enhanced with hill-shading to better highlight mountainous regions. **Right**: Histogram illustrating the frequency distribution of altitudes above sea level, using the same colourmap as employed in the map.

The DEM was provided with an initial spatial resolution of 0.04° in both latitude and longitude. Although this resolution is relatively low, it originates from NASA's ASTER Global Digital Elevation Model V003. The native resolution of the ASTER product is one arc-second, equivalent to about 30 meters at the equator. This indicates that the DEM has undergone some form of preprocessing to achieve its current resolution. For more details about the original data source, you can refer to the NASA ASTER website (`https://asterweb.jpl.nasa.gov/`).

The initial DEM image covered Southeast Asia but was cropped to emphasize Vietnam. We used a version of the DEM, aligned with the 4 km FY-4A grid and masked using a radar mask. Figure 2.7's left panel showcases this tailored DEM for Vietnam, while the right displays an altitude distribution. Vietnam's topography is mostly under 1000 meters, with exceptions reaching 2500 meters. Fansipan Mountain, the nation's tallest at 3,144 meters, isn't depicted likely due to the DEM's coarseness. The histogram also shows a peak at sea level, reflecting marine areas.

2.3 Analysis Workflow

This section is designed to outline the analytical process in a step-by-step manner. We begin with preliminary procedures that include quality assurance and outlier removal from the dataset. Subsequently, we delve into feature engineering, which prepares the data for the ML algorithms. The setup of the RF model, including both classification and regression tasks, is then elaborated upon. Finally, we evaluate the performance of these models using a variety of metrics. All the main analysis steps are reported in Figure 2.8 and will described in detail in the next sections.

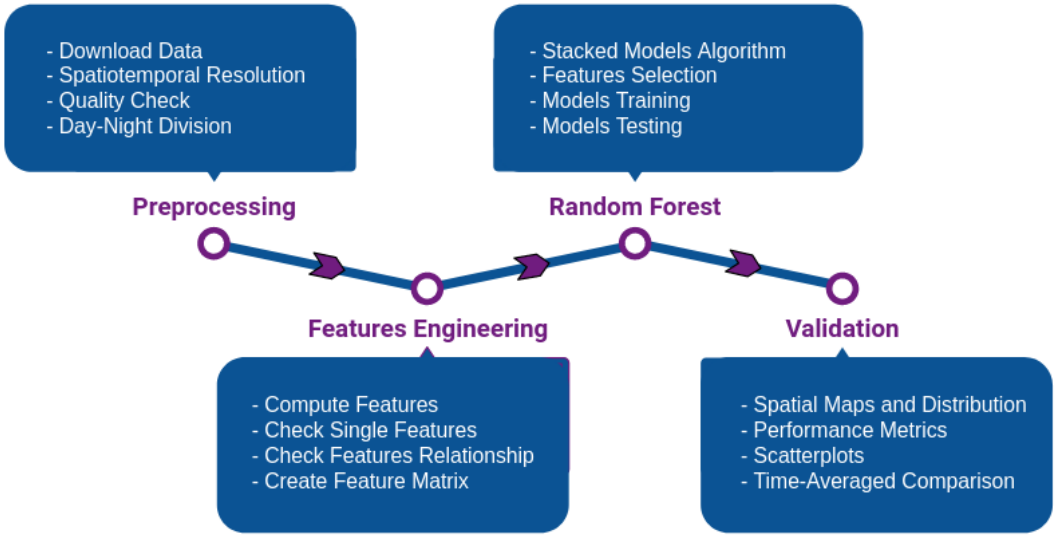

FIGURE 2.8: Diagram Illustrating the Key Phases of the Analysis Workflow: From Raw Data Acquisition to Quality Assessment, Feature Engineering, Random Forest Configuration, and Final Product Validation.

The analyses were performed on a workstation featuring a 64-bit architecture (x86_64) equipped with an Intel(R) Core(TM) i7-7700 CPU, consisting of 8 CPUs operating at a base frequency of approximately 800.819 MHz. This system boasted 15GB of RAM and operated on Ubuntu 18.04.6 LTS.

The primary software tools utilized were based on Python 3.10. Key among these was the Dask package (`https://docs.dask.org/en/stable/`), which facilitated efficient processing of sizable datasets without exhausting memory. The Satpy library (`https://satpy.readthedocs.io/en/stable/index.html`) was employed for FY-4A data manipulation, while the machine learning workflows leveraged the capabilities of scikit-learn (`https://scikit-learn.org/stable/index.html`). Additionally, general computations and data visualizations were achieved using the numpy, scipy, pandas, matplotlib, and seaborn libraries.

2.3.1 Preprocessing Steps

First, we downloaded the data files, checking and redownloading any corrupted files. We then assessed the dataset for missing temporal intervals and visualized data distribution to detect outliers, using a log-scaled density function and a temporal trend plot. We also produced spatial snapshots at certain time intervals. In the next subsection, we detail the preprocessing for our primary datasets, FY-4A and radar, which formed the foundation for our model training and testing.

The additional rainfall products used exclusively for validation (AWS, IMERG Early and Final, ERA5-Land, and the FY-4A and GK-2A rain products) were previously prepared for the study as cited in [59]. Therefore, no additional preprocessing was required for this work.

FY-4A

After downloading the October 2020 files, we assessed for any gaps. Figure 2.9 shows the missing files in red on the left. While we have data for most hourly 15-minute intervals, October 2nd and 3rd were exceptions with no available files. The right side of the figure showcases the file distribution within each hour. Most files fall within the hour's first quarter (00:00 to 14:59), with a notable absence in the third quarter (30:00 to 44:59).

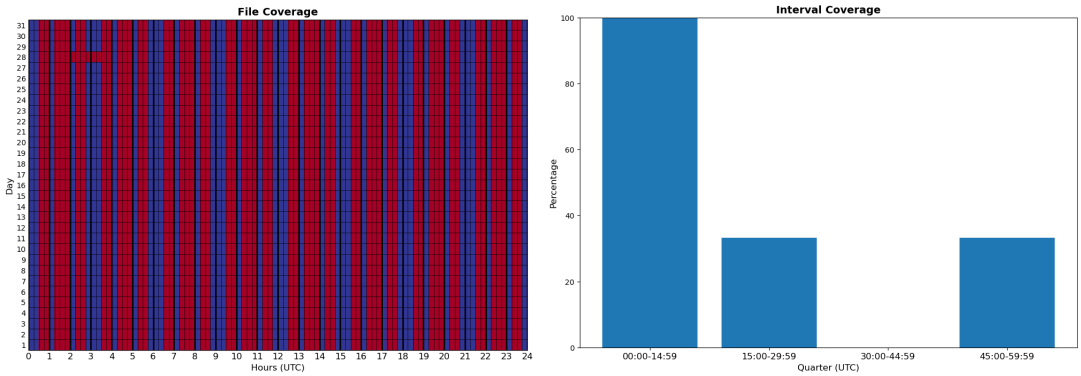

FIGURE 2.9: **Left**: File Coverage of FY-4A L1 data for October 2020. A red square indicates a missing 15-minute interval file. **Right**: Histogram illustrating the availability of files in each hourly quarter in October 2020.

Subsequently, the data were spatially restricted to encompass the area of Vietnam, specifically bounded by the longitude and latitude coordinates:

$$lon_{min} = 101^{\circ}E,\ lat_{min} = 7^{\circ}N,\ lon_{max} = 110^{\circ}E,\ lat_{max} = 24^{\circ}N \tag{2.1}$$

To achieve greater temporal consistency and in alignment with the methodology outlined in [59], we opted to rescale all data to a one-hour resolution by averaging the values for all channels within each hourly time span.

Afterwards, we assessed the frequency of missing values for each channel of the FY-4A data across all time steps within the Vietnam area. Bad-quality scan lines are automatically masked and replaced with NaN ('Not a Number') based on the quality flags provided by the dataset. The findings are illustrated in Figure 2.10. Notably, the proportion of missing values is generally quite low, consistently falling below 0.1. An exception occurred at a specific time—23:00 on October 8, 2020—where channels 13 and 14 exhibited approximately 20% missing values. However, as this anomaly was limited to a single time step, it is unlikely to impact the overall analysis, given that grid cells with NaN values are excluded from the dataset.

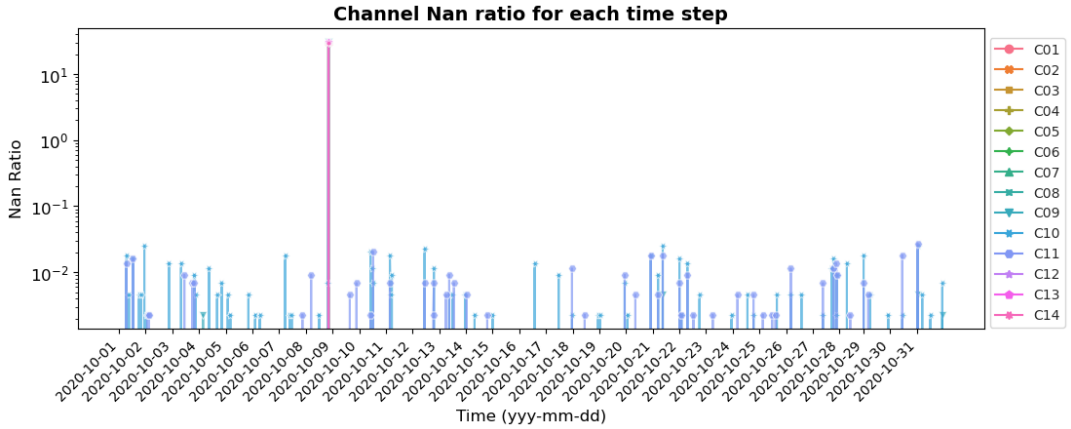

FIGURE 2.10: The graph displays the proportion of missing ('NaN') values for each FY-4A time step in October 2020. This proportion is derived by comparing the number of 'NaN' grid cells over Vietnam to the region's total cell count. This evaluation was conducted across all 14 satellite channels.

Then, we focused on the VIS channels, calculating their hourly average throughout the day. This helped determine when reflectance was under 5%, distinguishing between daytime (with VIS channels) and nighttime (without VIS channels). As shown in Figure 2.11, even the lowest average reflectance stayed above the 5% mark from 00:00 to 09:00 UTC. The 09:00 UTC value, however, reflects an average for the 09:00 to 09:59 interval. Thus, we've defined daytime as 00:00 to 10:00 UTC and nighttime from 10:00 UTC to 00:00 UTC. In local terms, this equates to 7:00 AM to 5:00 PM. An unexpected peak between 15:00 and 18:00 UTC seems to be a data anomaly, possibly due to an acquisition error. This doesn't affect our analysis since the VIS channels are excluded during that interval, which is considered nighttime.

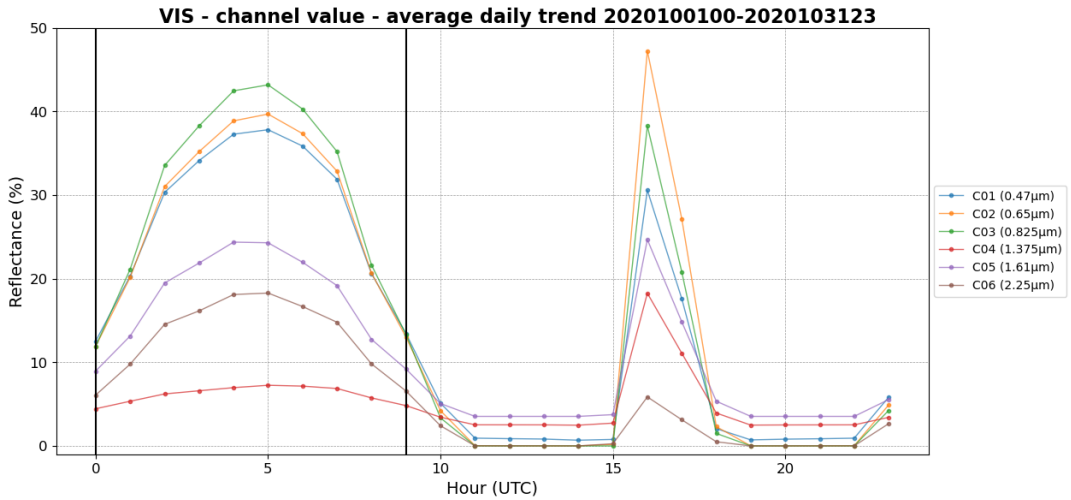

FIGURE 2.11: Daily Trend of VIS Channels: We averaged the values at each time step for the entire October 2020. The coloured curves show reflectance for each VIS channel (1-6), with black vertical lines marking the transition between day and night.

Radar

We reviewed the data for corrupted or missing files, finding 20 missing time steps. Thus, out of a month's 744 hours, 724 were processed successfully. We adjusted the data resolution from 1 km to 4 km to align with the satellite grid, using linear interpolation. The radar mask was also adapted to this grid for data filtering.

The histogram from the October 2020 rain gauge data, shown in Figure 2.12 (Left), reveals that most rainfall is under 100 mm/h, though events around 200 mm/h also occur. Given this distribution, setting a threshold of 150 mm/h for the radar rain product is justified. To avoid potential outliers in the radar data within this range, it's prudent to stay below this threshold.

In order to distinguish between dry and wet rain rates, we examined the probability density function of radar data for rates below 1 mm/h, as illustrated in the right-hand plot of Figure 2.12. We set the threshold at 0.1 mm/h, which serves as a pivotal point between two different distribution regimes: a flat distribution that likely represents noise and a decreasing distribution that is characteristic of actual rain rates.

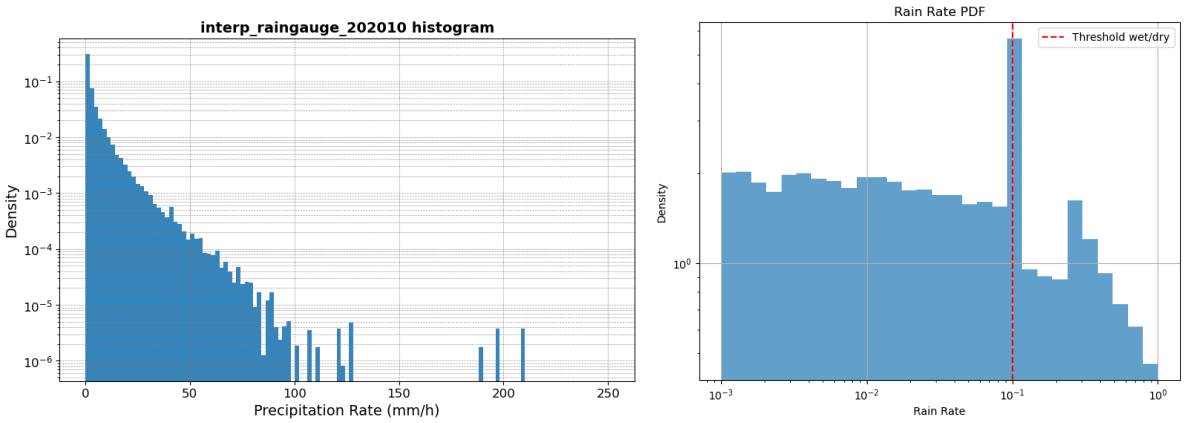

FIGURE 2.12: **Left**: Precipitation Rate Distribution from Rain Gauge Data, October 2020. **Right**: Precipitation Rate Distribution from Radar Data, October 2020 (rates < 1 mm/h). The vertical red line differentiates dry from wet rain rates.

2.3.2 Features Engineering

The features used in the RF model were carefully selected to capture a broad range of information. They span from spectral characteristics to spatial and temporal variations, providing a comprehensive picture that enhances the model's predictive capability. Together, these features create a multi-dimensional, comprehensive feature space that aims to provide the RF model with the best opportunity for accurate estimates.

- **Spectral Channels**:

  Reflectance Values for VIS Channels (1-6): As indicated in Table 2.1, the model utilizes values from the visible, near-infrared, and short-wave IR channels, which are available as reflectances. These values play a key role in capturing the characteristics of both the surface and the atmosphere during daylight hours.

Brightness Temperature for IR Channels (7-14): Brightness temperature from infrared channels is used for features, allowing the model to account for thermal variations. These are essential for understanding weather patterns and are applicable for both day and night.

- **Spatial Features**:

  Spatial Combination using a Sliding Window: To incorporate spatial context, a sliding window approach was implemented. A window size of 5x5 pixels was used to scan the data grid. From each window, two key statistics were computed:

  (1) Mean Value: The average of all the pixel values within the window.

  (2) Standard Deviation: A measure of the dispersion of the pixel values within the window.

  These spatial features enable the model to understand local variations, enriching the feature space.

- **Temporal Features**:

  For each pixel location, the difference between its value and the value at the same location in the previous time step was calculated. This feature captures temporal changes, adding a dynamic element to the model. Note that the first-time step and any time step following a file gap are discarded to ensure consistency in this calculation.

- **Channel Differences**:

  Differences between all possible combinations of the IR channels were calculated. These features can capture interactions between different thermal properties, adding another layer of complexity to the model.

- **Ancillary Data**:

  Digital Elevation Model. The DEM provides a static feature that represents the elevation of each pixel location. It's a crucial variable for understanding how topography affects weather patterns and other phenomena.

Features Type	Day	Night
Spectral Channels	14	8
Spatial Features	28	16
Temporal Features	14	8
Channel Differences	28	28
Ancillary Data	1	1
TOT	**85**	**61**

TABLE 2.3: Feature Count: Day vs. Night

Each feature was rigorously evaluated for its predictive utility. First, histograms and average temporal trends were analyzed to assess the distribution and behaviour of individual features over time. Second, to investigate the interdependence among features, a correlation matrix was computed. This allowed us to identify highly correlated features that might be redundant and could dilute the predictive power of the model. After the preliminary assessment, the RF model yielded insights via

its feature importance ranking, emphasizing the most pivotal features for precise estimates. Feature selection was then executed using this ranking. In the subsequent refinements of the RF model, only the top-ranked features were retained to enhance the model's efficiency, speed, and accuracy.

Details on the feature quality checks, the correlation matrix, and the feature importance rankings and selection are elaborated upon in Chapter 3.

2.3.3 Machine Learning Models Setup

ML stands at the convergence of statistics, artificial intelligence, and computer science, offering powerful tools for extracting meaningful insights from raw data. Within the domain of QPE, ML techniques have the potential to revolutionize our understanding and estimation capabilities. In this research, we deploy a combination of ML methods, specifically Random Forest Classifier (RFC) and Random Forest Regressor (RFR), to enhance the quality of precipitation forecasts.

We posit that these advanced computational techniques are well-suited to handle the inherent complexities of remote sensing data. This makes them a compelling choice for rigorously testing and improving the reliability of various precipitation measurement indices. The adaptability and robustness of these methods make them particularly relevant for handling the diverse challenges presented by QPE.

For a comprehensive understanding of machine learning algorithms and their implementation in Python through the Scikit-learn package, readers are referred to the referenced textbook [50]

Random Forest

In the scope of this work, we leverage the RF algorithm as our primary supervised ML model. RF is an ensemble-based algorithm built on **decision trees** (DT) [7, 6]. These trees are hierarchical models that recursively partition the feature space until each region—termed a leaf—contains predominantly one target class.

In the development of our model, a conscious effort was made to counteract overfitting, which is a common issue with individual DTs. By integrating an **ensemble** of DTs, we fortified our model's robustness. The ensemble approach of the RF offers a substantial advantage: it averages the predictions across multiple DTs, leading to smoother decision boundaries and enhanced generalizability, as illustrated in Figure 2.13.

Two pivotal elements underpin the effectiveness of the RF: bootstrapping and bagging. **Bootstrapping** involves the random selection of data subsets for each DT, ensuring diverse training sets. **Bagging** extends this concept to feature selection, where not only are data instances randomly chosen but so are the features for every DT. This dual randomness ensures that individual trees are distinct, contributing to the model's overall performance and resilience.

In the case of RFC, the **Gini Impurity** index serves as the default splitting criterion. The Gini Impurity measures the homogeneity within a dataset and is used to quantify how often a randomly chosen element from the set would be mislabeled if it was randomly labelled according to the label distribution in the set. Mathematically, the Gini Impurity G for a node containing N classes is calculated as follows:

$$G = 1 - \sum_{i=1}^{N} p_i^2 \tag{2.2}$$

Here, p_i is the probability of choosing an element belonging to class i within the node. A Gini Impurity value of 0 indicates perfect purity (all elements are of the same class), while higher values indicate increasing impurity. When constructing the DT, the algorithm aims to partition the data in a way that minimizes the Gini Impurity in the child nodes.

For RFR, the **Mean Square Error** (MSE) serves as the default splitting criterion. MSE is a measure of the average of the squares of the errors between the observed and predicted values. It is defined mathematically as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.3}$$

In this equation, n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value for the i^{th} observation. The goal is to minimize the MSE, which would indicate that the model's estimates closely match the actual data points. In the context of RFR, each split aims to create child nodes with lower MSE values compared to the parent node, thus refining the model's predictive power.

In RF, **feature importance** (FI) is commonly evaluated using the Gini Impurity Index or MSE method. This method gauges the utility of a feature by measuring how frequently a feature is utilized to split the data and how effectively these splits improve the model's purity. Specifically, for every DT in the forest, each split's contribution to model purity is noted based on the feature responsible for the split. The improvement is determined by assessing the difference in impurity (Gini impurity for classifiers or variance for regressors) before and after the split, with this difference being weighted by the number of samples the split affects. By aggregating these weighted improvements across all splits and all DTs where a particular feature is utilized, an average importance value for that feature is derived, providing an indication of its predictive power within the ensemble. It's important to note that a low FI score doesn't necessarily imply the feature is uninformative; it may indicate that another feature encapsulates similar information.

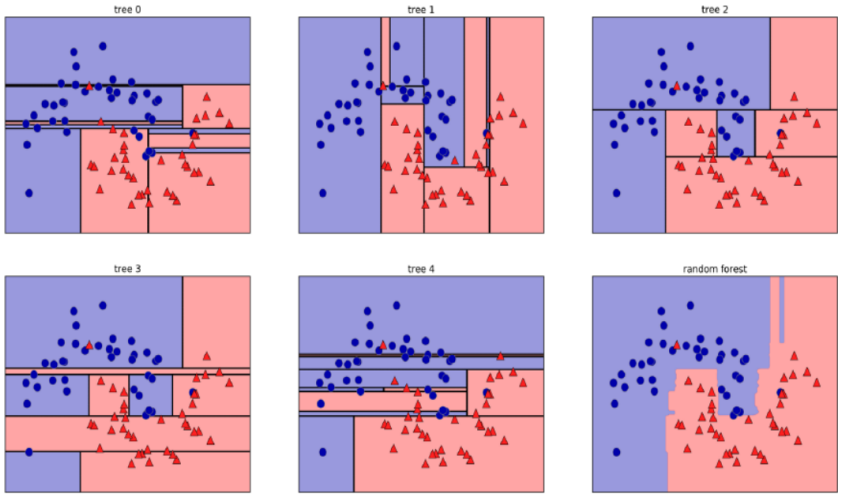

FIGURE 2.13: Decision boundaries found by averaging the predictions of five randomized DTs. Image taken from [50].

The **parameters** for the RF algorithm are adjustable to fine-tune the model's performance. We will adhere to the naming conventions established by scikit-learn's built-in functions for clarity and consistency.

- **N_estimators**

  This parameter specifies the number of trees in the RF ensemble. A higher number of trees generally leads to a more robust model but at the cost of longer training times.

- **Min_samples_leaf**

  This parameter specifies the minimum number of samples required for a leaf node. Smaller leaf sizes allow the model to capture more fine-grained patterns, but this could lead to overfitting.

- **Min_samples_split**

  This parameter sets the minimum number of samples required to split an internal node. Higher values for 'min_samples_split' make the decision tree more constrained, as it will then have fewer but purer nodes.

- **Max_depth**

  This parameter controls the maximum depth of each DT. If set to "None", the nodes are expanded until they contain fewer than 'min_samples_split' samples or until they are pure, meaning they contain samples from only one class.

- **Max_features**

  This parameter defines the number of features to consider when looking for the best split. For classification tasks, the default is the square root of the number of features, while for regression tasks, all features are considered by default.

- **Max_samples**

  This parameter sets the number of samples to be used for training each base estimator. If set to "None" all samples are used.

- **Criterion**

  This parameter specifies the function used to measure the quality of a split in the decision tree. For classification tasks, the Gini impurity is used by default, while for regression tasks, the MSE is the default criterion.

Parameter	Classifier	Regressor
n_estimators	100	100
Max_depth	None	None
Min_samples_leaf	1	1
Min_samples_split	2	2
Max_features	$\sqrt{n\_features}$	n_features
Max_samples	None	None
Criterion	Gini	MSE

TABLE 2.4: Default RF parameters from scikit-learn for our Classifiers and Regressors.

The initial settings for the RF parameters deployed in our model's first iteration are outlined in Table 2.4. During preprocessing, which encompassed FI ranking and feature selection, we adhered to the default RF model parameters. Our decision was bolstered by the observation that FI ranking remained consistent even when key

parameters were adjusted. In the final run, we fine-tuned a few parameters. We set 'min_samples_leaf' to 3 to mitigate overfitting, and adjusted both 'max_features' and 'max_samples' to 0.5 to capitalize on the bagging and bootstrap traits of RF.

For **feature selection**, we employed the FI ranking derived from the default RF parameters. To expedite computation, we trained using a subset of 10^5 randomly selected samples, except in the case of distinguishing heavy from extreme rainfall due to the scarcity of extreme values (see Table 2.5). Starting with the highest FI features, we incrementally added one feature per cycle, monitoring score variations on the test set. An optimal number of features for each RF model was then chosen based on a visual review of all metrics (refer to Section 2.3.4).

In summary, our RF implementation proves to be a robust and scalable choice for tackling the complexities inherent in QPE, offering an optimal balance between model complexity, speed and predictive power.

Stacked Models Algorithm

In the data preparation stage of this work, there are two primary focus areas: filtering the data and constructing the feature matrix. **Filtering** begins by selectively removing dry data points, which not only lightens the computational load but also helps to balance the dataset. Data points falling outside of a predefined radar mask are also excluded to ensure the geographical relevance of the study. To synchronize the temporal aspect, only time steps that are available in both satellite and radar datasets are considered for analysis. In addition, a NaN mask is generated by identifying the intersection of missing or undefined values in both radar and satellite data sources.

Turning to the **feature matrix**, the initial data dimensions are in the form '(number of features, number of times, number of latitudes, number of longitudes)'. This multidimensional data is reshaped into a 2D matrix of dimensions '(number of features, number of samples)' to fit conventional ML frameworks. Following the reshaping, the feature matrix is standardized by removing the mean and scaling the data to have a unit variance. This standardization is particularly important for algorithms that are sensitive to feature scaling, ensuring that no individual feature has an outsized impact on the ML model's learning process. These preprocessing measures lay the groundwork for the RF model, ensuring that it receives well-balanced and high-quality data for robust QPE.

Rain Classes (mm/h)	Training		Test	
	Day	Night	Day	Night
Dry $[0, 0.1)$	2332675	3348270	424284	617403
Wet $[0.1, 150)$	2423595	3725638	213863	359649
Light $[0.1, 1)$	1153884	1843756	96087	169813
Moderate $[1, 5)$	882732	1308203	79344	131786
Heavy $[5, 30)$	376463	555607	35240	54120
Extreme $[30, 150)$	10516	18072	3192	3930

TABLE 2.5: Rain Class Distribution for Train and Test Sets. Square brackets include the boundary value, while round brackets exclude it.

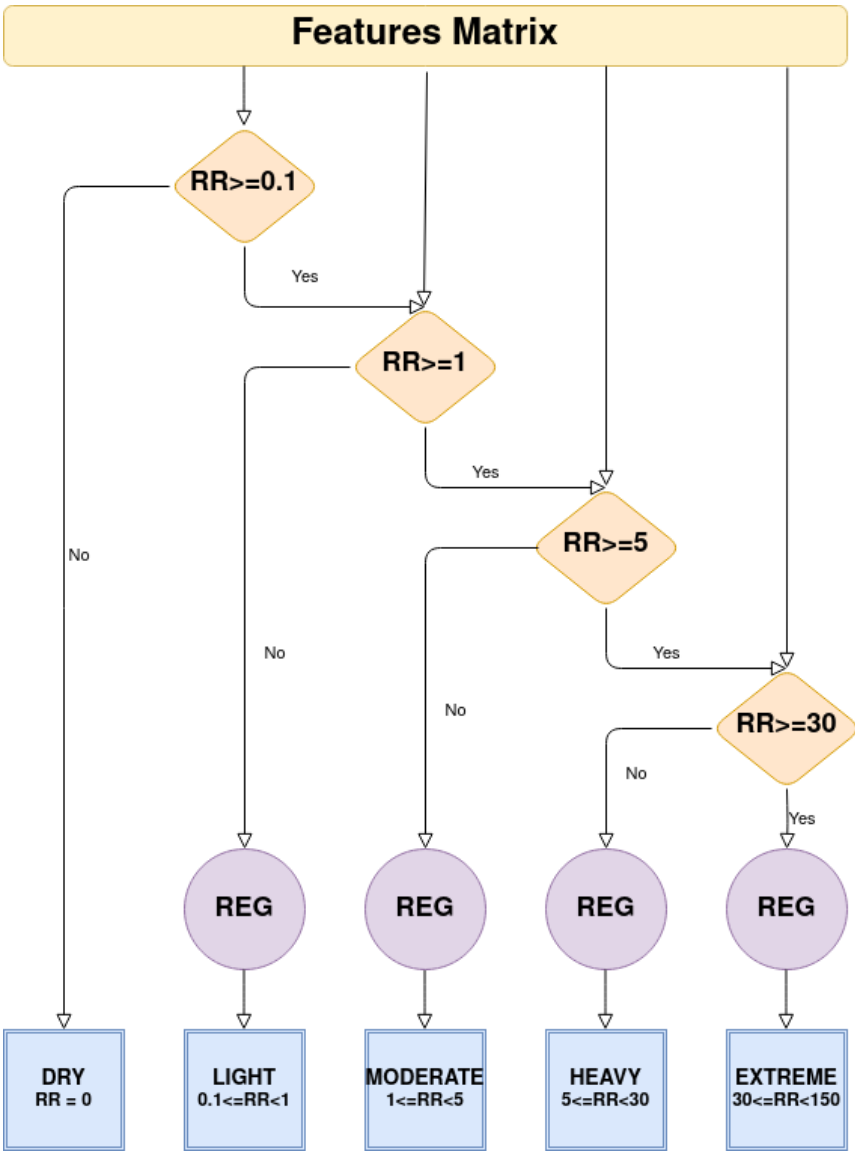

FIGURE 2.14: Flowchart Illustrating the Multi-Layered Model Algorithm: From Feature Matrix Construction to Predicting Rain Rates for Specific Rain Classes Using Regression Techniques.

Table 2.5 presents the sample distribution across various **rain classes** for both the training and test datasets during daytime and nighttime. These samples are derived from grid cells that pass through radar mask filtering and other preprocessing steps, and they are multiplied by the number of available time steps. The table differentiates between 'dry' and 'wet' classes based on a 0.1 mm/h threshold. Subsequent rows focus solely on 'wet' classes, which are further categorized into 'Light' (0.1-1 mm/h), 'Moderate' (1-5 mm/h), 'Heavy' (5-30 mm/h), and 'Extreme' (30-150 mm/h). For the Random Forest Classifier, binary classifications are made between the following class pairs: 'dry vs wet', 'light vs (moderate+heavy+extreme)', 'moderate vs (heavy+extreme)', and 'heavy vs extreme'. Our rain class delineations were inspired by literature which typically categorizes rainfall as light, moderate, or heavy [29, 49, 70]. However, we introduced an additional 'extreme' class to account for particularly severe rainfall events that can have significant impacts. While these classes are inherently imbalanced, reflecting the natural distribution of rainfall intensities, we refrained from performing class balancing. We removed some dry

pixels to ease computational demands but retained the natural skewness in rain rate distribution, aiming for a model attuned to real-world rainfall patterns.

We provide a detailed explanation of our **stacked models algorithm** implementation below. Once the Feature Matrix is prepared, the initial classification step comes into play, partitioning the data into two main categories: dry conditions and wet conditions. The data corresponding to dry are set to zero. Within the wet data subset, a subsequent classification occurs, distinguishing the data into "light rain" and the remaining instances.

Subsequently, the dataset labelled as "light rain" is employed to fine-tune the Feature Matrix, rendering it suitable for the subsequent regression analysis aimed at predicting light rain rates. Simultaneously, the rainfall instances categorized as exceeding the threshold for "light rain" become pivotal in filtering the Feature Matrix. These filtered instances then undergo another classification phase, discerning them into "moderate rain" and the remaining cases.

The outcomes attributed to the "moderate rain" classification are harnessed to facilitate the regression analysis tailored to predicting moderate precipitation rates. Consequently, the remaining data instances - those surpassing the "moderate rain" threshold will extract the corresponding samples from the Feature Matrix. This extraction yields the separation of data in "heavy rain" and "extreme rain". Then, dedicated regression analyses are undertaken for both the "heavy rain" and "extreme rain" categories.

Each step in Figure 2.14 represents an RF model, a classifier (diamond shape) and a regressor (round shape), respectively. They were previously trained independently using separate datasets.

2.3.4 Performance Assessment

A variety of statistical metrics were utilized following the in-depth analysis performed in [59]. To assess the accuracy of the RF precipitation product after classifications, we employed a suite of five categorical indices. Moreover, we used six continuous scores to evaluate the quality of the estimated rainfall fields at the locations of the rain gauges when regressions were employed.

Categorical Metrics

Categorical metrics are frequently employed to gauge the efficacy of a model or technique in identifying precipitation events. The dichotomy between rainy and non-rainy conditions or two rain classes was established by setting a threshold on both the observed and predicted rainfall data. Using this threshold, a 2x2 contingency table was constructed by comparing instances where the measured data and estimates either exceeded or fell short of the threshold. In Table 2.6 True positive (TP) or 'hits' and True Negative (TN) or 'correct rejections' occupy the principal diagonal, while False Negative (FN) or 'misses' and False Positive (FP) or 'false alarms' are situated elsewhere in the matrix. In our classifications, the negative class selections are: dry (for dry-wet), light (for light vs. moderate-heavy-extreme), moderate (for moderate vs. heavy-extreme), and heavy (for heavy vs. extreme).

In this study, we evaluated precipitation products using a suite of indices. These encompass the Probability of Detection (POD), which measures the proportion of true positives to all actual positive events; False Alarm Ratio (FAR), the fraction of false positives to predicted positives; Multiplicative Bias (BIAS), the ratio of predicted positives to actual positives; Critical Success Index (CSI), akin to POD but

with true positives normalized by all samples excluding true negatives; and Equitable Threat Score (ETS), a refined CSI adjusted for expected random hits (H_R) that is defined as H_R = [(TP+FN)x(TP+FP)]/[TP+FN+FP+TN]. Each of these metrics is calculated from different components of the contingency table. Table 2.7 offers a detailed guide to these indices, outlining their definitions, potential ranges, and ideal values [53].

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

TABLE 2.6: Confusion Matrix

Name	Equation	Range of Values	Optimum
Probability Of Detection	**POD** = TP/(TP+FN)	[0,1]	(1)
False Alarm Ratio	**FAR** = FP/(TP+FP)	[0,1]	(0)
Multiplicative Bias	**BIAS** = (TP+FP)/(TP+FN)	[0,∞]	(1)
Critical Success Index	**CSI** = TP/(TP+FN+FP)	[0,1]	(1)
Equitable Threat Score	**ETS** = (TP-H_R)/(TP+FN+FP-H_R)	$[-\frac{1}{3},1]$	(1)

TABLE 2.7: List of Categorical metrics together with their definition, range and ideal values.

Continuous Metrics

We also employed a suite of continuous indices to assess the precision of rainfall products, directly comparing predicted values to their corresponding ground truths. The comprehensive list is available in Table 2.8. Among these, the Correlation Coefficient (CC) gauges the linear relationship between predicted and actual values; the Coefficient of Variation (CV) measures the relative variability; Mean Error (ME) captures the average deviation of predictions from actuals; and Mean Absolute Error (MAE) represents the average magnitude of these deviations, irrespective of direction. These metrics are foundational for evaluating model performance, as discussed by Nurmi (2003) [53]. In addition to these, we also incorporated two less conventional metrics: the Modified Kling-Gupta Efficiency (mKGE) and P50.

The Modified Kling-Gupta Efficiency (mKGE) is an adaptation of the original KGE [14]. It calculates goodness-of-fit based on three dimensions: correlation, average ratio, and standard deviation ratio. The mKGE score is computed as one minus the Euclidean distance from an ideal point in this three-dimensional space, with coordinates (1, 1, 1) signifying perfect performance [27]. Notably, this modified version normalizes the standard deviations relative to global averages, enhancing the metric's interpretability by making its components orthogonal.

The P50 index offers a practical measure of a product's overall accuracy, calculating the likelihood that an estimated value will deviate from the actual measurement by no more than 50%. This metric accommodates the inherent variability in high-resolution precipitation products, offering a more lenient yet realistic evaluation compared to traditional metrics. By allowing for a reasonable range of deviation, P50 provides a more balanced assessment of a product's efficacy [59].

Name	Equation	Range of Values	Optimum
Correlation Coefficient	**CC** = $\sigma_{o,e}/\sigma_o\sigma_e$	[-1,1]	(1)
Coefficient of Variation	**CV** = $\sqrt{\frac{\sum(e_i-o_i)^2}{N}}/\mu_o$	[0,∞]	(0)
Normalized mean error, or bias	**ME** = $\frac{\sum_i(e_i-o_i)}{N}/\mu_o$	[-∞, ∞]	(0)
Normalized mean absolute error	**MAE** = $\frac{\sum_i \|\|e_i-o_i\|\|}{N}/\mu_o$	[0,∞]	(0)
Modified Kling-Gupta efficiency	**mKGE** = 1 - $\sqrt{(CC-1)^2+(\beta-1)^2+(\gamma-1)^2}$	[-∞, 1]	(1)
Probability to have e_i inside ±50% of o_i	$\mathbf{P}_{50} = n[\mathrm{e}_i \mid (0.5 \text{ x } o_i) \leq e_i \leq (1.5 \text{ x } o_i)]/\mathrm{N}$	[0,1]	(1)

TABLE 2.8: List of Continuous metrics together with their definition, range of values and ideal values. e: estimate; o: observed; σ: standard deviation: μ: mean; $\beta = \mu_e/\mu_o$: ratio of the averages; $\gamma = \frac{\sigma_e/\mu_e}{\sigma_o/\mu_o}$ is the ratio of the normalized standard deviation; N: total number of samples; n[x] is the number of the occurrences of x.

Chapter 3

Results & Discussion

3.1 Preliminary Analysis

This chapter is devoted to presenting the principal findings of the thesis, beginning with a preliminary analysis that serves as both a preprocessing step and a precursor to the application of the RF model. This initial analysis offers valuable insights into the data quality and sets the stage for subsequent discussions of the results. Specifically, the quality of the datasets was evaluated through visual inspection of feature distributions and temporal trends for both the predictor variables and target rain products. A correlation matrix is also presented to explore the interrelationships between features, thereby shedding light on their significance within the feature space, later revealed by the RF model. The Section concludes with an examination of the relationship between precipitation levels and Vietnam's topography.

3.1.1 Data Quality Check

VIS Channels

To commence our analysis, we initiate a quality control procedure for the features under study, leveraging both their distribution and temporal trends. This quality control encompasses the entirety of the data collected during October 2020, a period that includes both training and test datasets. The aim is to ensure the integrity and reliability of the features prior to employing them in subsequent analyses.

In Figure 3.1, we conduct a quality assessment of features associated with **VIS channels**. Subfigure A1 illustrates the distribution of the VIS channels themselves, ranging from C01 to C06. Notably, the first three channels—corresponding to wavelengths of 0.47, 0.65, and 0.825 micrometers—display a broader distribution extending into higher reflectance values. This suggests that these wavelengths are more effectively reflected by either the Earth's surface or its atmosphere. In contrast, the remaining VIS channels (C04, C05, C06) exhibit distributions that are more concentrated around lower reflectance values.

Several factors may contribute to these disparities. Shorter wavelengths like 0.47, 0.65, and 0.825 μm interact more variably with terrestrial features, such as vegetation and water bodies, leading to a broader distribution. Sensor sensitivity and variations in solar irradiance across these wavelengths could also influence these distributions. Furthermore, the impact of clouds and atmospheric particles varies across the spectrum, potentially skewing reflectance values.

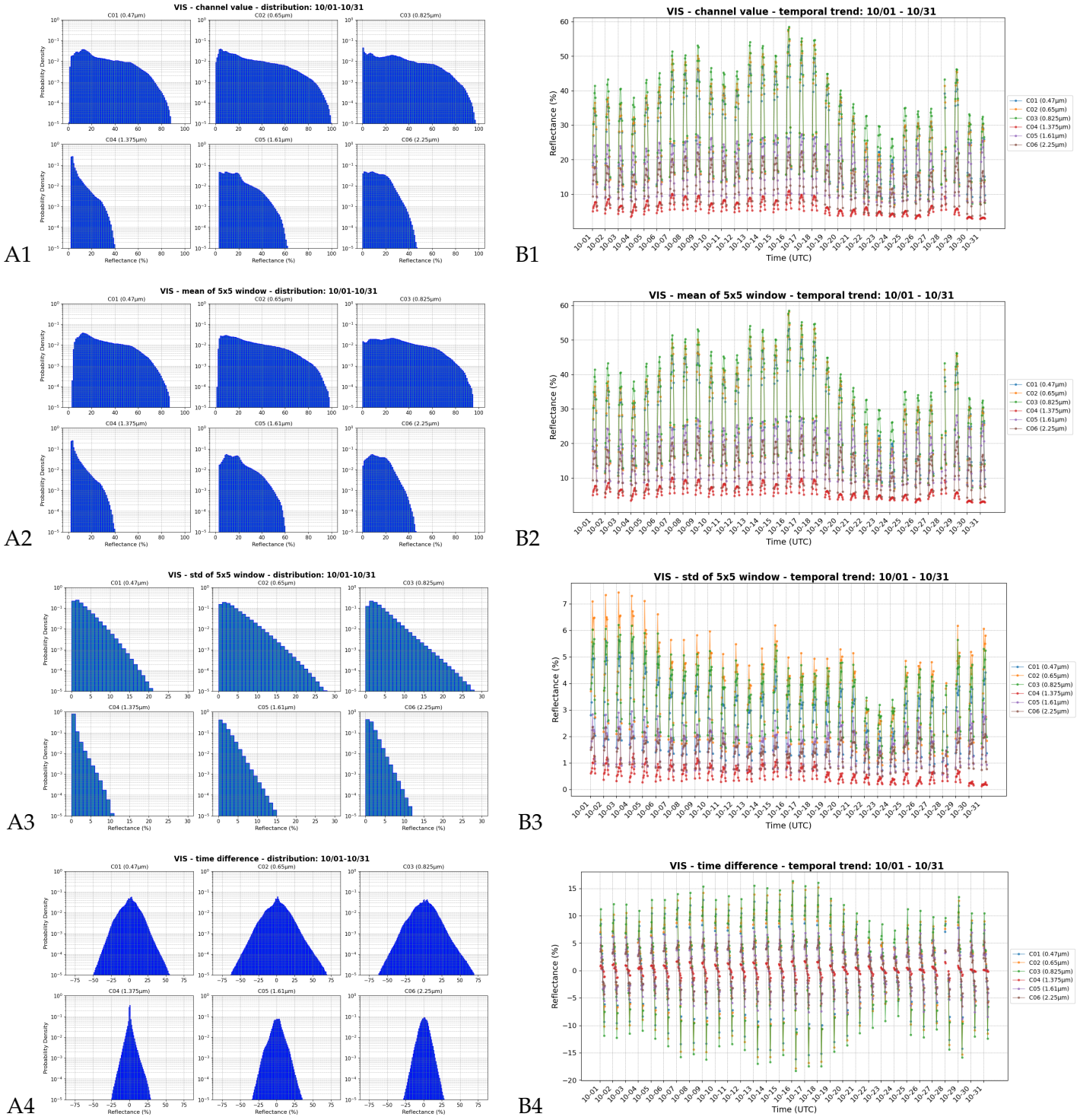

FIGURE 3.1: Data Quality Assessment for Features Related to Visible Channels: This figure presents four key metrics—channel values (1), mean of a 5x5 sliding window (2), standard deviation of a 5x5 sliding window (3), and time difference between consecutive time steps (4). The first column (A) depicts the probability density function on a logarithmic scale, while the second column (B) showcases the temporal trend of these metrics, averaged over the spatial domain for the entire time period considered (both training and test sets).

In Subplot B1, the temporal trend of the features is visualized, with each hourly time step representing the spatial mean across the Vietnam domain. A similar pattern to that in Subplot A1 emerges here as well; notably, channels C01, C02, and C03 exhibit higher reflectance values in comparison to the remaining channels. Intriguingly, the data reveal a noticeable decline in reflectance levels during the latter third of the month. This trend appears to correlate with a concurrent reduction in radar-detected rain rates, as will be further elucidated in Figure 3.4 (right panel). This synchronous decline in both reflectance and rain rates suggests a potential reduction in cloud cover during this period.

In Figure 3.1A2-B2, the **mean values** of the VIS channels are computed using a 5x5 sliding window. The resulting distributions bear a strong resemblance to those in Subplot A1 but exhibit a more refined, smoother profile. The use of this averaging technique acts as an effective spatial filter, dampening abrupt fluctuations while accentuating the primary characteristics of the data. While the same filtering principle applies to the temporal trends depicted in Subplot B2, its impact is less pronounced due to the inherent averaging already present in these trend plots.

In the third row of Figure 3.1, the **standard deviation** calculated within a 5x5 sliding window is presented. Subplot A3 reveals the distribution of standard deviations for each channel. Notably, the first three channels exhibit higher levels of standard deviation compared to the remaining channels, indicating greater variability within their respective windows. Subplot B3 reinforces the observations made in previous figures. Here, periods of lower reflectance correspond to lower standard deviations, suggesting that intervals with reduced cloud cover or precipitation also exhibit decreased spatial variability.

In Figure 3.1, Subplots A4 and B4 provide insights into the **temporal differences** between consecutive time steps for the VIS channels. The distributions exhibit a symmetrical shape centred around 0 %, which aligns with expectations: the reflectance may either increase or decrease over time, contingent on the evolution of cloud systems. Notably, channels C01, C02, and C03 display a broader distribution compared to the other channels, which are largely confined within the range of approximately -25 to 25 %. This observation is consistent with the distribution of channel values discussed earlier and is further corroborated by the temporal trends evident in Subplot B4.

The quality assessment of VIS channels reveals distinct differences in reflectance properties and variability across various wavelengths. Channels with shorter wavelengths (C01, C02, and C03) not only exhibit higher reflectance values but also display broader distributions and higher standard deviations, indicating greater variability influenced by factors like terrain and cloud cover. These channels also align with a noticeable temporal decline in reflectance, correlating with reduced rain rates, hinting at less cloud cover later in the month. The use of a 5x5 sliding window for averaging proves effective in refining the spatial profiles without significantly altering temporal trends. Overall, the findings underscore the need for nuanced interpretation of VIS channels, particularly when dealing with temporal and spatial variability, and suggest that channels with shorter wavelengths may offer more dynamic insights into environmental conditions.

IR Channels

In Figure 3.2, we examine the characteristics of the **IR channels**. Beginning with the basic attributes—channel values—Subplot A1 reveals the distribution of Tb for each channel. Notably, Channels 7 and 8, corresponding to 3.75 μm high and low resolution, respectively, exhibit minor peaks at low Tb around 200 K. As these channels operate at shorter wavelengths on the fringe of the IR spectrum, they are particularly susceptible to variations during twilight periods. This unique feature in the distribution is absent in the other channels and is likely a consequence of IR contamination due to reflected solar radiation during day/night transitions. While satellite sensors generally implement techniques to reduce such contamination, it's worth mentioning that we did not employ additional filtering in this study. Moreover, Channels 7 and 8 reach the highest Tb, going up to 320 K. Channels 11, 12, and 13 exhibit a similar range, extending up to 300 K, but with distributions that differ in width and shape. In contrast, Channels 9, 10, and 14 are more narrowly confined and shifted towards lower Tb. These observations are corroborated by Subplot B1, which delineates the temporal average trends. It becomes evident that Channels 7 and 8 typically register higher Tb, while Channels 9 and 10 consistently record lower values. Intriguingly, the period corresponding to diminished cloud activity during the third decade of the month—which coincides with decreased reflectance in the visible channels—reveals elevated Tb in the IR channels. This is likely because IR radiation is primarily emitted from the ground, which generally has a higher temperature than the cloud tops. A subsequent drastic reduction in Tb correlates with the onset of a typhoon featuring high convective clouds, reinforcing the complex interplay between cloud dynamics and Tb observations.

In Figure 3.2, the IR-derived features exhibit similarities to those of the VIS channels. Specifically, Subplot A2, which represents the **mean values** computed using a 5x5 sliding window, mirrors the distribution shape of Subplot A1 but with a smoother texture. Notably, the noisy peaks around 200 K are less pronounced in this representation, although the overall configuration and placement of the distributions remain consistent. Subplot B2 largely reiterates the temporal trends observed in Subplot B1.

Turning to the **standard deviation**, depicted in Subplots A3, the narrower distributions observed in Subplot A1 correspond to reduced standard deviations; for instance, Channel 09 demonstrates this characteristic. The temporal trend in Subplot B3 corroborates this observation, showing that periods of lower cloud cover align with lower standard deviation values, as anticipated.

Finally, Subplots A4 and B4 focus on the **temporal differences** between consecutive time steps for the IR channels. Interestingly, some distributions are not perfectly symmetrical around 0 K but exhibit a slight skew towards negative values. This suggests that the succeeding time step frequently exhibits a lower Tb than its predecessor, possibly indicating vertical cloud development. The distributions for Channels 7 and 8 (C07 and C08) exhibit peculiar shapes, likely due to spurious values around 200 K associated with contamination. This anomaly is particularly evident in Subplot B4, where Channel 7 displays a pronounced peak around October 25th, with a spatial average nearing 15 K, substantially higher than its closest counterparts, which do not exceed 5 K.

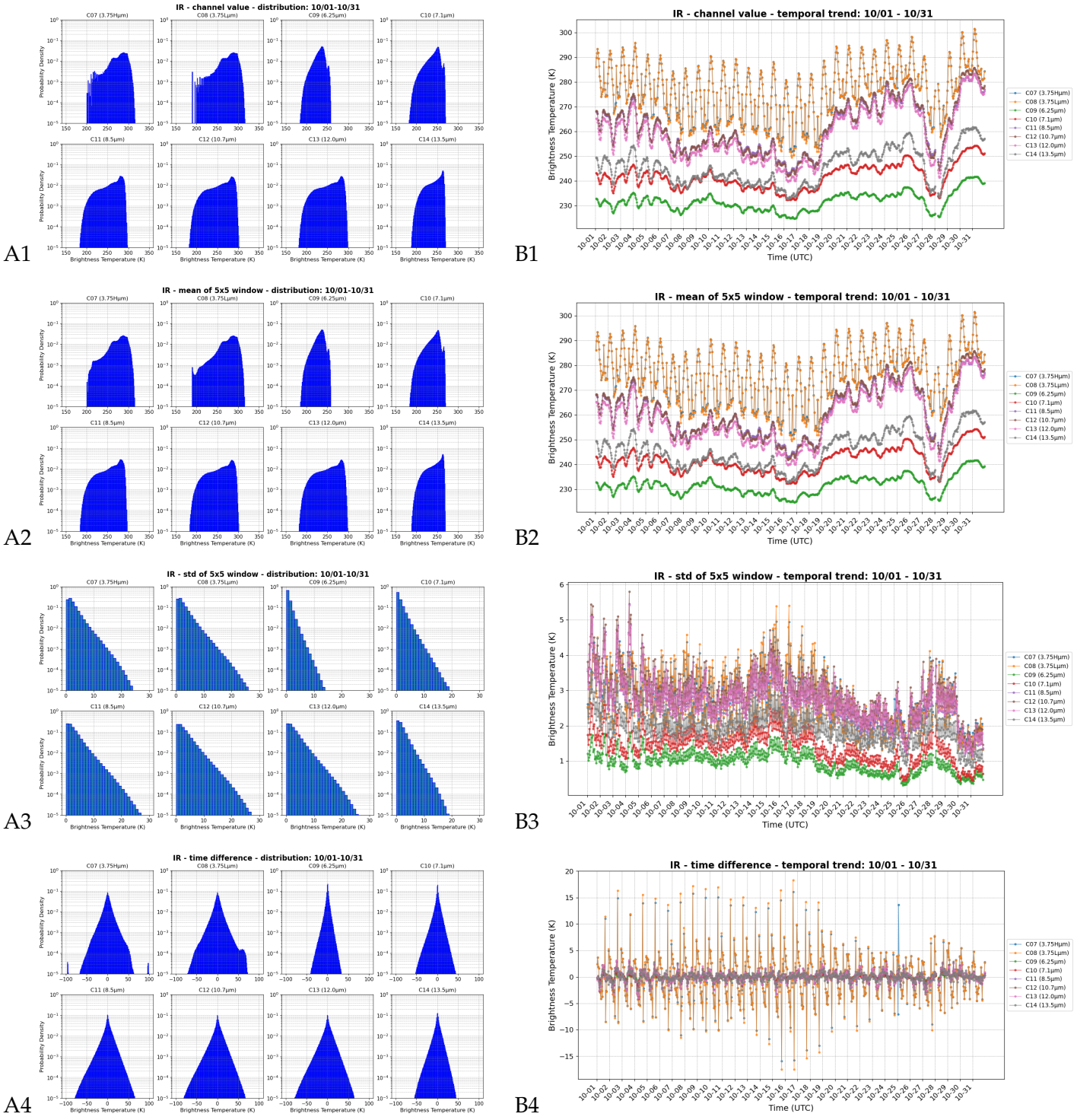

FIGURE 3.2: Data Quality Assessment for Features Related to Infrared Channels: This figure presents four key metrics—channel values (1), mean of a 5x5 sliding window (2), standard deviation of a 5x5 sliding window (3), and time difference between consecutive time steps (4). The first column (A) depicts the probability density function on a logarithmic scale, while the second column (B) showcases the temporal trend of these metrics, averaged over the spatial domain for the entire time period considered (both training and test sets).

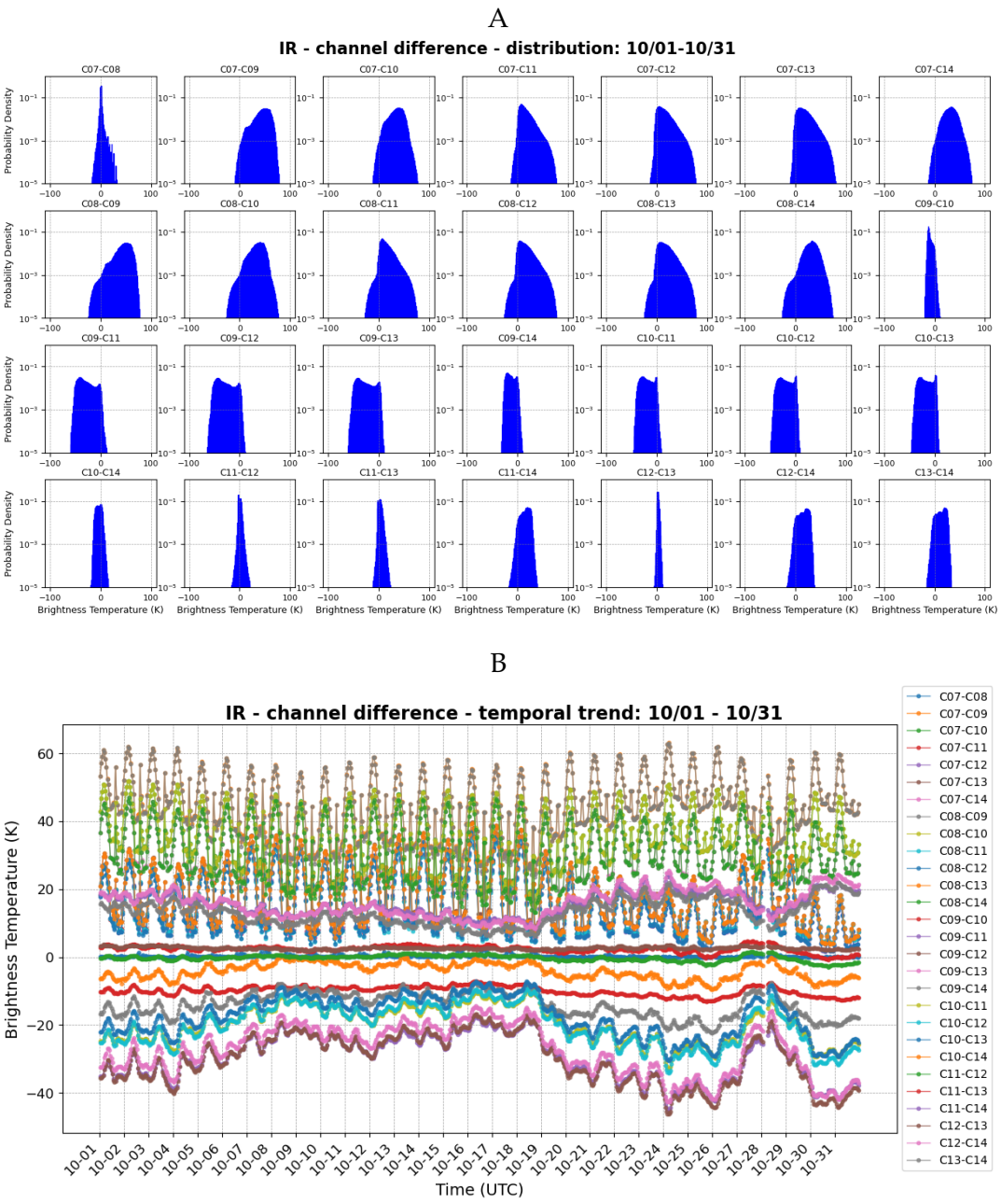

FIGURE 3.3: Data Quality Assessment for Infrared Channels differences. The upper plot (A) depicts the probability density function on a logarithmic scale, while the lower plot (B) showcases the temporal trend of these metrics, averaged over the spatial domain for the entire time period considered (both training and test sets)

In Figure 3.3, the distributions and temporal trends of **IR channel differences** are presented. These distributions vary significantly depending on the channel pair in question. For instance, broader distributions, as seen in the C08-C09 pair, imply disparate Tb values between the channels. In contrast, narrower distributions like C12-C13 suggest closely aligned Tb measurements.

Furthermore, some distributions are centralized around 0 K, such as the C10-C14 pair, indicating that neither channel consistently registers higher or lower Tb values than the other. Conversely, other channel pairs exhibit a more pronounced trend: for example, the C07-C13 pair is predominantly positive, while the C09-C11 pair is

mainly negative. These skewed distributions result from our choice to retain the original values rather than use absolute differences, in order to preserve the distinguishable characteristics of each channel pair.

Much like the previously discussed temporal trends for other features, here too we observe varying trends for each channel pair. Some pairs display more stable values over time, while others exhibit oscillating behavior, influenced by both short-term weather fluctuations and the diurnal cycle.

The analysis of IR channels presents intricate patterns of Tb, with Channels 7 and 8 standing out for their susceptibility to variations during twilight and for reaching the highest Tb up to 320 K. This study highlights that IR channels are influenced by a range of factors, including cloud activity, diurnal cycles, and possibly contamination due to reflected solar radiation. Elevated Tb during periods of diminished cloud activity corroborates that IR radiation is mainly emitted from the Earth's surface, which is generally warmer than cloud tops. The findings also show a complex relationship between cloud dynamics, such as the onset of typhoons, and Tb observations. Standard deviations and temporal trends further confirm these observations, revealing that reduced cloud cover is associated with lower standard deviations. The study of IR channel differences offers additional nuance, showing varying Tb measurements depending on the channel pair, which can be influenced by short-term weather changes and diurnal cycles. These complexities in IR channel characteristics underscore the importance of cautious interpretation and potential filtering strategies for more accurate estimates.

Radar Rain Product

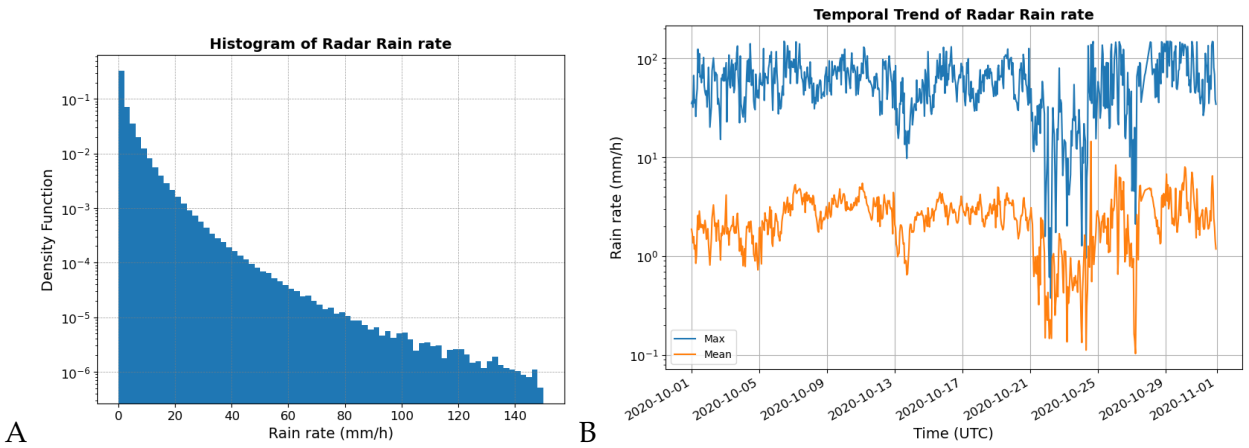

FIGURE 3.4: **A**: Probability Density Function (PDF) of radar-derived rain rates for October 2020, with outliers exceeding 150 mm/h removed. **B**: Mean (orange) and Max (blue) Temporal trend of rain rates as captured by radar during October 2020.

In an initial step, we conducted an analysis of the rain rate values for the designated time frame, which is October 2020. Figure 3.4 presents two aspects of this data. Firstly, it displays the Probability Density Function (PDF) of the rain rates, truncating values that exceed 150 mm/h. This outlier threshold was determined based on the empirical observations from rain gauge data (refer to Section 2.3.1 for further details). Secondly, the figure illustrates the temporal trend of rain rates throughout the specified period. It is noteworthy that the rain rate distribution does not conform to any unconventional pattern; it exhibits a high concentration of lower values,

which precipitously diminish as we move toward higher rain rates. Additionally, the metrics of mean and maximum rain rates are fairly typical, except for a noticeable reduction in precipitation between October 21 and 25.

3.1.2 Features Correlation Matrix

To complement the discussion on feature importance (that will be discussed in Section 3.2.1), we also present the correlation matrix calculated from the samples used in the training phase for each RF model. The correlation coefficients serve as indicators of the relationships between features. High correlation (or anti-correlation) between features could imply redundancy, meaning that one feature may not add additional value to the RF model. By examining this matrix, we aim to better understand the interdependencies among features and assess whether any of them can be omitted without sacrificing model performance.

In subsequent sections, we will present plots exclusively from the wet/dry pixel classification for both day and night scenarios. This decision stems from the observation that the correlation matrices for other classification and regression cases exhibited similar patterns. To maintain brevity while adequately illustrating our findings, we've chosen these representative examples. The consistency across matrices is likely attributed to the correlation coefficients being influenced more by the number of features shaping the samples rather than the specific case in question.

The upper plot of Figure 3.5 displays the correlation matrices for the **daytime** models. As anticipated, we observe a significant correlation among features within the same category, exemplified by the red squares along the diagonal—particularly among the VIS channels for both the primary and derived features. However, there are exceptions to this general trend. For example, within the VIS category, C05 shows a mild anticorrelation with C04. When examining the derived features, discrepancies are evident in the correlation levels among pairs of features like mean, standard deviation, and time differences.

As previously discussed in Section 3.1.1, the correlation between channel values and their corresponding mean within a 5x5 window closely resembles the correlation among channel values or just the mean values themselves. This similarity in correlation patterns suggests that including both channel values and their 5x5 window means may not contribute additional information for classification tasks. In contrast, the correlation between channel values and standard deviations is less pronounced, as is the case with time differences.

Interestingly, the VIS and IR channels generally show anticorrelation. This phenomenon can be attributed to the likelihood that high reflectance levels may indicate elevated cloud formation, resulting in lower Tb in the IR channels. An exception here is C05, which shows a slight positive correlation with the IR channels. Additionally, the standard deviations of the IR and VIS channels are positively correlated, suggesting that both types of spectra similarly capture changes in atmospheric conditions.

Examining the IR channel differences reveals distinct patterns of correlation, both within the channel differences themselves and with other features. To begin with intra-channel correlations, pairs like C07-C011, C07-C012, and C07-C013 exhibit strong correlations with C08-C011, C08-C012, and C08-C013 pairs. On the other hand, certain pairs display negative correlation—specifically, C12-C14 and C12-C13 are negatively correlated with all pairs involving C09 and C10. Additionally, some pairs demonstrate moderate levels of correlation. For example, C08-C09 has a moderate correlation with C07-C11, C07-C12, and C07-C13.

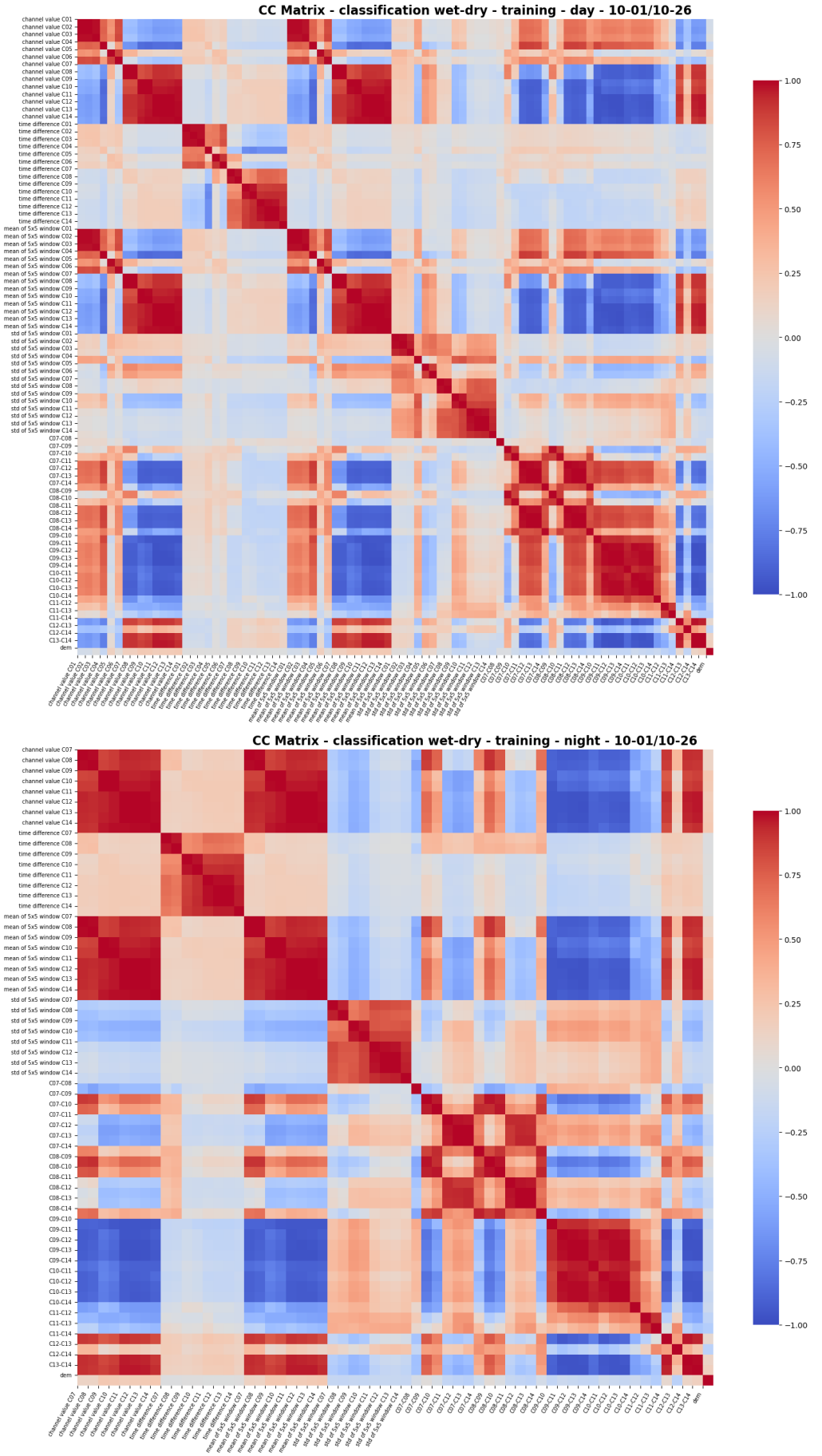

FIGURE 3.5: Correlation Matrices for daytime (upper) and nighttime (lower) samples in the wet vs. dry classification training.

Separately, the DEM features stand out for their low correlation with the rest of the features. This is a promising observation, as the low correlation suggests that DEM could serve as an important, independent variable in classification tasks.

The correlation matrix obtained with **nighttime** data is presented in the lower plot of Figure 3.5. In this case, the matrix is altered because it lacks features related to VIS channels.

Noticeably, high positive correlation exists among certain types of features — specifically, channel values, time differences, means, and standard deviations are closely interrelated, as indicated by intense red clusters in the correlation matrix. The correlation between time differences and channel values (or their means) is only modest. Conversely, standard deviation appears to be slightly anticorrelated with other features, including channel values, time differences and means.

When examining the correlations among IR channel differences, the matrix structure becomes more intricate, displaying a variety of behaviours among the pairs without any discernible overarching pattern. Notably, all differences involving C09 and C10 are highly correlated with each other and are highly anticorrelated with the channel values and their means. In other instances, each feature exhibits unique correlation characteristics relative to the others.

As for the DEM feature, it continues to show a low correlation with the rest of the features, further highlighting its potential as an independent variable in classification tasks.

In summary, the analysis of correlation matrices reveals a high degree of intra-category correlation, particularly among VIS channels. This suggests limited additional benefit in using both channel values and their corresponding 5x5 window means, offering the potential for feature reduction. Standard deviations and time differences show less correlation with other features, indicating their unique value in the models. The DEM consistently exhibits a low correlation with other variables, highlighting its potential as an independent variable. Overall, the matrices underscore the importance of a diverse set of features to capture the complexity of weather phenomena.

3.1.3 Elevation-Precipitation Interaction

In addition to utilizing spectral features derived from the VIS and IR channels of the FY-4A satellite, we also included an ancillary feature: the DEM. Prior to incorporating DEM into the feature set, we conducted an initial analysis to investigate its relationship with precipitation patterns. This was done to ascertain whether orography plays a significant role in influencing rainfall in Vietnam. The results of this study are presented in Figure 3.6.

Examining the boxplot in Figure 3.6, it becomes evident that as altitude increases, there's a subtle decline in precipitation levels. This trend is observable both in the mean and maximum values of precipitation over the specified time period for both training and testing data sets. The decline is illustrated through the distribution of these values across different elevation categories: 0-200 meters (flat land), 200-600 meters (hills), 600-1000 meters (low mountains), and 1000+ meters (high mountains).

Upon scrutinizing the lower section of Figure 3.6, it becomes clear that data samples are less abundant for the higher elevation categories. This limited dataset for elevated regions poses challenges for generating reliable statistics in those cases. Furthermore, a gradual decline in the wet ratio can be observed as one moves from flat land areas (slightly above 0.2) to high mountain regions (around 0.15).

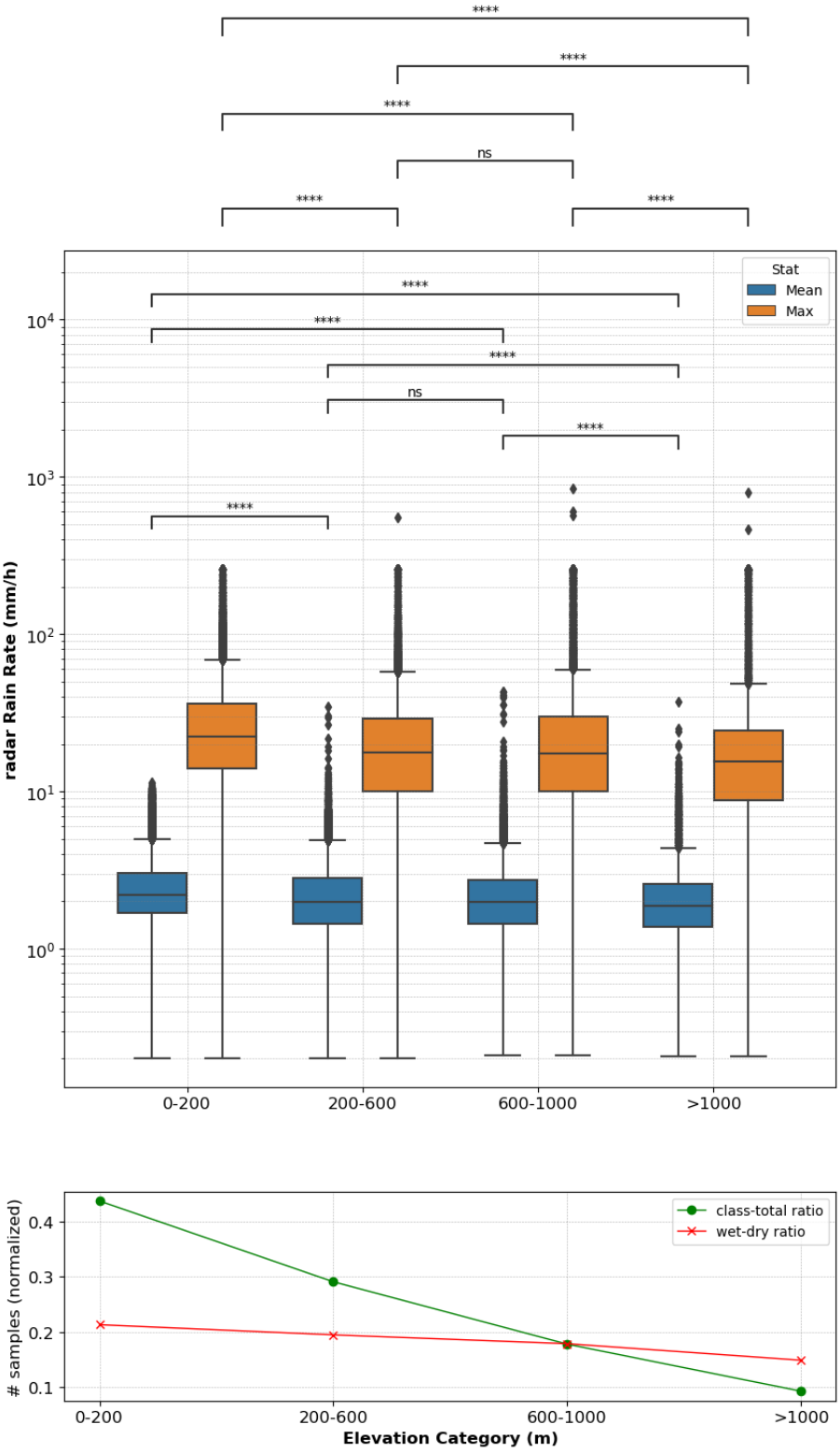

FIGURE 3.6: Upper Plot: Box plot illustrating the distribution of radar-derived rain rate across various elevation classes, with mean values in blue and maximum values in orange. Each pair of distributions was evaluated using a Wilcoxon-Mann-Whitney test. A notation of '****' indicates statistical significance with a p-value less than or equal to 1.00×10^{-4}, whereas 'ns' denotes non-significance. Lower Plot: Depicts both the ratio of samples and the proportion of wet conditions within each elevation class.

As shown in the notations, there's a statistically significant difference between nearly all paired distributions in both the mean and maximum boxplots. This difference is substantiated by the results of the Wilcoxon-Mann-Whitney test, where almost all of the assessed pairs display high statistical significance, registering a p-value lower than 10^{-4}. The only exceptions are the elevation pairs 200-600 and 600-1000, for which both mean and maximum distributions do not exhibit statistical significance.

The trend of lower rainfall in mountainous areas, which may seem counterintuitive, can be explained by the characteristics of typhoons that dominate the wet seasons. These storms gather moisture and energy from the Pacific Ocean but expend much of it as rainfall upon making landfall on Vietnam's coast. As a result, by the time they reach the inland mountainous regions, their intensity is significantly reduced, leading to less rainfall despite the orographic factors that might typically enhance precipitation.

Given these findings, incorporating DEM as a feature in our model becomes strategically important. The statistical evidence supports the notion that there is a significant relationship between rainfall and elevation, making DEM a valuable addition to our feature set.

3.2 Random Forests Outcomes

This section delves into the results of our RF models, applied to both daytime and nighttime scenarios. We begin by emphasizing the most critical features, then proceed with feature selection based on FI ranking. Upon determining the optimal features for each RF model, we undertake model training and testing. The model's efficacy is gauged through various metrics for classification and regression tasks. Finally, we validate our model's rain estimates using Vietnam-specific pluviometers and other rain datasets.

3.2.1 Features Importance Rankings

In this section, we examine the FI generated by RF models, utilizing default hyperparameters (see Table 2.4), across various classification and regression tasks. We focus on the top half most influential features, and for easier comparison, these importance scores are normalized against the highest-ranking feature. This implies that when all features have high normalized scores, they are relatively equally important in the RF model. On the other hand, if only the top features exhibit high scores while the rest have low scores, it indicates that just a handful of features predominantly contribute to the model's training. Therefore, it's crucial to consider both the ranking and the relative scores of the features, rather than focusing solely on their absolute values, which could be misleading when assessed in isolation.

It's crucial to note that FI doesn't equate to causality. While a high importance score indicates a feature's usefulness in accurate estimation, it doesn't imply that the feature causes a particular outcome in the target variable. However, understanding these importances aids in model interpretation, and feature selection, and provides valuable insights into the data and the issue being addressed.

Classifications

Let's begin with Figure 3.7, which displays the FI rankings for classification tasks. Subplot 1.A focuses on the first half of all features for the **'wet vs dry'** classification using **daytime** data. Notably, many top-ranking features are differences between IR channels, such as C09-C12, C09-C11, and C10-C11. The features in focus are differences between the WV channels (Ch 9 and 10) and long-wave IR channels (Ch 11 and 12). Literature references these differences for retrieving cloud depths [42] and for traditional rain estimates, such as the 'Kurino' technique [30]. Notably, these differences have also been leveraged in RF-based QPE studies to ascertain cloud top height concerning the tropopause level [31, 46]. Interestingly, single-channel features like the 5x5 window mean for C08 and channel values for C07 and C08, corresponding to the spectral channel 3.75 μm, are also prominent. The first VIS-related feature, the time difference of C02, ranks 13th, while the ancillary DEM feature is 29th. The top 12 features have marked importance with scores above 0.4, but the relevance wanes for the subsequent ones.

Moving on, we can contrast these findings with their **nighttime** equivalents, as shown in Figure 3.7.1.B. Here, the most impactful features are the means of the 5x5 windows for channels C07 and C08, followed by their individual channel values. These features hold a normalized score higher than 0.6. Subsequent features display diminishing normalized scores below 0.5, tapering off to approximately 0.2 in the first half of the ranking. Similar to the daytime scenario, prominent IR channel differences like C10-C12 and C10-C13 make the list, as do time differences for C07 and C08, reaffirming the significance of these channels. Notably, DEM holds a relatively high rank, appearing in the 11th position. In comparison to the daytime scenario, the significance of the differences between WV and IR channels diminishes, while the 3.75 μm channel becomes more influential for the RF model. Although primarily categorized as mid-wave IR, this channel also captures reflected solar radiations during twilight periods, as evidenced in Section 3.1.1. It's occasionally employed for cloud phase differentiation during nighttime, especially when paired with other IR channels [31]. Despite its limited conventional use in QPE, this channel is predominantly utilized for fire detection [23].

Subplot 2.A of Figure 3.7 reveals the feature importance rankings for the classification between **light rainfall** and higher rain rates for the **daytime** case. The top five features in this case are all mean values from 5x5 windows of the following channels: C01, C06, C02, C011, and C05. These are followed by the time difference of C01, DEM, and the IR channel difference C09-C11. In contrast to the earlier scenario, neither IR nor VIS channels distinctly dominate in this context. Both types offer substantial contributions. In particular, the VIS/NIR channels 1 and 2 provide significant input, as does the short-wave channel 6. Channel around 0.6 μm can provide information on the cloud water path [49]. Moreover, as reported in [43] visible reflectances can help determine the cloud top glaciation. Additionally, the difference between WV and IR channels that gives information on the cloud top height is still noteworthy. In this instance, more than half of the features retain high importance levels, well above a normalized score of 0.5. However, only the top five features have importance scores exceeding 0.8, after which the scores gradually decline.

In the **nighttime** context depicted in Figure 3.7.2.B, the predominant feature is the standard deviation obtained from a 5x5 window of channel C09. The mean from the same 5x5 window of this channel also emerges as the third most significant feature. Given the absence of VIS information during the nighttime, IR channels associated with high- and middle-level water vapour become notably influential in the RF

model. In this instance, DEM emerges as the second most important feature, playing a critical role in the classification task. The list is further populated by other feature derivatives such as standard deviations, means, and time differences across various IR channels. Generally, the feature importance remains notably high, consistently exceeding a normalized score of 0.6, with the top four features surpassing 0.8.

Figure 3.7.3.A presents the feature importance rankings for distinguishing **moderate rainfall** from heavy and extreme categories during **daytime**. Notably, the top features primarily originate from the mean values of various channels, with C06, C05, C01, and C11 taking the first four spots. Out of the top 20 features, 13 are derived from the mean of 5x5 windows. Also included within these top features are standard deviations from 5x5 windows for channels C05, C04, C06, and C09. Additional features include time differences for channels C01 and C02, as well as the value for channel C06. Interestingly, most of these influential features come from VIS channels. In this classification task, DEM ranks relatively low, appearing at the 29th position. Regarding feature scores, the first half of the listed features show high normalized importance scores, all exceeding 0.6. While the top ten features display a sharp drop in importance—ranging from close to 1 down to below 0.8—the remaining features maintain a fairly stable score, hovering around 0.7.

Unlike the daytime scenario, the **nighttime** feature ranking (Figure 3.7.3.B) is not solely dominated by mean-derived attributes. While the mean from channels related to WV level, namely C09 and C10, occupy the first and third positions respectively, the initial ranks are primarily filled by standard deviations, specifically from channels 9, 10, 14, 8, 7, and 11. In addition to means and standard deviations, time differences also make an appearance among the top 10 features, particularly for channels C08 and C07. DEM is ranked 15th in this case. The overall normalized importance scores remain high, consistently exceeding 0.7, and the significance levels taper off gradually, without showing stark differences between the top-ranked and subsequent features.

The final subplot, 4.A, in Figure 3.7 outlines the **daytime** feature importance rankings for classifying between **heavy and extreme** rainfall. Predominantly, standard deviations of channels such as C05, C07, and C08 lead the rankings. The first half of the list includes standard deviations from other VIS and IR channels as well. Other notable features that perform well in this classification task include means from channels like C08, C01, and C05, along with some time differences, such as that of C08. DEM appears quite low in the rankings, securing the 35th position. In terms of normalized importance scores, most features maintain a high score consistently above 0.7. Starting from the top features, which hover around a 0.9 score, there is a gradual decline in importance as we move down the list.

Figure 3.7.4.B presents the **nighttime** equivalent. Here, it's evident that standard deviations are pivotal in training the RF model; eight out of the first 10 features are standard deviations of various channels such as C14, C07, and C11. These channels differ from those highlighted in the daytime scenario. Mean values from 5x5 windows also play a significant role, particularly for channels C08 and C07. In the remaining top-ranked features, time differences and IR channel contrasts emerge. Notably, in this instance, DEM does not make it into the top half of the feature rankings. Similar to the daytime case, normalized scores generally remain high, tapering off slightly to just above 0.7.

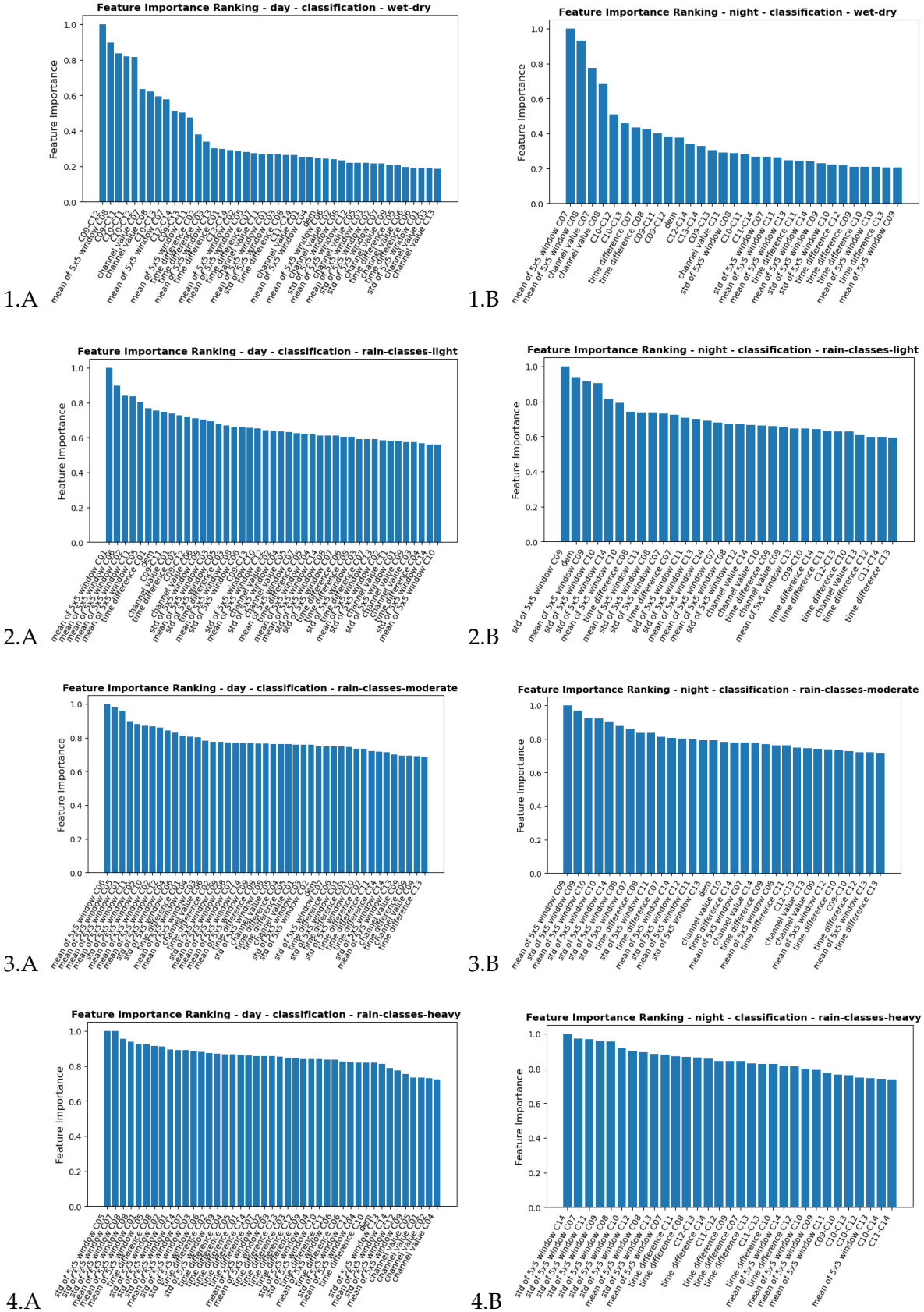

FIGURE 3.7: Feature Importance Ranking for Classification Tasks: 1) Wet vs. Dry, 2) Light vs. Moderate+Heavy+Extreme, 3) Moderate vs. Heavy+Extreme, 4) Heavy vs. Extreme. Displayed for daytime (A) and nighttime (B) training data, showing only the top half of features.

From our comprehensive analysis of FI across daytime and nighttime classification tasks, we can deduce several crucial observations. Primarily, the prominence of features varies based on the specific classification objective, highlighting that distinct rainfall categories demand a specialized set of features for optimal distinction. In essence, VIS channels are pivotal in categorizing varying intensities of rainfall when accessible, though they aren't as vital in detecting rain areas. For IR channels, the differences between WV-IR occasionally stood out, an anticipated outcome given their correlation with cloud depth and consequently, cloud convectivity. Furthermore, individual channels, particularly those associated with WV, along with mid-wave IR channels, consistently proved to be of significance. The DEM rarely emerges as significant, with the notable exception being in the detection of light precipitation during nighttime. Features consistently hold importance across classifications, but their significance varies. In 'light vs rest' and 'moderate vs heavy and extreme', importance is more evenly spread, suggesting a diverse feature set for the RF model. However, in 'wet vs dry', the model leans more on a few top-ranking features. The utility of means from 5x5 windows is evident in both daytime and nighttime classifications, whereas standard deviations from 5x5 windows mainly help discern heavy and extreme rain classes. Individual channel values often play a secondary role, likely because they're correlated with their corresponding means, as indicated in the Correlation Matrix (Section 3.1.2). Hence, the models favour mean values that capture information on a broader area, seeing individual channels as redundant. Overall, these findings not only affirm the complexity of rainfall classification tasks but also underline the necessity for a nuanced approach that considers a range of features, varying both in nature and significance, for accurate and robust modelling.

Regressions

Moving on to the regression model feature rankings within our RF framework, Figure 3.8.1.A presents the importance of various features for predicting **light rain** rates during the **daytime**. The top two features are mean values derived from channels C01 and C08, with DEM occupying the third spot. The remaining features represent a blend of both VIS and IR channels, covering means, standard deviations, and time differences. Notably, in the upper half of the rankings, pure channel values and IR channel differences are relatively scarce. This suggests that spatiotemporal combinations of channel values are predominantly crucial for this regression task. In terms of feature importance scores, there's a pronounced emphasis on the top few features; the first three maintain scores above 0.9, after which there's a gradual decline from 0.8 to approximately 0.4.

The **nighttime** scenario, as illustrated in Figure 3.8.1.B, offers a slightly different landscape of feature importance. The most crucial feature in this case is the standard deviation derived from a 5x5 window of channel C09. Additional standard deviation features also occupy prominent positions in the ranking. Notably, DEM secures a high rank, coming in at the second position. Other key features include the time differences of channels C08 and C07, ranked 3rd and 8th respectively. IR channel differences are somewhat more prevalent in the nighttime case, with the first one (C12-C13) appearing at the 9th position. This difference corresponds to 10.7-12.0 μm and it is sometimes used to gain information on cloud phase [31]. As for the importance scores, they notably peaked; the second feature scores above 0.9, but there is a significant drop to below 0.8 at the third position, gradually tapering off to around 0.4.

The task of regression for **moderate rain** is next in line for examination. Figure 3.8.2.A highlights the **daytime** scenario, where standard deviations of VIS channels (C04, C05, and C06) jointly with DEM dominate the FI rankings. Subsequent to these, other spatial combinations of channels, including both means and standard deviations, contribute substantially. These contributions come from both VIS and IR channels. While temporal differences and IR channel differences do make an appearance in the upper half of the ranking, they are not found among the initial top-ranking features. In terms of FI scores, there is a gradual decline, dropping to below 0.6; however, there is no pronounced peak of highly important features at the beginning of the list.

The **nighttime** situation is depicted in Figure 3.8.2.B. The top positions are primarily occupied by features related to standard deviation, specifically from channels 9, 14, 07, 8, and 10. A mean feature from channel C10 also ranks highly, claiming the 3rd position. DEM is situated at the 7th spot on the list. Additionally, the rankings include some time differences and IR channel differences (such as the C12-C13). Notably, no individual channel values are present in the upper half of the feature ranking. Regarding normalized importance scores, the top feature attains a score of 1, which then drops to below 0.8 for the second most important feature and continues to decline to below 0.4. In this scenario, it appears that a single feature plays a significantly prominent role in the training process.

Figure 3.8.3.A presents the FI rankings for the regression task concerning **heavy rain** rates during **daytime**. The top two features are the means derived from 5x5 windows of channels C09 and C05, followed by DEM. Subsequent high-ranking features consist mainly of standard deviations and time differences across both VIS and IR channels, along with additional mean values. Notably, IR channel differences and individual channel values don't appear until after the 28th position in the ranking. In terms of normalized importance scores, there is a steady decline, dropping to below 0.4 without any abrupt shifts. However, it's worth noting that the score falls below 0.8 after just the first five features.

In the **nighttime** scenario, Figure 3.8.3.B reveals a feature ranking that diverges considerably from its daytime counterpart. The top four features are standard deviations of channels C09, C08, C14, and C07, followed by an individual channel value for C10, and then DEM. The remainder of the list is populated by spatiotemporal features across various channels, as well as some differences in IR channels, like the C12-C13 which is again the best-ranked among the IR channel differences. Notably, there is a sharp drop in the normalized importance score from the first to the second feature, with the latter hovering around 0.8. The scores then steadily decline, tapering off to just below 0.5.

The final RF model under discussion focuses on the regression of **extreme rain** classes. FI rankings for **daytime** samples are illustrated in Figure 3.8.4.A. The most striking observation is that DEM emerges as the most critical feature. Following DEM, we find the mean and standard deviation of a 5x5 window for channel C05. After these top features, there is a noticeable decline in normalized importance scores, dropping below 0.7. The subsequent rankings include a blend of spatiotemporal features from both VIS and IR channels, as well as a handful of differences in IR channels (e.g. C07-C08). The importance scores for these features continue to gradually diminish, ultimately nearing 0.3.

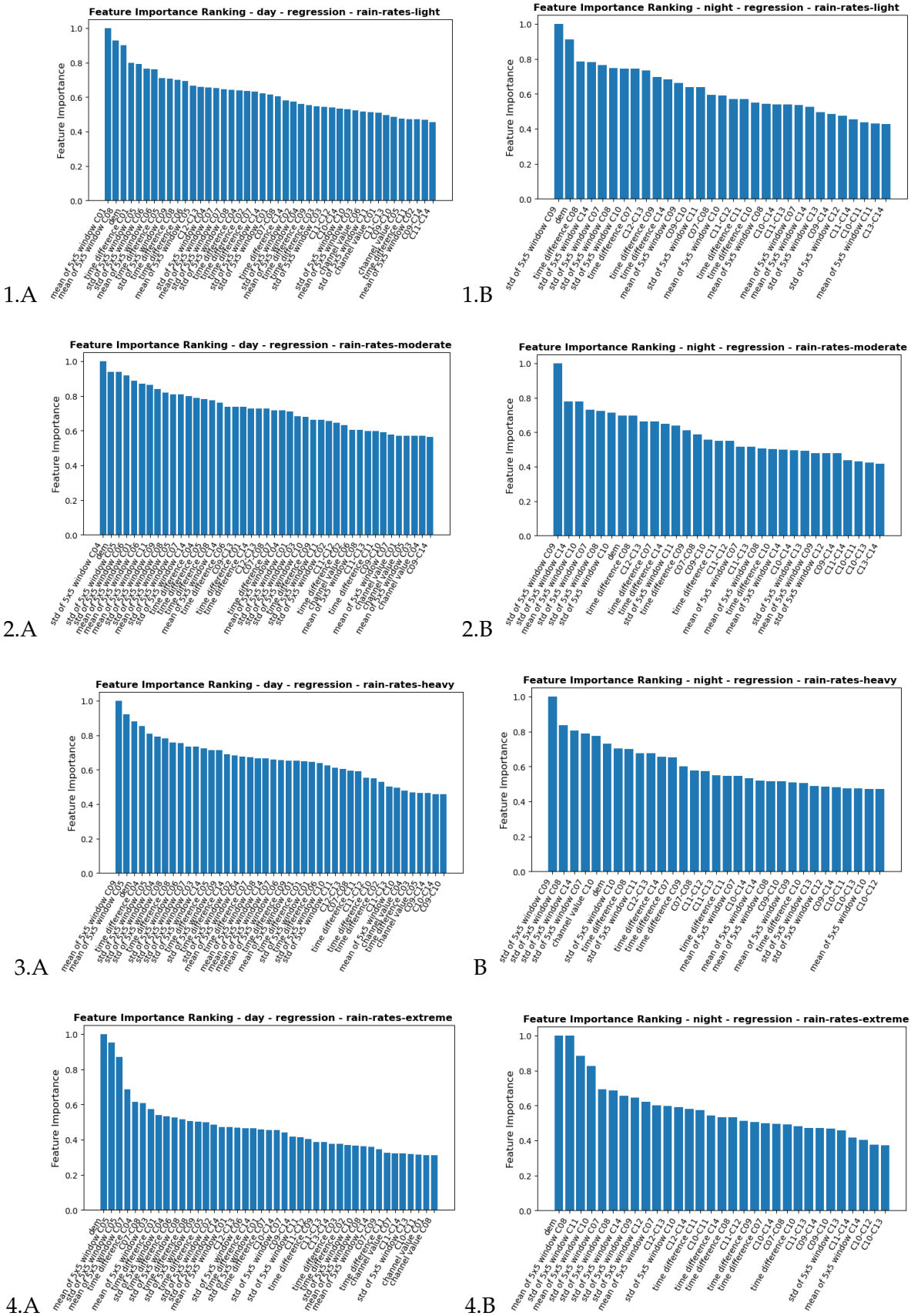

FIGURE 3.8: Feature Importance for Regression Tasks: 1) Light, 2) Moderate, 3) Heavy, 4) Extreme. Displayed for daytime (A) and nighttime (B) training data, showcasing the top half of features.

In the corresponding **nighttime** ranking, presented in Figure 3.8.4.B, DEM retains its position as the most critical feature. Closely following DEM is the mean of a 5x5 window for channel C08, which holds a score just shy of 1. The next two features—namely, the standard deviation of C11 and the mean of C10—have importance scores ranging between 0.8 and 0.9. Subsequent features show a decline in their importance scores, falling below 0.7 and gradually decreasing to nearly 0.3. The list primarily consists of standard deviations initially, later diversifying to include a mixture of time differences, IR channel differences, and other spatiotemporal features.

The analysis of FI rankings in the RF regression models for various classes of rain—light, moderate, heavy, and extreme—reveals a complex interplay of influential variables. Consistently across all classes and time periods, the DEM stands out as a critical factor, highlighting its importance in rain estimation tasks. For estimates of extreme rainfall, the DEM emerges as the primary predictor, underscoring the growing relevance of topographical elements in extreme weather phenomena. Generally, IR channel differences variations hold less significance compared to rain detection cases. Additionally, the RF models show a preference for spatial data over straightforward channel values.

The disparity in FI scores highlights the model's reliance: in certain scenarios, it predominantly depends on a select few features (notably during nighttime), while in others, the influence is more evenly spread, albeit less so for rainfall classification. In essence, achieving a precise rain estimate model necessitates a varied feature set, from DEM and channel spatial and temporal information. This underscores the importance of amalgamating diverse data types for a deeper grasp of rainfall trends.

In the literature, various studies employing RF for rain/no-rain classification sometimes present differing rankings of influential variables. For example, Min et al. 2019 [46] identified the IR channel difference 11.2-12.3 μm as the most predictive variable, followed by the IR difference 7.3-12.3 μm and individual values of IR channels 8.6 and 10.4 μm. These channels also outperformed other non-spectral features, including meteorological variables and second-tier satellite retrievals. When Min et al. conducted RF regressions, the prominence of these channels slightly diminished, ceding ground to modelled weather variables. Yet, the top-ranking feature remained the 7.3-12.3 μm difference.

Turini et al. 2019 [66] also highlighted the significance of the IR 10.8 μm in RF classification, noting its utility in gauging cloud top temperatures and, consequently, cloud heights. Other noteworthy channels included differences like 8.7-10.8 and 10.8-12.0, as well as combinations of water vapour with other IR channels. These selections were associated with cloud water path data, correlating with deep convective clouds that produce heavy rainfall. However, a paramount observation from Turini et al. was the high impact of terrain-related features.

Zhang et al. 2021 discussed FI rankings but found satellite channels generally less influential than features derived from weather modelling for both rain area delineation and precipitation grade estimation. Interestingly, the Satellite Zenith Angle emerged as a prominent feature. Among satellite channels, the IR difference 11.2-8.6 μm was highlighted as particularly significant.

3.2.2 Feature Selection

Using the FI rankings from section 3.2.1, we derived the feature selection plots in this section. From Figure 3.9, we observed the number of chosen features for each **classification** task. While all categorical scores are provided for reference in Table 2.7, our primary metric for determining optimal feature count was the stringent ETS. We aimed to avoid selecting too few features, especially when some metrics showed high scores with minimal features. When scores stabilized, we selected the fewest features to avoid overcomplication.

For the wet-dry classification, we chose 41 and 25 features based on the plateau in the ETS growth. While most metrics exhibit similar trends, BIAS appears to improve with fewer features. For the classification distinguishing light from higher rain classes, we selected 15 and 19 features for day and night, respectively. ETS, like most other metrics, reached a plateau. Notably, some metrics, such as POD and BIAS for the nighttime scenario, declined with more features. For the classification of moderate versus heavy and extreme rain, we chose 29 features for day and 30 for night. The daytime selection was more straightforward, with ETS, CSI, POD, and FAR reaching their optimal points. However, BIAS continued to rise; despite this, we opted not to further increase the feature count to avoid added complexity. For the nighttime scenario, only ETS and FAR plateaued, while other metrics improved with more features. For the sake of simplicity, we capped the feature count early. For the extreme versus heavy rainfall classification, the scarcity of extreme instances made it more challenging. Our primary metric, ETS, was consistently zero, preventing its use for decision-making. While BIAS began high with one feature and continued to grow with the addition of more features, it never surpassed reasonable limits. Consequently, we turned to FAR—which measures heavy rainfall misclassified as extreme, normalized to the predicted heavy counts. Based on the lowest FAR values, we selected 22 features for daytime and 5 for nighttime scenarios.

For the **regressions**, feature selection results are depicted in Figure 3.10. Among all metrics in this study (Table 2.8), mKGE stands out as our primary choice due to its comprehensive and stringent nature. However, our selection wasn't solely based on this score. For instance, in the light rainfall regression, the optimal feature count is 30 for day and 20 for night. While mKGE indicates better results with fewer features, other metrics like CC, CV, and MAE didn't align. Thus, we opted for the feature count where these metrics stabilized. For the moderate regression, features were chosen at 16 and 29 for day and night scenarios, respectively. In the day setting of the heavy regression, the mKGE peaked at its maximum, i.e., 17. This also aligned with the highest CC and favourable values for CV and MAE. Even though other metrics indicated better outcomes with fewer features, we opted against minimizing feature count excessively. For the night scenario, 18 features were selected, striking a balance between ME, CC, CV, and mKGE. While mKGE indicated a preference for fewer features, it wasn't consistent with other metrics, leading us to choose a more median feature count. For the extreme regression, barring the scores for very few features (up to 3), the metrics displayed a stable trend, particularly for p50, CV, MAE, and ME. Given this, we primarily considered mKGE and CV for decision-making. We chose 18 and 32 features for the day and night scenarios, respectively, as these counts corresponded to points slightly above the average on their respective curves.

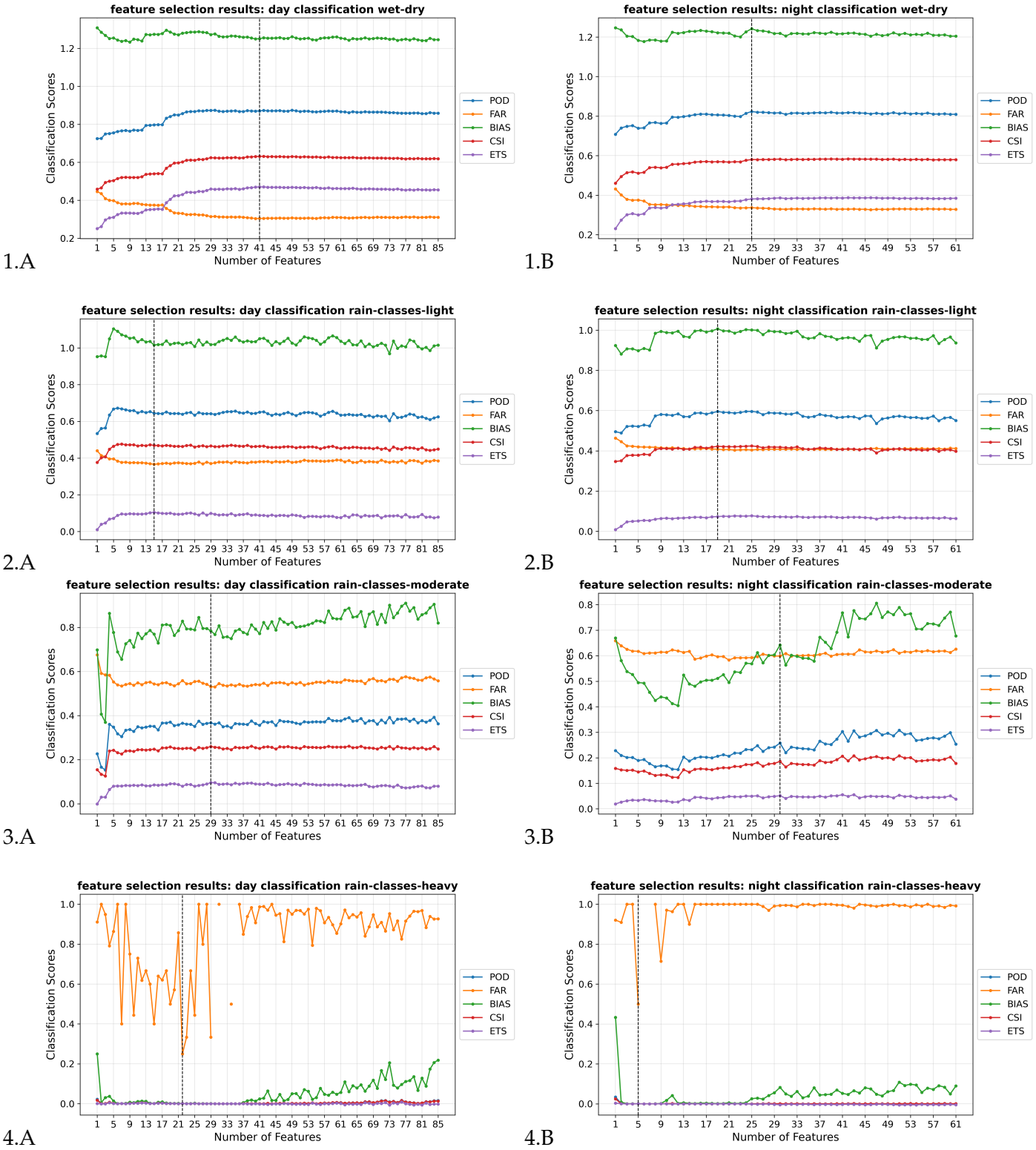

FIGURE 3.9: Feature selection is based on the FI rankings from section 3.2.1. Using this ranking, RF models were progressively trained by incrementally adding features, with performance measured on the test set (See Table2.7). The optimal feature count is indicated by a vertical black line. Columns A and B depict day and night scenarios, respectively. Rows represent classification tasks: 1) wet vs. dry, 2) light vs. the rest, 3) moderate vs. heavy+extreme, and 4) heavy vs. extreme.

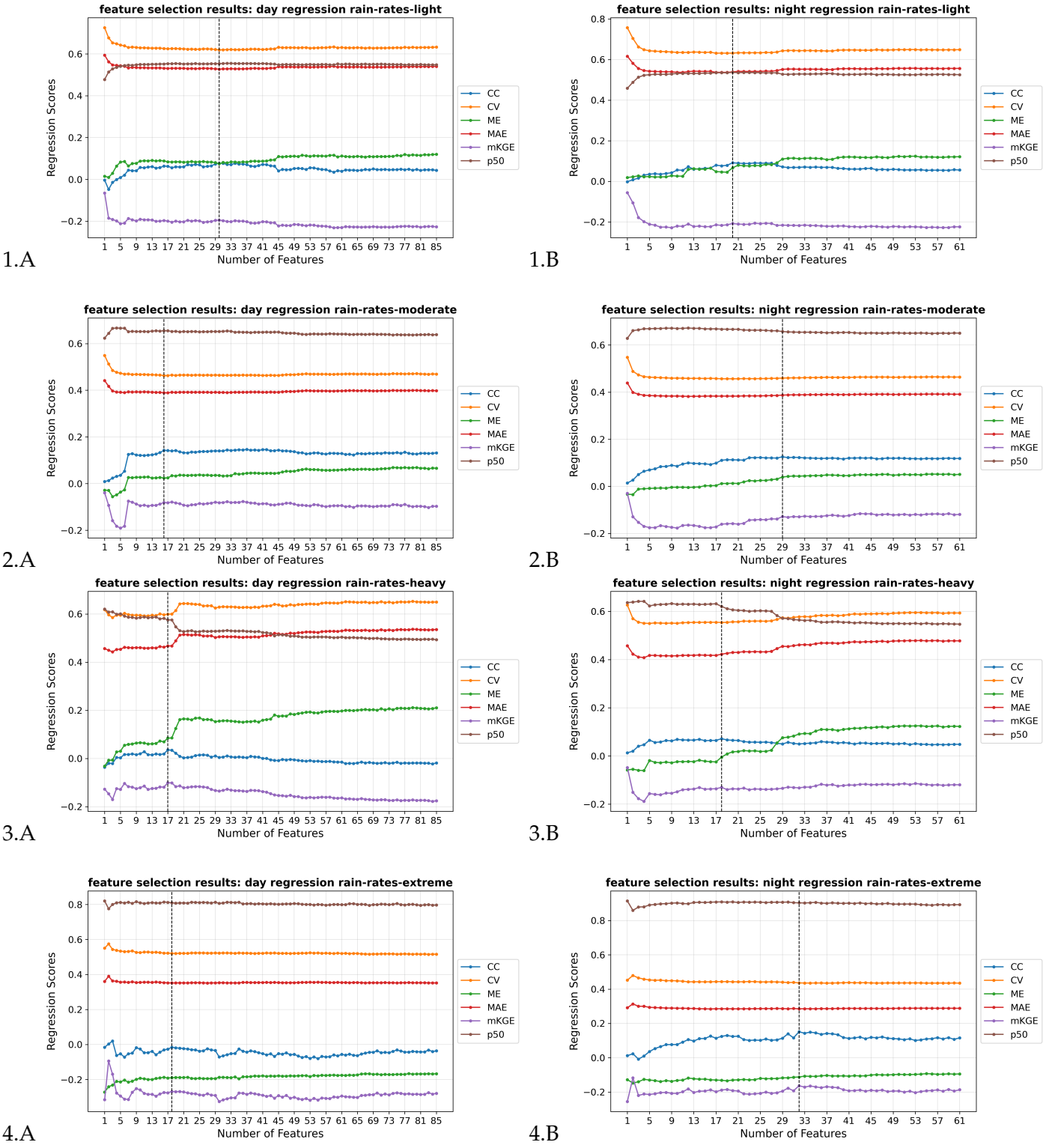

FIGURE 3.10: Feature selection is based on the FI rankings from section 3.2.1. Using this ranking, RF models were progressively trained by incrementally adding features, with performance measured on the test set (See Table2.7). The optimal feature count is indicated by a vertical black line. Columns A and B depict day and night scenarios, respectively. Rows represent regression tasks on these rainfall classes: 1) light, 2) moderate, 3) heavy, and 4) extreme.

3.2.3 Training and Test Results

The next section presents the model's performance on training and test datasets. Using metrics from Section 2.3.4, we evaluate its classification and regression efficacy. We also compare training and test metrics to check for overfitting.

Classifications

Based on the classification performance metrics presented in Table 3.1, we observe a detailed comparison of scores for each task across both daytime and nighttime scenarios. The table also includes scores derived from the training dataset for context. The evaluation metrics—POD, FAR, BIAS, CSI, and ETS—approach optimal values as referenced in Table 2.7. While this suggests really good performance during the training phase (except for the heavy vs extreme classification), it is important to exercise caution in interpreting these results as evidence of generalization. Upon examining the scores obtained from the test dataset, which consists of unseen data, we find that the model's performance tends to degrade, sometimes significantly. These disparities between training and test performance underscore the need for further scrutiny, especially with regard to the model's ability to generalize to new data.

Time	**Classification**	**POD**		**FAR**		**BIAS**		**CSI**		**ETS**	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Day	Wet vs Dry	0.97	0.87	0.06	0.28	1.03	1.22	0.91	0.64	0.83	0.49
Day	Light vs Moderate-Heavy-Extreme	0.95	0.68	0.05	0.37	1.00	1.06	0.90	0.49	0.80	0.11
Day	Moderate vs Heavy-Extreme	0.83	0.33	0.01	0.53	0.84	0.70	0.82	0.24	0.76	0.09
Day	Heavy vs Extreme	0.11	0.00	0.00	0.29	0.11	0.00	0.11	0.00	0.11	0.00
Night	Wet vs Dry	0.95	0.83	0.08	0.33	1.03	1.23	0.88	0.59	0.77	0.40
Night	Light vs Moderate-Heavy-Extreme	0.96	0.62	0.05	0.40	1.01	1.04	0.91	0.44	0.83	0.08
Night	Moderate vs Heavy-Extreme	0.78	0.22	0.01	0.60	0.79	0.56	0.78	0.17	0.71	0.04
Night	Heavy vs Extreme	0.00	0.00	0.00	NaN	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 3.1: Classification Performance Metrics for Day and Night: The 'Train' columns show scores from October 1-26, 2020, and the 'Test' columns represent scores from October 27-31, 2020. Details of categorical metrics can be found in Table 2.7.

For instance, in the case of **wet vs. dry** classification, a notable divergence exists between the training and test metrics. The POD, which gauges the ability of the model to correctly predict observed events, exhibits a drop to 0.87, which, while not optimal, remains respectable. Similarly, FAR increased from 0.06 to 0.28. Turning to BIAS, the value above 1, i.e. 1.22, hints at a tendency to under-predict the 'dry' class, a conclusion further supported by the class distribution depicted in Figure 3.11.1.A. The CSI, calculated as the ratio of 'hits' to all positive samples (both actual and predicted), registers at 0.64. This value is adjusted in ETS to account for random

chance, resulting in a reduced score of 0.49. In nighttime scenarios, we observe a similar trend, albeit with slightly worse scores compared to daytime conditions. Specifically, the POD drops to 0.83, the FAR rises to 0.33, BIAS further decreases to 1.23, CSI settles at 0.59, and ETS falls to 0.40. Figure 3.11.1.A presents the distribution of samples for each class in 'dry vs. wet' classification task, comparing the actual and predicted distributions based on the daytime test dataset. The data indicates that the model slightly underestimates the 'dry' class and overestimates the 'wet' class, thereby tending to equalize the class distribution. Notably, the actual distribution is skewed towards the 'dry' class (with a ratio of approximately 0.6 to 0.4 for 'dry' and 'wet,' respectively), whereas the predicted distribution leans towards a more balanced ratio. Similar behaviour is observed in the nighttime scenario, as shown in Figure 3.11.1.B.

In the task of classifying **light rain from higher-rate categories**, the performance metrics reveal notable shortcomings, particularly in daytime scenarios. For instance, the POD accounts for merely 68% of true labels correctly predicted, reflecting a significant area for improvement. Additionally, the FAR, which quantifies the proportion of falsely predicted light rain events, escalates to 37%. On a somewhat positive note, the BIAS score is relatively commendable, hovering near 1, which suggests that the distribution of the 'light' rain class is well-estimated—a point further corroborated by Figure 3.11.2.A. However, the CSI, representing the proportion of 'hits' relative to all samples except false negatives (encompassing moderate, heavy, and extreme categories), falls below 50%. Moreover, when adjusted for random chance, the ETS experiences a drastic decline to 11%. Similar trends manifest in nighttime scenarios, albeit with slightly inferior performance metrics. Specifically, the POD drops to 62%, and the FAR rises to 40%. Similar to the daytime scenario, the BIAS score is slightly above 1 (1.04), indicating an underestimation of light precipitation. Both the CSI and ETS scores further decline to 0.44 and 0.08, respectively. In the daytime scenario, as illustrated in Figure 3.11.2.A, the actual distribution has a marginally higher prevalence of 'moderate-heavy-extreme' cases. However, the RF model's estimates appear to broaden this gap by underestimating the 'light' category and slightly overestimating the 'moderate-heavy-extreme' category. This trend is consistent with what is observed in the nighttime case, although with more similar class distributions, as depicted in Figure 3.11.2.B.

In the task of differentiating **moderate rain from heavy and extreme categories**, the model exhibits markedly worse performance metrics. For example, the POD scores are quite bad, registering at 0.33 and 0.22 for daytime and nighttime scenarios, respectively. The FAR also shows scores of 0.53 and 0.60 for daytime and nighttime. Furthermore, the BIAS scores are quite far from the ideal value of 1 for both time frames (0.70 for daytime and 0.56 for nighttime), indicating an unbalanced distribution between the true and predicted labels across classes. The CSI sets to a value of 0.24 for daytime and 0.17 for nighttime scenarios. However, it is important to note that when adjusted for random hits, the ETS experiences an even more significant drop, falling to 0.09 for daytime and 0.04 for nighttime. The class distribution for the 'moderate vs. heavy+extreme' classification task is depicted in Figure 3.11.3.A for daytime scenarios. In this case, the true labels show a higher ratio for 'moderate' rain rates (approximately 0.7) compared to 'heavy+extreme' rates (around 0.3). The RF model's estimates appear to exacerbate this gap: the predicted ratio for the 'moderate' class increases to almost 0.8, while it decreases for the 'heavy+extreme' class to approximately 0.2. A slightly worse behaviour is observed in the nighttime case, as shown in Figure 3.11.3.B.

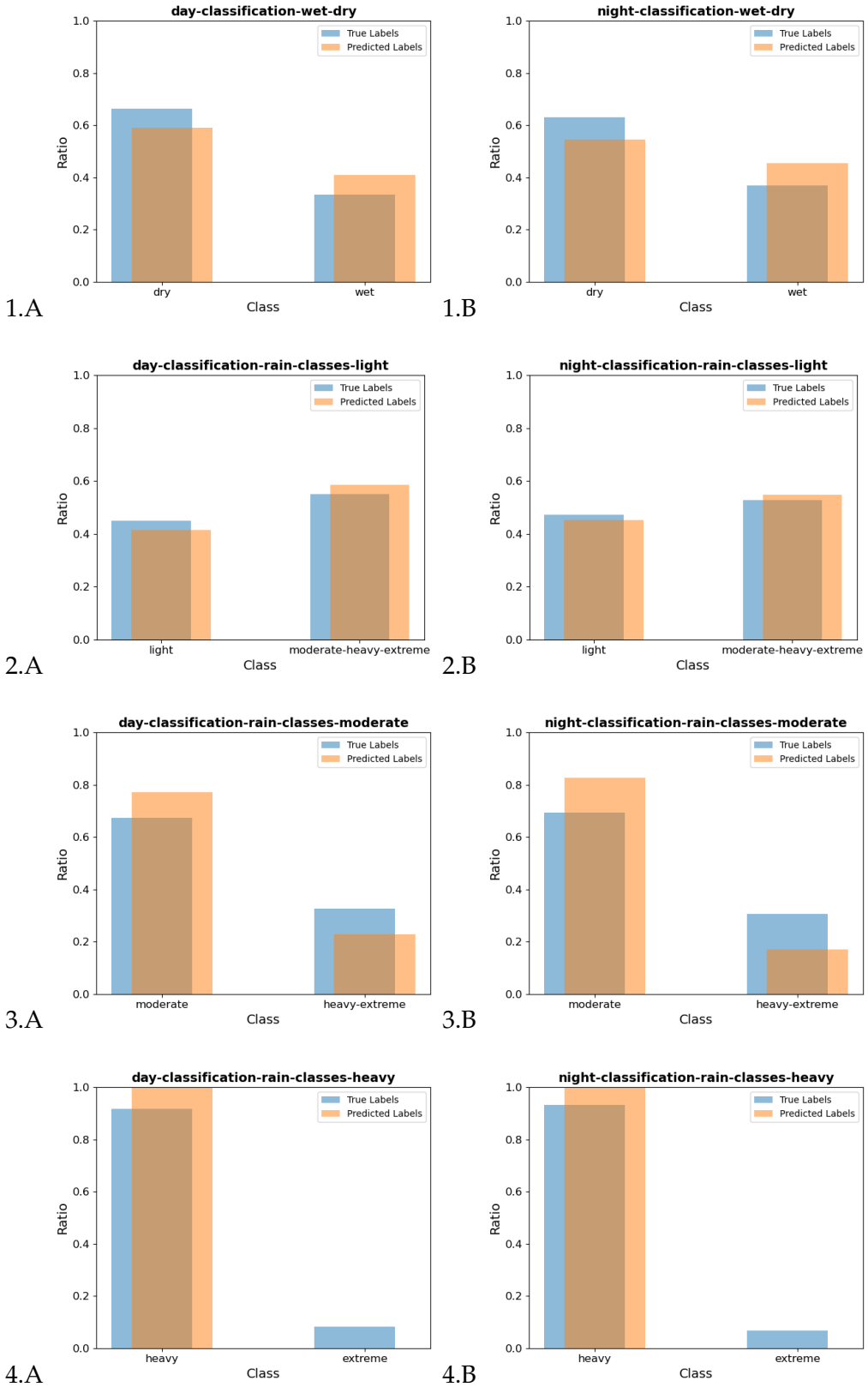

FIGURE 3.11: Class Distribution for Daytime (A) and Nighttime (B) Test Datasets: Blue bars show the percentage of actual labels in each class, while orange bars reflect Random Forest estimates. Results are presented for conditions: 1) Dry vs. Wet, 2) Light vs. Moderate-Heavy-Extreme, 3) Moderate vs. Heavy-Extreme, and 4) Heavy vs. Extreme.

In the classification task aimed at **heavy and extreme rainfall**, the performance metrics display unsatisfactory results. This can likely be attributed to the limited data available for these specific rain rate categories. The FAR records a value of 0.29, indicating a low tendency for the model to classify heavy rainfall pixels as extreme. Strangely, for the nighttime scenario, the FAR displays a NaN value, which means the predicted class for extreme rainfall is absent. Consequently, during nighttime, the RF model fails to predict any extreme precipitation events. The rest of the scores touch the lowest possible value of zero in the test dataset. This brings into focus the model's reliability concerns for heavy and extreme rainfall estimates, underscoring the urgency to refine it, especially considering the critical importance of accurately forecasting such extreme conditions. Figures 3.11.4.A and 3.11.4.B indicate a pronounced imbalance between the two classes in both daytime and nighttime datasets. Notably, the ratio of true labels for heavy rainfall stands close to 0.9, whereas for extreme rainfall, it hovers around a mere 0.1. Such a significant disparity poses challenges in classification. Regrettably, the RF model amplifies this problem, correctly classifying extreme labels to only a handful of pixels during the day and none during nighttime. As a result, the model considerably underestimates extreme rainfall events, a phenomenon that can be attributed to their sparse representation in the dataset.

Our detailed assessment of the RF model across diverse rainfall classification tasks reveals a spectrum of both strengths and weaknesses, offering insightful guidance for subsequent improvements. While the training data yields encouraging initial results, with performance metrics approaching ideal values, the model's effectiveness reveals complexities when tested on unseen data, especially at different times of the day. For tasks like 'wet vs. dry', the results remain commendable. However, the performance gradually declines as the model classifies increasing rainfall intensities, culminating in the least satisfactory outcomes when dealing with the sparse class of extreme rainfall.

Regressions

For regression tasks, we detail the performance metrics for continuous variables in Table 3.2. While the scores from the training dataset are predictably high, given they originate from the calibration data, there's a noticeable decline in the model's performance when evaluated on unseen data.

Time	**Regression**	**CC**		**CV**		**ME**		**MAE**		**mKGE**		**P_{50}**	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Day	Light	0.90	0.09	0.35	0.63	0.00	0.08	0.29	0.53	0.50	-0.18	0.73	0.55
Day	Moderate	0.90	0.15	0.28	0.46	0.00	0.01	0.23	0.39	0.47	-0.07	0.85	0.66
Day	Heavy	0.89	0.04	0.30	0.58	0.01	0.04	0.22	0.45	0.48	-0.12	0.90	0.60
Day	Extreme	0.85	0.07	0.20	0.52	0.01	-0.21	0.13	0.34	0.48	-0.24	0.99	0.83
Night	Light	0.90	0.10	0.38	0.63	0.00	0.07	0.31	0.54	0.45	-0.20	0.72	0.54
Night	Moderate	0.91	0.13	0.29	0.46	0.00	0.02	0.24	0.38	0.42	-0.13	0.83	0.67
Night	Heavy	0.88	0.07	0.33	0.55	0.01	-0.04	0.25	0.41	0.42	-0.15	0.87	0.64
Night	Extreme	0.86	0.16	0.20	0.44	0.01	-0.13	0.14	0.28	0.50	-0.19	1.00	0.92

TABLE 3.2: Regression Performance in Diurnal and Nocturnal Conditions: The 'Train' columns show scores from October 1-26, 2020, and the 'Test' columns from October 27-31, 2020. Continuous metrics are detailed in Table 2.8.

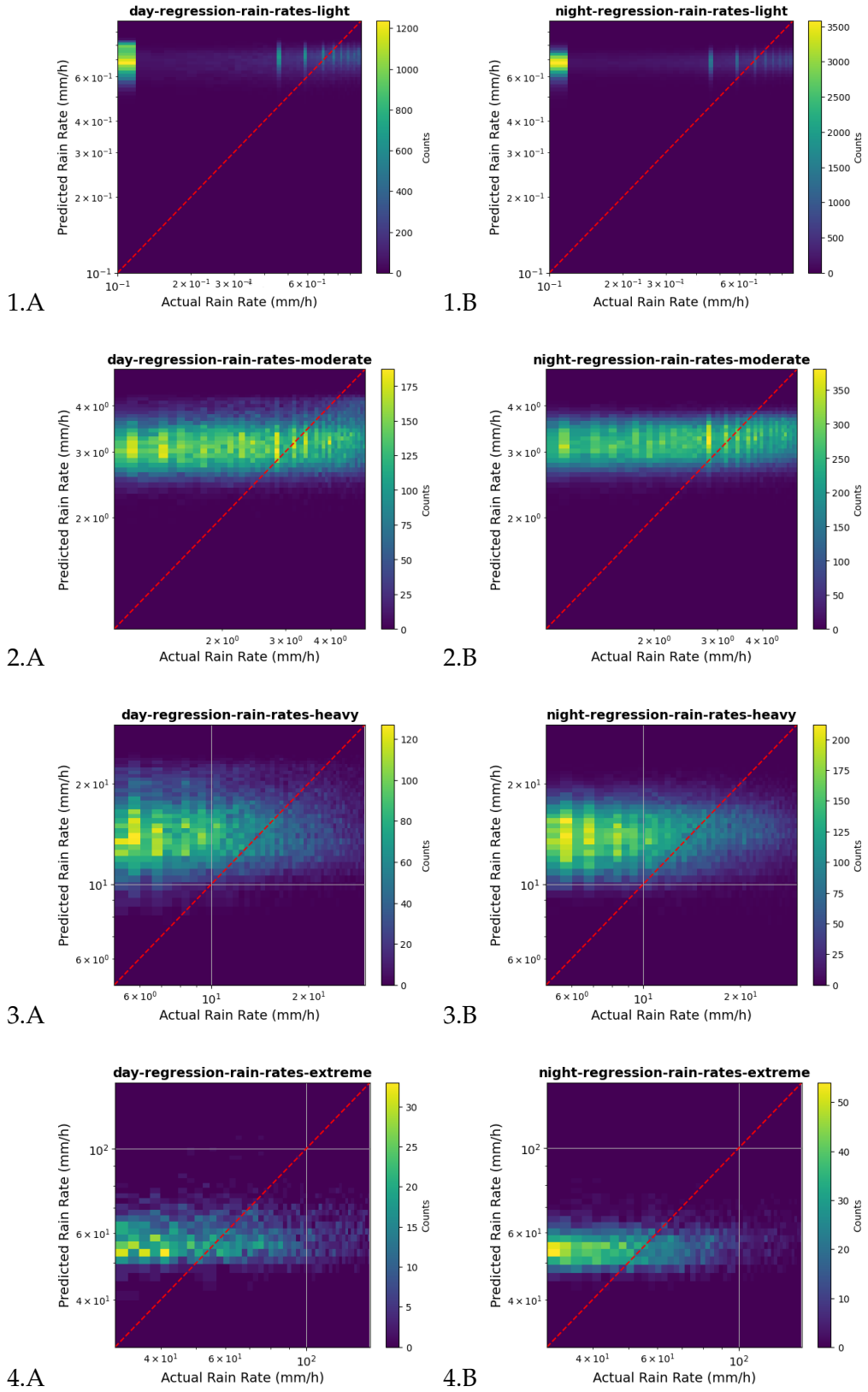

FIGURE 3.12: 2D Histogram of Predicted vs. Actual Values for Daytime (A) and Nighttime (B) Test Datasets: Logarithmic scale aids clarity, with color denoting bin frequency. The red diagonal line represents the ideal model. Rows show regression results for: 1) Light, 2) Moderate, 3) Heavy, and 4) Extreme rain conditions.

In the regression analysis for **light rainfall**, the CC is relatively low, registering at 0.09 for daytime and 0.10 for nighttime scenarios. These figures suggest a lack of strong correlation between actual and predicted labels. Conversely, the CV exhibits values close to zero—specifically, 0.63 for both daytime and nighttime—indicating relatively low data dispersion. The ME is also noteworthy, showing values near the ideal zero mark at 0.08 and 0.07 for daytime and nighttime, respectively. This suggests a slight overestimation, given the positive sign of the ME. However, the MAE, which employs the absolute difference and therefore does not cancel out positive and negative deviations, reveals a somewhat higher score. This suggests that the observed and estimated rain rates may not be as closely aligned as implied by the ME, with MAE values standing at 0.53 for daytime and 0.54 for nighttime. The mKGE presents negative values of -0.18 and -0.20, which are not so good. The implications of these scores will be discussed further in Section 3.2.4. Lastly, the P50 metric, which is generally considered a more lenient measure, also yields mediocre performance, with scores of 0.55 and 0.54 for daytime and nighttime scenarios, respectively. One effective way to evaluate the performance of a regression model is to juxtapose actual and predicted values on a scatter plot. Given the large number of data points in our study, we employ a 2D histogram with logarithmic binning to enhance the visual clarity of the plot. The daytime regression for light precipitation is displayed in Figure 3.12.1.A. Here, it's apparent that the predicted values are confined to the range of 0.6-0.8 mm/h, while the actual values for light precipitation vary widely from 0.1 to 1 mm/h. This suggests that the model tends to concentrate the distribution of predicted rain rates around central values. Interestingly, a distinct vertical striping pattern is also evident, likely stemming from the fine-grained distribution of low rain rates in the radar product (see Figure 2.12.B). A similar trend is evident in the nighttime scenario, as illustrated in Figure 3.12.1.B.

The regression analysis for **moderate rainfall** displays modest improvements in key metrics compared to its light rainfall counterpart. Specifically, the CC shows a marginal increase, standing at 0.15 and 0.13 for daytime and nighttime scenarios. Likewise, the CV is reduced to 0.46 for both time periods, suggesting a more correlated and less dispersed relationship between the true and predicted labels. The ME and the MAE also show enhancements, registering scores of 0.01 and 0.02 for ME and 0.39 and 0.38 for MAE, respectively. These improved scores indicate reduced normalized differences between observed and predicted values in the moderate rainfall scenario, while the positive sign of the ME suggests a slight overestimation. The P50 metric also exhibits a minor improvement, with scores of 0.66 for daytime and 0.67 for nighttime. The mKGE records values of -0.07 and -0.13. Figures 3.12.2.A and 3.12.2.B display the 2D histograms for the regression of moderate rain rates during daytime and nighttime scenarios, respectively. Both figures exhibit similar patterns. Similar to the case for light rain, the RF model tends to narrow the distribution towards the central values, specifically between 2.5 and 4 mm/h, even though the actual moderate rain rates span from 1 to 5 mm/h. Vertical striping patterns are also observed in these graphs, albeit less prominently than in the light rain case. This reduced emphasis on striping could be attributed to the larger logarithmic bins, which may not capture the fine structure of the distribution as effectively.

The regression results for **heavy rainfall** are less than impressive. Specifically, the CC registers at 0.04 for daytime and 0.07 for nighttime, indicating a near absence of correlation between predicted and actual values. The CV scores are 0.58 for daytime and 0.55 for nighttime, suggesting a not-so-high degree of data dispersion. Then, we observe 0.04 and -0.04 for ME, and 0.45 and 0.41 for MAE, respectively. The mKGE records values of -0.12 and -0.15, while the P50 metrics are 0.60 for daytime and 0.64

for nighttime. Although nighttime performance is somewhat better than daytime, these metrics should be interpreted alongside the visual data presented in Figure 3.12.3.A-3.12.3.B for a more comprehensive understanding. Both figures are quite similar in structure, allowing for a unified discussion. Like the previous cases, the data points are mostly off the diagonal line, which represents the ideal location for perfectly predicted values. The model primarily captures the central values of the distribution, approximating rain rates only within the 10 to 25 mm/h range. Rates within 5-10 mm/h and 25-30 mm/h are largely treated as noise, indicating that the model fails to accurately represent these values.

The final regression task focuses on **extreme rainfall**. Notably, the CC for daytime is 0.07, while it registers at 0.16 for nighttime. The CV scores are relatively moderate, recording at 0.52 for daytime and 0.44 for nighttime. It is important to highlight the ME values of -0.21 and -0.13, which indicate an underestimation of rain rates for this category. Although the MAE shows the lowest values among the tasks at 0.34 for daytime and 0.28 for nighttime, the mKGE presents scores of -0.24 and -0.19, which are not superior when compared to other regression tasks. The P50 metrics, however, display promising performance, with scores of 0.83 for daytime and 0.92 for nighttime. Generally, the model performs better during nighttime. However, it's essential to highlight that the dataset for extreme events is notably limited. If the initial classification fails to detect extreme rainfall, even a reasonably accurate regression becomes largely irrelevant. The comparison between the actual and predicted labels for extreme rain is displayed in Figures 3.12.4.A and 3.12.4.B. Both daytime and nighttime scenarios exhibit similar patterns, allowing for a joint discussion. Once more, the RF model struggles to accurately represent values at the extreme ends of the rain rate distribution. Predominantly, the predicted rain rates are clustered within the 50-75 mm/h range. The earlier-observed striping patterns are no longer discernible, likely because the log-bin size has increased to the point where it obscures finer details in the distribution of rain rates.

In summary, while our RF model shows promise in predicting rainfall intensities, it faces challenges with unseen data and tends to bias estimates around median values. It struggles to capture patterns like vertical striping in 2D histograms and often misses extreme rainfall events due to limited data. For applications demanding precision, refinements are essential, particularly in capturing extreme events and detailed patterns.

3.2.4 Validation Results

The subsequent stage in evaluating the effectiveness of our RF models involves comparing their outputs with data from rain gauges and other available rainfall products in Vietnam during the test period, from October 27 to 31, 2020. To start, we amalgamate the outcomes from multiple regression models to create a comprehensive spectrum of rain rates, ranging from 0.1 mm/h to 150 mm/h, while accounting for the continuous time period covering both day and night. Rain rates falling below the 0.1 mm/h threshold are set to zero. This aggregated data forms the basis of our rainfall product, which we have designated as **FY-4A_RF**. The evaluation presented in this section follows the validation methodology outlined in the study by Roversi et al. 2023, which is currently under review [59]. The figures included here have been modified from that study to reflect the specific test period examined in this work, as well as to incorporate the FY-4A_RF product, which is unique to this study.

Spatial Distribution of Rain Gauges Data

To begin, let's delve into a comprehensive analysis of the reference data, which in this case is sourced from rain gauges. While Chapter 2 already covered the histogram of rain rate distribution in Figure 2.12, we further enrich this overview with spatial statistics depicted in Figure 3.13. Consistent with Vietnam's climatology, the figure presents the spatial distribution of four key time-aggregated metrics, created using the nearest neighbour algorithm for seamless spatial mapping. Over the 5-day test period, the data reveals distinct regional trends: Central Vietnam experienced the highest frequency of wet hours, while the northeast and northwest mountainous areas were comparatively drier. There are also specific zones in the southern part of the country that registered higher levels of precipitation. Moreover, the spatial distribution of maximum recorded rain rates and time-based standard deviations aligns well with the patterns for average rain rates and frequency of wet hours. Specifically, the highest values are concentrated in the central lowlands, which also coincide with the typhoon's landfall area. These regions additionally exhibit greater variability, transitioning rapidly between dry conditions and extreme rainfall events.

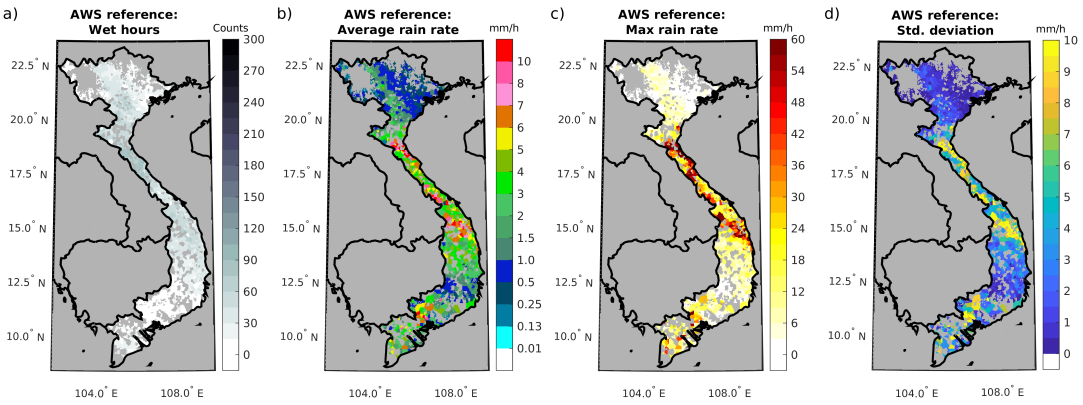

FIGURE 3.13: Map displaying assorted statistics related to the AWS reference dataset. (a) Count of wet hours in each grid cell. (b) Mean rain rate (mm/h) during wet periods. (c) Peak rain rate (mm/h) recorded during wet hours. (d) Variability in rain rate (mm/h) throughout wet periods. Image adapted from [59].

Rain Products Snapshot

In our evaluation, let's turn our attention to the spatial distribution of all the products involved in this validation stage. Figure 3.14 presents the rainfall rates for October 28, 2020, at 08:00 UTC. Notably, the figure reveals substantial variations in how each data source captures the meteorological conditions. Specifically, this snapshot captures the landfall of Typhoon Goni in Vietnam in late October 2020, as elaborated in Section 2.1.2.

Using the **rain gauge** data in panel a as a reference, we observe the heaviest rainfall in central Vietnam, featuring intense spikes just above latitude 15°N. **Radar** data in panel b exhibit similar trends but notably miss the peak values exceeding 40 mm/h. Geostationary satellites FY-4A and GK-2A (panels d and e, respectively) display contrasting features. **FY-4A** underestimates the core rainfall intensity of the cyclone, failing to capture the peak red zone, and also indicates regions with zero rainfall. In contrast, **GK-2A** portrays the cyclone's centre as intensely wet while

estimating light rain over almost the entire observed area, resulting in a reduced moderate rain zone compared to FY-4A.

The two versions of **IMERG** (panels f-g) exhibit closely related patterns, although the Final run captures more intense rainfall near the cyclone's centre. **ERA5-Land** (panel h) is characterized by its smoothed transitions between different rainfall regimes, rendering the areas affected by the typhoon more rounded compared to other products.

Finally, our focus shifts to panel c, showcasing the product developed in this study, the **FY-4A_RF**. The cyclone's shape is vividly represented, including a small centralized area of intense rainfall. However, this product diverges notably from FY-4A. It captures more heavy rainfall surrounding the cyclone's centre and detects lighter rain at a greater distance, showing a broader reach compared to even the radar data.

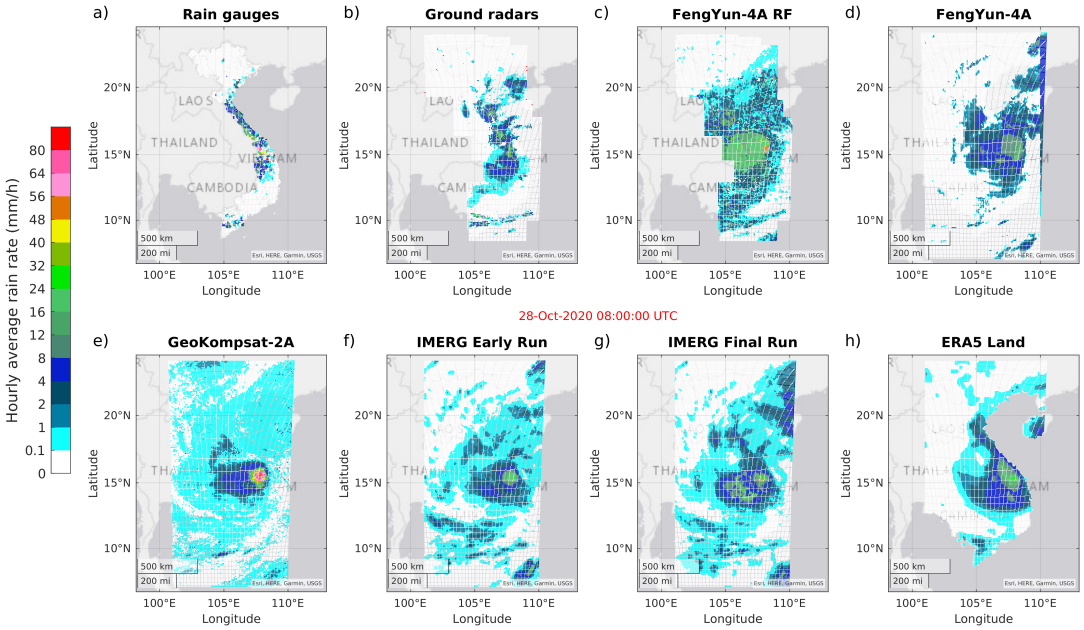

FIGURE 3.14: Map displaying a spatial snapshot of all the available products for the 28th of October 2020 at 08:00 UTC. (a) Rain gauges. (b) Radars (c) FY-4A_RF (d) FY-4A (e) GK-2A (f) IMERG Early (g) IMERG final (h) ERA5-Land. Image adapted from [59].

Probability Density Function

To initiate the comparative analysis with other rain products, we evaluated the probability density functions (PDFs) of all eight products' estimations at AWS locations. The PDFs were derived over a span of 5 days, encompassing all grid points with at least one active AWS. We utilized a bin size of 2 mm/h, and limited the PDF to 100, as subsequent rain rates lacked sufficient samples to construct statistically reliable curves. The curves begin at approximately 1.2, as it represents the centroid of the initial bin. The results are displayed in Figure 3.15.

The **AWS** data, indicated by a black line, maintains a consistent curve up to 30 mm/h. Minor inconsistencies at higher rainfall rates are likely due to noise, given the smaller sample size. **Radar** data, marked by a green line, closely mirrors the AWS measurements across a range of intensities and exhibits similar noise levels. However, the radar data tend to slightly overestimate AWS measurements between

1 and 20 mm/h, becoming too noisy to determine their relative positions at higher rates.

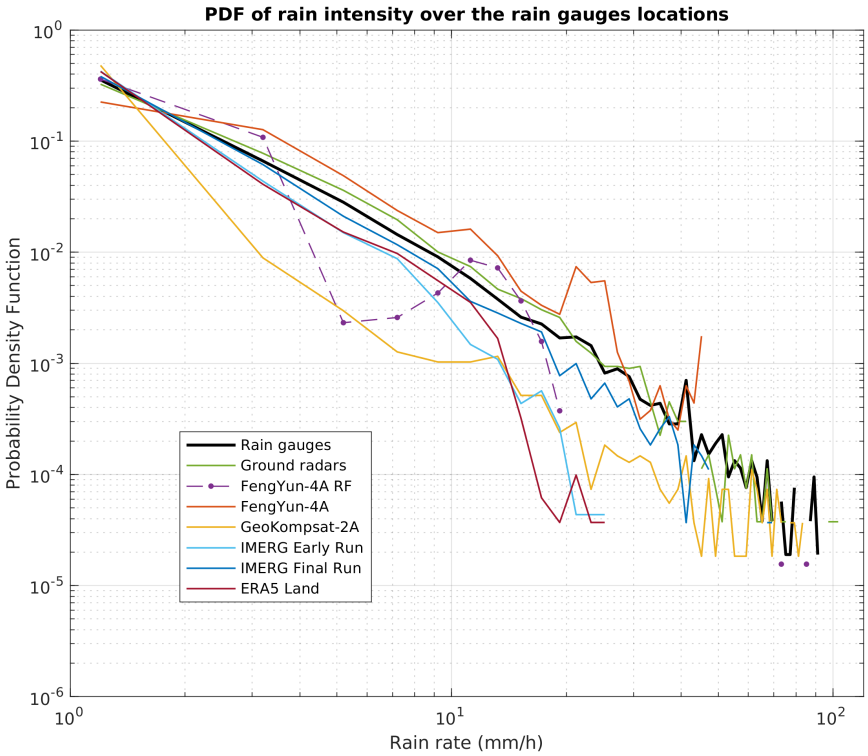

FIGURE 3.15: The distributions of both observed and forecasted rainfall intensities are depicted as probability density functions (PDFs), with the areas under the curves standardized to one. The distribution related to our FY-4A_RF product is indicated by individual points connected with a dashed line. The Rain Gauges distribution is represented by a bold black line, while other products are illustrated with simple coloured lines. Both the x and y-axes are displayed on a logarithmic scale. This image is adapted from [59].

The products from the geostationary satellites **FY-4A** and **GK-2A**, coloured in orange and yellow respectively, exhibit a smooth trend at lower rates up to 10 mm/h, beyond which the curve displays increased noise likely due to scarce data points. Notably, FY-4A tends to underestimate rainfall rates up to approximately 2 mm/h and then overestimate up to around 30 mm/h. In contrast, GK-2A exhibits the opposite behaviour, overestimating low rates and then underestimating higher ones. GK-2A also extends its detection to higher rates, similar to AWS, whereas the FY-4A PDF ceases around 50 mm/h.

The multi-platform **IMERG** products (Early run in light blue and Final run in dark blue) align well with AWS measurements at low rates. However, they start to diverge at 1.5 mm/h for the Early run and 3 mm/h for the Final run, consistently underestimating higher rates. The Final run extends its estimates up to 50 mm/h, whereas the Early run maxes out around 30 mm/h.

The model-based **ERA5-Land** product, depicted in red, initially parallels the IMERG Early run up to 7 mm/h but then consistently underestimates rainfall compared to AWS, capping at 30 mm/h.

Our product, **FY-4A_RF**, exhibits an anomalous trend, initially overestimating rainfall up to 2.5 mm/h and then underestimating up to 10 mm/h. After a further overestimation range at 10-11 mm/h, it underestimates higher rates, displaying a

noisy and inconsistent curve. This wavering pattern is likely attributed to the 'stripe' estimation approach used for various rainfall categories (light, moderate, heavy, extreme), which concentrates the distribution around the midpoint of each class. Due to the sparsity of extreme rainfall events, the PDF for this category exhibits only a couple of points, suggesting that the model is not well-calibrated for such conditions.

Hourly-based Scatterplots

In the following analysis, each of the seven rainfall products is scrutinized against the benchmark measurements taken from rain gauges. We've preserved the highest possible level of spatial and temporal detail within the scope of our study, with scatterplots represented in Figure 3.16. Subplot 'a' uses logarithmic binning to detail the distribution of reference rain rates. Subplots 'b' to 'h' provide scatterplots that juxtapose the performance of each product (depicted on the y-axis) against the reference measurements (shown on the x-axis).

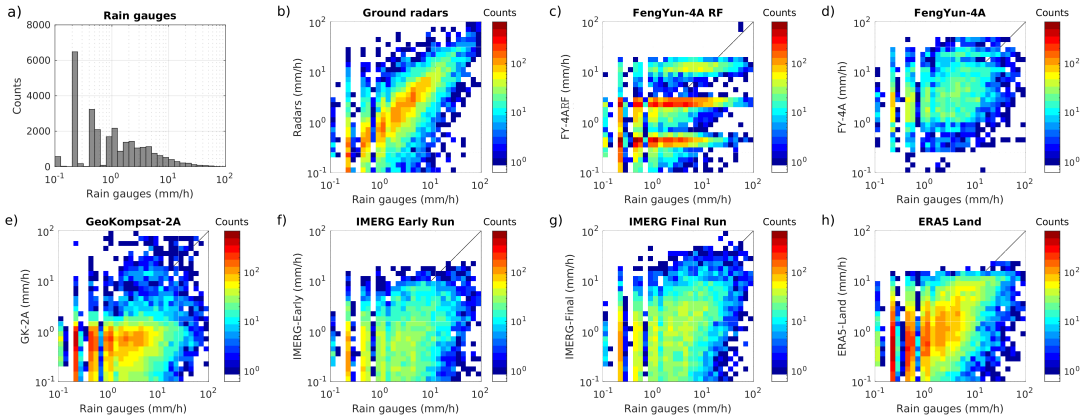

FIGURE 3.16: Distribution of rain rates across all evaluated hours and chosen grid boxes. (a) Frequency chart based on AWS reference data. (b-h) Density plots comparing various products (y-axis) with matching AWS observations (x-axis): Radars (b); FY-4A (c); FY-4A_RF (d); GK-2A (e); IMERG-Early (f); IMERG-Final (g); and ERA5-Land (h). All axes are set to a logarithmic scale. Image adapted from [59].

Logarithmic binning offers a refined x-axis resolution compared to AWS rain gauges, which predominantly have a 0.2 mm/h resolution, with a few having a finer 0.1 mm/h resolution. This leads to distinct peaks in the histogram displayed in Figure 3.16.a, resulting in observable vertical bands on the left margins on the rest of the subplots.

The ground **radar** data are the only product demonstrating a strong correlation, with most data points situated along the main diagonal. Nonetheless, there are some outliers near the axes, indicating false alarms and missed detections. Subplots 'd' and 'e', which represent the geostationary products, show varying degrees of estimation accuracy. The **FY-4A** data points are dispersed and not well-aligned with the diagonal, displaying an inclination to overestimate precipitation rates. The **GK-2A**, in contrast, generally underestimates rainfall, as confirmed by the PDF in Figure 3.15.

Both versions of **IMERG** reveal significant discrepancies from the benchmark data at an hourly scale, as illustrated in Figures 3.16.f-g. The Final run of IMERG shows data points more evenly distributed around the diagonal, particularly for rain

rates below 10 mm/h. However, it struggles with correlation, especially at higher rainfall rates. The Early run presents additional issues, most notably underestimating higher rainfall rates and seeming to cut off above 30 mm/h.

ERA5-Land also leans toward underestimation, particularly for rates exceeding 1 mm/h, and as the IMERG Early run, no estimates go beyond 30 mm/h.

Our own product, **FY-4A_RF**, exhibits the horizontal striping pattern previously noted in its PDF. The data show limited correlation, with a predominance of vertical and horizontal structures in the scatterplot. Evidently, data points tend to cluster around the mid-range of each rain category: around 0.6-0.8 mm/h for light rain, 3-4 mm/h for moderate rain, and approximately 10 mm/h for heavy rain. This pattern underscores the overall lack of correlation in the data. Additionally, the RF model appears to cap the rain rate at the heavy class, with a few extreme points near 100 mm/h.

Performance Metrics

In Table 3.3, we present a comprehensive set of validation metrics for both categorical and continuous evaluation. The **categorical metrics** specifically focus on the classification task of distinguishing between dry and wet samples, using the AWS product as a reference. It should be noted that our FY-4A_RF is a composite outcome generated from prior regression models covering various precipitation intensities ranging from light to extreme. Samples identified as dry were assigned a value of zero.

Beginning with the FAR, ground radars exhibit the lowest score at 0.16, followed by IMERG-Final with 0.39. All other products have values exceeding 0.40. Our FY-4A_RF product does not fare the worst, registering at 0.47, but both GK-2A and ERA5-Land scored 0.49, indicating nearly half of their classified 'wet' samples were actually 'dry' according to the rain gauges.

For the POD, FY-4A_RF ranks second with a score of 0.65, surpassed only by ERA5-Land at 0.78. Aside from GK-2A at 0.52, the remaining products all scored below 0.5. Specifically, FY-4A scored a meagre 0.18, revealing that our FY-4A_RF product substantially outperforms its Fengyun product counterpart in correctly predicting wet samples.

BIAS scores mirror these trends. FY-4A_RF, GK-2A, and ERA5-Land have scores slightly above 1 (1.22, 1.03, and 1.54, respectively), indicating a slight overestimation of actual wet samples. In contrast, the remaining products all score under 1, suggesting they underestimate wet instances and are subpar in precipitation detection. Notably, FY-4A registered the worst score at 0.30, while even ground radars posted a mediocre 0.51.

The CSI and the ETS offer additional insights. ERA5-Land leads in CSI with 0.45, followed by FY-4A_RF and ground radars at 0.41. GK-2A is also noteworthy with a score of 0.35, while IMERG Early and Final, along with FY-4A, display poor scores ranging from 0.16 to 0.26. In terms of ETS, which accounts for random hits, the ranking sees minor alterations. Ground radars emerge as the top performer with a score of 0.30, succeeded by ERA5-Land at 0.26 and our FY-4A_RF product at 0.25. FY-4A continues to underperform, registering only 0.09.

In summary, when it comes to precipitation detection—i.e., the ability to distinguish between rainy and non-rainy conditions—ERA5-Land and radars are the most reliable products. However, the Random Forest-based FY-4A_RF estimates also hold their own, outperforming several other products, including FY-4A, in specific metrics.

In evaluating the accuracy of estimated rain rates, we relied on **continuous metrics** as outlined in Section 2.3.4. This analysis was confined to 'wet' samples, defined as those with rain rates exceeding 0.1 mm/h, and aimed to quantitatively substantiate observations made in Figure 3.16. The associated scores for these metrics are detailed in Table 3.3.

Starting with the CC, Radar products exhibit the highest value of 0.69, corroborating graphical evidence. Following the Radar, IMERG Early, IMERG Final, and ERA5 also report commendable scores of 0.36, 0.42, and 0.43, respectively. Our FY-4A_RF product trails with a score of 0.36, while the geostationary products report the lowest correlation coefficients of 0.24 and 0.21.

In terms of the ME, our FY-4A_RF product performs relatively well, coming with a score of -0.40. Only ERA5 behaves better with a ME of -0.42. All other products exhibit more negative ME values, indicating that the estimated rain rates are, on average, lower than the observed rates. Thus, underestimations are more prevalent than overestimations. The least favourable ME scores are attributed to GK-2A and IMERG Early, registering at -0.73 and -0.84, respectively.

Product	FAR	BIAS	POD	CSI	ETS	CC	ME	MAE	CV	p50	mKGE
Radars	0.16	0.51	0.42	0.39	0.30	0.69	-0.45	0.66	1.68	0.74	0.33
FY-4A_RF	**0.47**	**1.22**	**0.65**	**0.41**	**0.25**	**0.36**	**-0.44**	**0.89**	**2.13**	**0.56**	**0.22**
FY-4A	0.41	0.30	0.18	0.16	0.09	0.24	-0.60	1.03	2.41	0.68	-0.15
GK-2A	0.49	1.03	0.52	0.35	0.19	0.21	-0.73	0.92	2.40	0.50	-0.50
IMERG-Early	0.43	0.44	0.25	0.21	0.11	0.36	-0.84	0.92	2.25	0.53	-0.13
IMERG-Final	0.39	0.52	0.31	0.26	0.16	0.42	-0.67	0.89	2.13	0.53	-0.04
ERA5-Land	0.49	1.54	0.78	0.45	0.26	0.43	-0.42	0.79	2.05	0.21	0.22

TABLE 3.3: Comparative Analysis of both Categorical and Continuous Metrics Scores: This table evaluates the performance of the random forest-based rain product against other existing products, using rain gauge data as the ground truth.

When considering the MAE, which focuses on the magnitude of the error irrespective of its direction, Radar leads with a score of 0.66. The following products in ascending order are ERA5 (0.79), IMERG Final and FY-4A_RF (0.89), GK-2A and IMERG Early (0.92). Although not outstanding, our product outperforms the original FY-4A, which scored 1.03.

In the assessment of CV, which measures the dispersion as depicted in Figure 3.16, the Radar showcases the least variance with a score of 1.68. All other products surpass a CV of 2.0. Among them, ERA5 fares marginally better with a score of 2.05 compared to our FY-4A_RF product, which registers at 2.13. Both IMERG variants log scores of 2.25 (for the Early run) and 2.13 (for the Final run), while the two geostationary products performed with the highest scores of 2.41 and 2.40.

Regarding the P50 score, some inconsistencies are evident. While Radar maintains its superior performance with a score of 0.74, ERA5 disappointingly registers at 0.21. Our FY-4A_RF scores 0.56, surprisingly falling short of the original FY-4A product, which scored 0.68. GK-2A and both IMERG versions have comparable scores of around 0.5.

The **mKGE index** warrants special attention as a comprehensive metric that encapsulates multiple measurements into a singular value. As detailed in source [27], the mKGE index can be subdivided into three distinct components. These are represented on the x, y, and z axes of a three-dimensional graph: the CC, the ratio of global mean values between the test and reference datasets (μ_e/μ_o, equivalent to 1

+ ME), and the ratio of normalized standard deviations for each dataset ($\frac{\sigma_e/\mu_e}{\sigma_o/\mu_o}$). Calculations for these metrics are carried out in comparison to the AWS as a reference standard. An optimal point in this 3D representation occurs at coordinates (1, 1, 1) when evaluating the reference against itself. This multi-faceted breakdown of the mKGE index for the seven assessed products is visualized in Figure 3.17.

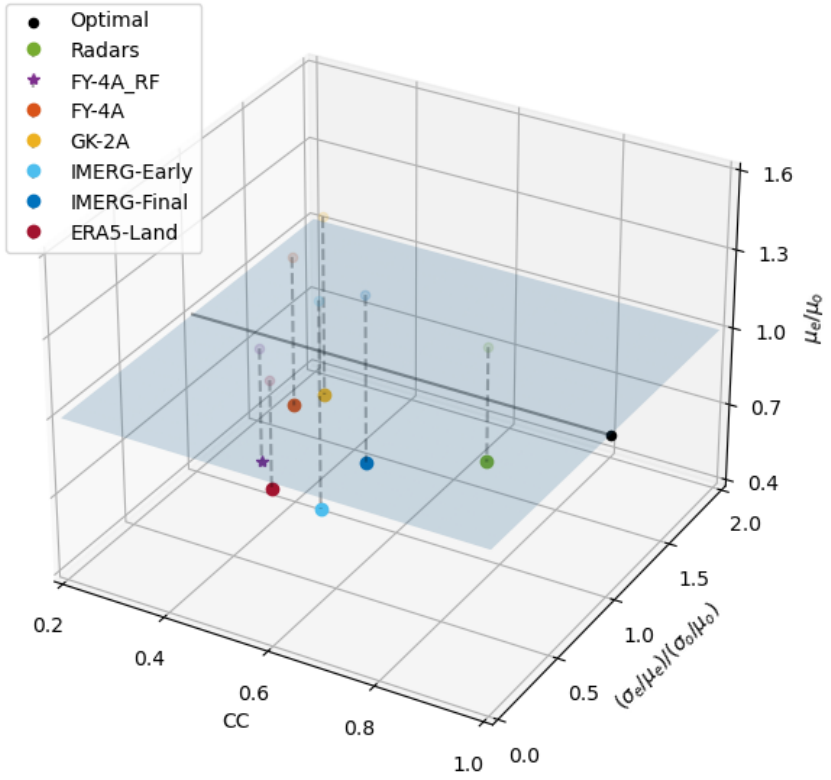

FIGURE 3.17: Visualization of mKGE decomposition in a 3D Euclidean space. The optimal point at coordinates (1,1,1) is indicated in black. The axis labelled (CC, 1, 1) is highlighted with a black line. A plane representing $\mu_e/\mu_o = 1$ is shaded in semi-transparent blue. Lighter, semi-transparent points indicate the projections of data points onto this plane, and these projections are connected by dashed lines. The scales for the three axes are not uniform. This image is adapted from [59].

A broad observation shows that all product points lie below the plane where $\mu_e/\mu_o = 1$, signifying that the mean values of all products are less than those of the reference. Upon examining the distance from the black line (located at coordinates CC,1,1), it's apparent that all dots, except for ERA5 and FY-4A_RF, lie to their right when viewed from above. ERA5 uniquely has a normalized standard deviation lower than the AWS reference. Despite having a lower normalized standard deviation than the reference, our product (FY-4A_RF) remains relatively close to the black line. The most disparate product in terms of normalized standard deviation is GK-2A. In terms of the CC, radar products exhibit the best performance, reaffirming previous observations. The geostationary products are among those with the lowest CC scores. While our product, the IMERG versions and ERA5 have similar CC values. Specific mKGE values, as tabulated in Table 3.3, reveal that radar leads the pack with an outstanding score of 0.33. Following radar, ERA5 and FY-4A_RF secure

the second position with a score of 0.22. All remaining products recorded negative mKGE scores, with GK-2A performing the worst at -0.50.

In summary, our comprehensive evaluation of rain rate estimation products, grounded in the metrics laid out in Section 2.3.4, reveals a number of important findings. Firstly, Radar products consistently outperform others across multiple metrics, confirming their efficacy for estimating rain rates, and justifying its use as ground truth for training our RF model. This aligns with its highest CC and lowest MAE and CV scores. Second, ERA5 generally performs well but shows some inconsistencies, particularly in the P50 score. Thirdly, our FY-4A_RF product demonstrates a promising yet varied performance. With its low ME and MAE, it surpasses the primary competitor, FY-4A, in several metrics and secures the second position in the mKGE index, a comprehensive measure. However, there is room for improvement. The geostationary products, along with our FY-4A_RF, yield the lowest CC scores, pointing to limited accuracy in capturing the correlation with observed rain rates. Furthermore, all tested products consistently underestimate rain rates, as evidenced by the negative ME scores and positions below the $\mu_e/\mu_o = 1$ plane in the mKGE decomposition. In terms of the mKGE index, a nuanced evaluation revealed Radar as the best-performing product, with our FY-4A_RF not far behind. Notably, GK-2A lags significantly in multiple metrics, most critically with a negative mKGE score, highlighting its limitations in accurate rain rate estimations.

Time-averaged Analysis

The subsequent findings delve into an extended temporal analysis of the test dataset, encompassing the full-time scale under consideration. Time-averaged rain rates were computed for each grid point to facilitate this in-depth analysis. The resulting map, visualized in Figure 3.18, displays these **average rain rates** in units of mm/h. To produce a comprehensive spatial representation, we expanded the AWS data from its designated grid box to the complete study area using a nearest-neighbour algorithm. Areas where AWS stations are present but data is missing are marked in grey on the map.

To enhance interpretability, Figure 3.19 graphically illustrates the **biases** in all tested products, normalized against the AWS reference data. A colour gradient transitions from lighter to darker red to signify increasing overestimation, while shades of blue indicate underestimation. The mean field derived from the AWS data during the test period is displayed in both Figure 3.18.a and Figure 3.19.a, consistent with the representation previously shown in Figure 3.13.b.

AWS data indicate that the highest average precipitation levels are centrally located, exceeding 10 mm/h, with additional peaks in the southern region and the driest areas in the northeast. Most products follow a similar spatial distribution but have unique characteristics. For instance, **radar** product displays high central rain rates but falls short of reaching the elevated levels observed in the south. This is substantiated by the predominantly blue bias map, suggesting that radar generally underestimates precipitation.

The geostationary satellites, **FY-4A** and **GK-2A**, show high central rain rates but with a slight southeastern shift compared to AWS data. This localized overestimation near the coast is particularly prominent for GK-2A, as evidenced by the bias map in Figure 3.19.d-e. In contrast, **IMERG** products, displayed in Figure 3.18f-g, indicate the highest average precipitation in central areas, albeit slightly more towards the south. The bias maps, shown in Figure 3.19f-g, corroborate this behaviour.

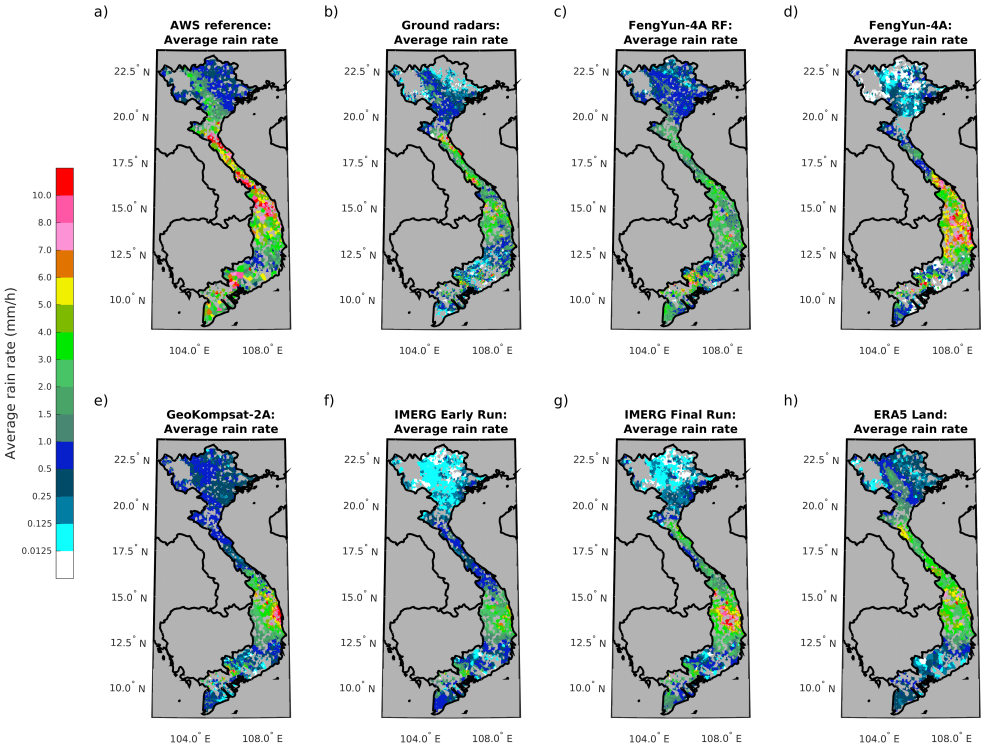

FIGURE 3.18: Mean maps of rain rates from these products: (a) AWS; (b) Radar; (c) FY-4A_RF; (d) FY-4A; (e) GK-2A; (f) IMERG-Early; (g) IMERG-Final; (h) ERA5-Land. Adapted from [59].

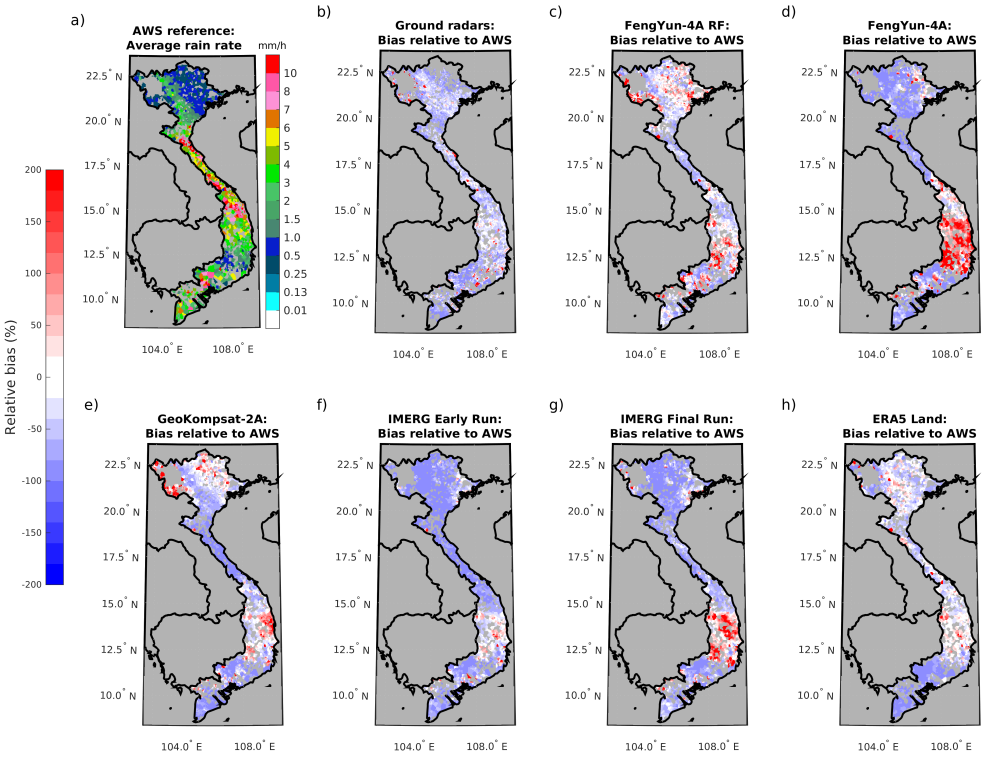

FIGURE 3.19: BIAS maps of rain rates from these products: (a) AWS; (b) Radar; (c) FY-4A_RF; (d) FY-4A; (e) GK-2A; (f) IMERG-Early; (g) IMERG-Final; (h) ERA5-Land. Adapted from [59].

ERA5-Land reports elevated average rain rates in the central areas but falls short when compared to rain gauge measurements, particularly in the southern regions. The bias map reveals this strong underestimation in the south and a milder trend in the central areas, interspersed with pockets of overestimation.

Our **FY-4A_RF** product, depicted in Figure 3.18.c, mirrors the patterns seen in radar products, though it indicates elevated precipitation levels in the south. The associated bias maps in Figure 3.19.b reveal a prevalent inclination to overestimate rainfall, especially in the southern areas and certain northern regions, with a mild underestimation evident in the central part of the country.

Figure 3.20 showcases **scatterplots** that juxtapose each product against AWS measurements, restricting the analysis to grid boxes that were selected for quantitative assessment—namely, those with at least one rain gauge. Panel 'a' illuminates the average reference distribution through the use of logarithmic bins, while panels 'b' to 'h' depict the seven products in the order previously illustrated in Figure 3.16.

Radar products remain the most closely aligned with the reference data, as indicated by the preponderance of points along the scatterplot's diagonal. Nevertheless, they exhibit a tendency to underestimate higher rainfall intensities. **IMERG** variants also tend to underestimate, with the Final run showing slightly better performance at higher intensities.

Geostationary products generally exhibit a weaker correlation with the AWS reference at this scale. In regions of the scatterplots where lower rainfall intensities are common, these products manifest distinct behaviours. For instance, **GK-2A** displays a narrow value range with little variability, translating into minimal correlation with AWS data, evident from its horizontally oriented distribution. Conversely, **FY-4A** has a more expansive distribution of points that align better with the scatterplot's diagonal, indicating improved average correlation. Both satellites share similar patterns for higher intensities but are limited by a certain threshold. Notably, the tail end of FY-4A's distribution is more densely populated than that of GK-2A.

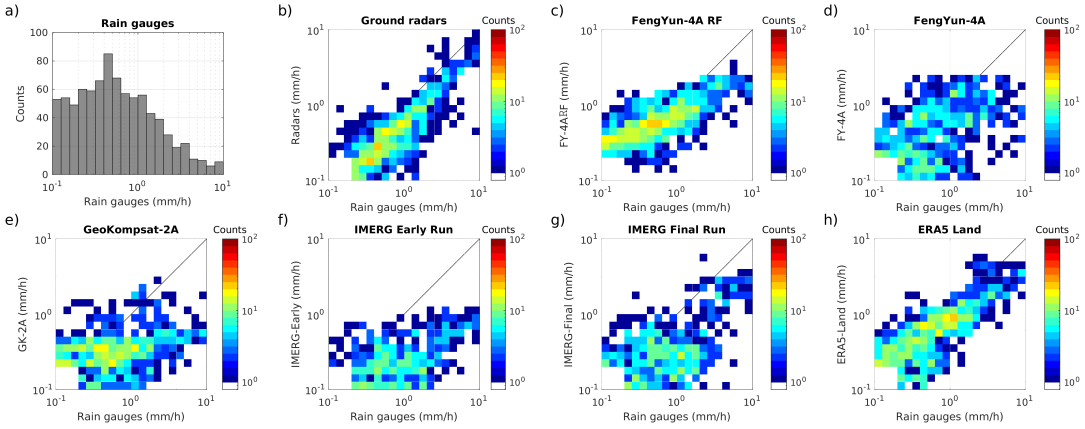

FIGURE 3.20: Time-averaged rain rate distributions across grids. (a) AWS frequency chart. (b-h) Density plots comparing products (y-axis) to AWS observations (x-axis): Radars (b); FY-4A (c); FY-4A_RF (d); GK-2A (e); IMERG-Early (f); IMERG-Final (g); ERA5-Land (h). Logarithmic scale used. Adapted from [59].

ERA5-Land shows noticeable improvements in correlation, as evidenced by a higher number of points aligning along the diagonal and reduced variance. However, challenges remain in capturing very high-intensity events.

Regarding our product, **FY-4A_RF**, the averaging process yields an overall improvement, especially for lower rainfall rates. Points are more clustered along the diagonal, and instances of missed detections are substantially reduced. However, underestimation at higher rainfall rates persists.

In conclusion, the extensive temporal analysis of the test dataset has provided critical insights into the performance and biases of various rainfall measurement products in comparison to AWS reference data. Radar products consistently display the strongest alignment with AWS data but fall short in estimating higher rainfall intensities. Geostationary products like FY-4A and GK-2A show specific regional biases, with FY-4A performing slightly better in terms of correlation. IMERG products also generally underestimate rain, although the Final run shows improvement at higher intensities. ERA5 exhibits significant strides in correlation, particularly in the central regions, but still struggles to capture extreme rainfall events accurately.

Our proprietary product, FY-4A_RF, benefits from the averaging process, displaying improved alignment for lower rainfall rates and fewer instances of missed detections. However, it still tends to underestimate higher intensities. The spatial maps and bias representations further complement these findings, providing a nuanced understanding of each product's strengths and limitations. Overall, while no single product perfectly mirrors the AWS reference, each has its unique merits and areas for improvement, underscoring the need for continual refinement and validation.

Conclusion

Precise precipitation measurements are essential for understanding Earth's hydrological cycle, its impact on ecosystems, and human activities. Despite its importance, accurate measurement is challenging, especially in areas with uneven rainfall or limited infrastructure. With climate change altering precipitation patterns, traditional methods sometimes fall short. However, satellite remote sensing offers potential solutions, providing more comprehensive data on precipitation trends.

This study aims to improve Quantitative Precipitation Estimation (QPE) by integrating multispectral satellite data with a Random Forest machine learning model. We calibrated our model using radar data, which correlates closely with rain gauge measurements. Though rain gauges are the benchmark, radars offer broader coverage. To address radar limitations, especially during light rainfall or in challenging terrains, we combined radar and satellite data. Intense rainfall can also impact radar accuracy, emphasizing the importance of this integration.

Early training phases emphasized the importance of evaluating feature rankings to discern model behaviours. Each task, whether classification or regression, exhibited unique patterns. Both the visible and infrared channel groups were essential, particularly when considering their spatial combinations through metrics such as standard deviations and means of neighbouring pixels. The IR channel differences also played a significant role in model training, specifically those IR combinations that are related to cloud-top height, cloud phase and water vapour content. An important ancillary feature in our study was the Digital Elevation Model (DEM). Preliminary analyses showcased an inverse relationship between precipitation and elevation in Vietnam. Specifically, higher altitudes displayed a statistically significant decline in rainfall compared to lower plains. This pattern likely emerges from typhoons making landfall near coasts, which typically coincide with flatlands. In terms of feature importance, DEM consistently scored high, particularly in regression tasks, often securing top positions.

Our research employed Random Forest models, initialized using the default values for most of the parameters of the Scikit-learn package. Despite demonstrating commendable performance during its training phase for both classification and regression tasks, it experienced a noticeable decline when faced with unseen data. This contrast suggests a potential overfitting issue, indicating the model's potential difficulty in generalizing beyond its training data. During validation against rain gauge measurements, our model's rain area classification was suboptimal, mispredicting nearly half of the 'wet' pixels with a False Alarm Ratio of 0.47. Moreover, The regression results showed a notable issue, with the model often estimating precipitation around the median of each rainfall class, leading to a modest correlation coefficient of 0.36 with the actual precipitation.

However, not all results were on the lower spectrum. Our model showcased a performance that, in certain metrics, surpassed other rainfall products available for Vietnam. Notably, those derived from the geostationary satellites FY-4A and, to some extent, the GK-2A and IMERG versions. Specifically, our model achieved a Probability of Detection score of 0.65, substantially outperforming the FY-4A's

0.18 and even radar-derived results, which stood at 0.42. Additionally, our model recorded a Bias Score of 1.22 and an Equitable Threat Score of 0.25, presenting notable improvements over FY-4A's respective scores of 0.30 and 0.09. Further, when examining regression metrics, our model's performance appeared to be at least on par with, if not better than, other prevalent products in the domain. This was particularly evident when compared to the original product from FY-4A. To exemplify, the Modified Kling-Gupta Efficiency for our model, termed FY-4A_RF, stood at 0.22. This contrasts positively with several other products, excluding ERA5-Land and radars, which showcased suboptimal, often negative, performances. In wrapping up, this preliminary study underscores the potential viability of our model, especially when positioned against existing precipitation estimation methodologies for the region.

Looking ahead to the future prospects of this study, it's evident that our model is in its nascent stages and warrants further refinements. A logical starting point involves fine-tuning the Random Forest parameters through a systematic hyperparameter search. This will allow us to focus on key parameters that influence the model's complexity.

It's also worth noting that our current model is based on a mere month of data. To enhance its robustness, it's prudent to expand the dataset across a more extended time frame. A more extensive dataset will encapsulate a wider variety of rainfall scenarios, inherently strengthening the model. Another issue observed was the difficulty in estimating the class of extreme rainfall due to the scarcity of this data. To address this challenge, collecting more data from intense precipitation events and implementing some form of class balancing might be beneficial.

Our ultimate ambition is the development of an operational platform capable of delivering precise, high-resolution, and near-real-time rainfall estimates. Such a system would offer dependable forecasts, even in remote regions bereft of advanced weather instrumentation. Not only would this greatly aid early warning systems for extreme weather phenomena, but it would also significantly enrich data assimilation for extended forecasts and climatic models.

Through this study, we endeavour to contribute significantly to the domain of QPE using ML, leveraging the power of Big Data and advanced remote sensing techniques. These innovations are central to refining our understanding, estimation and prediction of rainfall patterns.

Acknowledgement

This research is a component of the INDRA Project, an initiative focusing on INtegrateD RAinfall measurements platform for applications spanning agriculture, hydro - meteorological hazard prevention and mitigation and water management. The funding for the project was provided by the Italian Ministry of Foreign Affairs and International Cooperation (MAECI) within the framework of their Major Projects in the Scientific and Technological Collaboration Executive Programmes. Additionally, the Vietnam Ministry of Science and Technology contributed support under the bilateral project code NDT/IT/22/07 between Italy and Vietnam.

Furthermore, I extend my heartfelt gratitude for the invaluable ground-based data—comprising rain gauge and radar main products—furnished by our Vietnamese collaborators: Thanh Thi Nhat Nguyen from the University of Engineering and Technology at the Vietnam National University (VNU) of Hanoi and Thu Vinh Nguyen from the National Centre for Hydro-Meteorological Network (NCN), and the Vietnam Meteorological and Hydrological Administration (VNMHA).

I am really thankful to my supervisor, Prof. Federico Porcù, and co-supervisor Dr. Giacomo Roversi, both of whom were consistently available for guidance and offered invaluable advice. Last but not least, a special thanks goes to my family and friends who have been unwavering pillars of support, especially during moments of doubt and fatigue.

A few years ago, I thought my academic endeavours had culminated. Yet, an insatiable curiosity and a profound urge to address global challenges kept the flame alive and propelled me towards new horizons.

Just like travelling, education is a never-ending journey of exploration where the journey itself is the true reward, enhanced by the remarkable individuals I've met along the way. A sincere appreciation goes out to everyone who has made this possible.

Nomenclature

AMW Active Microwave

AGRI Advanced Geosynchronous Radiation Imager

AMI Advanced Meteorological Imager

AI Artificial Intelligence

ANN Artificial Neural Network

AWS Automatic Weather Stations

CV Coefficient of Variation

CC Correlation Coefficient

CSI Critical Success Index

DT Decision Trees

DEM Digital Elevation Model

DSD Drop Size Distribution

DPR Dual-frequency Precipitation Radar

ETS Equitable Threat Score

$\mathbf{H}_R$ Expected Random Hits

FAR False Alarm Ration

FN False Negative

FP False Positive

FI Feature Importance

FY-4A Fengyun-4A

GK-2A GeoKompsat-2A

GEO Geostationary

GPM Global Precipitation Measurement

GB Gradient Boosting

IR Infrared

ITCZ Intertropical Convergence Zone

KNN K-Nearest-Neighbour

P50 Likelihood that an estimated value will deviate from the actual measurement by no more than 50%

LEO Low Earth Orbit

ML Machine Learning

MAE Mean Absolute Error

ME Mean Error

MSE Mean Square Error

MW Microwave

mKGE Modified Kling-Gupta Efficiency

IMERG NASA Integrated Multi-satellite Retrievals for GPM

NRT Near Real Time

NaN Not a Number

PMW Passive Microwave

PERSIANN-CCS Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System

PDF Probability Density Function

POD Probability of Detection

QPE Quantitative Precipitation Estimation

RADAR RAdio Detection And Ranging

RG Rain Gauge

RFC Random Forest Classifier

RFR Random Forest Regressor

RF Random Forest

SVM Support Vector Machines

Tb Temperature of Brightness

CHIRPS The Climate Hazards Group InfraRed Precipitation with Station data

TRMM Tropical Rainfall Measuring Mission

TN True Negative

TP True Positive

VIS Visible

WV Water Vapour

WMO World Meteorological Organization

Bibliography

[1] Hylke Beck et al. "Present and future Köppen-Geiger climate classification maps at 1-km resolution". In: *Scientific Data* 5 (Oct. 2018), p. 180214. DOI: 10.1038/sdata.2018.214.

[2] Noureddine Bensafi, Lazri Mourad, and Ameur Soltane. "Novel WkNN-based technique to improve instantaneous rainfall estimation over the north of Algeria using the multispectral MSG SEVIRI imagery". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 183 (Dec. 2018). DOI: 10.1016/j.jastp.2018.12.004.

[3] Lea Beusch et al. "Satellite-Based Rainfall Retrieval: From Generalized Linear Models to Artificial Neural Networks". In: *Remote Sensing* 10 (June 2018), p. 939. DOI: 10.3390/rs10060939.

[4] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. 1st ed. 2006. Corr. 2nd printing. Information science and statistics. Springer, 2006. ISBN: 9780387310732; 0387310738.

[5] Marco Borga, Francesco Marra, and Marco Gabella. "Rainfall estimation by weather radar". In: Jan. 2022, pp. 109–134. ISBN: 9780128225448. DOI: 10.1016/B978-0-12-822544-8.00016-0.

[6] L Breiman. "Random Forests". In: *Machine Learning* 45 (Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010950718922.

[7] M. Buhmann et al. "Random Decision Forests". In: Jan. 2010. DOI: 10.1007/978-0-387-30164-8_694.

[8] Davide Capacci and Federico Porcu. "Evaluation of a Satellite Multispectral VIS-IR Daytime Statistical Rain-Rate Classifier and Comparison with Passive Microwave Rainfall Estimates". In: *Journal of Applied Meteorology and Climatology - J APPL METEOROL CLIMATOL* 48 (Feb. 2009), pp. 284–300. DOI: 10.1175/2008JAMC1969.1.

[9] Daniele Corradini et al. "Challenges in the Use of Artificial Intelligence for Prostate Cancer Diagnosis from Multiparametric Imaging Data". In: *Cancers* 13 (Aug. 2021), p. 3944. DOI: 10.3390/cancers13163944.

[10] Spandan Das et al. "A Comprehensive Machine Learning Study to Classify Precipitation Type over Land from Global Precipitation Measurement Microwave Imager (GPM-GMI) Measurements". In: *Remote Sensing* 14 (July 2022), p. 3631. DOI: 10.3390/rs14153631.

[11] Renato Morbidelli (editor). *Rainfall: Modeling, Measurement and Applications*. 1st ed. Elsevier, 2022. ISBN: 0128225440; 9780128225448.

[12] Ramage C. S. (Eds.) *Monsoon Meteorology*. International Geophysics 15. Academic Press, 1971. ISBN: 9780125766500; 0125766505.

[13] Massimo Guarascio et al. "A Machine Learning Approach for Rainfall Estimation Integrating Heterogeneous Data Sources". In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Nov. 2020). DOI: 10.1109/TGRS.2020.3037776.

[14] Hoshin Gupta et al. "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling". In: *Journal of Hydrology* 377 (Oct. 2009), pp. 80–91. DOI: 10.1016/j.jhydrol.2009.08.003.

[15] Minh Cuong Ha et al. "Machine Learning and Remote Sensing Application for Extreme Climate Evaluation: Example of Flood Susceptibility in the Hue Province, Central Vietnam Region". In: *Water* 14 (May 2022), pp. 1–24. DOI: 10.3390/w14101617.

[16] Slimane Hameg, Lazri Mourad, and Ameur Soltane. "Using naive Bayes classifier for classification of convective rainfall intensities based on spectral characteristics retrieved from SEVIRI". In: *Journal of Earth System Science* 125 (July 2016). DOI: 10.1007/s12040-016-0717-7.

[17] Dennis L Hartmann. *Global Physical Climatology*. 2nd ed. Elsevier Science, 2016. ISBN: 0123285313; 9780123285317.

[18] Hans Hersbach et al. "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* (May 2020). DOI: 10.1002/qj.3803.

[19] Hitoshi Hirose et al. "High Temporal Rainfall Estimations from Himawari-8 Multiband Observations Using the Random-Forest Machine-Learning Method". In: *Journal of the Meteorological Society of Japan. Ser. II* 97 (Mar. 2019). DOI: 10.2151/jmsj.2019-040.

[20] George Huffman et al. "Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG)". In: Apr. 2020, pp. 343–353. ISBN: 978-3-030-24567-2. DOI: 10.1007/978-3-030-24568-9_19.

[21] George Huffman et al. "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales". In: *Journal of Hydrometeorology - J HYDROMETEOROL* 8 (Feb. 2007). DOI: 10.1175/JHM560.1.

[22] Robert A. Houze Jr. *Cloud Dynamics, Volume 104, Second Edition*. 2nd ed. International Geophysics. Academic Press, 2014. ISBN: 0123742668; 9780123742667.

[23] Yogesh Kant, Krishna Vadrevu, and K.V.S Badarinath. "Algorithm for detection of active fire zones using NOAA AVHRR data". In: *Infrared Physics Technology* 41 (Feb. 2000), pp. 29–34. DOI: 10.1016/S1350-4495(99)00053-5.

[24] Chris Kidd and Vincenzo Levizzani. "Satellite rainfall estimation". In: Oct. 2022, pp. 135–169. ISBN: 978-0-12-822544-8. DOI: 10.1016/B978-0-12-822544-8.00005-6.

[25] Chris Kidd et al. "So, How Much of the Earth's Surface Is Covered by Rain Gauges?" In: *Bulletin of the American Meteorological Society* 98 (June 2016). DOI: 10.1175/BAMS-D-14-00283.1.

[26] Chiho Kimpara et al. "Quantitative Precipitation Estimation by Combining Rain gauge and Meteorological Radar Network in Viet Nam". In: *Vietnam Journal of Hydrometeorology* 5 (Aug. 2020), pp. 36–50. DOI: 10.36335/VNJHM.2020(5).36-50.

[27] Harald Kling, Martin Fuchs, and Maria Paulin. "Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios". In: *Journal of Hydrology* s 424–425 (Mar. 2012), 264–277. DOI: 10.1016/j.jhydrol.2012.01.011.

[28] Venkatesh Kolluru et al. "Secondary Precipitation Estimate Merging Using Machine Learning: Development and Evaluation over Krishna River Basin, India". In: *Remote Sensing* 12 (Sept. 2020), p. 3013. DOI: 10.3390/rs12183013.

[29] Ashish Kumar et al. "A Machine Learning Approach for Improving Near-Real-Time Satellite-Based Rainfall Estimates by Integrating Soil Moisture". In: *Remote Sensing* 11 (Sept. 2019), p. 2221. DOI: 10.3390/rs11192221.

[30] Toshiyuki Kurino. "A satellite infrared technique for estimating "deep/shallow" precipitation". In: *Advances in Space Research* 19.3 (1997). Proceedings of the A0.1 Symposium of COSPAR Scientific Commission A, pp. 511–514. ISSN: 0273-1177. DOI: https://doi.org/10.1016/S0273-1177(97)00063-X.

[31] Meike Kühnlein et al. "Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI". In: *Remote Sensing of Environment* 141 (Feb. 2014), pp. 129–143. DOI: 10.1016/j.rse.2013.10.026.

[32] Luca Lanza, Arianna Cauteruccio, and Mattia Stagnaro. "Rain gauge measurements". In: Jan. 2022, pp. 77 –108. ISBN: 9780128225448. DOI: 10.1016/B978-0-12-822544-8.00002-0.

[33] Douglas LeComte. "International Weather Highlights 2020: Record Atlantic Tropical Season, Historic Flooding in Asia and Africa". In: *Weatherwise* 74.3 (2021), pp. 26–35. DOI: 10.1080/00431672.2021.1896936.

[34] Xinyan Li et al. "Leveraging machine learning for quantitative precipitation estimation from Fengyun-4 geostationary observations and ground meteorological measurements". In: (July 2021). DOI: 10.5194/amt-2021-175.

[35] Zhi Li et al. "Advancing Satellite Precipitation Retrievals with Data Driven Approaches: Is black box model explainable?" In: *Earth and Space Science* 8 (Feb. 2021). DOI: 10.1029/2020EA001423.

[36] Kangwen Liu, Jieying He, and Haonan Chen. "Precipitation Retrieval from Fengyun-3D Microwave Humidity and Temperature Sounder Data Using Machine Learning". In: *Remote Sensing* 14 (Feb. 2022), p. 848. DOI: 10.3390/rs14040848.

[37] Chinh Luu, Bui Duy Quynh, and Jason von Meding. "Mapping direct flood impacts from a 2020 extreme flood event in Central Vietnam using spatial analysis techniques". In: *International Journal of Disaster Resilience in the Built Environment* ahead-of-print (Nov. 2021). DOI: 10.1108/IJDRBE-07-2021-0070.

[38] Lei Ma et al. "Deep learning in remote sensing applications: A meta-analysis and review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (Apr. 2019), pp. 166–177. DOI: 10.1016/j.isprsjprs.2019.04.015.

[39] Ziqiang Ma, Siyu Zhu, and Jun Yang. "FY4QPE-MSA: An All-Day Near-Real-Time Quantitative Precipitation Estimation Framework Based on Multispectral Analysis From AGRI Onboard Chinese FY-4 Series Satellites". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (Jan. 2022), pp. 1–1. DOI: 10.1109/TGRS.2022.3159036.

[40] Y. Makihara. "A method for improving radar estimates of precipitation by comparing data from radars and raingauges". In: *Journal of the Meteorological Society of Japan* 74.4 (1996), pp. 459–480. DOI: https://doi.org/10.2151/jmsj1965.74.4_459.

[41] J. S. Marshall and W. Mc K. Palmer. "THE DISTRIBUTION OF RAINDROPS WITH SIZE". In: *Journal of Atmospheric Sciences* 5.4 (1948), pp. 165 –166. DOI: https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.

[42] John Mecikalski et al. "Cloud-Top Properties of Growing Cumulus prior to Convective Initiation as Measured by Meteosat Second Generation. Part I: Infrared Fields". In: *Journal of Applied Meteorology and Climatology - J APPL METEOROL CLIMATOL* 49 (Mar. 2010), pp. 521–534. DOI: 10.1175/2009JAMC2344.1.

[43] John Mecikalski et al. "Cloud-Top Properties of Growing Cumulus prior to Convective Initiation as Measured by Meteosat Second Generation. Part II: Use of Visible Reflectance". In: *Journal of Applied Meteorology and Climatology* 49 (Dec. 2010). DOI: 10.1175/2010JAMC2480.1.

[44] Hanna Meyer et al. "Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique". In: *Remote Sensing Letters* 8 (Apr. 2017), pp. 647–656. DOI: 10.1080/2150704X.2017.1312026.

[45] Eugen Mihulet et al. "Enhancing the Performance of Quantitative Precipitation Estimation Using Ensemble of Machine Learning Models Applied on Weather Radar Data". In: *Atmosphere* 14 (Jan. 2023), p. 182. DOI: 10.3390/atmos14010182.

[46] Min Min et al. "Estimating Summertime Precipitation from Himawari-8 and Global Forecast System Based on Machine Learning". In: *IEEE Transactions on Geoscience and Remote Sensing* 57 (May 2019), pp. 2557–2570. DOI: 10.1109/TGRS.2018.2874950.

[47] Juhar Mohammed, Yenesew Mengiste, and Vijay Singh. "Improving spatio-temporal precipitation estimates in data scarce river basins: an application of machine learning-based multi-source data merging". In: *Stochastic Environmental Research and Risk Assessment* 37 (Nov. 2022). DOI: 10.1007/s00477-022-02346-4.

[48] Lazri Mourad, Ameur Soltane, and Yacine Mohia. "Instantaneous rainfall estimation using neural network from multispectral observations of SEVIRI radiometer and its application in estimation of daily and monthly rainfall". In: *Advances in Space Research* 53 (Jan. 2014), 138–155. DOI: 10.1016/j.asr.2013.10.005.

[49] Lazri Mourad et al. "Improving satellite rainfall estimation from MSG data in Northern Algeria by using a multi-classifier model based on machine learning". In: *Journal of Hydrology* 584 (Feb. 2020), p. 124705. DOI: 10.1016/j.jhydrol.2020.124705.

[50] Andreas C. Müller and Sarah Guido. *Introduction to Machine Learning with Python. A Guide for Data Scientists*. O'Reilly Media, 2016. ISBN: 9781449369897.

[51] Quang Nguyen, James Renwick, and J. Mcgregor. "Variations of surface temperature and rainfall in Vietnam from 1971 to 2010". In: *International Journal of Climatology* 34 (Jan. 2014). DOI: 10.1002/joc.3684.

[52] Quang Nguyen et al. "Settling with forests in Thua Thien Hue (Vietnam)". In: *Trees, Forests and People* 7 (Dec. 2021), p. 100181. DOI: 10.1016/j.tfp.2021.100181.

[53] Pertti Nurmi. "Recommendations on the verification of local weather forecasts". In: *ECMWF Technical Memorandum* 430 (Jan. 2003).

[54] World Meteorological Organization. *Guide to meteorological instruments and methods of observation 7th Edition*. ISBN: 9789263100085; 926310008X.

[55] Andrea Ortiz-Vargas and Zita Sebesvari. "Technical Report: Floods in Central Viet Nam". In: *Interconnected Disaster Risks 2020/2021. United Nations University - Institute for Environment and Human Security (UNU-EHS)* (2021). DOI: 10.53324/TNAE9416.

[56] Fethi Ouallouche, Lazri Mourad, and Ameur Soltane. "Improvement of rainfall estimation from MSG data using Random Forests classification and regression". In: *Atmospheric Research* 211 (May 2018). DOI: 10.1016/j.atmosres.2018.05.001.

[57] Chandrasekar Radhakrishnan et al. "Rainfall Estimation From TEMPEST-D CubeSat Observations: A Machine-Learning Approach". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (May 2022). DOI: 10.1109/JSTARS.2022.3170835.

[58] Cynthia Rosenzweig et al. "Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison". In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (Dec. 2013). DOI: 10.1073/pnas.1222463110.

[59] Giacomo Roversi et al. "High-resolution intercomparison of multi-platform precipitation products over Vietnam (under Review)". In: *Remote Sensing MDPI* (2023).

[60] Mounir Sehad, Lazri Mourad, and Ameur Soltane. "Novel SVM-based technique to improve rainfall estimation over the Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery". In: *Advances in Space Research* 59 (Dec. 2016). DOI: 10.1016/j.asr.2016.11.042.

[61] Mounir Sehad and Ameur Soltane. "A multilayer perceptron and multiclass support vector machine based high accuracy technique for daily rainfall estimation from MSG SEVIRI data". In: *Advances in Space Research* 65 (Nov. 2019). DOI: 10.1016/j.asr.2019.11.018.

[62] B. Sevruk, M. Ondrás, and B. Chvíla. "The WMO precipitation measurement intercomparisons". In: *Atmospheric Research* 92.3 (2009). 7th International Workshop on Precipitation in Urban Areas, pp. 376–380. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2009.01.016.

[63] Kyuhee Shin et al. "Quantitative Precipitation Estimates Using Machine Learning Approaches with Operational Dual-Polarization Radar Data". In: *Remote Sensing* 13 (Feb. 2021), p. 694. DOI: 10.3390/rs13040694.

[64] Roger A. Pielke Sr. *Climate Vulnerability: Understanding and Addressing Threats to Essential Resources*. 1st ed. Academic Press, 2013. ISBN: 0123847036; 9780123847034.

[65] Kevin Trenberth et al. "The Changing Character of Precipitation". In: *Bull. Amer. Meteor. Soc.* 84 (Sept. 2003), pp. 1205–1217. DOI: 10.1175/BAMS-84-9-1205.

[66] Nazli Turini, Boris Thies, and Jörg Bendix. "Estimating High Spatio-Temporal Resolution Rainfall from MSG1 and GPM IMERG Based on Machine Learning: Case Study of Iran". In: *Remote Sensing* 11 (Oct. 2019), p. 2307. DOI: 10.3390/rs11192307.

[67] Nazli Turini et al. “Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data”. In: *European Journal of Remote Sensing* 54 (Feb. 2021), pp. 117–139. DOI: 10.1080/22797254.2021.1884002.

[68] Dalia B. Kirschbaum Christian D. Kummerow Kenji Nakamura F. Joseph Turk Vincenzo Levizzani Christopher Kidd. *Satellite Precipitation Measurement: Volume 1*. Springer, 2020. DOI: 10.1007/978-3-030-24568-9.

[69] Gaohong Yin, Jongjin Baik, and Jongmin Park. “Comprehensive analysis of GEO-KOMPSAT-2A and FengYun satellite-based precipitation estimates across Northeast Asia”. In: *GIScience Remote Sensing* 59 (Dec. 2022), pp. 782–800. DOI: 10.1080/15481603.2022.2067970.

[70] Yushan Zhang et al. “Estimating Rainfall with Multi-Resource Data over East Asia Based on Machine Learning”. In: *Remote Sensing* 13 (Aug. 2021), p. 3332. DOI: 10.3390/rs13163332.