ORTA DOĞU TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ŞABLONU ONAY FORMU
THESIS TEMPLATE CONFIRMATION FORM

1. Şablonda verilen yerleşim ve boşluklar değiştirilmemelidir.
2. **Jüri tarihi** Başlık Sayfası, İmza Sayfası, Abstract ve Öz'de ilgili yerlere yazılmalıdır.
3. İmza sayfasında jüri üyelerinin unvanlarıdoğru olarak yazılmalıdır. Tüm imzalar **mavi pilot kalemle** atılmalıdır.
4. **Disiplinlerarası** programlarda görevlendirilen öğretim üyeleri için jüri üyeleri kısmında tam zamanlı olarak çalıştıkları anabilim dalı başkanlığının ismi yazılmalıdır. Örneğin: bir öğretim üyesi Biyoteknoloji programında görev yapıyor ve biyoloji bölümünde tam zamanlı çalışıyorsa, İmza sayfasına biyoloji bölümü yazılmalıdır.İstisnai olarak, disiplinler arası program başkanı ve tez danışmanı için disiplinlerarası program adı yazılmalıdır.
5. Tezin **son sayfasının sayfa** numarası Abstract ve Öz'de ilgili yerlere yazılmalıdır.
6. Bütün chapterlar, referanslar, ekler ve CV sağ sayfada başlamalıdır. Bunun için **kesmeler** kullanılmıştır. **Kesmelerin kayması** fazladan boş sayfaların oluşmasına sebep olabilir. Bu gibi durumlarda paragraf (¶) işaretine tıklayarak kesmeleri görünür hale getirin ve yerlerini **kontrol edin**.
7. Figürler ve tablolar kenar boşluklarına taşmamalıdır.
8. Şablonda yorum olarak eklenen uyarılar dikkatle okunmalı ve uygulanmalıdır.
9. Tez yazdırılmadan önce PDF olarak kaydedilmelidir. Şablonda yorum olarak eklenen uyarılar PDF dokümanında yer almamalıdır.
10. Tez taslaklarının kontrol işlemleri tamamlandığında, bu durum öğrencilere METU uzantılı öğrenci e-posta adresleri aracılığıyla duyurulacaktır.
11. Tez yazım süreci ile ilgili herhangi bir sıkıntı yaşarsanız, Sıkça Sorulan Sorular (SSS) sayfamızı ziyaret ederek yaşadığınız sıkıntıyla ilgili bir çözüm bulabilirsiniz.

1. Do not change the spacing and placement in the template.
2. Write**defense date** to the related places given on Title page, Approval page, Abstract and Öz.
3. Write the titles of the examining committee members correctly on Approval Page. **Blue ink** must be used for all signatures.
4. For faculty members working in **interdisciplinary programs**, the name of the department that they work full-time should be written on the Approval page. For example, if a faculty member staffs in the biotechnology program and works full-time in the biology department, the department of biology should be written on the approval page. Exceptionally, for the interdisciplinary program chair and your thesis supervisor, the interdisciplinary program name should be written.
5. Write **the page number of the last page** in the related places given on Abstract and Öz pages.
6. All chapters, references, appendices and CV must be started on theright page.**Section Breaks** were used for this. **Change in the placement** of section breaks can result in extra blank pages. In such cases, make the section breaks visible by clicking paragraph (¶) mark and **check their position.**
7. All figures and tables must be given inside the page. Nothing must appear in the margins.
8. All the warnings given on the comments section through the thesis template must be read and applied.
9. Save your thesis as pdf and Disableall the comments before taking the printout.
10. This will be announced to the students via their METU students e-mail addresses when the control of the thesis drafts has been completed.
11. If you have any problems with the thesis writing process, you may visit our Frequently Asked Questions (FAQ) page and find a solution to your problem.

☒ **Yukarıda bulunan tüm maddeleri okudum, anladım ve kabul ediyorum. / I have read, understand and accept all of the items above.**

Name :
Surname :
E-Mail :
Date :
Signature : _____

MACHINE LEARNING-BASED APPROACH FOR BIAS CORRECTION OF SATELLITE-BASED PRECIPITATION PRODUCTS USING ENVIRONMENTAL PARAMETERS AND GROUND TRUTH DATA IN TURKIYE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖKHAN SEVİNÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
GEOLOGICAL ENGINEERING

SEPTEMBER 2024

Approval of the thesis:

MACHINE LEARNING-BASED APPROACH FOR BIAS CORRECTION OF SATELLITE-BASED PRECIPITATION PRODUCTS USING ENVIRONMENTAL PARAMETERS AND GROUND TRUTH DATA IN TURKIYE

submitted by **GÖKHAN SEVİNÇ** in partial fulfillment of the requirements for the degree of**Master of Science** i**n Geological Engineering, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, **Graduate School of Natural and Applied Sciences** _____

Prof. Dr. Erdin Bozkurt
Head of the Department, **Geological Engineering, METU** _____

Assoc. Prof. Dr. Koray Kamil Yılmaz
Supervisor, **Geological Engineering, METU** _____

Examining Committee Members:

Prof. Dr. M. Zeki Çamur
Geological Engineering, METU _____

Assoc. Prof. Dr. Koray K. Yılmaz
Geological Engineering, METU _____

Prof. Dr. M. Lütfi Süzen
Geological Engineering, METU _____

Prof. Dr. M. Tuğrul Yılmaz
Civil Engineering, METU _____

Assoc. Prof. Dr. Bedri Kurtuluş
Geological Eng., Muğla Sıtkı Koçman University _____

Date:06.09.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name: Gökhan Sevinç

Signature:

ABSTRACT

MACHINE LEARNING-BASED APPROACH FOR BIAS CORRECTION OF SATELLITE-BASED PRECIPITATION PRODUCTS USING ENVIRONMENTAL PARAMETERS AND GROUND TRUTH DATA IN TURKIYE

Sevinç, Gökhan
Master of Science, Geological Engineering
Supervisor: Assoc. Prof. Koray Kamil Yılmaz

September 2024, 108 pages

Satellite precipitation data are very important in hydrological studies, but contain bias. In this study, XGBoost and Random Forest machine learning algorithms are used to correct the bias with ground observations and environmental parameters such as distance to the coast and elevation.

The machine learning models were trained daily from 2015 to 2022 with optimal hyperparameters to obtain the most accurate and robust results and results are filtered to be more representative on rainy days. Although machine learning models are generally considered as black box, SHAP values were utilized in this study in an effort to explain and interpret their behavior by showing the contribution of each feature to the model prediction and how these contributions change as a function of space and time.

The performance of the models was examined using different metrics to clearly explain their strengths and weaknesses. Average RMSE scores of filtered IMERG (7.08), Random Forest (4.00), and XGBoost (4.33) showing machine learning models provide a much more accurate prediction because they reduce the average RMSE of filtered IMERG by about 3 mm/day. Average KGE scores of filtered IMERG (-0.28), Random Forest (0.46) and XGBoost (0.47) and their positive

improvements of KGE values indicating machine learning models perform better in capturing precipitation variability and accuracy of predictions. The Average Mean Bias Error scores of filtered IMERG (0.197) indicates, overestimation of observations, while Random Forest (-0.068) and XGBoost (-0.071) models slightly underestimate the observed values. These results shows that the accuracy and reliability of the prediction performance are improved. It was found that XGBoost models are better to capture variability in data and predicting extreme precipitation events (10 mm/day or higher events). Random forest is better at predicting lower threshold events such as 1 mm/day and 2 mm/day.

The overall behavior of the models is visualized by merging their daily SHAP values. Machine learning models are consistent with their feature importance scores (FI) each year and adapt their behavior seasonally. The SHAP analysis further emphasizes that the models successfully capture the aridity over the Mediterranean and Central Anatolian regions by providing low summer precipitation at these latitudes, while positive SHAP values at higher latitudes in summer translate into increased precipitation in the Black Sea region. The clear positive correlation of precipitation with elevation is evident in the models, while the effect of distance from the coast in summer is minimal due to generally dry climatic conditions. The SHAP analysis also shows that the models capture the high winter precipitation in the Mediterranean and Black Sea regions, as well as the dry conditions of the Central Anatolian Plateau. In addition, the models show a strong seasonal influence of the distance to coast feature on precipitation, with a superior ability to capture coastal precipitation in winter, demonstrating their ability to adapt to seasonality.

Keywords: Satellite-based Precipitation, Bias Correction, Machine Learning, SHAP, Explainable AI

ÖZ

TÜRKİYE'DE ÇEVRESEL PARAMETRELER VE YER GERÇEĞİ VERİLERİ KULLANILARAK UYDU TABANLI YAĞIŞ ÜRÜNLERİNİN HATA DÜZELTMESİ İÇİN MAKİNE ÖĞRENME TABANLI YAKLAŞIM

Sevinç, Gökhan
Yüksek Lisans, Jeoloji Mühendisliği
Tez Yöneticisi: Doç. Dr. Koray Kamil Yılmaz

Eylül 2024, 108 sayfa

Uydu yağış verileri hidrolojik çalışmalar için çok önemli olmakla birlikte hatalar içermektedir. Bu çalışmada, XGBoost ve Random Forest makine öğrenimi algoritmaları, yer gözlemleri ve kıyıya uzaklık, yükseklik gibi çevresel parametreleri hata düzeltmek için kullanılmaktadır.

Makine öğrenimi modelleri, en doğru ve sağlam sonuçları elde etmek için optimum hiperparametrelerle 2015'ten 2022'ye kadar günlük olarak eğitilmiştir ve sonuçlar yağmurlu günlerde daha temsili olacak şekilde filtrelenmiştir. Makine öğrenimi modelleri genellikle kara kutu olarak kabul edilse de, her bir özelliğin model tahminine katkısını açıklamak ve yorumlamak amacıyla bu çalışmada SHAP değerleri kullanılmıştır.

Modellerin performansı, güçlü ve zayıf yönlerini net bir şekilde açıklamak için farklı metrikler kullanılarak incelenmiştir. Filtrelenmiş IMERG (7,08), Rastgele Orman (4,00) ve XGBoost'un (4,33) ortalama RMSE puanları, filtrelenmiş IMERG'in ortalama hatasını yaklaşık 3 mm/gün azalttıkları için makine öğrenimi modellerinin çok daha doğru bir tahmin sağladığını göstermektedir. filtrelenmiş

IMERG (-0,28), Random Forest (0,46) ve XGBoost'un (0,47) ortalama KGE puanları ve KGE değerlerindeki olumlu gelişmeler, makine öğrenimi modellerinin yağış değişkenliğini ve tahminlerin doğruluğunu yakalamada daha iyi genel performans gösterdiğini ortaya koymaktadır. Filtrelenmiş IMERG (1.578), Random Forest (-0.068) ve XGBoost'un (-0.071) Ortalama Bias Hataları filtrelenmiş IMERG'nin Ortalama Bias Hatası (0.197) tahminlerini daha yüksek belirttiğine işaret etmektedir. Rastgele Orman (-0,068) ve XGBoost (-0,071) modelleri gözlenen değerleri biraz daha düşük tahmin etmektedir, bu da tahmin performansının doğruluğu ve güvenilirliğinin arttığı anlamına gelmektedir. XGBoost modelleri, verilerdeki değişkenliği yakalamada ve aşırı yağış olaylarını (10 mm/gün veya daha yüksek olaylar) tahmin etmede daha iyidir. Rastgele orman, 1 mm/gün ve 2 mm/gün gibi daha düşük eşik olaylarını tahmin etmede daha iyidir.

Modellerin genel davranışı, günlük SHAP değerleri birleştirilerek görselleştirilmiştir. Makine öğrenimi modelleri her yıl özellik önem puanları (FI) ile tutarlıdır ve davranışlarını mevsimsel olarak düzenlerler.SHAP analizi ayrıca, modellerin Akdeniz ve İç Anadolu bölgelerindeki kuraklığı, bu enlemlerde düşük yaz yağışları sağlayarak başarılı bir şekilde yakaladığını, yaz aylarında daha yüksek enlemlerde pozitif SHAP değerlerinin Karadeniz bölgesinde artan yağışlara dönüştüğünü vurgulamaktadır. Yağışın yükseklikle pozitif korelasyonu modellerde açıkça görülürken, yaz aylarında kıyıdan uzaklığın etkisi genel olarak kuru iklim koşulları nedeniyle minimum düzeydedir. SHAP analizi ayrıca modellerin Akdeniz ve Karadeniz bölgelerindeki yüksek kış yağışlarının yanı sıra Orta Anadolu Platosu'nun kuru koşullarını da yakaladığını göstermektedir. Buna ek olarak, modeller, kıyıya uzaklık özelliğinin yağış üzerinde güçlü bir mevsimsel etkisi olduğunu ve kış aylarında kıyı yağışlarını yakalamada üstün bir kabiliyete sahip olduklarını göstererek mevsimselliğe uyum sağlama yeteneklerini ortaya koymaktadır.

Anahtar Kelimeler: Uydu Tabanlı Yağış, Hata Düzeltme, Makine Öğrenimi, SHAP, Açıklanabilir Yapay Zeka

This Thesis is Dedicated to My Beloved Family

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to my supervisor, Assoc. Prof. Koray Kamil Yilmaz. His encouragement and motivation have played an important part in this research. His constructive criticism was very significant during my research and helped me gain new perspectives. He showed great patience and was always supportive during my academic journey. I'm really indebted to his mentorship and guidance.

I would like extend my deepest thanks to my beloved colleagues Dr. Hatice Kılıç Germeç and Ceren Yazıgülü Tural for their continuous support throughout the research and writing process for my thesis. Their positive and understanding attitudes helped me a lot in this journey and made it more productive and enjoyable.

I would like to express my sincere gratitude to the Scientific and Technological Research Council of Turkiye (TÜBİTAK) for its very important financial support throughout the research project under Project TUBİTAK 121Y468. The value of research and investment in innovation is immeasurable for me.

I would also like to thank to the General Directorate of Meteorology (GDM) for providing necessary ground observation data for this study.

I am deeply thankful to my beloved family for being very supportive and encouraging throughout the journey. They have believed in me and never left me alone. I thank them very much for the love and support extended to me in all ways in life.

I will also need to extend my appreciation to all of my friends for their support and encouragement in this journey.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI: Artificial Intelligent

CSI: Critical Success Index

DMC: Double Mass Curve

ETH: Effective Terrain Height

FAR: False Alarm Ration

FI: Feature Importance

GBDT: Gradient Boost Decision Tree algorithm

IMERG: Integrated Multi-satellitE Retrievals for GPM

KGE: Kling Gupta Efficiency

MBE: Mean Bias Error

MSE: Mean Squared Error

NSE: Nash-Sutcliffe Efficiency

POD: Probability of Detection

RMSE: Root Mean Square Error

SPPs: Satellite Precipitation Products

TMPA: TRMM Multi-Satellite Precipitation Analysis

TRMM: The Tropical Rainfall Measuring Mission

XGBoost: The eXtreme Gradient Boosting

CHAPTER 1

INTRODUCTION

The aim of this chapter is to introduce the general importance of precipitation and to inform the reader about the precipitation products and the bias correction to highlight the background of bias correction of satellite precipitation products with environmental parameters and ground observations. The importance and objectives of this study are also given in this chapter.

1.1 Precipitation: Measurement and Applications

Precipitation plays a critical role in the climate system and the water cycle (Kidd and Huffman, 2011).Precipitation datasets are playing a vital role in the evaluation of flood risks, conditions of drought, and also the availability of water resources-a matter concerning both urban and rural areas (Hui-Mean et al., 2018; Kidd et al., 2009; Thornes et al., 2010; Kidd and Huffman, 2011; Zhou et al., 2022).Measuring precipitation with high precision is indeed a difficult task, as it is a very complex process and interacts with so many other parameters. Precipitation has spatio-temporal heterogeneities that make estimations challenging, due to their variability in space and time and environmental factors like elevation (Kucera et al., 2013; Herold et al., 2016). However, despite all these difficulties, the accuracy of precipitation data is really important from the point of view of disciplines like hydrology, hydrogeology, climate science, and meteorology (Hobbs, 1989; Ren et al., 2021).

Various methods are used to estimate precipitation, including gauges, radars and satellites, all of which have their advantages and limitations. It is clear that direct

point-scale measurements obtained from ground-based observations are limited in capturing spatial variability (He et al., 2023; Zeng et al., 2018b), while radar and satellite data are indirect and have their own limitations in terms of accuracy (Derin and Yilmaz, 2014). As a result, there is increasing interest in algorithms that would optimally normalize data from these different sources (Zeng et al., 2021; Wang et al., 2020). Machine learning algorithms have emerged as effective tools to improve rainfall estimation and prediction (Aghelpour et al., 2020; Guan et al., 2020; Mohammadi et al., 2021; Hadadi et al., 2022).Some of the recent works apply tree-based AI algorithms that synthesize ground-based observations with environmental parameters and satellite-collected precipitation data into enhanced precipitation fields, incorporating strengths in direct precipitation measurements, spatial coverage and environmental controls (Başağoğlu et al., 2022; Chakraborty et al., 2021a; Chang et al., 2016; Dumitrescu et al., 2021). Among other things, quality control and correction of data processing bias are essential steps for these datasets. In this regard, machine learning algorithms that can handle complex, linear and nonlinear relationships have shown great potential to improve the accuracy of precipitation product production (Chivers et al., 2020; Abualigah et al., 2021; Zhang et al., 2021; Agushaka et al., 2022; Oyelade et al., 2022).

1.1.1 Precipitation Products

In the recent years, with the advance in the technology of remote-sensing, especially with the advances of weather radars and satellites, quantitative precipitation assessment has become significantly improved (Tang et al., 2020; Shi et al., 2020). Satellite precipitation products (SPPs) and radar products offer data with indirect measurement, hence with inherent bias. Radar precipitation products have higher spatial and temporal resolution within a limited radius and have improved significantly over the last decades, typically to 1 km and 5 min (Overeem, Holleman, & Buishand, 2009; Serafin & Wilson, 2000).However, radars have limited coverage and constructing a homogeneous radar field by combining

neighboring radars is challenging due to several factors, including inter-radar calibration differences (Patakchi et al., 2023), and they do not have global coverage (Li et al., 2021a). On the other hand, satellites provide quasi-global coverage and provide consistent measurements of precipitation across the globe in space and time (Derin and Yilmaz, 2014).The resolution of satellite-based products varies depending on the product, for example, the IMERG Late product provides precipitation with a spatial resolution of 0.1° × 0.1° and a temporal resolution of 30 min (Huffman., 2019; Huffman et al., 2018).The value of SPPs lies in their ability to very accurately identify broader patterns and trends in weather.

Satellite precipitation products (SPPs) are effective for monitoring global precipitation, but they provide reduced accuracy at the local scale. They are essential for regional and global monitoring of precipitation and flood warning, especially for developing countries where lack of detailed precipitation data is an issue (Setiawati and Miura, 2016; Hossain and Lettenmaier, 2006). There is an absolute need for these products for regional and global tracking of precipitation. SPPs are useful for hydrological research in areas where gauge networks are inadequate (Xue et al., 2013; Saouabe et al., 2020). Despite their promising application potential, SPPs usually have many shortcomings in providing accurate data on hydrology and water resource management at the basin scale (Yong et al., 2010; Zhou et al., 2022). The strength of SPPs lies in the highly accurate representation of the spatial distribution of precipitation; however, the ability to capture fine-scale features and small-scale variability in local rainfall intensity is limited (Mastrantonas et al., 2019; Zhou et al., 2022). Robust calibration and bias correction based on ground observations are essential to improve the accuracy of the SPP. However, bias correction reveals the limitations of these SPPs to accurately describe precipitation at the local scale.

1.1.1.1 Accuracy of Rain Gauges

Rain gauges are instruments used to quantify the amount of precipitation at a specific location. The most common method of ground-based precipitation measurement is the use of rain gauges to estimate punctual precipitation at a single location (New et al., 2001). These instruments are mainly located outdoors and either collect precipitation in a container or use a tipping bucket mechanism to measure precipitation intensity. Precipitation observations from rain gauges are characterized by exceptional accuracy at the point scale. At the regional scale, the accuracy of ground stations is compromised by the limited distribution and density of stations (He et al., 2023; Zeng et al., 2018b). Precipitation is a meteorological process and the spatial and temporal distribution of precipitation can vary considerably (Marani, 2005). The complexity of the topography can also influence this phenomenon. Insufficient distribution of rain gauges hinders the ability to spatially represent precipitation. Rain gauges provide detailed measurements of precipitation at the point scale (New et al., 2001). Gauges provide ground point measurements that promote site validation and bias correction of SPPs (Yang et al., 2016; Tapiador et al., 2020; Zhou et al., 2022; Tian and Peters-Lidard, 2010).

1.1.1.2 Errors in Rain Gauges

The rainfall measurement using rain gauges is prone to many types of errors, affecting both meteorological factors and the actual design of the gauges themselves. Meteorological conditions such as evaporation, temperature variability, air turbulence, and wind have been reported to seriously affect the accuracy of rainfall capture (Robinson and Rodda, 1969; Constantinescu et al., 2007; Sieck et al., 2007). Also, the type of design of the rain gauge used is critical in measurement accuracy since some gauges are not designed to capture all forms of precipitation with equal effect. It is also subjected to bias from air turbulence, flow deflection and wind effects that may cause smaller raindrops to miss the gauge or be tilted

hence introducing inaccuracies. As each rain gauge measurement contains some uncertainty due to speed, raindrop size, and finally the design of the gauge itself (Mueller and Kidder, 1972; Neff, 1977; Folland, 1988; Hanna, 1995; Nešpor and Sevruk, 1999; Chang and Harrison, 2005, Sieck et al., 2007).

Regionally, the accuracy of gauges is reduced due to limitations in distribution and density (He et al., 20-23; Zeng et al., 2018b). In regions with sparse gauge distribution, a single gauge often represents thousands or even tens of thousands of square kilometers (Ibrahim et al., 2015, Zhou et al., 2022). Poor observation networks affect the quality of rain gauge data. The estimation of precipitation distribution using a network of ground observations is also limited for the same reasons (He et al., 2023; Zeng et al., 2018b; Wang et al., 2021).

The improvement in the reliability of gauge data will include the elimination and correction of the uncertainties related to rain gauge measurements. Further, the errors can be minimized using the correction factors that consider the various meteorological variables, such as wind velocity, temperature, and precipitation intensity and type (Stisen et al., 2012; Hoffmann et al., 2016). The use of a number of rain gauges in the same area increases the knowledge of the local precipitation since the precipitation is determined by complicated environmental factors such as elevation.

1.1.1.3 Satellite Precipitation Products

Satellite-based products can be considered as an alternative to ground-based observations for precipitation data. Over the last three decades, satellite data have become a valuable tool for global precipitation monitoring (Levizzani and Cattani, 2019). Various versions of satellite-based precipitation products have been developed and improved (e.g: Adler et al. 2003; Ashouri et al. 2015; Hong et al. 2004; Hsu et al. 1997; Hsu et al. 1999; Huffman, 2019; Huffman et al. 2018; Huffman et al. 2010; Huffman et al. 2007). They provide globally consistent

measurements of precipitation in space and time with quasi-global coverage (Derin and Yilmaz, 2014). Several researchers have compared satellite precipitation products and found that multi-sensor ensemble products provide the highest quality data (Beck et al., 2020; Derin and Yilmaz, 2014; Gehne et al., 2016; Sun et al., 2014; Sun et al., 2018a; Zeng et al., 2018; Zhu et al., 2015). Precipitation patterns are easier to understand in SPPs, showing various precipitation characteristics like amount, probability, and type. Satellites provide very important data on precipitation, particularly at high altitude areas (Derin and Yilmaz, 2014) where access to gauges is very poor. Their coverage is high in space, but SPPs can perform near real-time monitoring; often, they do not resolve correctly the intensity of precipitation at the local scale (Mastrantonas et al., 2019; Zhou et al., 2022). These are the uncertainties in sampling, indirect observation errors, difficulties in estimating precipitation intensity, especially in regions of complex topography and convective systems (Lo Conti et al., 2014; Zhou et al., 2022). Despite many advantages offered by SPPs, the systematic mistakes are being made in the cases of lack of ground-based measurements as references (Yang et al., 2016), making them not sufficient to serve the general purposes of hydrology and water resources (Yong et al., 2010; Zhou et al., 2022).

1.1.1.4 Bias Corrections of Satellite Precipitation Products

SPPs normally exhibit bias and errors over areas of complex topography and diversified climates. This was pointed out by several studies such as: Chaudhary and Dhanya (2019), Kidd and Huffman (2011), Prakash et al. (2015, 2016), Tang et al. (2015), Tian et al. (2009), Xu et al. (2017). Such inaccuracies in precipitation estimation can also lead to errors in the precise distribution of precipitation over different areas (Mastrantonas et al., 2019; Zhou et al., 2022). Bias correction is applied as part of the process of tuning the satellite data to better match the gauge observations, thus improving the accuracy and reliability of the SPPs. Various bias correction methods exist (Maraun, 2016) and have been practiced to improve the

quality of SPP data. The simple scaling method is one of the common approaches (Boushaki et al., 2009; Lin and Wang, 2011; Tesfagiorgis et al., 2011; Vila et al., 2009). This approach follows additive or multiplicative bias factors to reduce the discrepancy between the satellite estimates and the reference data, hence enhancing the quality of SPPs. (Yang et al., 2016). Bias correction is quite vital for any sort of accurate climate prediction and reliable hydrological data (Yong et al., 2010; Zhou et al., 2022, Chaudhary and Dhanya, 2019).

The combination of SPPs with ground gauge observations is fundamental to complement each other (Rasmy et al., 2014; Zhou et al., 2022). Indirect applications of satellite sensors can introduce errors. Direct measurements of precipitation at a specific point can be used to detect and reduce these errors. In many cases, the accuracy of SPPs increases significantly when combined with gauged observations (Tapiador et al., 2020; Zhou et al., 2022). To illustrate, in one study, calibration of IMERG products with assimilation of ground observations can reduce errors by 47%-63% in the US (Tian and Peters-Lidard, 2010). Improvements in precipitation accuracy are common (Akinyemi et al., 2020; Chen et al., 2021; Jafarpour et al., 2022; Jiang et al., 2021; Yu et al., 2020; Zhang et al., 2022). However, studies considering the effects of complex topography on precipitation are much less (Amjad et al., 2020; Lei et al., 2021; Ward et al., 2011; Yu L.et al., 2020).

Over these years, progress has been excellent in meteorological satellites and satellite-based quantitative precipitation estimation (QPE) technologies (Tang et al., 2015; Yang et al., 2018; Zheng et al., 2021). As a matter of fact, the integration of SPPs into ground observations enhances their reliability and continuity in space. This study combines ground observations with environmental data through a machine learning approach to IMERG satellite data, focusing bias correction over highly variable topography in Turkiye. This represents the first daily precipitation study using ancillary environmental features in a bias correction framework supported by an explainable artificial intelligence with SHAP.

1.2 Importance of Study

1.2.1 Goals and Motivation of the Study

SPPs, though presenting high spatial resolution, are not as widely applied in hydrology and water resources due to their relatively lower accuracy in terms of representing the spatial distribution of precipitation (Yong et al., 2010; Zhou et al., 2022; Mastrantonas et al., 2019). In contrast, the well-documented good point-scale accuracy of the gauge data (He et al., 2023; Zeng et al., 2018b) increases the spatial reliability when integrated into the SPPs. Many previous types of studies have estimated the output performance of model and satellite-based products only for complex topography (e.g., Derin et al., 2016; Derin and Yilmaz, 2014; Gampe and Ludwig, 2017; Hobouchian et al., 2017; Mei et al., 2014; Xu et al., 2017). Several studies focused on complex and non-complex topographic features simultaneously (e.g. Beck et al., 2019; El Kenawy et al., 2015; Mayor et al., 2017; Sharifi et al., 2016). Some of the previously mentioned studies (e.g., Derin and Yilmaz, 2014; Mei et al., 2014) were completed prior to the release of the promising precipitation product called IMERG.

Traditional statistical methods tend to be less capable of dealing with the non-linear conditions and high-dimensional variability of environmental variables (Donoho, 2000; J. Fan & Li, 2006; Johnstone & Titterington, 2009).The popularity of machine learning promises huge improvements, thanks to big data, advanced algorithms such as deep learning and ensemble methods, and increased computational power (Li, 2022).Random Forest algorithm significantly enhances satellite-based precipitation estimation (Li et al., 2021b; Lao et al., 2021; He et al., 2016).Numerous studies have underlined that XGBoost proves its efficiency in regression related tasks, showing outstanding performance even with complex and high-dimensional data (e.g., Zhang et al. 2018; Zhong et al. 2018; Nguyen et al. 2019; Feng et al., 2021).This study focused on the bias correction of the IMERG precipitation record by investigating how environmental features and their

interactions impact the performances of different models. SHAP is used to enable model-agnostic interpretability, showing the importance of the features and their contribution to the result (Li, 2022). The motivation of this study is to uncover relationships between environmental factors and satellite precipitation data in order to obtain robust local precipitation data. Within this motivation, the goals of this study are;

- Improving the accuracy of satellite precipitation products (SPPs) in Turkiye.
- Solving the spatial representation problem of IMERG by using ground observation data
- Training machine learning models using environmental features and apply them to ungauged areas
- Interpreting the control of complex environmental features, such as topography, on precipitation patterns over Turkiye.
- Utilizing the SHAP library to explain and interpret machine learning models to understand model decision mechanisms
- Producing robust and accurate precipitation products with using only one satellite product with related environmental features

CHAPTER 2

STUDY AREA AND DATASETS

The goal of this chapter is to inform the reader about the study area and the datasets. An examination of the dataset is given to evaluate the models and their training steps. This chapter also provides information about preprocessing, which is used to improve the quality and relevance of the data and thus optimize model performance.

2.1 Study Area

In the Turkiye, topographical varieties from high mountains to arid plateaus to coastal influences give a characteristic shape to different precipitation patterns and climatic conditions (Amjad et al., 2020). Starting from the north to the south, these parts contain major mountain ranges on the ridge orographic and a flat plateau in the middle (Amjad et al., 2020). Topography is among the key factors affecting the precipitation and precipitations pattern of Turkiye (Amjad et al., 2020). The country has a diversified climatic condition, with its coastal areas usually mild and interior Anatolian Plateau experiencing extremely hot summers with cold winters and minor annual precipitation (Sensoy, 2004; Amjad et al., 2020).

Selection of the study area was performed to capture the unique peninsular shape and diversified climate and complex topographical characteristics of Turkiye. The meaning of "complex topography" varies according to different literatures: To illustrate, some refer to it in terms of high elevation (e.g., Dinku et al., 2002; Hirpa et al., 2010; Milewski et al., 2015), while others describe it as the standard deviation of elevation (e.g., Chiaravalloti et al., 2018). Among studies, a clear

explanation for the complex topography is lacking. The reason in Turkiye is high elevation, steep slopes, and the resulting impact on variable climatic conditions (Amjad et al., 2020). Due to topography, the amount of precipitation is certainly influenced by such factors because air, warmed through the influence of topography while ascending, leads to condensation along slopes and results in higher precipitation where the gradients are steep (Hughes et al., 2009). These would be the topographic and climatic conditions that will seriously challenge the accurate measurement of precipitation with the use of the Satellite Precipitation Products.

In Turkiye, the long-term average annual precipitation was recorded as 573.4mm (1991-2020 period), During 2023, the average annual precipitation was approximately 12% higher than the long-term average reaching 641.5 mm (General Directorate of Meteorology (GDM), 2023). The selection of the area depends on the number of rain gauges. The gauges should be representative of the area. Figure 2.1 shows the distribution of quality controlled rain gauges in the area. 301 gauges distributed around the study area represent different climatic and topographical parts of Turkiye, of which 244 are green, linear gauges and the remaining 57 are purple, triangular test gauges. Coastal and inland gauge total precipitation amounts change because of mountain blockage (Amjad et al., 2020). There exist a number of studies performed to characterize the accuracy of precipitation products over Turkiye (e.g., Bıyık et al., 2009, Demir et al., 2018, Derin and Yilmaz, 2014, Toros et al., 2018, Yucel, 2015, Yucel et al., 2011, Yucel and Onen, 2014).

The study area defined and presented in Figure 2.1 introduces diversified topographical and climatic features. The highest annual precipitation in Turkiye occurs in the Black Sea region, due in part to the barrier effect from the surrounding mountains (GDM 2023; Gottardi et al. 2012). The cross-section in Figure 2.2 shows the relatively dry plateau and contrasting orographic features on either side. The area selected for showing the complex topography conditions influence in precipitation patterns of Turkiye.

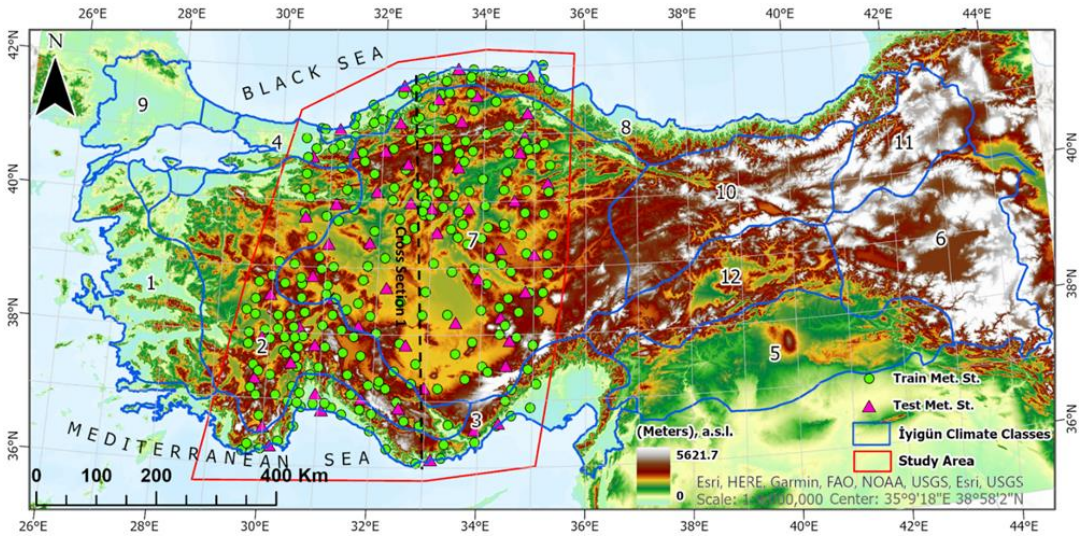


Figure 2.1. Study Area

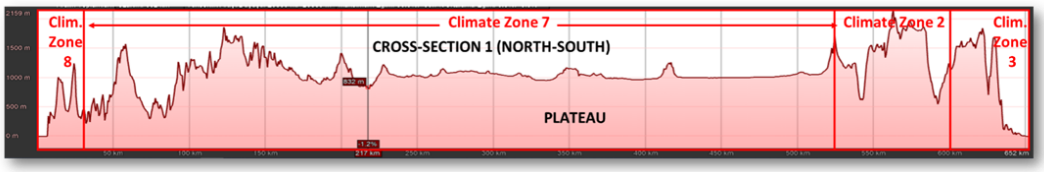


Figure 2.2. Cross Section of Study Area

2.1.1 Bias Correction of Satellite Precipitation Products, Importance and Literature Review

2.2 Datasets

2.2.1 Ground Based Precipitation Data

Ground-based precipitation observations are essential for understanding precipitation patterns. It is measured by an instrument called a gauge. Gauges measure points of precipitation on the surface. They are located in different environmental areas to collect information on the intensity of rainfall over time. The ground-based gauge dataset for this study was provided by the General Directorate of Meteorology (GDM). The most common instrument used to measure

precipitation is the rain gauge. These instruments aggregate rainfall at a specific point and measure the amount of precipitation. A total of 301 qualities controlled and complete rain gauges were used in the study. Figure 2.1 shows a map of the study area and the distribution of gauges. The data covers the period between January 2015 and December 2022.

2.2.2 IMERG/Satellite-based Precipitation Product

Integrated Multi-Satellite Retrievals for GPM (IMERG) is an algorithm developed by NASA that has been evaluated for estimating surface precipitation from satellite retrievals with global coverage (Huffman, 2019; Huffman et al., 2018). NASA applied the IMERG algorithm to both TRMM-era and GPM-era data, creating a relatively long (more than 20 years), high spatial (0.1 degrees) and temporal (30 minutes) resolution satellite-based precipitation record with near global coverage. The Global Precipitation Measurement (GPM) mission is an international satellite mission launched by NASA and JAXA (on February 27, 2014) following the success of the Tropical Rainfall Measuring Mission (1997-2015; Gebregiorgis et al., 2018). The main component of the GPM mission is the "Core Observatory" satellite, which carries an advanced radar/radiometer system to measure precipitation from space and serves as a reference to unify precipitation measurements from a constellation of satellites. IMERG products are available for free download (https://gpm.nasa.gov/data/imerg). The IMERG algorithm consists of three products. They are Early, Late, and Final. The Early and Late IMERG products are satellite-only products with a lag of 3 hours and 12 hours, respectively. They are near real-time products. IMERG Final Run, on the other hand, uses MERRA2 for the vertically integrated vapor, GPCC monthly monitoring analysis for the gauge, and revised precipitation retrievals that depend on ERA-5 to further correct the satellite-based precipitation retrievals. The integration of these additional data sets requires a latency of about 3.5 months for the final run.

These products used in the IMERG algorithm to update data from TRMM, which outperforms the TRMM Multi-Satellite Precipitation Analysis (TMPA) algorithm (Gebregiorgis et al., 2018). The GPM core satellite is a multi-channel, dual-polarization passive microwave (PMW) sensor with an active scanning radar. Compared to the previous system, the TRMM satellite, this system has several improvements: the orbital inclination has been increased from 35° to 65° for better coverage; the radar has been upgraded from single to dual frequency, while "high frequency" channels have been added to the PMW, allowing and enabling the detection of light and solid precipitation, respectively. (Huffman et al., 2015; Hou et al., 2014; Ramsauer et al., 2018). The IMERG algorithm collects data from multiple satellites and merges them to obtain global precipitation.

The IMERG Late product is a near real-time, gridded, multi-satellite global precipitation estimate with quasi-Lagrangian time interpolation provided every 30 minutes at 0.1° × 0.1° (Huffman., 2019; Huffman et al., 2018).

2.2.3 Digital Elevation Model (DEM)

This study used the Copernicus European Union Digital Elevation Model (EU-DEM 1.1). This DEM is part of the European Union Earth Observation datasets under the Copernicus program. The spatial resolution of the DEM is 25 m with vertical accuracy: +/- 7 m RMSE (EU-DEM 1.1). The DEM data were used in this study to illustrate the spatial distribution of precipitation as a function of topography and elevation change. The effects of topography and related features are considered to provide a more reliable and accurate representation of precipitation. DEM data were also used to obtain Effective Terrain Height (ETH), Distance to Coast, and Facet features to represent not only elevation but also other related feature effects on precipitation. These topographic features were selected considering the Precipitation-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008; Daly, 2006; Daly et al. 2002).

2.2.4 Climate Zones

Climate regions are important for developing effective strategies in meteorology, hydrology, forestry, and agriculture. Iyigün et al. (2013) used a hierarchical clustering method called Ward's method to classify climate regions in Turkiye according to temperature, precipitation and humidity data from meteorological stations for the period 1970 to 2010 (Iyigün et al., 2013). This study used 224 meteorological stations across Turkiye and identified 12 climate zones (Iyigün et al., 2013). Figure2.1 shows the climate classes with numbers. In the study area, 5 different climate classes influence our models. These climate classes are presented and defined below:

The Dry-subhumid Mid-Western Anatolia Region is described as a transition zone between the semi-humid Aegean and the humid Mediterranean to the dry subhumid/semiarid continental central Anatolia region (Iyıgün et al., 2013). This region is located in the western part of the study area. This climatic region is symbolized as number 2 and is shown in figure 2.3.

The Dry Summer Subtropical Humid Coastal Mediterranean Region is coastal desert in winter experiences a distinct season of mid-latitude cyclones and tropical high pressure systems in summer. This climate class is seasonal in that there is always high rainfall in the winter season and low rainfall in the summer. This climate is also characterized by Mediterranean forests and scrub (Iyıgün et al., 2013). These are symbolized by the number 3 and are shown in Figure 2.3.

Semihumid Eastern Marmara Transition Sub-region is specified as transition climate region in between the Mediterranean, and the Black Sea climate regions. The vegetation of this climate is mainly mixed or pure dry forests (both conifers and broad-leaved deciduous and red pine and oaks) and maquis (Iyıgün et al., 2013). The northern part of the study area has this climate region and is shown in figure 2.3 as number 4.

Dry-subhumid/Semiarid Continental Central Anatolia Region is characterized by a mesothermal and microthermal Continental Central Anatolia climate. The indicated vegetation consists of dry forests and expansive steppe lands across its huge plains, plateaus, and highlands (Iyıgün et al., 2013). This climate region is presented in figure 2.3 as number 7. The central part of the Anatolian region and the study area belong to this climate region.

Mid-latitude Humid Temperate Coastal Black Sea Region occurs in the mid-latitude region of the Black Sea coastal belt. The Black Sea coast, with the exception of the part in the Marmara region, is considered to be within the humid temperate region of the mid-latitude Black Sea coast. This region is subject to precipitation throughout the year. Due to the effects of mid-latitude cyclones and the orographic uplift of polar air masses, the heaviest rainfall occurs in the autumn season. The vegetation is characterized by humid boreal mixed forests It is shown as number 8 in Figure 2.3.

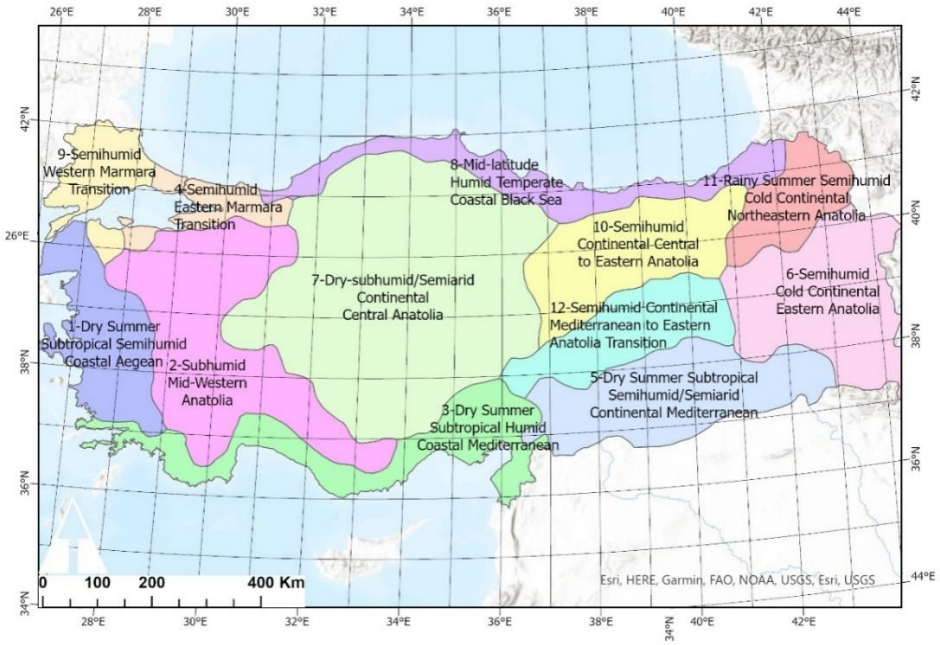


Figure 2.3. Map of Iyigün Climate Classes

2.3 Pre-processing of Data

2.3.1 Pre-processing

SPP needs to be integrated with ground gauge observations (Rasmy et al., 2014; Zhou et al., 2022). The bias in the satellite precipitation data is corrected using the gauge data. The structure of precipitation is complex and therefore highly variable in space and time (Marani, 2005). Satellites do not account for topography, and this feature improves model performance (Senocak et al., 2023), so precipitation-related environmental features were obtained to train the models.

Eliminating biased gauges improves the quality of the target variables. Following the PRISM methodology of Daly et al. (2008), distance to shore facet and elevation were selected to improve model accuracy. The IMERG data are pre-processed to a daily value and synchronized with gauge data. The pre-processed input data were converted to tabular form. Categorical variables converted to one-hot encoded features to make them easier for machine learning models to understand. Table 2.1 shows an example of an input table used to train the model (F: Facet, CR: Climate Region).

Table 2.1. Input table sample

St_No	Obs	IMERG	PLP	Lat	Long	Elev	Dist_Coast	ETH	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	CR_2	CR_3	CR_4	CR_7	CR_8
17015	0	0.00	1	41.0895	31.1374	10.00	715.891	314.88	0	0	0	0	0	0	1	0	0	0	0	0	1
17602	0	0.35	1	41.7526	32.3827	41.07	50	241.55	0	0	0	0	0	0	1	0	0	0	0	0	1
17652	0.6	2.58	1	40.9787	34.8011	419.00	91416.1	548.84	0	0	0	0	1	0	0	0	0	0	0	1	0
17723	0	0.17	1	39.3659	31.0209	900.00	178095	199.03	1	0	0	0	0	0	0	0	0	0	0	1	0
17744	0	5.48	1	38.7191	32.175	1002.00	228828	176.33	0	0	0	0	0	0	1	0	0	0	0	1	0
17745	0	2.15	1	39.3038	34.3421	1127.00	279380	421.54	0	1	0	0	0	0	0	0	0	0	0	1	0
17823	0	11.84	1	38.5529	31.173	1150.00	192155	353.22	1	0	0	0	0	0	0	0	0	0	0	1	0
17826	7.8	1.92	1	38.1047	30.5577	959.00	138525	636.20	1	0	0	0	0	0	0	0	1	0	0	0	0
17882	0	2.72	1	37.8377	30.872	920.00	108865	803.98	1	0	0	0	0	0	0	0	1	0	0	0	0
17892	0.9	14.10	1	37.3161	29.7792	1142.00	92447.1	781.36	0	0	1	0	0	0	0	0	1	0	0	0	0
17906	7.6	1.46	1	37.548	34.4867	1453.00	85153	982.22	0	0	0	0	0	0	0	1	0	0	0	1	0
18004	0	1.09	1	38.8768	30.7551	1153.00	218376	310.56	0	0	0	0	1	0	0	0	0	0	0	1	0
18037	1.1	22.17	1	37.9309	29.8667	900.00	139278	464.65	0	0	1	0	0	0	0	0	1	0	0	0	0
18047	12.2	23.01	1	37.0468	31.7971	1063.00	43717.9	1259.99	0	0	0	0	0	1	0	0	0	1	0	0	0
18062	2.1	0.76	1	36.5839	33.9267	1204.00	23311.5	1063.01	0	0	0	1	0	0	0	0	0	1	0	0	0
18074	0	0.26	1	39.9194	32.9944	1013.00	196399	516.29	0	0	0	0	0	0	0	1	0	0	0	1	0

2.3.1.1 Quality Control of Gauges

The quality of the target data is directly related to the quality and reliability of the results. Calibration issues can affect the quality of rain gauge data (e.g., Robinson and Rodda, 1969; Constantinescu et al., 2007; Sieck et al., 2007). Therefore, it is critical to account for errors in gauge data that create the potential for poor generalization of models. In this study, to ensure the quality of the target variable, gauges with more than 10% missing (null) values were eliminated to reduce noise in the data. Then, the cumulative precipitation of each station is compared with nearby stations by considering topographic effects with double mass curves (DMC) to understand the consistency and harmony of rain gauges. Anomalies are detected and related hydrographs are also examined to eliminate gauges with bias. After eliminating insufficient gauges, the sufficient number of gauges is 301 to represent the spatio-temporal variability of precipitation within the study area.

The gauges with a high amount of missing data or data with bias introduce noise into the models, which can distort the model outputs and reduce the reliability of the models. In addition, using only gauges with rich and accurate records reduces bias and improves the overall performance of the Random Forest and XGBoost models. Figure 2.4 shows an example of the DMCs used in the study plots (a-coherent, b-non-coherent rain gauge data). Each station was compared with its neighbors to find errors in the record. Topographic factors were meticulously considered. Plots of DMCs, along with hydrographs and vital statistics (R^2, slope, bias ratio, and number of missing data) were drawn to clearly assess station performance. Incoherent plots have vertical or horizontal line structures. Hydrographs are checked after observing these anomalies, then stations with missing records were eliminated (Figure 2.4 plot b:x axis station was eliminated).

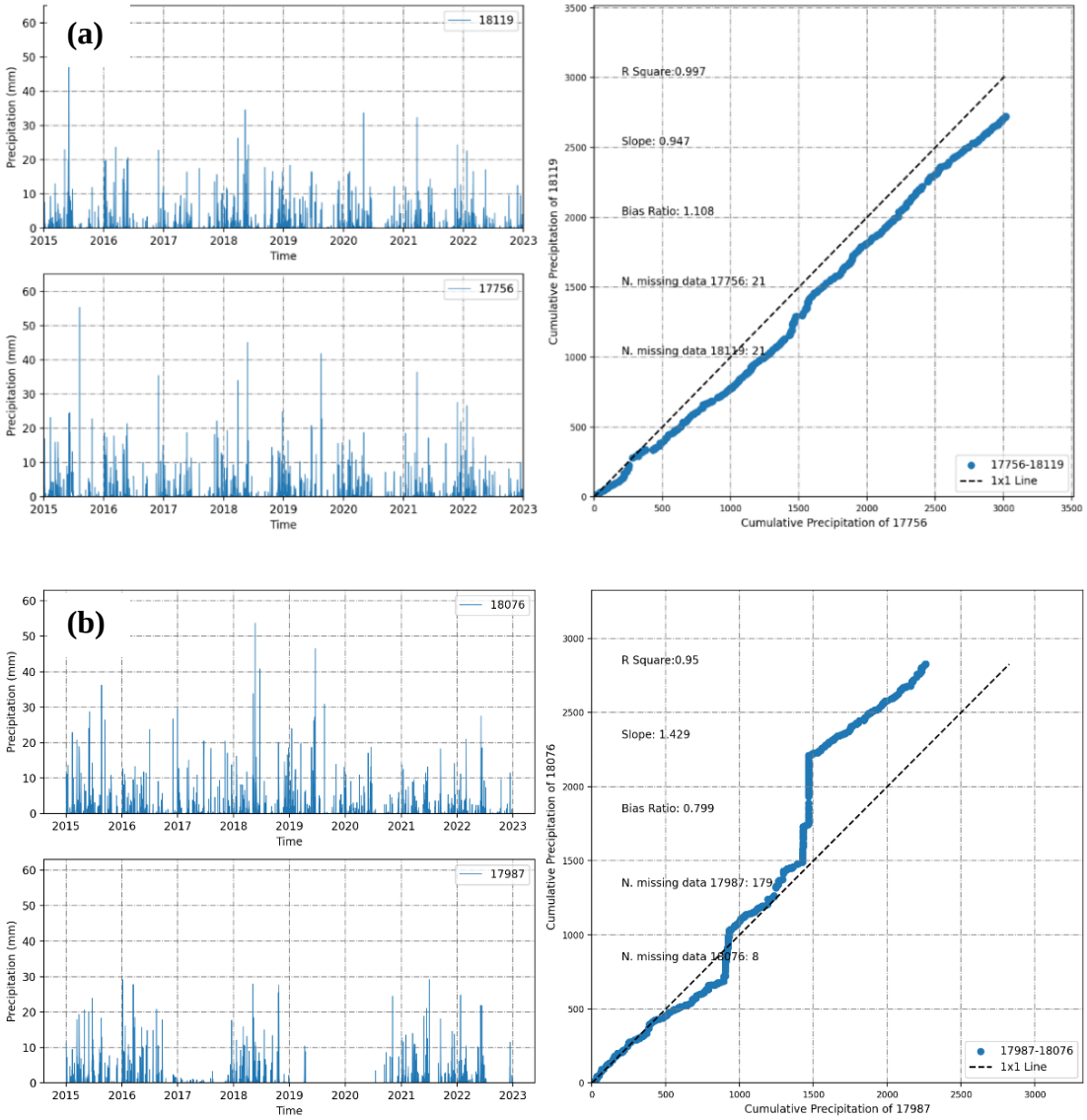


Figure 2.4. Double Mass Curve Graphs to Compare Coherence Between Gauges

CHAPTER 3

METHODOLOGY

The purpose of this chapter is to inform the reader of the methodological processes. Bias correction algorithms and their working principles are also covered to ensure that the reader understands the whole process. In addition, this study uses advanced bias correction algorithms and explainable artificial intelligence (XAI) techniques, mainly from the SHAP library, to provide transparency in model decisions.

3.1 Features

Satellite precipitation estimates are derived from measurements of meteorological quantities by highly specialized sensors on satellite platforms. However, this approach often fails to adequately represent topographic effects on precipitation, as factors such as slope and elevation changes have the potential to influence local precipitation pattern characteristics. Higher slopes and elevations receive more precipitation due to condensation or orographic lift (Hughes et al., 2009). There are numerous studies showing that SPPs contain significant uncertainties over complex topography because they are unable to represent the effect of topography on precipitation when the influence of topography is high (e.g., Derin et al. 2016, Hirpa et al., 2010; Krakauer et al., 2013; Mantas et al. 2015; Thouret et al. 2013,). Some studies show that SPPs contain large errors when orographic precipitation is highly effective over the area (Derin et al., 2016; Derin and Yilmaz, 2014; Dinku et al., 2007; El Kenawy et al., 2015). These studies were considered in the feature selection.

Proper feature selection is key to quality bias correction and model development. A feature should represent a region with due clarity and relevance to the task. Table 3.1 shows the selected features. There are two groups of features in this study: static and dynamic. Static features are Iyigün Climate Regions, Facet, Elevation, Distance to Coast, Effective Terrain Height (ETH), Longitude, and Latitude. The IMERG products represent dynamic features such as daily precipitation and probability of liquid precipitation.

Table 3.1. Machine Learning Algorithm Feature Classification

Features	
Static	**Dynamic**
Iyigün Climate Classes (Iyigün)	IMERG
Facet	Probability of Liquid Precipitation (PLP)
Elevation (Elev)	
Distance to Coast (Dist_Coast)	
Effective Terrain Height (ETH)	
Longitude (Long)	
Latitude (Lat)	

In this study, Facet, Effective Terrain Height, and Distance to Coast features were created based on the Precipitation-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008; Daly, 2006; Daly et al. 2002) in order to gain a deeper understanding of the effect of complex topography on the precipitation distribution in Turkiye.

3.1.1.1 Facet

Facet is the relationship between the slope of the terrain and its frequent orientation. Facets group different landforms, ranging in size from gigantic

mountains to gentle slopes. (Daly et al. 2008). A Gaussian filter was applied in the Matlab environment to generate facet data from the DEM. Wavelengths with different values are applied to the original DEM data for smoothing. The filter wavelengths applied to the DEM data have 3 values (λ= 5, 10, 12.5 km). The original DEM resolution is 25 m. Based on the DEM resolution, the filter wavelengths are determined by value differences to obtain better results. The specific orientation of a cell is determined by the orientations of neighboring cells within a radius equal to the wavelength. This approach ensures that the applied filter does not calculate orientation based only on the applied cell and its neighbors. It is determined by elevation variations over a larger surrounding area. The orientation value of each cell in the DEM data was then determined by comparing its elevation to its 8 neighboring cells, with 4 neighboring cells given double weight. Differences between cell values were taken into account to assign numbers between 1 and 8, corresponding to N, NE, E, SE, S, SW, W, and NW compass orientations, respectively (Daly et al. 2008).

3.1.1.2 Distance to Coast

The Distance to Coast feature is generated from the coastline and DEM data in ArcGIS using the Euclidean Distance tool. The generalized coastline was used because bays and inlets were not considered to be important sources of moisture for precipitation compared to the open sea. The purpose of calculating the distance to the coastline feature is to examine precipitation occurrence in relation to proximity to large bodies of water (Daly et al. 2008). Coastal areas may have different precipitation patterns due to interactions between sea and land. Turkiye has some sea affecting precipitation such as the Mediterranean Sea, and the Black Sea.

3.1.1.3 Effective Terrain Height

Using the Copernicus DEM data, the Effective Terrain Height (ETH) feature was obtained in the ArcGIS environment. Firstly, the minimum elevation within the 40km radius is determined for each cell. The spatial average of the minimum values of each cell is calculated to smooth the DEM data. This smoothed elevation data was then extracted from the original DEM data to obtain the ETH value. Finally, this value is smoothed by averaging within a 20 km radius (Daly et al. 2008).

3.1.1.4 Probability of Liquid Precipitation (PLP)

Satellite precipitation products tend to contain more errors when estimating precipitation, including snowfall, and the performance of the products decreases (Derin and Yilmaz, 2014). The active radar on board the GPM mission allows the identification of the phase of precipitation by focusing on different wavelengths. In this study, the probability of liquid precipitation (PLP) of the IMERG product is also used to understand the effect of the precipitation phase on the satellite products and to correct the bias accordingly. This feature takes values between 0 and 1. Higher PLP values indicate rain. A value of 0.5 represents a mixture of rain and snow. Snow and snow pellets are represented as equal to or less than 0.5. A value of 0 represents ice. Dew and frost are not events directly related to precipitation, so the IMERG PLP data do not include information on them (Huffman., 2019).

3.1.1.5 Correlation of Features

Features of any machine learning model can contain similarities (Nohara et al., 2022), which in turn can reduce the effectiveness of machine learning models. Maximum relevance and minimum redundancy are desired to make machine learning model results more accurate and robust (Zhao et al., 2019). The correlation

matrix of features is constructed to determine pairs of related features. Related features provide the same information to machine learning models called as feature redundancy and negatively affect the performance of the model because models may not learn related information from other features (Zhao et al., 2019). Correlations of features can prevent interpretation. Therefore, correlated features are usually removed before training the models (Nohara et al., 2022). Thus, feature correlations are examined before model training to increase efficiency, and these correlations between static features are shown in Figure 3.1. Only time-independent feature correlations are considered to provide an overall view of the features. The feature correlations are not high. Therefore, these features are suitable for machine learning algorithms. Positive correlations show tendency of features to increase together while negative value of correlations indicates that increase in one feature resulted as decrease for other one (Nicodemus & Malley, 2009). Correlation value of 0 indicates that there is no correlation between features (Nicodemus & Malley, 2009).

Highly correlated features provide redundant information to the model in the training step (Zhao et al., 2019). With feature redundancy, models become overconfident and produce results without considering other features, which reduces the generalization ability of models (Zhao et al., 2019). To avoid redundancy of features, this study only considers IMERG, a merged satellite precipitation product, in the training process. Using other satellite precipitation products along with IMERG creates redundancy and lessens the effectiveness of the model. It further removes the effects of environmental features that the satellite precipitation products cannot represent.

Shapley values also produce unrealistic results when features are correlated (Salih et al., 2024). High correlations require feature engineering to construct new features Feature engineering categorizes and converts existing data or provides new information (Chollet, 2017; Senocak et al., 2023).

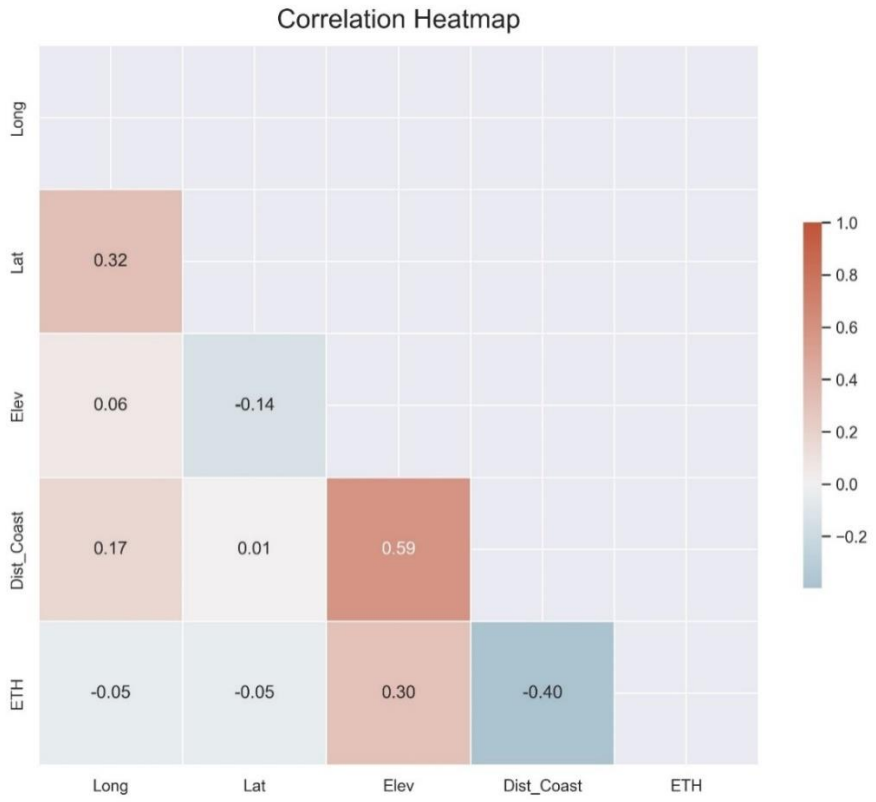


Figure 3.1. Feature Correlation Matrix of Static Feature

3.2 Bias Correction Algorithms and Explainable Artificial Intelligence (XAI)

Satellite precipitation products have inherent errors, but correcting their bias leads to significant improvements, increasing their value in hydrological modelling studies (Derin & Yilmaz, 2014; Yilmaz et al., 2005; Su et al., 2008; Thiemig et al., 2013). Tree-based AI models such as random forest and XGBoost provide more interpretable with higher prediction accuracy in precipitation than traditional statistical models (Başağoğlu et al., 2022; Chakraborty et al., 2021a; Chang et al., 2016; Dumitrescu et al., 2021). On tabular data containing independent and meaningful features, tree-based machine learning models outperform neural network-based deep learning models (Lundberg et al., 2020; Feng et al., 2021). The

advantages of tree-based algorithms lie in their capability of understand intricate patterns between parameters resulting in robust output data.

The inclusion of environmental parameters increases the accuracy and quality of the corrected data. In areas where environmental features are especially effective drivers of precipitation (such as topographic complexity). Multiple features related these drivers improve performance of models.

3.2.1 Model Selection

The model selection process is a critical part of the bias correction of IMERG precipitation products. Several factors come into play when selecting a model, including data characteristics (type, volume, linearity), accuracy, interpretability, and effectiveness of the model for the specific subject. Each model has advantages and disadvantages. To illustrate, high-dimensional nonlinear data may not be handled by simple models. Simple models often fail to learn the complex internal patterns of the given data in a meaningful manner. Therefore, they memorize the data, a phenomenon called overfitting. More complex models may struggle with interpretability and transparency (Lin et al., 2023).

The machine learning models evaluated in this study are random forest (RF) (Breiman, 2001) and extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016) for satellite-based precipitation estimates. Models are selected based on high prediction accuracy in multivariate nonlinear problems in different domains (e.g., Chakraborty et al., 2021b; Ben Jabeur et al., 2021; Qiu et al., 2020; Geurts and Louppe, 2011; Acosta et al., 2020; Jabeur et al., 2022). The effectiveness of machine learning has been widely demonstrated under conditions of complicated problems with non-linear datasets. These conditions make it difficult to build physical models using conventional mathematical and statistical analyses (El-Alfy and Mohammed, 2020).

3.2.2 Evaluation of Model Training Process

The selected gauge target variable and environmental features have been carefully prepared in daily tabular form as input to the machine learning models (Random Forest and XGBoost) to correct for the IMERG precipitation bias in Turkiye

To achieve robust precipitation data in Turkiye, XGBoost and Random Forest models were trained separately for each day. Random Forest model was implemented from Python scikit learn random forest regressor models (Pedregosa et al., 2011), XGBoost python package was used to fit XGBoost models (developed by Chen & Guestrin, 2016).

Several metric scores calculated to understand and evaluate the model performance, detailed information given in Metric used for model evaluation part and metric results and evaluations are provided in the Results chapter. Models run on a daily basis, so their performance varies. Models with the highest scores are generally found on days when more gauges' report precipitation (higher coverage).

3.2.3 Transparency of Machine Learning Models

3.2.3.1 Difference between Interpretable Machine Learning (IML) and Explainable Artificial Intelligence (XAI)

Interpretable Machine Learning (IML) and Explainable Artificial Intelligence (XAI) are concepts about the human understandability of machine learning models. However, these concepts differ from each other. IML includes models with transparent decision mechanism and understandable by humans without considering other explainers, therefore called as white box (Ersöz et al., 2022). Examples of IML models are linear regression and decision trees. On the other hand, Explainable AI focuses on explaining the results of more complexes, often "black box" models by using explainers to make their decision mechanisms

transparent by showing process of decision (Ridley, 2022; Senocak et al., 2023). Examples of models that require XAI are deep neural networks or ensemble methods such as random forest and XGboost.

3.2.3.2 Concept of XAI

The concept of Explainable AI aims to shed light on the inner workings of machine learning models, which are often seen as impenetrable black boxes. This black-box nature can lead to unexpected performance and also makes human inspection of such models impossible. (Castelvecchi, 2016; Senocak et al., 2023). As artificial intelligence touches more and more areas of our lives, the importance of Explainable AI is growing.

The Random Forest and XGBoost algorithms are ensemble methods that build numerous decision trees, creating complex structures that reduce the understandability and trustworthiness of the models. Fortunately, they can be interpreted globally or locally using XAI techniques. Global explainers provide insight into the overall behavior of a model, helping to understand how it makes decisions across the dataset through general patterns and feature importance. Examples of global explainers include permutation feature importance (PFI; Breiman, 2001), accumulated local effects (ALEs; Apley & Zhu., 2020), and SHapley additive explanations (SHAP; Lundberg and Lee, 2017; Lundberg et al., 2020). On the other hand, local explainers focus on detailed explanations of individual predictions. Examples of local explainers are local interpretable model-agnostic explanations (LIME; Ribeiro et al., 2016) and SHapley additive explanations (SHAP; Lundberg and Lee, 2017; Lundberg et al., 2020).

One of the XAI techniques used in this study is SHAP (Shapley Additive explanations) (Lundberg and Lee, 2017; Lundberg et al., 2020). SHAP is a local and global explainer that finds the exact contribution of traits by considering all possibilities (Lundberg and Lee, 2017), whereas other local explainers such as

LIME are not exact values. Therefore, merging SHAP values of different predictions is possible while other local explainers introduce bias in the results.

In this study, Tree explainer (Lundberg et al., 2020), which is suitable for tree-based algorithms such as random forest and XGBoost, is utilized.

3.2.4 Random Forest

Random forest is a supervised machine learning algorithm (Breiman, 2001) that efficiently solves both regression and classification problems. During the training process, this algorithm creates an ensemble of decision trees (Gong et al., 2020), a resembling technique called bootstrapping generates random subsets from the original data (Efron 2000; Sushanth et al., 2023; Stef et al., 2023) to build each tree. Each subset is randomly drawn and may have duplicate rows to build different decision trees. This allows the model to better handle complex relationships, avoid overfitting, and improve generalization.

While performing classification, it aggregates the majority votes from trees (Mod), regression models calculate the average of each tree (Sushanth et al., 2023). The performance and functions of the models are determined by hyperparameters (Rong et al., 2020; Sam et al., 2020; Wang et al., 2021). The random forest algorithm is advantageous for satellite-based precipitation estimation (Li et al., 2021b; Lao et al., 2021). Random forest models from the ScikitLearn library are used in this study (Pedregosa et al., 2011).

Machine learning algorithms have become increasingly popular for bias correction (Zeng et al., 2021; Wang et al., 2020). To illustrate, He et al. (2016) developed a random forest method to improve precipitation in the United States. They used numerous spatial resolutions for downscaling and tested the performance of the product. They concluded that the spatial distribution is the main problem of SPPs, and this problem can be reduced by a random forest algorithm.

3.2.5 XGBoost

XGBoost is designed to extend the machine learning algorithm of the gradient boosting framework by using an ensemble learning method (Chen & Guestrin, 2016). Ensemble means building more than one base model to compute results. It constructs predictive decision tree models as weak learners by iteratively correcting the errors of previous models using a method called gradient descent to minimize a given hyper function (Chen & Guestrin, 2016; Chollet, 2017; Li et al., 2023; Stef et al., 2023; Senocak et al., 2023). As an improvement of the gradient-boost decision tree algorithm (GBDT) (Han et al., 2024; Hancock & Khoshgoftaar, 2020; Friedman, 2001), XGBoost is known for its high accuracy, speed in processing time, capture of complex nonlinear relationships, and computational efficiency (Ali et al., 2023; Zhu et al., 2023). XGBoost is well suited for bias correction of satellite precipitation products because the technique can effectively model the complex nonlinear relationships and thus correct systematic bias to improve the accuracy and reliability of satellite precipitation estimates (Ali et al., 2023; Zhu et al., 2023).

XGBoost provides efficient tree pruning, regularization to reduce the complexity of decision trees to improve performance and versatility. XGBoost is also a highly compatible model with SHAP and tree explainer can efficiently provide SHAP values (Lundberg et al., 2018). In a study comparing four ML-based algorithms and three interpolation methods, XGBoost was selected as the best performer in evaluating the downscaling of precipitation data via other parameters (Zhu et al., 2023). Other regression-related studies have also utilized XGBoost and recorded its success (e.g., Zhang et al., 2018; Zhong et al., 2018; Nguyen et al., 2019; Feng et al., 2021).

3.2.6 Hyperparameter Tuning

Model performance and accuracy depend on hyperparameters that define how the model operates to optimize accuracy (Chollet, 2017; Rong et al., 2020; Sam et al.,

2020; Wang et al., 2021; Verma et al., 2022; Senocak et al., 2023). Proper tuning ensures that the model captures the most important patterns in the data without being overly complex or biased. Effective hyperparameters tuning improves the model's ability to generalize to unseen data, resulting in better performance metrics and more reliable predictions by selecting optimal hyperparameters that maximize model performance and accuracy. This requires adjusting parameters such as number of trees and tree depth to prevent overfitting and improve model robustness.

Bayesian optimization improves performance and efficiency by quickly finding the optimal hyperparameters space (Klein et al., 2016; Stuke et al., 2020; Wang et al., 2021). Bayesian optimization emerges as a probabilistic model-based approach to effectively navigate the hyperparameters space through iterative processes. Using the Bayesian algorithm to optimize hyperparameters in random forest and extreme gradient boosting decision tree models has been shown to be highly effective (e.g., Wang et al., 2021). Different combinations of hyperparameters are explored by Bayesian search to find the combination with minimum bias and maximum efficiency and accuracy. The BayesianSearchCV function in sci-kit-learn (Pedregosa et al., 2011) is used to tune the hyperparameters. By systematically exploring different combinations of hyperparameters, Bayesian search selects the set of parameters that minimizes bias and maximizes the performance efficiency and accuracy of the models.

In the region with complex topography, the tuning of hyperparameters becomes particularly important due to the complicated relationships between parameters and precipitation caused by the complex topography and associated climate changes. It selects the most appropriate parameters by finding the minimum of objective functions in large problem spaces (Peter, 2018) so that environmental descriptors benefit with the highest efficiency. Understanding feature changes and topographic influences to improve the accuracy of satellite-based precipitation data.

3.3 Metrics Used for Model Evaluation

Metrics used for model evaluation and comparisons are Coefficient of Determination (R^2), Kling–Gupta efficiency (KGE), Alpha, Beta, r, Root Mean Square Error (RMSE), Probability of detection (POD), False Alarm Ration (FAR), Critical Score Index (CSI), Bias Error and Mean Bias Error (MBE).

Firstly, the accuracy of the model was analyzed using the Root Mean Square Error (RMSE), KGE and its components which are commonly employed in the literature (e.g. Zhu et al., 2023). All metrics reported in this study were strictly derived from the test data. Original filtered IMERG vs. ground gauge data statistics were also calculated to represent the harmony between satellite precipitation products and ground observations. The study aims to improve the simulation of precipitation data by exploiting the superior capabilities of advanced Random Forest and XGBoost algorithms for better fits to ground observations.

3.3.1 Filter Usage and Its Importance

In this study, the models were trained daily for each day, but since the model predictions for days with no precipitation are meaningless and distort the global feature importance scores, the models were selected by filtering all metric and SHAP plots. The models are filtered by the number of stations in the train data with 2 mm/day or more precipitation greater than or equal to 25 and the corresponding daily SHAP values of the test data in order to understand and represent the feature contributions in a meaningful way. The choice of this filter was based on the need to understand not only the days with high precipitation, but also the spatial distribution of the precipitation and how the models perceive this distribution. For example, since it is quite possible to observe precipitation in only one area in Turkiye, summer precipitation is only dominant in the Black Sea region of the study area (Sensoy et al., 2008). In order to keep the variability of the data above a certain level, 2 mm/day of precipitation was considered appropriate.

3.3.2 Coefficient of Determination (R^2)

The scikit-learn metric library includes numerous model selection and evaluation tools according to model types such as classification, regression, and clustering (Pedregosa et al., 2011). In this study, R2 (coefficient of determination) in scikit-learn is used to evaluate the performance of models and original filtered IMERG data from ground observations.

A value of 1 represents perfect representation. Negative values (-Inf is the minimum) are also possible for models that are worse than a base model. The output of a base model is always the average of the target variable (ground observation), ignoring the inputs. A model that always predicts the average target variable receives an R2 value of 0.0 (Pedregosa et al., 2011). In this study, days without rain or low precipitation are predicted by the models as 0 mm/days or close to 0 mm/days, and the average of the model predictions is almost the same as the average of the observations, so the R2 score is calculated as 0 close to 0. In this situation, there is no feature that affects the target variable, so it is pointless to use different features.

The mathematical representation of R^2 is given as:

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

$\hat{y}_i$ is the predicted value of the i-th sample

y_i is the corresponding true value for total n samples

$\overline{y}$ is the mean of all true values (Pedregosa et al., 2011).

The coefficient of determination is a very useful but limited metric, especially for complex real-world applications. Other topic-related evaluations must also be considered to define the overall performance of machine learning models.

3.3.3 KGE and KGE Components

Kling-Gupta Efficiency (KGE) is an important unitless metric in hydrological modeling recognizing that calibration of hydrological models is a multi-objective task (Gupta et al., 1998). KGE decomposes and obtains different components (correlation, bias and variability) from the Nash-Sutcliffe Efficiency (NSE) in the concept of hydrology (Gupta et al., 2009). In addition, each component of the KGE provides unique insights into the objective performance of models, highlighting their strengths and weaknesses:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \qquad (2)$$

Where:

r: correlation coefficient.

α: Variability ratio (standard deviation ratio of simulated vs observed).

β: Bias ratio (mean ratio of simulated vs observed).

Understanding the KGE and its components of correlation coefficient (r), bias ratio (beta), and variability ratio (alpha) is critical for effective water resource management. The correlation coefficient (r) shows how well the model output matches the observations over time. The bias ratio (beta) shows how the model overestimates or underestimates the overall magnitude of the observations, and the variability ratio (alpha) shows how the model overestimates or underestimates variability. The ideal value of each component - r, α, β - is 1.

3.3.4 Root Mean Squared Error

Root Mean Squared Error (RMSE) is a common metric used to evaluate model performance to show how well the model is working. RMSE is derived from the root mean square error. In this study, RMSE is calculated from Scikit-learn metrics (Pedregosa et al., 2011). The better the prediction of the model, the lower the

RMSE. The higher the value, the larger the error margin. The unit of the RMSE is the same as the unit of the dependent variable y (mm/day).

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}{n_{\text{samples}}}} \quad (3)$$

In equation 3;

$\hat{y}_i$ is the predicted value of the i-th sample

$y_i$is the corresponding true value (Pedregosa et al., 2011).

3.3.5 Metrics for Precipitation Events and Extreme Precipitation Events

Traditional metrics such as Kling-Gupta efficiency or other conventional performance metrics tend to be inadequate to adequately capture model performance with respect to extreme precipitation events. Therefore, in this study, precipitation events were grouped according to precipitation event thresholds that provide an accurate representation of model event performance. The general precipitation event thresholds are 1 mm/day, 2 mm/day, and 5 mm/day, and the extreme precipitation event thresholds are 10 mm/day, 20 mm/day, and 50 mm/day. The purpose of the threshold is to provide an accurate representation of model performance and to define true positive (TP), true negative (TN), false negative (FN) and false positive (FP) values.

3.3.5.1 Probability of Detection (POD)

Probability of Detection (POD) (unitless) is the first measure of model performance for precipitation and extreme events with event thresholds. The POD score represents the ability of the model to capture real events as they occur (Sharifi et al., 2016). POD is the proportion of correctly identified events out of all observed events:

$$POD = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

TP: True Positive: Events that correctly predicted

FN: False Negative: Events that actually occurred but models missed

The best POD score is 1, indicating that all events are predicted.

3.3.5.2 False Alarm Ratio (FAR)

False alarm ratio (FAR) (unitless) is the proportion of reported but not observed events over all reported events:

$$FAR = \frac{\text{FP}}{\text{TP+FP}} \quad (5)$$

FP: False Positive: Event actually not occurred but models predicted

TP: True Positive: Events that correctly predicted

The perfect score for FAR is 0 indicating that all predictions are true (Sharifi et al., 2016).

3.3.5.3 Critical Score Index (CSI)

The Critical Score Index (CSI) or Threat Score (TS) (unitless) represents the ability of the model to correctly identify events of interest, taking into account false alarms and missed events. In other words, the proportion of correct model predictions out of all observed and simulated events.

$$CSI = \frac{\text{TP}}{\text{TP+FP+FN}} \quad (6)$$

TP: True Positive: Events that correctly predicted

FP: False Positive: Event actually not occurred but models predicted

FN: False Negative: Events that actually occurred but models missed

A perfect score of CSI is 1, indicating that all predictions are actual events (Sharifi et al., 2016).

3.3.5.4 Bias Score

The bias score (unitless) is the ratio of model predicted events to observed events, representing the tendency of the model to over- or underestimate the occurrence of events:

$$Bias = \frac{\text{TP+FP}}{\text{TP+FN}} \quad (7)$$

TP: True Positive: Events that correctly predicted

FP: False Positive: Event actually not occurred but models predicted

FN: False Negative: Events that actually occurred but models missed

The perfect value of the bias score is 1, which represents all predicted events occurring. Bias scores greater than 1 indicate overestimation, with models predicting more events than actually occur. Similarly, a bias score less than 1 is an underestimate, with models predicting fewer events than actually occur.

3.3.5.5 Mean Bias Error (MBE)

After analyzing the machine learning models performance according to precipitation thresholds. IMERG and the overall model behavior in terms of overestimation or underestimation is analyzed by Mean Bias Error (MBE).

$$MBE = \frac{1}{N}\sum_{i=1}^{N} P_{s,i} - P_{o,i} \quad (8)$$

$P_{s,i}$ is simulated precipitation data

$P_{o,i}$ is observed precipitation data.

A positive MBE indicates overestimation and a negative MBE indicates underestimation. The unit of the MBE is the same as the unit of the dependent variable y (mm/day).

3.4 Explainable AI and SHAP

Explainable AI (XAI) is used to find out the decision mechanisms of models by showing the relationships between input and output values and which features are most used by the model in the prediction step (Li, 2022). Machine learning models are generally considered to be black boxes (Li, 2022). XAI algorithms also remove this black box consideration. The main reasons for the importance of XAI are:

- Figuring out the reasoning behind output decisions,
- Contribution of each feature to the result,
- Checking bias and errors in models,
- Making decision steps clear and transparent,
- Enhancing trust in machine learning models by making them user-friendly and easy to explain,

There is no standard definition in the literature for the concept of "explainability" in AI models (Adadi and Berrada, 2018; Linardatos et al., 2020; Stef et al., 2023). However, if the output of the model can be clearly explained and the logic behind the decisions can be understood by humans, then the model is considered explainable (Miller, 2019; Stef et al., 2023).

Shapley scores are derived from game theory concepts to quantify and explain the contribution of each feature to the output result (Shapley, 1953; He et al., 2023). SHAP is one of the additive feature attribution methods (Lundberg and Lee, 2017). SHAP values are effective in elucidating the contribution of each feature to the model output and understanding the model behavior (Sushanth et al., 2023; Wang et al., 2023). The most common SHAP explicators are kernel SHAP (Lundberg and Lee, 2017) and tree SHAP (Lundberg et al., 2020). Tree SHAP, an interpretable

method for tree-based algorithms (Feng et al., 2021), improves Kernel SHAP by ignoring feature dependencies (Mi et al., 2020) and focusing on individual feature contributions. Tree SHAP does not evaluate every possible combination of features. Instead, it focuses on a set of computations specific to each leaf in the tree (Lundberg et al., 2020; Feng et al., 2021), and this approach makes the computational step faster and even more accurate.

In this study, Tree SHAP was used as an XAI to improve the interpretability and reliability of model results. The SHAP (Shapley Additive exPlanations) Python package helps to understand the impact of each feature on model predictions by calculating SHAP values (Štrumbelj and Kononenko, 2014; Li, 2022) and providing visualizations through various graphs, such as summary plots and individual feature importance ('force') plots (Wang et al., 2023).

3.4.1 Model Agnostic and Model Specific Interpretations

There are two different approaches to interpretability, model-specific and model-agnostic (Molnar, 2020). Model-specific interpretations are designed specifically for models and only for that type of model, with the goal of representing the internal structure. Some basic illustrations are; linear regression coefficients and smooth operating principles in a generalized additive model (Li, 2022). More complex models also take advantage of model agnostic interpretation, such as examining the activation of neurons and layers to see which parts of the image contribute to the final classification (e.g., Shrikumar, Greenside, & Kundaje, 2017; Simonyan, Vedaldi, & Zisserman, 2013; Yosinski, Clune, Nguyen, Fuchs, & Lipson, 2015). Model agnostic interpretation means that this methodology can be applied to any machine learning model, and generally does not represent the internal structure of the model. A deep understanding of the internal structure of the model is not required to understand the results of the model agnostic approach (Li, 2022).

SHAP's model-agnostic behavior allows it to be applied to any machine learning model, making it a valuable tool for increasing model transparency and understanding feature contributions. SHAP represents the direct influence of features on the prediction with the same unit as the output. Several studies have started to take advantage of SHAP (Ribeiro et al., 2016; Štrumbelj and Kononenko, 2014; Parsa et al., 2020). SHAP also has great potential for machine learning model discovery, especially for spatial data. Each observation in the training and test datasets has a unique geographic location (Li, 2022).

3.4.2 Understanding the Meaning of SHAP Features Importance

SHAP importance is quantified at the row level and reflects the contribution of a feature in making a particular prediction, considering other given features for that prediction, relative to the average prediction. This measure includes both the direction and the magnitude of the influence of the feature. In practice, SHAP importance is often presented in absolute terms for interpretation and comparison during model training (Molnar, 2020). SHAP scores highly effected by variation of given feature therefore reducing variation will reduce the effects of given feature.

3.4.3 Feature Importance Heatmap

The method of assessing the essential features in a machine learning model output is called Feature Importance (Musolf et al., 2022; Khan et al., 2020). Visual representations of the filtered SHAP Feature Importance (FI) scores for each daily model are provided as heatmaps for each year, showing similar patterns. XGBoost assigns zero importance to irrelevant features, so XGBoost's heatmaps contain a large amount of dark blue color (Appendix A SHAP Feature Importance scores as heatmap).

3.4.4 Merging SHAP Absolute Mean Plots

In this study, the RF and XGBoost models are trained on a daily basis. Absolute mean SHAP values are calculated to obtain an overall importance of each feature over the entire study period, i.e., 2015-2022, to provide a complete overview of the collective impact of the features in the RF and XGBoost models. The average importance of the filtered daily SHAP values is calculated to represent the importance of the features.

3.4.5 Variation of Descriptor Importance in Space

Filter used to pre-select the data prior to generating the descriptor importance variation plots. The following key static features were selected to examine their spatial variation in descriptor importance: Elevation, Distance to Coast, Latitude, and Longitude. Then, for all models from 2015 to 2022, the feature values and their corresponding SHAP values were sorted and plotted with respect to the magnitude of the feature value and the associated season.

SHAP values are used to explain the positive and negative contributions of each feature. A SHAP value of 0 indicates no effect and is highlighted in red along the y-axis. The median SHAP value is shown as a line extending from the 25th to the 75th percentile SHAP value. It describes the central tendency and variability of feature importance.

In order to investigate the seasonal dynamics in more detail, the present analysis focused on the two most important seasons: summer and winter. For this reason, the importance of selected static features was evaluated in terms of seasonal representativeness. This can be used as a basis for a fine-grained interpretation of how static features contribute to model performance in different seasonal contexts. This approach allows for a more detailed interpretation of the spatial and temporal dynamics of feature contribution to model prediction.

3.5 Summary of the Methodology Utilized in This Study

Satellite Precipitation Products (SPPs) have exhibited bias that have seriously limited their ability to reflect the true patterns and spatial distribution of precipitation. On the other hand, ground observations are very precise data at a specific location. This study aims to correct the bias in IMERG late precipitation data over Turkiye using environmental data and ground observations with machine learning algorithms.

The supervised machine learning algorithms Random Forest and XGBoost, known for their success in bias correction, are used to capture the complex relationships between precipitation and various other environmental parameters. The selected features included in the models are: climate class of a given location, facet, elevation, distance from the coast, effective terrain height, IMERG data, liquid precipitation probability, and geographic coordinates (longitude and latitude). In this study, 5-fold cross-validation with Bayesian optimization was used to fine-tune the models, improve model generalization, and more reliably capture complex relationships in the input data.

Evaluations and comparisons are then made based on metrics such as KGE, R2, CSI, POD, FAR, and SHAP values. SHAP explainers are used to improve model reliability by providing a better understanding of the contribution of features and behaviors in the model. By combining the SHAP values of the daily models, the overall effects and the effects of spatial and temporal features are also analyzed through different types of plots.

CHAPTER 4

RESULTS AND DISCUSSION

This chapter evaluates the performance of machine learning algorithms in bias correction of the IMERG precipitation product. The contributions of each feature were analyzed with different SHAP graphs.

4.1 Evaluation of Bias Correction Algorithms

Different metrics are calculated for each model, then their mean, median and standard deviation (Std) values between 2015 - 2022 are listed in Table 4.1. All metrics are filtered and calculated from test data.

Table 4.1. Effectiveness Metrics Evaluation between 2015-2022

R2			KGE			RMSE		
Filtered IMERG	Mean	-2.09	Filtered IMERG	Mean	-0.28	Filtered IMERG	Mean	7.08
	Median	-0.16		Median	0.07		Median	5.90
	Std	8.69		Std	1.29		Std	4.70
RF	Mean	0.47	RF	Mean	0.46	RF	Mean	4.00
	Median	0.49		Median	0.48		Median	3.34
	Std	0.25		Std	0.22		Std	2.60
XGBoost	Mean	0.33	XGBoost	Mean	0.47	XGBoost	Mean	4.33
	Median	0.40		Median	0.51		Median	3.58
	Std	0.63		Std	0.28		Std	2.86
Alpha			**Beta**			**r**		
Filtered IMERG	Mean	1.26	Filtered IMERG	Mean	1.28	Filtered IMERG	Mean	0.48
	Median	0.98		Median	0.94		Median	0.50
	Std	1.15		Std	1.22		Std	0.24
RF	Mean	0.62	RF	Mean	0.99	RF	Mean	0.71
	Median	0.61		Median	0.97		Median	0.74
	Std	0.19		Std	0.22		Std	0.17
XGBoost	Mean	0.77	XGBoost	Mean	0.99	XGBoost	Mean	0.65
	Median	0.75		Median	0.97		Median	0.68
	Std	0.31		Std	0.23		Std	0.19

4.1.1 Coefficient of Determination

Table 4.1 shows the coefficient of determination values of the filtered IMERG and the machine learning models. The average R^2 for IMERG data from 2015 to 2022 was -2.09, which shows a very high inconsistency between IMERG estimates and ground observations, suggesting that filtered IMERG is underperforming in terms of using the mean of observed precipitation. This also reflects the inability of IMERG to accurately capture the spatial distribution of precipitation. The standard deviation of 8.69 explains that there is a high inconsistency and variability in the predictions given by IMERG. In contrast, the mean R^2 values for Random Forest and XGBoost are 0.47 and 0.33, respectively, indicating that, on average, the inclusion of environmental features and the use of machine learning models significantly improve the accuracy of satellite-based precipitation estimates over the stand-alone IMERG estimate.

4.1.2 KGE and Its Components

Table 4.1 shows the KGE and its components with the mean, median, and standard deviation of each daily model and the corresponding filtered IMERG using test data between 2015 and 2022.

The overall mean (median) KGE of IMERG, Random Forest, and XGBoost are -0.28(0.07), 0.46(0.48), and 0.47(0.51), respectively. This indicates that the machine learning algorithms significantly improve the precipitation estimates over the original IMERG product. The positive KGE scores of the machine learning models indicate that these models are better at capturing precipitation variability while being able to improve forecast accuracy. Similarly, the higher median KGE of XGBoost indicates that it generally outperforms Random Forest, although the mean score is slightly lower due to some outliers. It should be noted that the complex structure of precipitation and the lower amount of precipitation on some days

lowers the score. The models are found to be very efficient on rainy days, achieving KGE values of around 0.7 - 0.9.

Table 4.2 shows R2 and KGE values of some daily models with corresponding precipitation amounts to further investigate the sensitivity of the model results. Day column of the table shows year and day number of that year starting from 0 (0 corresponds to January 1). Train column shows the number of stations with 2 mm/day or more precipitation on that day. The Obs column shows the total amount of precipitation in mm/day for that day. Other columns show scores for IMERG, Random Forest, and XGBoost, respectively. The KGE scores also show that the models perform better for high amounts of stations with 2mm/day or more.

Table 4.2. R2 and KGE scores and corresponding precipitation amounts

	Days	Train (>2mm/day)	Obs (mm/day)	R2			KGE			r			Alpha			Beta		
				IMERG	RF	XGBoost	IMERG	RF	XGBoost	IMERG	RF	XGBoost	IMERG	RF	XGBoost	IMERG	RF	XGBoost
1	2018_87	243	4695.7	-2.079	0.754	0.643	-0.414	0.732	0.817	0.811	0.876	0.818	2.255	0.762	0.979	1.624	0.986	0.994
2	2016_74	242	2379.6	-0.272	0.508	0.403	0.395	0.725	0.664	0.583	0.740	0.684	1.072	0.927	0.915	0.568	1.053	1.077
3	2016_334	241	5798.4	-0.071	0.645	0.789	0.244	0.553	0.698	0.544	0.856	0.914	0.599	0.587	0.720	0.549	0.910	0.927
4	2016_36	241	2533.3	-0.206	0.482	0.406	0.341	0.533	0.579	0.351	0.697	0.649	0.920	0.647	0.770	0.915	0.962	0.968
5	2016_7	240	3656	0.407	0.589	0.579	0.489	0.540	0.638	0.680	0.802	0.772	0.711	0.626	0.763	0.727	0.819	0.851
6	2017_324	239	3687.2	-0.116	0.276	0.315	0.438	0.362	0.615	0.466	0.544	0.629	0.968	0.562	0.905	0.827	1.085	1.037
7	2016_264	238	3239.5	-2.048	0.556	0.442	-0.387	0.581	0.489	0.476	0.750	0.668	1.674	0.664	0.613	2.092	1.004	0.973
8	2016_18	238	1986.8	0.174	0.666	0.690	0.403	0.609	0.799	0.544	0.841	0.840	0.757	0.652	0.928	0.701	1.081	1.098
9	2018_55	237	1901	0.185	0.553	0.525	0.462	0.610	0.641	0.569	0.753	0.736	0.776	0.710	0.774	0.769	0.919	0.907
10	2021_149	103	651.1	0.458	0.606	0.407	0.499	0.553	0.684	0.707	0.798	0.709	0.793	0.666	1.010	0.650	0.782	0.878
11	2021_114	103	306.7	-8.888	0.805	0.890	-2.397	0.682	0.793	0.757	0.924	0.952	3.547	0.715	0.843	3.235	0.883	0.873
12	2021_356	103	370	-0.085	0.690	0.599	-0.383	0.606	0.655	0.161	0.848	0.782	0.199	0.705	0.842	0.247	0.787	0.784
13	2022_82	103	201.8	-1.785	0.315	0.290	-0.129	0.295	0.306	0.053	0.579	0.543	1.372	0.440	0.481	0.511	1.074	1.060
14	2022_134	103	260.4	-3.500	0.217	0.126	-0.548	0.388	0.415	0.365	0.511	0.488	2.232	0.690	0.803	1.689	1.200	1.205
15	2021_106	43	24.6	-100.324	0.320	-0.447	-14.555	0.522	0.331	-0.012	0.590	0.529	7.370	0.756	1.369	15.154	0.993	1.299
16	2021_169	43	133.2	-0.549	0.213	-0.052	0.262	0.302	0.221	0.319	0.473	0.329	1.121	0.565	0.717	1.258	1.145	1.275
17	2021_207	43	214.7	0.079	0.177	-0.521	0.274	0.037	0.030	0.384	0.486	0.058	0.643	0.260	0.782	0.855	0.660	0.913
18	2022_351	43	15	-0.176	0.035	0.141	-0.755	-0.113	0.197	-0.072	0.198	0.388	0.025	0.237	0.484	0.010	0.882	0.939
19	2015_287	42	85.2	-0.044	0.178	0.214	-0.667	0.154	0.048	0.027	0.426	0.493	0.046	0.435	0.362	0.040	0.740	0.508
20	2015_115	42	66.9	-0.159	0.103	0.018	-0.184	0.142	0.273	0.287	0.361	0.364	0.710	0.504	0.699	1.900	0.715	0.815
21	2020_360	13	8.9	-12.730	0.662	0.390	-2.237	0.456	0.313	0.137	0.928	0.679	3.685	0.483	0.416	2.589	0.846	0.832
22	2021_225	13	44.8	-18.327	-162.509	-1018.491	-2.689	-14.063	-32.189	0.026	0.007	-0.047	4.283	12.337	31.587	2.371	10.869	13.837
23	2021_260	13	13.6	-1.839	-0.773	-0.071	-0.275	-1.461	-0.484	-0.040	-0.081	-0.125	1.311	0.711	0.157	1.668	3.192	0.523
24	2022_249	13	10.9	-0.143	0.053	-0.122	-0.407	-0.115	-0.093	-0.009	0.234	0.070	0.339	0.222	0.426	0.275	0.774	1.002
25	2022_3	13	4.3	-228.854	0.135	0.213	-26.534	0.077	0.130	0.166	0.371	0.500	13.077	0.331	0.322	25.730	0.911	0.783
26	2022_144	13	5.2	-42.210	-0.223	-0.161	-5.722	-0.813	-0.557	-0.039	0.112	0.030	6.399	0.462	0.357	4.868	2.486	2.034

The KGE scores of the models were compared on a daily basis using only the test stations in figure 4.1. showing filtered models test KGE values. the x-axis is for Random Forest scores and the y-axis is for XGBoost scores. It has a red line showing 1:1 correspondence. The plot here shows the overall superior performance of XGBoost since most points lie above the red line.

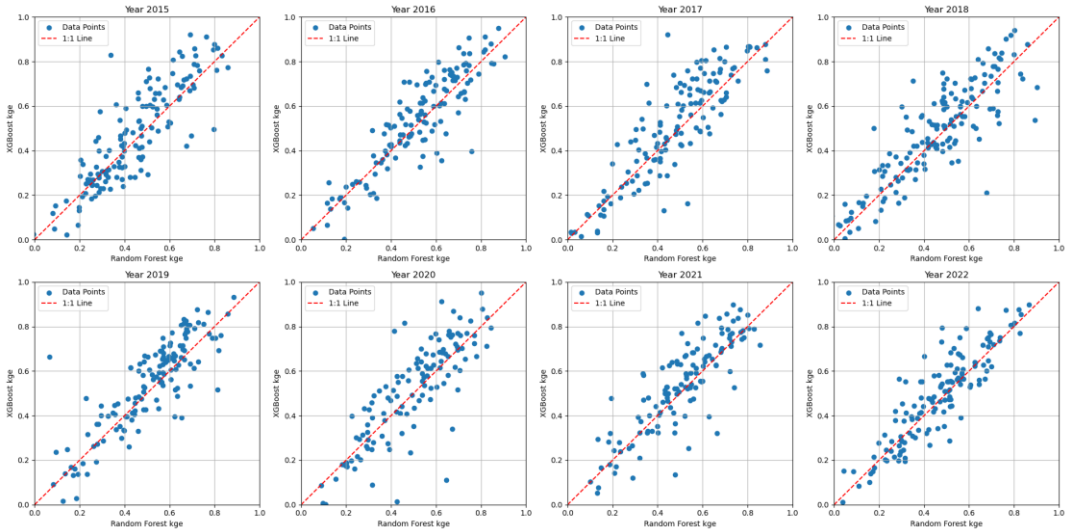


Figure 4.1. KGE score comparison of Random Forest and XGBoost models for each year by selecting days only having more than 25 station with 2mm/day or higher precipitation value

Figure 4.2 shows the violin plots of the filtered KGE and its components over the period 2015-2022.Comparisons of the KGE scores show that, on average, XGBoost is associated with a higher distribution and a higher mean score compared to Random Forest, and with a greater dispersion of anomalies. The higher anomalies observed with XGBoost can be better explained by its increased sensitivity to change and the fact that it is much more responsive to the IMERG data. As a result, XGBoost is able to capture certain complex patterns.

The correlation scores compare the two models, and Random Forest has a better mean and overall correlation, with a smaller distribution. XGBoost understands more about variability (alpha) than Random Forest, while both have almost the

same bias correction (beta). The KGE graph shows that the XGBoost bias correction simulation outperforms Random Forest, mainly due to the improvement in variability. The main reason for XGBoost's exceptional performance is that it captures the variability (alpha) in the data, which improves model performance in fluctuating precipitation patterns and shows that models are better at learning complicated relationships.

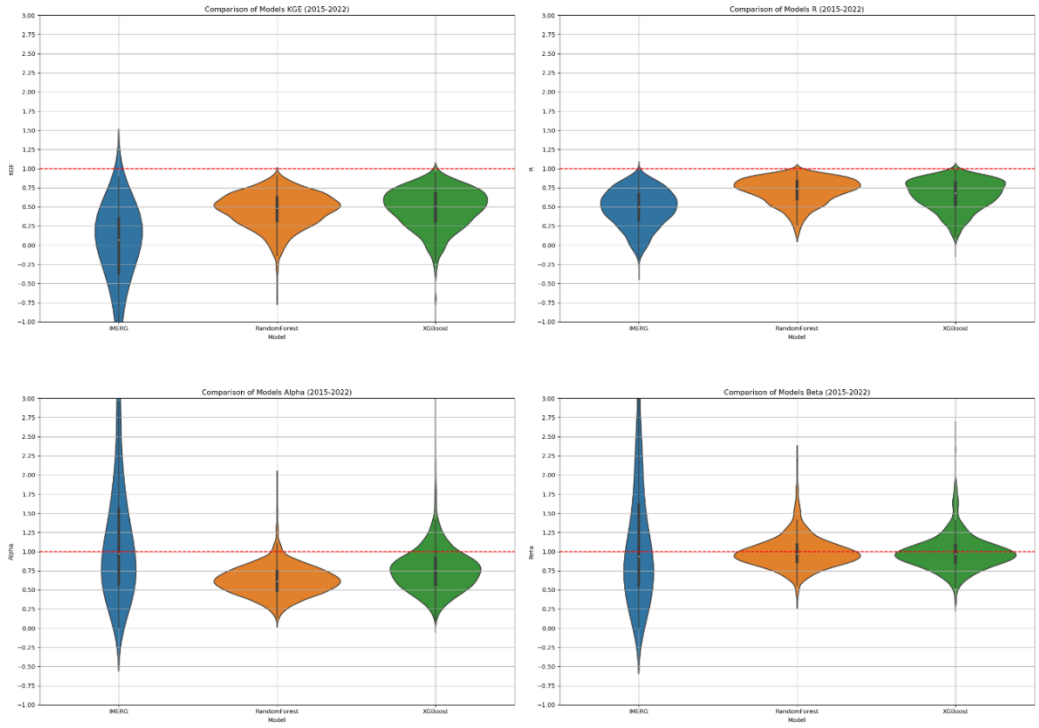


Figure 4.2. KGE and component analyses between 2015-2022

4.1.3 Root Mean Squared Error

Table 4.1 shows the RMSE values, where the mean (median) values for IMERG, RF, and XGBoost models are 7.08 (5.90), 4.00 (3.34), and 4.33 (3.58), respectively, with Random Forest and XGBoost performing almost identically. The performance comparison underscores the superiority of machine learning models in terms of predictive accuracy, reducing the average RMSE by approximately 3 mm/day. The minimal performance difference between Random Forest and

XGBoost further underscores the efficiency of each model, given the significant performance difference for both models with respect to the filtered IMERG.

4.1.4 Evaluation of Models' Performance for Precipitation Events and Extreme Precipitation Events

Tables 4.3 and 4.4 show the POD scores for filtered IMERG, Random Forest and XGBoost. Filtered IMERG scores decrease significantly with increasing precipitation thresholds, indicating that struggles to predict extreme events. Random Forest performs better at lower precipitation thresholds (POD1 and POD2). The POD5 scores of both models are very similar and XGBoost is better at predicting extreme events. All scores decrease as the precipitation threshold increases. Overall, these two machine learning algorithms, Random Forest and XGBoost, provide a significant improvement over filtered IMERG in predicting precipitation events.

Table4.3. Probability of Detection scores of filtered IMERG and machine learning models in general precipitation events (thresholds 1,2, and 5 mm/day)

Year	POD1			POD2			POD5		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.476	0.590	0.550	0.430	0.507	0.484	0.391	0.408	0.389
2016	0.488	0.649	0.601	0.467	0.624	0.587	0.430	0.492	0.483
2017	0.480	0.654	0.614	0.446	0.591	0.579	0.384	0.450	0.457
2018	0.604	0.713	0.665	0.548	0.601	0.571	0.482	0.431	0.449
2019	0.576	0.692	0.645	0.563	0.652	0.615	0.446	0.457	0.480
2020	0.526	0.668	0.628	0.490	0.627	0.613	0.397	0.476	0.500
2021	0.502	0.711	0.680	0.449	0.638	0.600	0.377	0.462	0.470
2022	0.459	0.685	0.628	0.407	0.593	0.566	0.343	0.427	0.418
Average	**0.514**	**0.670**	**0.626**	**0.475**	**0.604**	**0.577**	**0.406**	**0.450**	**0.456**

Table 4.4. Probability of Detection scores of filtered IMERG and machine learning models in extreme precipitation events (thresholds 10,20, and 50 mm/day)

Year	POD10			POD20			POD50		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.337	0.270	0.310	0.286	0.214	0.305	0.282	0.204	0.296
2016	0.327	0.353	0.381	0.287	0.260	0.313	0.156	0.252	0.235
2017	0.304	0.306	0.344	0.297	0.204	0.234	0.288	0.136	0.136
2018	0.440	0.324	0.374	0.400	0.217	0.253	0.269	0.051	0.237
2019	0.368	0.338	0.402	0.275	0.289	0.311	0.294	0.184	0.476
2020	0.331	0.307	0.365	0.255	0.179	0.312	0.370	0.282	0.442
2021	0.313	0.362	0.395	0.282	0.247	0.279	0.194	0.194	0.222
2022	0.221	0.302	0.327	0.162	0.246	0.273	0.078	0.170	0.170
Average	**0.330**	**0.320**	**0.362**	**0.281**	**0.232**	**0.285**	**0.241**	**0.184**	**0.277**

Tables 4.5 and 4.6 show the FAR scores of the filtered IMERG and machine learning models. At all thresholds, filtered IMERG FAR scores are high, ranging from a low of 0.522 at the lowest threshold (FAR1) to a high of 0.867 at the highest threshold (FAR50). Filtered IMERG over predicts precipitation especially at high thresholds indicating errors in precipitation reports. For both models, the correction for false reports is very significant. Random Forest FAR scores are lower than XGBoost for general precipitation events (FAR1, FAR2, and FAR5) and almost equal for extreme precipitation events (FAR10, FAR20, and FAR50). Random forest reduces false alarms better than XGBoost.

Table 4.5 False Alarm Ratio scores of filtered IMERG and machine learning models in general precipitation events (thresholds 1, 2, and 5 mm/day)

Year	FAR1			FAR2			FAR5		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.537	0.422	0.439	0.562	0.409	0.410	0.580	0.328	0.382
2016	0.510	0.373	0.389	0.518	0.326	0.370	0.536	0.277	0.357
2017	0.530	0.390	0.391	0.523	0.335	0.400	0.566	0.288	0.440
2018	0.510	0.393	0.418	0.534	0.377	0.410	0.600	0.344	0.397
2019	0.543	0.395	0.392	0.542	0.354	0.383	0.592	0.318	0.361
2020	0.541	0.409	0.447	0.539	0.359	0.368	0.590	0.307	0.402
2021	0.507	0.372	0.398	0.506	0.353	0.372	0.533	0.284	0.362
2022	0.493	0.434	0.484	0.517	0.373	0.425	0.579	0.355	0.460
Average	**0.522**	**0.399**	**0.420**	**0.530**	**0.360**	**0.392**	**0.572**	**0.313**	**0.395**

Table 4.6 False Alarm Ratio scores of filtered IMERG and machine learning models in extreme precipitation events (thresholds 10,20, and 50 mm/day)

Year	FAR10			FAR20			FAR50		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.653	0.013	0.014	0.775	0.002	0.003	0.896	0.000	0.000
2016	0.670	0.015	0.017	0.762	0.003	0.004	0.928	0.000	0.000
2017	0.646	0.012	0.013	0.739	0.002	0.003	0.855	0.000	0.000
2018	0.680	0.015	0.019	0.766	0.003	0.004	0.889	0.000	0.001
2019	0.677	0.011	0.014	0.740	0.003	0.005	0.782	0.000	0.001
2020	0.671	0.009	0.013	0.794	0.002	0.003	0.834	0.000	0.000
2021	0.600	0.012	0.014	0.706	0.002	0.005	0.870	0.000	0.000
2022	0.684	0.014	0.016	0.795	0.003	0.004	0.880	0.000	0.001
Average	**0.660**	**0.013**	**0.015**	**0.760**	**0.002**	**0.004**	**0.867**	**0.000**	**0.000**

The CSI scores of filtered IMERG with machine learning models are shown in Tables 4.7 and 4.8. Filtered IMERG has a very poor ability to predict precipitation events and its accuracy decreases as the precipitation threshold increases. In comparison, Random Forest has a better ability to predict precipitation events, especially for the lower precipitation thresholds (CSI1 and CSI2). The CSI5 scores between Random Forest and XGBoost are almost identical. XGBoost actually performs slightly better for extreme precipitation events. Both models significantly improve the CSI scores of filtered IMERG and precipitation event prediction.

Table 4.7 Critical Success Index of filtered IMERG and machine learning models in general precipitation events (thresholds 1,2, and 5 mm/day)

Year	CSI1			CSI2			CSI5		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.297	0.383	0.374	0.262	0.349	0.350	0.232	0.323	0.304
2016	0.299	0.450	0.429	0.274	0.463	0.429	0.248	0.398	0.378
2017	0.285	0.446	0.440	0.273	0.448	0.433	0.223	0.353	0.335
2018	0.355	0.470	0.449	0.309	0.426	0.405	0.255	0.332	0.334
2019	0.327	0.461	0.456	0.316	0.462	0.448	0.238	0.356	0.362
2020	0.300	0.435	0.412	0.276	0.448	0.439	0.220	0.375	0.375
2021	0.303	0.491	0.467	0.273	0.469	0.442	0.231	0.372	0.363
2022	0.283	0.440	0.398	0.253	0.422	0.392	0.193	0.322	0.303
Average	**0.306**	**0.447**	**0.428**	**0.280**	**0.436**	**0.417**	**0.230**	**0.354**	**0.344**

Table 4.8 Critical Success Index of filtered IMERG and machine learning models in extreme precipitation events (thresholds 10,20, and 50 mm/day)

Year	CSI10			CSI20			CSI50		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	0.180	0.225	0.243	0.122	0.185	0.251	0.073	0.184	0.255
2016	0.176	0.292	0.290	0.118	0.232	0.251	0.044	0.237	0.193
2017	0.163	0.250	0.277	0.124	0.179	0.198	0.088	0.091	0.091
2018	0.197	0.270	0.284	0.148	0.188	0.186	0.078	0.038	0.145
2019	0.179	0.276	0.308	0.129	0.236	0.228	0.137	0.181	0.388
2020	0.168	0.260	0.274	0.097	0.149	0.214	0.114	0.268	0.374
2021	0.167	0.310	0.322	0.147	0.219	0.210	0.089	0.194	0.179
2022	0.130	0.242	0.247	0.079	0.211	0.205	0.030	0.154	0.127
Average	**0.170**	**0.266**	**0.281**	**0.120**	**0.200**	**0.218**	**0.082**	**0.169**	**0.219**

The bias scores of the filtered IMERG and the machine learning models are shown in Tables 4.9 and 4.10. Filtered IMERG scores greater than 1 represent very high event overestimation up to the highest precipitation threshold. Overestimation decreases with increasing precipitation thresholds, indicating less accuracy at lower precipitation levels. For general precipitation events (Bias1, Bias2, and Bias5), the bias scores for XGBoost and Random Forest generally have similar values, indicating that their effects are quite close within these ranges. On the contrary, Random Forest shows a strong underestimation for extreme precipitation events (Bias10, Bias20 and Bias50) and the bias score decreases as the thresholds increase to the lowest of 0.238 for Bias50, meaning that it strongly underestimates extreme event occurrence compared to XGBoost.

Table 4.9 Bias Score of filtered IMERG and machine learning models in general precipitation events (thresholds 1,2, and 5 mm/day)

Year	BIAS1			BIAS2			BIAS5		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	1.479	1.138	1.043	1.302	0.916	0.855	1.149	0.625	0.637
2016	1.376	1.109	1.036	1.177	0.987	0.980	1.252	0.715	0.801
2017	1.586	1.142	1.043	1.304	0.910	0.964	1.063	0.668	0.797
2018	1.531	1.264	1.204	1.461	1.007	1.005	1.347	0.697	0.774
2019	1.657	1.268	1.097	1.664	1.048	1.031	1.400	0.684	0.810
2020	1.314	1.183	1.059	1.397	1.006	0.973	1.184	0.725	0.814
2021	1.324	1.234	1.138	1.080	1.014	0.976	1.060	0.683	0.785
2022	1.432	1.321	1.174	1.119	0.999	0.956	1.144	0.694	0.778
Average	**1.462**	**1.207**	**1.099**	**1.313**	**0.986**	**0.968**	**1.200**	**0.686**	**0.774**

Table 4.10 Bias Score of models of filtered IMERG and machine learning models in extreme precipitation events (thresholds 10,20, and 50 mm/day)

Year	BIAS10			BIAS20			BIAS50		
	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost	Filtered IMERG	RF	XGBoost
2015	1.318	0.399	0.568	1.319	0.287	0.442	0.889	0.249	0.396
2016	1.207	0.501	0.663	1.230	0.359	0.476	0.700	0.252	0.310
2017	1.010	0.451	0.601	1.034	0.273	0.381	0.848	0.273	0.386
2018	1.785	0.468	0.642	1.460	0.301	0.455	0.788	0.090	0.449
2019	1.407	0.480	0.659	1.133	0.423	0.548	0.757	0.188	0.647
2020	1.132	0.456	0.682	1.127	0.271	0.607	0.724	0.318	0.523
2021	1.001	0.497	0.589	0.841	0.333	0.500	0.630	0.306	0.333
2022	0.933	0.502	0.655	0.788	0.332	0.557	0.195	0.225	0.291
Average	**1.224**	**0.469**	**0.632**	**1.116**	**0.322**	**0.496**	**0.691**	**0.238**	**0.417**

Table 4.11 shows the MBE scores of the filtered IMERG and machine learning models. Filtered IMERG overestimates its predictions, while Random Forest and XGBoost have similar performance with very low underestimation, indicating overall model success.

Table 4.11 Mean Bias Error scores of filtered IMERG and machine learning models (mm/day)

Year	MBE		
	Filtered IMERG	RF	XGBoost
2015	0.298	-0.129	-0.129
2016	0.169	-0.045	-0.046
2017	0.103	-0.070	-0.079
2018	0.665	-0.063	-0.054
2019	0.278	-0.061	-0.066
2020	0.201	-0.053	-0.051
2021	-0.019	-0.079	-0.084
2022	-0.117	-0.043	-0.058
Average	**0.197**	**-0.068**	**-0.071**

In summary, the Random Forest and XGBoost models significantly improved the detection performance of IMERG. Although the performance of Random Forest and XGBoost decreases for extreme events (Senocak et al., 2023), XGBoost is better at predicting extreme events. Random forest significantly underestimates extreme events.

4.2 Evaluation of Machine Learning Algorithms, Environmental Features effects on Complex Topography and Shapley Related Graphs

The SHAP library provides an invaluable feature of plotting different graphs, making models user-friendly, easy to understand, and accountable. After calculating Shapley scores by Tree SHAP (Lundberg et al., 2020), plotting these scores in different ways creates new perspectives for users and is more reliable.

4.2.1 Feature Importance Heatmap

Figure 4.3 shows an example heatmap of Random Forest for the year 2018. (Appendix A: SHAP features importance scores as heatmap for Random Forest and

XGBoost between 2015-2022). The top blue and white heatmap represents rainy days. Observation data are checked for precipitation values greater than 2mm/day, and then the number of stations greater than 2mm/day are counted and grouped to be displayed in a heatmap. Dark blue-purple color represents more than 200 stations with more than 2 mm/day, blue color represents nearly 100 stations, and white color represents 0 or nearly 0 stations. The lower heatmap reflects the order of importance of each feature, color coded to clearly show their impact on the daily models, with red representing the most important features, cyan the medium and dark blue the least important features. The lower heatmap is also filtered to 25 or more stations with 2mm/day or more precipitation to represent the temporal distribution of the models used in the metric evaluations and to get more representative results about the importance of the model features.

Topographic features are highly effective on precipitation-related model predictions (Senocak et al., 2023) and IMERG, Latitude, Longitude, Elevation, Distance to Coast, and Effective Terrain Height are the most important features for each day (red) throughout the year, showing the importance of topographic features. Other features such as facet effects are very small in the model output. IMERG is selected as one of the most important features for almost all models. Senocak et al (2023) show in their forecasting study that different machine learning models, including XGBoost and Random Forest, consider numerical weather prediction (NWP) models as the most important feature.

In winter and near-winter periods, the probability of liquid precipitation becomes essential, although precipitation isn't always in liquid form. Conversely, in summer, the probability of liquid precipitation does not vary (always 1), making it the least important variable. As can be seen in Figure 4.3, the climate classes and facet values do not significantly change the prediction results. However, the Random Forest and XGBoost models reduce the dimensions for unrelated features by assigning them an importance of 0, thus mitigating any potential problems these features could pose to the model results. Daily changes in feature importance for different models are evaluated by this graph.

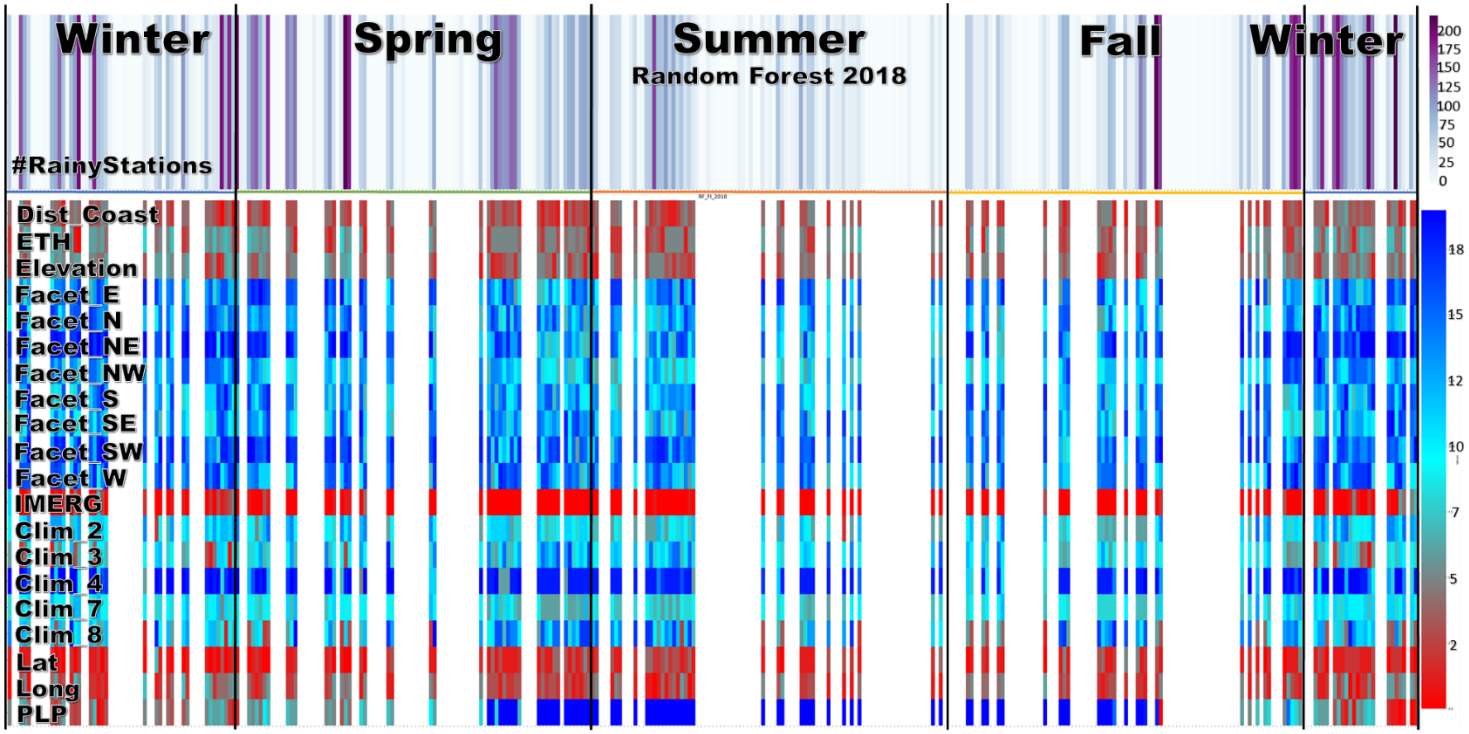


Figure 4.3. SHAP RF Feature Importance scores as heatmap 2018

4.2.2 SHAP Absolute Mean Graphs

SHAP values quantify the impact of each feature on the model's prediction, including both positive and negative contributions. Figure 4.4 shows the average feature values of all samples generated for the day 2018-11-16. This indicates how significant they are in the model. Consistent results between training and testing confirm the reliability of the models. Most notably, Random Forest and XGBoost give the highest priority to IMERG precipitation data, which is likely to signal logical predictions. Topographic features such as elevation, ETH, latitude, and longitude have significant effects that correlate with the Senocak et al. (2023) study, suggesting that model bias are corrected according to environmental relationships.

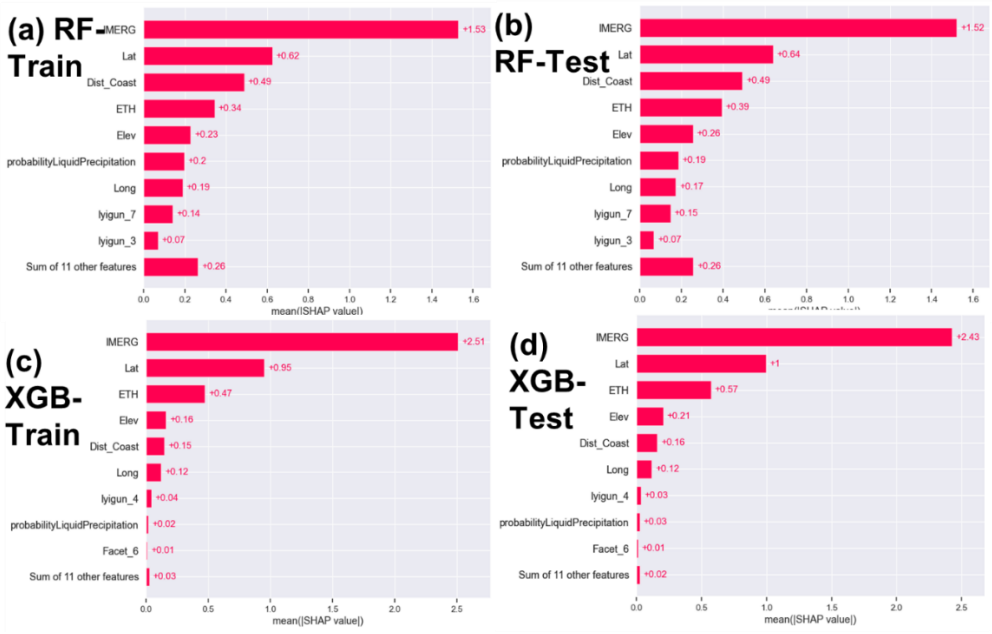


Figure 4.4. SHAP absolute mean plots (Feature Importance) for Random Forest (a) Train dataset, (b) Test dataset and XGBoost (c) Train Dataset (d) Test Dataset (selected day: 16/11/2018)

4.2.3 Merging SHAP Absolute Mean Plots

The average of the daily absolute mean SHAP scores for filtered days is calculated from the test scores for each feature to represent the overall effect of the features. This usage also reduces the daily noise that may affect an individual model. Studies show that provided features may be used differently by different models. (e.g., Senocak et al., 2023). Figure 4.5 shows the absolute mean SHAP plot from 2015 to 2022, plotted as a bar graph to provide a representation of the overall effect of features on model outputs and to show how models use these features together. Features are generally static and their variability does not change from day to day.

The most important features are IMERG, Distance to Coast (DC), Latitude (Lat), Longitude (Long), Elevation (Elv), and Effective Terrain Height (ETH). In general, the contribution of the features is slightly different. XGBoost gives more weight to IMERG SHAP when calculating output values. The overall high importance of XGBoost SHAP is due to the strong influence of IMERG on XGBoost predictions.

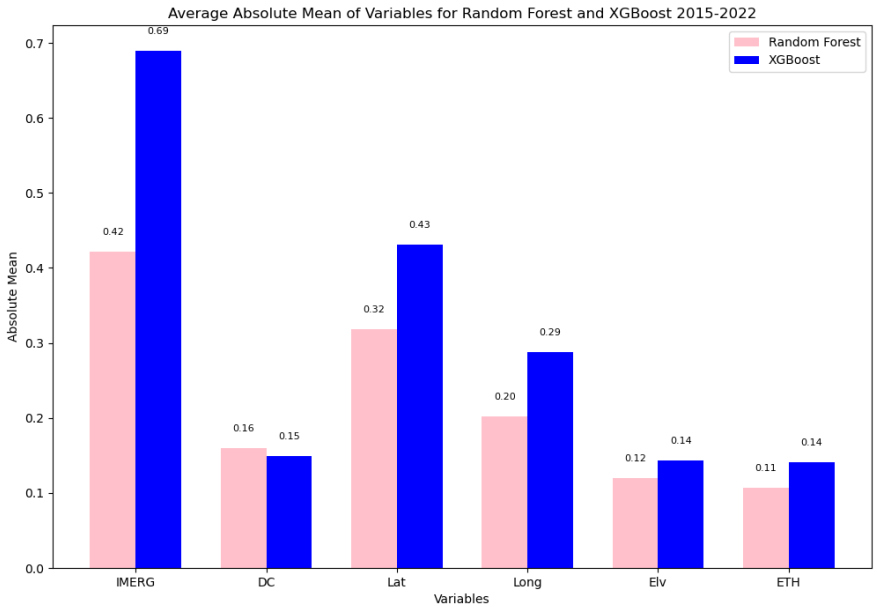


Figure 4.5: Absolute Mean SHAP values for most important features and model comparison from 2015 – 2022

4.2.4 Variation of Descriptor Importance in Space

Figure 4.6 and Figure 4.7 show the variation of descriptor importance in space for Random Forest and XGBoost models separately for filtered days corresponding summer season.

Summers in the Mediterranean and Central Anatolian regions are hot and dry with minimal precipitation (Sensoy et al., 2008). Latitude SHAP plot analysis shows that the models correctly capture this aridity in these regions, as indicated by their low precipitation predictions in the summer season. The SHAP values for these latitudes are slightly negative, indicating a decrease in predicted precipitation. On the other hand, the positive SHAP values at higher latitudes indicate that the models have captured the generally high amount of precipitation in the Black Sea region.

The SHAP plots for longitude show that the models tend to increase their precipitation over the western area during the summer months. More specifically, the models give positive SHAP values for the western parts of the study area, increasing model predictions of precipitation, and negative SHAP values for the eastern regions, corresponding to a reduced effect in that area.

The SHAP plots for elevation show that the predictions for both models increase with increasing elevation. Overall, this suggests that the models are capturing a lot of variation that is positively correlated between precipitation and elevation, indicating that the models are successfully correlating precipitation and elevation values.

Distance to coast values have mostly similar contribution in predicting precipitation during summer months as shown in the SHAP plot, this could perhaps be explained by the dry climatic conditions that normally characterize the Mediterranean region in summer, thus reducing the overall impact of distance to coast on summer precipitation.

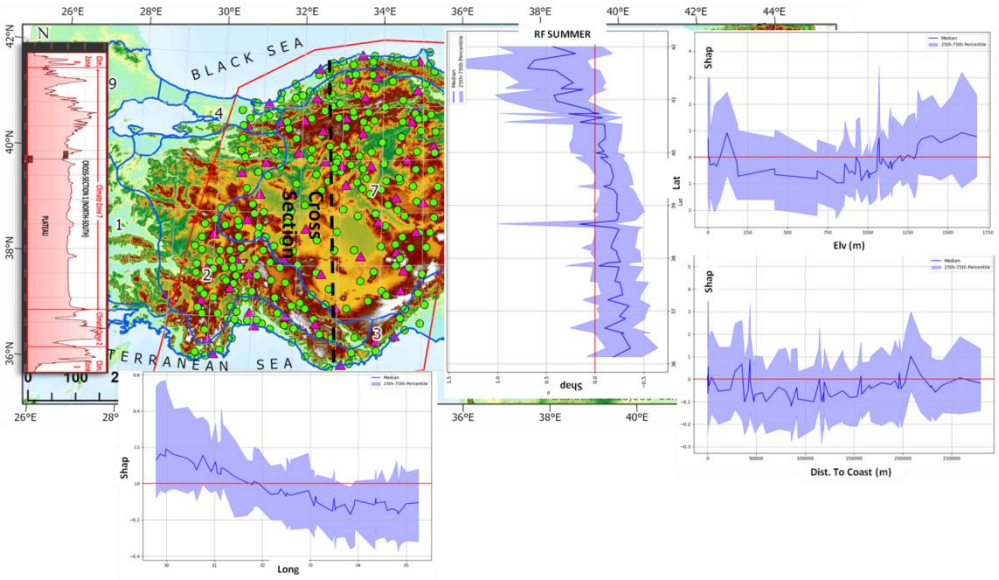


Figure4.6. Variation of Descriptor Importance in Space (Summer Season) for Random Forest Model

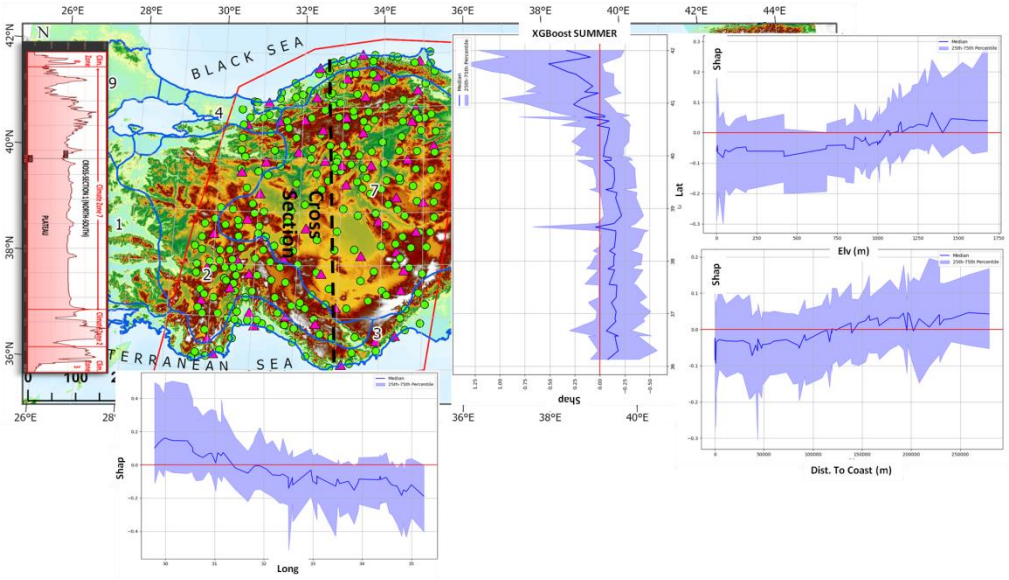


Figure 4.7. Variation of Descriptor Importance in Space (Summer Season) for XGBoost Model

Figure 4.8 and figure 4.9 show variation of descriptor importance in space for Random Forest and XGBoost models separately for filtered days corresponding winter season.

The Mediterranean and Black Sea regions receive high amounts of precipitation during the winter season (Sensoy et al., 2008). This can be easily seen from the latitude SHAP plots of the models. The models also detect the dry plateau in the Central Anatolian region by giving a negative effect to values corresponding to this plateau, indicating the ability of the models to learn complex relationships in the data.

Sensoy et al. (2008) showed with the seasonal precipitation distribution map of Turkiye that the study area does not have a significant change in the east-west directions of precipitation in winter. Considering the longitude SHAP values in winter indicates that the precipitation falling in the study area does affect the longitude significantly.

Analysis of the Elevation SHAP plots in different seasons shows that the models can capture the correlation between precipitation and elevation without distinguishing between summer and winter seasons.

Winter is a season of precipitation for the Black Sea and Mediterranean regions (Sensoy et al., 2008). Therefore, the distance to coast feature ended up being more informative for the models in winter seasons. The machine learning models are able to capture more precipitation along the coasts and assign them positive SHAP values. The fact that the models can play with the influence of the Distance to Coast feature according to seasonal variations further proves their ability to efficiently handle the complex relationships within the data.

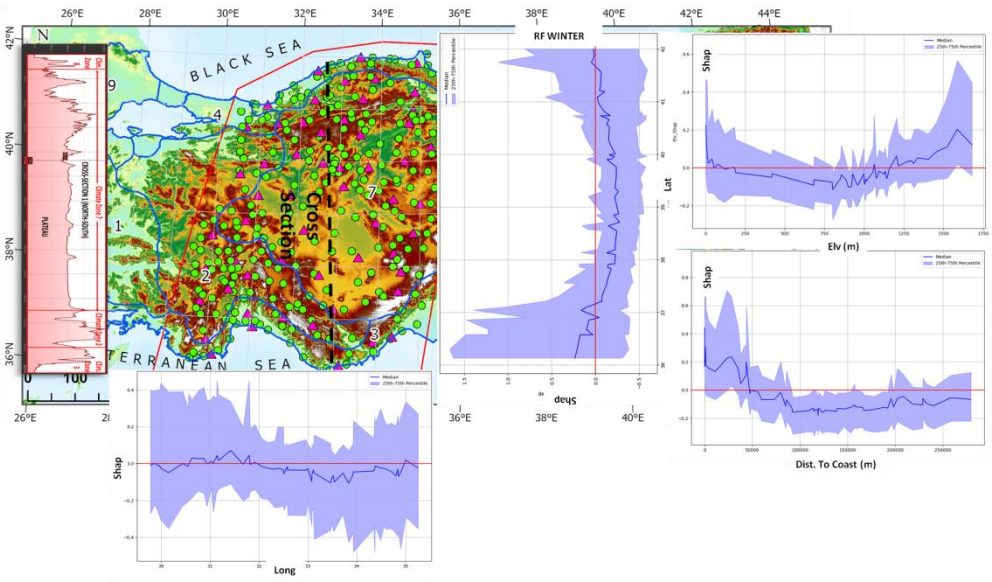


Figure 4.8. Variation of Descriptor Importance in Space (Winter Season) for Random Forest Model

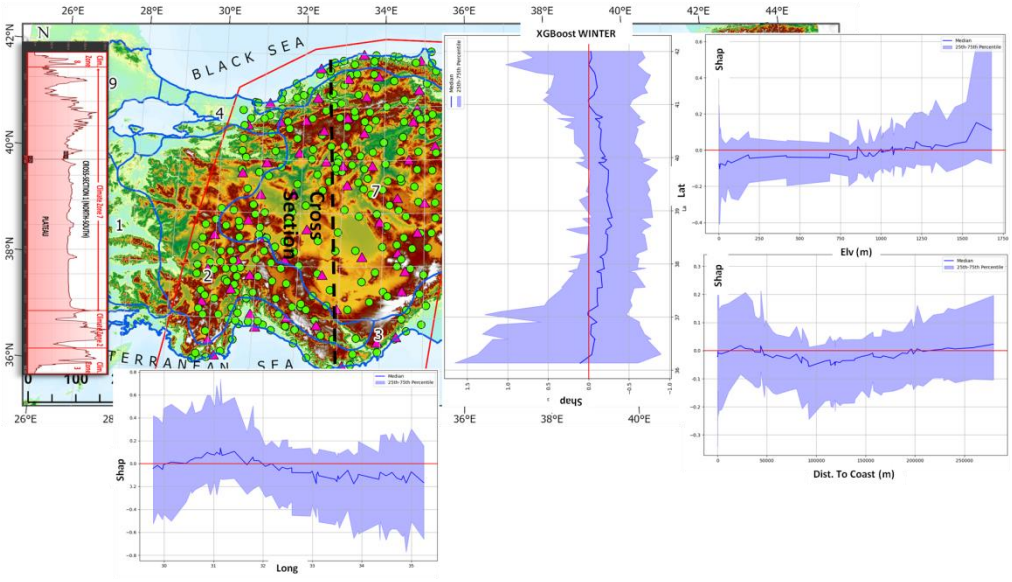


Figure 4.9. Variation of Descriptor Importance in Space (Winter Season) for XGBoost Model

4.2.5 Potential SHAP Limitations and Their effects in this study

Representing feature importance with SHAP scores has some inherent limitations that can lead to misleading results if not handled carefully. High correlations of complex features and training set bias distort the interpretation of SHAP. Looking at feature dependencies helps, but does not solve the problem that SHAP scores are affected by complex feature combinations. These challenges are partly overcome by the intentional selection of the model in this study to be able to handle correlated data efficiently; thus, this study narrows its focus to tree-based models such as regression trees, random forests, and XGBoost, which are better able to handle such complexity (Rabinowicz & Rosset, 2022). Furthermore, high correlations between features can mislead SHAP scores and lead to misinterpretation of feature importance. The first method used in this area is to fool SHAP algorithms by creating perturbations in the background distribution (Baniecki & Biecek, 2022). They used genetic algorithms to manipulate the background data and influence the calculation and interpretation of SHAP scores to illustrate how correlations can affect SHAP scores and their reliability. This study addresses these issues through data preparation, noise reduction, correlation testing, and strategic model selection of tree-based models to ensure the representational and reliable nature of SHAP scores.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Precipitation is a tremendous gear in the climate and water cycle system (Kidd and Huffman, 2011). The sustainability of water resource management and agricultural studies depends on precipitation measurements (Hui-Mean et al., 2018; Kidd et al., 2009; Thornes et al., 2010; Kidd and Huffman, 2011; Zhou et al., 2022). Accurate precipitation data contribute to a better representation of precipitation patterns and water availability.

This study suggests that bias correction IMERGLate satellite product with ground-based observations and environmental parameters such as elevation, effective terrain height, latitude, longitude, distance to coast to improve its spatial representation. Bias correction is an important issue, and there is increasing interest in using machine learning algorithms for bias correction due to their efficiency (Zeng et al., 2021; Wang et al., 2020; et al., 2021b; Lao et al., 2021). Tree-based machine learning models propose more interpretable and predictive accuracy compared to traditional models (Başağoğlu et al., 2022; Chakraborty et al., 2021a; Chang et al., 2016; Dumitrescu et al., 2021). In this study, random forest (RF) (Breiman, 2001) and extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016) tree-based machine learning algorithms are used to train the daily model. Machine learning models have been proven to recognize complex structures and patterns from large amounts of structured and unstructured data (Li, 2022). The eXtreme Gradient Boosting (XGBoost) method is characterized by its remarkable ability to capture complex and nonlinear relationships between precipitation and environmental features (Ali et al., 2023; Zhu et al., 2023).

The performance of the machine learning models is evaluated in chapter 4. The models were filtered to show their performance meaningfully. The models

outperformed on negative R2 values of filtered IMERG. XGBoost performed slightly better on KGE due to the alpha value. Mean Bias Error of Random Forest (-0.068 mm/day) and XGBoost (-0.071 mm/day) shows both models slightly underestimate precipitation events. While filtered IMERG (0.197 mm/day) overestimates. The overall performance of the models is ensured with these metrics, and then the models are evaluated with thresholds to analyze their specific event performance. The Probability of Detection scores show that Random Forest is better at predicting small events (1 mm/day, 2 mm/day). The XGBoost models are more effective at higher thresholds (10 mm/day and above), demonstrating their reliability in detecting larger peak events. IMERG achieves high FAR scores, which increase at higher thresholds. The machine learning models reduce the scores. The bias score of extreme thresholds (Bias10, Bias20 and Bias50) shows that Random Forest (0.469, 0.322, and 0.238) underestimates extreme event occurrence compared to XGBoost (0.632, 0.496, and 0.417). IMERG overestimates precipitation events. Both models improve the CSI. Random Forest performs better for the lower precipitation thresholds (CSI1 and CSI2). XGBoost performs better for extreme precipitation events (CSI10 and above).

Machine learning models are referred to as black boxes due to their complex decision-making processes (Li, 2022). The concept of Explainable Artificial Intelligence makes model decisions transparent and explains how input data is transformed into output (Li, 2022). Models with SHAP (Shapley Additive Explanation) (Lundberg and Lee, 2017; Lundberg et al., 2020) are among the best approaches to explain feature contributions in models (Li, 2022). SHAP draws a relationship between input and output values, showing how features are mostly used by the model during the prediction step and influence the outcome (Li, 2022).

The SHAP feature importance scores of the daily models are plotted annually in the form of a heat map to show their impact. From the heatmaps, the most important features were IMERG, latitude, longitude, elevation, distance to coast, and effective terrain height. The least important features are facets and climate zones.

The probability of liquid precipitation affected the models only in winter seasons, because in summer its value is always 1.

The interpretation of important features by the models was analyzed using the combination of absolute mean SHAP plots. The mean SHAP plots of the XGBoost models had higher importance scores than the IMERG. To compensate for this, the mean SHAP values of XGBoost are assumed to be slightly higher than those of Random Forest.

Series of combinations of 2015-2022 SHAP values with variation of descriptor plots are created separately for the summer and winter seasons of the Random Forest and XGBoost models to show how the models evaluate features in different seasons. Note that these procedures are performed first for the models with 25 or more stations with 2 mm/day or more precipitation. The results are corrected to remove the dry day bias. These features were chosen to be static during the combination of SHAP values to avoid the addition of bias and noise. This again shows that the complex relationships between features and precipitation data are understood by the models.

Evaluation of variation in descriptor plots. The models can correlate precipitation with elevation data regardless of the season. The models understand that the summer seasons are dry in the Mediterranean and Central Anatolian regions (Sensoy et al., 2008), and this is clearly seen in the latitude and distance to coast SHAP plots. In summer, only at high latitude values, which corresponds to the Black Sea region, the models detect the amount of precipitation and give positive SHAP values, but the models cannot make a high interpretation for the distance to coast data in summer. In the winter seasons, we observe that the models increase the precipitation in these regions in the latitude SHAP plot, which is understood by the fact that precipitation is effective in the Black Sea and the Mediterranean. In addition, in the Latitude SHAP plot, the models provide more meaningful predictions by decreasing precipitation in locations corresponding to the Central

Anatolian region. The Distance to Coast graph has become more meaningful in the winter months.

The results confirm that model training effectively improves the performance of IMERG satellite precipitation products with satisfactory statistical results. New methods to reduce bias in satellite precipitation products can be further explored. More features can be added, or feature engineering will even tune the input variables. To illustrate, for further research, feature engineering techniques can be used to improve the contribution of these features. Creating synthetic parameters by combining these features with elevation, like other contributing numerical features, can increase the contribution of these features. Alternatively, another numerical feature such as elevation can be considered categorical, including groups as low, medium, and high, and then converted to a one-hot coding to use facets and climate zones with them.

Random forests work well for low precipitation and XGBoost works better for high precipitation. The combination of both model results may provide a very promising approach toward increasing the accuracy of satellite-based precipitation estimates.

REFERENCES

Acosta, M.R.C., Ahmed, S., Garcia, C.E., Koo, I., 2020. Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks. IEEE Access 8, 19921–19933. https://doi.org/10.1109/ACCESS.2020.2968934.

Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., & Gandomi, A. H. (2021). The arithmetic optimization algorithm. Computer methods in applied mechanics and engineering, 376, 113609.

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160. https://doi.org/10.1109/ access.2018.2870052.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., ... & Nelkin, E. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). Journal of hydrometeorology, 4(6), 1147-1167.

Aghelpour, P., Guan, Y., Bahrami-Pichaghchi, H., Mohammadi, B., Kisi, O., & Zhang, D. (2020). Using the MODIS sensor for snow cover modeling and the assessment of drought effects on snow cover in a mountainous area. Remote Sensing, 12(20), 3437.

Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2022). Dwarf mongoose optimization algorithm. Computer methods in applied mechanics and engineering, 391, 114570.

Ali, S., Khorrami, B., Jehanzaib, M., Tariq, A., Ajmal, M., Arshad, A., ... & Khan, S. N. (2023). Spatial downscaling of GRACE data based on XGBoost model for improved understanding of hydrological droughts in the Indus Basin Irrigation System (IBIS). Remote Sensing, 15(4), 873.

Akinyemi, D.F., Ayanlade, O.S., Nwaezeigwe, J.O., Ayanlade, A., 2020. A comparison of the accuracy of multi-satellite precipitation estimation and ground meteorological records over southwestern Nigeria. Remote Sens. Earth Syst. Sci. 3, 1–12.

Amjad, M., Yilmaz, M. T., Yucel, I., & Yilmaz, K. K. (2020). Performance evaluation of satellite-and model-based precipitation products over varying climate and complex topography. Journal of Hydrology, 584, 124707.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 82(4), 1059-1086.

Ashouri, H., Hsu, K. L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., ... & Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bulletin of the American Meteorological Society, 96(1), 69-83.

Baniecki, H., Kretowicz, W., & Biecek, P. (2022, September). Fooling partial dependence via data poisoning. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 121-136). Cham: Springer Nature Switzerland.

Başağoğlu, H., Chakraborty, D., Do Lago, C., Gutierrez, L., S ¸ahinli, M.A., Giacomoni, M., Furl, C., Mirchi, A., Moriasi, D., S ¸eng¨ or, S.S., 2022. A review on interpretable and explainable artificial intelligence in hydroclimatic applications. Water 14, 1230. https://doi.org/10.3390/w14081230.

Beck, H.E., et al., 2020. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. In: Satellite Precipitation Measurement. Springer, pp. 625–653.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23, 207–224. https://doi.org/10.5194/hess-23-207-2019.

Ben Jabeur, S., Khalfaoui, R., Ben Arfi, W., 2021. The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: evidence from explainable machine learning. J. Environ. Manag. 298, 113511 https://doi. org/10.1016/j.jenvman.2021.113511.

Bıyık, G., Unal, Y., & Onol, B. (2009, September). Assessment of Precipitation Forecast Accuracy over Eastern Black Sea Region using WRF-ARW. In 11th Plinius Conference on Mediterranean Storms (p. Plinius11).

Boushaki, F. I., Hsu, K. L., Sorooshian, S., Park, G. H., Mahani, S., & Shi, W. (2009). Bias adjustment of satellite precipitation estimation using ground-based measurement: A case study evaluation over the southwestern United States. Journal of Hydrometeorology, 10(5), 1231-1242.

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. doi:10.1023/A:1010933404324.

Castelvecchi, D. (2016). Can we open the black box of AI?. Nature News, 538(7623), 20.

Chakraborty, D., Başağoğlu, H., Gutierrez, L., Mirchi, A., 2021a. Explainable AI reveals new hydroclimatic insights for ecosystem-centric groundwater management. Environ. Res. Lett. 16, 114024 https://doi.org/10.1088/1748-9326/ac2fde.

Chakraborty, D., Başağoğlu, H, Winterle, J., 2021b. Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. Expert Syst. Appl. 170, 114498 https://doi.org/10.1016/j.eswa.2020.114498.

Chang, M., and L. Harrison (2005), Field assessments on the accuracy of spherical gauges in rainfall measurements, Hydrol. Processes., 19, 403–412.

Chang, Y.-C., Chang, K.-H., Chu, H.-H., Tong, L.-I., 2016. Establishing decision tree-based short-term default credit risk assessment models. Commun. Stat. Theory Methods 45, 6803–6815. https://doi.org/10.1080/03610926.2014.968730.

Chaudhary, S., & Dhanya, C. T. (2019, October). Investigating the performance of bias correction algorithms on satellite-based precipitation estimates. In Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI (Vol. 11149, pp. 279-285). SPIE.

Chen, C., Hu, B., Li, Y., 2021. Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy. Hydrol. Earth Syst. Sci. 25, 5667–5682.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794).

Chiaravalloti, F., Brocca, L., Procopio, A., Massari, C., Gabriele, S., 2018. Assessment of GPM and SM2RAIN-ASCAT rainfall products over complex terrain in southern Italy. Atmos. Res. 206, 64–74. https://doi.org/10.1016/j.atmosres.2018.02.019.

Chivers, B. D., Wallbank, J., Cole, S. J., Sebek, O., Stanley, S., Fry, M., & Leontidis, G. (2020). Imputation of missing sub-hourly precipitation data in a large sensor network: A machine learning approach. Journal of Hydrology, 588, 125126.

Chollet, F. (2021). Deep learning with Python. Simon and Schuster.

Constantinescu, G. S., W. F. Krajewski, C. E. Ozdemir, and T. Tokyay (2007), Simulation of airflow around rain gauges: Comparison of LES with RANS models, Adv. Water Resour., 30, 43–58.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge based approach to the statistical mapping of climate. Climate Res., 22, 99–113.

Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. Int. J. Climatol, 26, 707–721.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. Int. J. Climatol., 28, 2031-2064.

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C. H., Delrieu, G., Hong, Y., Hsu, Y.C., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E.I., Sahlu, D., Salerno, F., Rodríguez-Sánchez, J.-P., Vergara, H.J., Yilmaz, K.K., 2016. Multiregional satellite precipitation products evaluation over complex Terrain. J. Hydrometeorol. 17, 1817–1836. https://doi.org/10.1175/JHM D-15-0197.1.

Derin, Y., Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. J. Hydrometeorol. 15, 1498–1516. https://doi.org/10. 1175/JHM-D-13-0191.1.

Demir, İ., Kılıç, G., & Coşkun, M. (2008). Türkiye ve bölgesi için PRECIS bölgesel iklim modeli çalışmaları. İklim Değişikliği ve Çevre, 1(1), 11-17.

Dinku, T., Anagnostou, E.N., Borga, M., 2002. Improving radar-based estimation of rainfall over complex terrain. J. Appl. Meteorol. 41, 1163–1178. https://doi.org/10. 1175/1520-0450(2002) 0412.0.CO;2.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., & Ropelewski, C. F. (2007). Validation of satellite rainfall products over East

Africa's complex topography. International Journal of Remote Sensing, 28(7), 1503-1526.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. AMS math challenges lecture, 1(2000), 32

Dumitrescu, E., Hu´ e, S., Hurlin, C., Tokpavi, S., 2021. Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2021.06.053.

Efron B (2000) The bootstrap and modern statistics. J Am Stat Assoc 95(452):1293–1296. https://doi.org/10.1080/01621459.2000.10474333.

El-Alfy, E. S., and Mohammed, S. (2020). A review of machine learning for big data analytics: Bibliometric approach. Technol. Analysis Strategic Manag. 32, 984–1005. doi:10. 1080/09537325.2020.1732912.

El Kenawy, A. M., Lopez-Moreno, J. I., McCabe, M. F., & Vicente-Serrano, S. M. (2015). Evaluation of the TMPA-3B42 precipitation product using a high-density rain gauge network over complex terrain in northeastern Iberia. Global and Planetary Change, 133, 188-200.

Ersöz, B., Sağıroğlu, Ş. E. R. E. F., & Bülbül, H. A. L. İ. L. (2022). Methods of Explainable Artificial Intelligence (XAI), Trustworthy Artificial Intelligence (TAI) and Interpretable Machine Learning (IML) in Renewable Energy. *icsmartgrid journal*, *6*(4).

EU-DEM v1.1. Available online: https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata.

Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. arXiv preprint math/0602133. https://doi.org/10.48550/arXiv.math/0602133.

Feng, D. C., Wang, W. J., Mangalathu, S., & Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. Journal of Structural Engineering, 147(11), 04021173.

Folland, C. K. (1988), Numerical models of the raingauge exposure problem, field experiments and an improved collector design, Q. J. R. Meteorol. Soc., 114, 1485–1516.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Ann. Stat., 29, 1189–1232, https://doi.org/10.1214/aos/1013203451, 2001.

Gampe, D., Ludwig, R., 2017. Evaluation of gridded precipitation data products for hy drological applications in complex topography. Hydrology 4, 53. https://doi.org/10. 3390/hydrology4040053.

General Directorate of Meteorology (GDM), 2023. 2023 yağış değerlendirmesi. Received from https://www.mgm.gov.tr/FILES/arastirma/yagisdegerlendirme/2023yagisdegerlendirmesi.pdf

Gebregiorgis, A. S., Kirstetter, P. E., Hong, Y. E., Gourley, J. J., Huffman, G. J., Petersen, W. A., ... & Schwaller, M. R. (2018). To what extent is the day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT?. Journal of Geophysical Research: Atmospheres, 123(3), 1694-1707.

Gehne, M., Hamill, T.M., Kiladis, G.N., Trenberth, K.E., 2016. Comparison of global precipitation estimates across a range of temporal and spatial scales. J. Clim. 29 (21), 7773–7795.

Geurts, P., Louppe, G., 2011. Learning to rank with extremely randomized trees. In: Proceedings of the Learning to Rank Challenge. PMLR, pp. 49–61.

Guan, Y., Mohammadi, B., Pham, Q. B., Adarsh, S., Balkhair, K. S., Rahman, K. U., ... & Tri, D. Q. (2020). A novel approach for predicting daily pan evaporation in the coastal regions of Iran using support vector regression

coupled with krill herd algorithm model. Theoretical and Applied Climatology, 142, 349-367.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resources Research 34 (4), 751–763.

Gong, J., Ou, J., Qiu, X., Jie, Y., Chen, Y., Yuan, L., et al. (2020). A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin. Infect. Dis. 71, 833–840. doi: 10.1093/cid/ciaa443.

Gottardi, F., Obled, C., Gailhard, J., Paquet, E., 2012. Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. J. Hydrol. 432, 154–167.

Han, Y., Wei, Z., & Huang, G. (2024). An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing. Scientific Reports, 14(1), 10151.

Hancock, J. T. & Khoshgoftaar, T. M. Catboost for big data: An interdisciplinary review. J. Big Data 7(1), 94 (2020).

Hanna, E. (1995), How effective are tipping-bucket rain gauges? A review, Weather, 50, 336–342.

He, Z., Yang, Y., Fang, R., Zhou, S., Zhao, W., Bai, Y., ... & Wang, B. (2023). Integration of shapley additive explanations with random forest model for quantitative precipitation estimation of mesoscale convective systems. Frontiers in Environmental Science, 10, 1057081.

He, X., Chaney, N. W., Schleiss, M., & Sheffield, J. (2016). Spatial downscaling of precipitation using adaptable random forests. Water resources research, 52(10), 8217-8237.

Heidinger, H., Yarlequé, C., Posadas, A., Quiroz, R., 2012. TRMM rainfall correction over the Andean Plateau using wavelet multi-resolution analysis. Int. J. Remote Sens. 33, 4583–4602. https://doi.org/10.1080/01431161.2011.652315.

Herold, N., Alexander, L.V., Donat, M.G., Contractor, S., Becker, A., 2016. How much does it rain over land? Geophys. Res. Lett. 43, 341–348. https://doi.org/10.1002/ 2015GL066615.

Hirpa, F. A., Gebremichael, M., & Hopson, T. (2010). Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. Journal of Applied Meteorology and Climatology, 49(5), 1044-1051.

Hobbs, P. V. (1989). Research on clouds and precipitation: Past, present, and future, part I. Bull. Amer. Meteor. 70 (3), 282–285. doi:10.1175/1520-0477-70.3.282.

Hobouchian, M.P., Salio, P., García Skabar, Y., Vila, D., Garreaud, R., 2017. Assessment of satellite precipitation estimates over the slopes of the subtropical Andes. Atmos. Res. 190, 43–54. https://doi.org/10.1016/j.atmosres.2017.02.006.

Hoffmann, M., Schwartengräber, R., Wessolek, G., & Peters, A. (2016). Comparison of simple rain gauge measurements with precision lysimeter data. Atmospheric Research, 174, 120-123.

Hong, Y., Hsu, K. L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. Journal of Applied Meteorology, 43(12), 1834-1853.

Hosking, J. G., C. D. Stow, and S. G. Bradley (1985), Corrections for horizontal winds and wind shear in raindrop size spectrometers, J. Atmos. Oceanic Technol., 2, 181–189.

Hossain, F., Lettenmaier, D.P., 2006. Flood prediction in the future: Recognizing hydrologic issues in anticipation of the Global Precipitation Measurement mission. Water Resour. Res. 42 (11).

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., ... & Iguchi, T. (2014). The global precipitation measurement mission. Bulletin of the American meteorological Society, 95(5), 701-722.

Hsu, K. L., Gao, X., Sorooshian, S., & Gupta, H. V. (1997). Precipitation estimation from remotely sensed information using artificial neural networks. Journal of Applied Meteorology and Climatology, 36(9), 1176-1190.

Hsu, K. L., Gupta, H. V., Gao, X., & Sorooshian, S. (1999). Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. Water Resources Research, 35(5), 1605-1618.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Nelkin, E. J., Hossain, F., & Gebremichael, M. (2010). Satellite rainfall applications for surface hydrology. M. Gebremichael & F. Hossain (Eds.), 3-22.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., ... & Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of hydrometeorology, 8(1), 38-55.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., & Yoo, S. H. (2015). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm theoretical basis document (ATBD) version, 4(26), 2020-05.

Huffman, G. (2019). IMERG V06 quality index. NASA, Available for download from: https://gpm. nasa. gov/sites/default/files/2020-02/IMERGV06_QI_0. pdf.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., Xie, P., 2018. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 5.2.

Hughes, M., Hall, A., Fovell, R.G., 2009. Blocking in areas of complex topography, and its influence on rainfall distribution. J. Atmos. Sci. 66, 508–518

Hui-Mean, F., Yusop, Z., Yusof, F., 2018. Drought analysis and water resource availability using standardised precipitation evapotranspiration index. Atmos. Res. 201, 102–115.

Islam, T., Rico-Ramirez, M.A., Han, D., Srivastava, P.K., Ishak, A.M., 2012. Performance evaluation of the TRMM precipitation estimation using ground-based radars from the GPM validation network. J. Atmos. Solar-Terrestrial Phys. 77, 194–208. https://doi. org/10.1016/j.jastp.2012.01.001.

Ibrahim, B., Wisser, D., Barry, B., Fowe, T., Aduna, A., 2015. Hydrological predictions for small ungauged watersheds in the Sudanian zone of the Volta basin in West Africa. J. Hydrol.: Reg. Stud. 4, 386–397.

Iyigun, C., Türkeş, M., Batmaz, İ., Yozgatligil, C., Purutçuoğlu, V., Koç, E. K., & Öztürk, M. Z. (2013). Clustering current climate regions of Turkey by using a multivariate statistical method. Theoretical and applied climatology, 114, 95-106.

Jabeur, S. Ben, Ballouk, H., Mefteh-Wali, S., Omri, A., 2022. Forecasting the macrolevel determinants of entrepreneurial opportunities using artificial intelligence models. Technol. Forecast. Soc. Chang. 175, 121353 https://doi.org/10.1016/j. techfore.2021.121353.

Jafarpour, M., Adib, A., Lotfirad, M., 2022. Improving the accuracy of satellite and reanalysis precipitation data by their ensemble usage. Appl. Water Sci. 12, 1– 15.

Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., Wang, J., 2021. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. (Vol. 367, pp. 4237–4253). The Royal Society Publishing.

Khan, N. M., Madhav C, N., Negi, A., & Thaseen, I. S. (2020). Analysis on improving the performance of machine learning models using feature selection technique. In Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 69-77). Springer International Publishing.

Klein, A., Falkner, S., Bartels, S., Hennig, P., and Hutter, F. (2016). Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets. Neural Netw. 106, 294–302. doi:10.1093/obo/9780195389661-0226

Krakauer, N. Y., Pradhanang, S. M., Lakhankar, T., & Jha, A. K. (2013). Evaluating satellite products for precipitation estimation in mountain regions: A case study for Nepal. Remote Sensing, 5(8), 4107-4123.

Kidd, C., and Huffman, G. “Global precipitation measurement. Meteorological Applications”, 18(3), 334-353 (2011).

Kidd C, Levizzani V, Turk J, Ferraro R. 2009. Satellite precipitation measurements for water resource monitoring. Journal of the American Water Resources Association 45: 567–579.

Kucera, P.A., Ebert, E.E., Turk, F.J., Levizzani, V., Kirschbaum, D., Tapiador, F.J., Loew, A., Borsche, M., 2013. Precipitation from space: advancing earth

system science. Bull. Am. Meteorol. Soc. 94, 365–375. https://doi.org/10.1175/BAMS-D-11-00171.1.

Lao, P., Liu, Q., Ding, Y., Wang, Y., Li, Y., and Li, M. (2021). Rainrate estimation from FY-4A cloud top temperature for mesoscale convective systems by using machine learning algorithm. Remote. Sens. 13, 3273. doi:10.3390/rs13163273.

Lei, H., Li, H., Zhao, H., Ao, T., Li, X., 2021. Comprehensive evaluation of satellite and reanalysis precipitation products over the eastern Tibetan plateau characterized by a high diversity of topographies. Atmos. Res. 259, 105661.

Levizzani, V., Cattani, E., 2019. Satellite remote sensing of precipitation and the terrestrial water cycle in a changing climate. Remote Sens. 11 (19), 2301.

Li, H., Zhang, Y., Lei, H., & Hao, X. (2023). Machine Learning-Based Bias Correction of Precipitation Measurements at High Altitude. Remote Sensing, 15(8), 2180.

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Computers, Environment and Urban Systems, 96, 101845.

Li, X., Wu, C., Meadows, M. E., Zhang, Z., Lin, X., Zhang, Z., et al. (2021a). Factors underlying spatiotemporal variations in atmospheric PM2.5 concentrations in Zhejiang Province, China. Remote Sens. 13 (15), 3011. doi:10.3390/rs13153011.

Li, X., Yang, Y., Mi, J., Bi, X., Zhao, Y., Huang, Z., et al. (2021b). Leveraging machine learning for quantitative precipitation estimation from Fengyun-4 geostationary observations and ground meteorological measurements. Atmos. Meas. Tech. 14, 7007–7023. doi:10.5194/amt-14-7007-2021.

Lin, X., Fan, J., Hou, Z. J., & Wang, J. (2023). Machine learning of key variables impacting extreme precipitation in various regions of the contiguous United

States. Journal of Advances in Modeling Earth Systems, 15(3), e2022MS003334.

Lin, A., & Wang, X. L. (2011). An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada. Journal of Geophysical Research: Atmospheres, 116(D21).

Linardatos, P., Papastefanopoulos, V., Kotsiantis, S., 2020. Explainable ai: a review of machine learning interpretability methods. Entropy 23, 18. https://doi.org/ 10.3390/e23010018.

Lo Conti, F., Hsu, K., Noto, L.V., Sorooshian, S., 2014. Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. Atmos. Res. 138, 189–204.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.

Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Mantas, V. M., Liu, Z., Caro, C., & Pereira, A. J. S. C. (2015). Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes. Atmospheric Research, 163, 132-145.

Marani, M., 2005. Non-power-law-scale properties of rainfall in space and time. Water Resour. Res. 41 (8).

Mastrantonas, N., Bhattacharya, B., Shibuo, Y., Rasmy, M., Espinoza-D´avalos, G., Solomatine, D., 2019. Evaluating the Benefits of Merging Near-Real-

Time Satellite Precipitation Products: A Case Study in the Kinu Basin Region, Japan. J. Hydrometeorol. 20 (6), 1213–1233.

Mayor, Y.G., Tereshchenko, I., Fonseca-Hernández, M., Pantoja, D.A., Montes, J.M., 2017. Evaluation of error in IMERG precipitation estimates under different topographic conditions and temporal scales over Mexico. Remote Sens. 9, 1–18. https://doi.org/ 10.3390/rs9050503.

Mei, Y., Anagnostou, E.N., Nikolopoulos, E.I., Borga, M., 2014. Error analysis of satellite precipitation products in mountainous basins. J. Hydrometeorol. 15, 1778–1793. https://doi.org/10.1175/JHM-D-13-0194.1.

Mi, J. X., Li, A. D., & Zhou, L. F. (2020). Review study of interpretation methods for future interpretable machine learning. IEEE Access, 8, 191969-191985.

Milewski, A., Elkadiri, R., Durham, M., 2015. Assessment and comparison of TMPA sa tellite precipitation products in varying climatic and topographic regimes in Morocco. Remote Sens. 7, 5697–5717. https://doi.org/10.3390/rs70505697.

Miller, T., 2019. Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Molnar, C. (2020). Interpretable machine learning. Lulu. com.

Mousa, S. R., Bakhit, P. R., & Ishak, S. (2019). An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: a naturalistic driving study. Canadian Journal of Civil Engineering, 46(8), 712-721.

Mohammadi, B., Moazenzadeh, R., Christian, K., & Duan, Z. (2021). Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. Environmental Science and Pollution Research, 28, 65752-65768.

Mueller, C. C., and E. H. Kidder (1972), Rain gage catch variations due to airflow disturbances around a standard rain gage, Water Resour. Res., 8, 1077–1082.

Musolf, A. M., Holzinger, E. R., Malley, J. D., & Bailey-Wilson, J. E. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Human Genetics, 141(9), 1515-1528.

Nabika, H., Itatani, M., & Lagzi, I. (2019). Pattern formation in precipitation reactions: The Liesegang phenomenon. Langmuir, 36(2), 481-497.

Neff, E. L. (1977), How much rain does a rain gage gage? J. Hydrol., 35, 213–220.

Nešpor, V., and B. Sevruk (1999), Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation, J. Atmos. Oceanic Technol., 16, 450–464.

New, M., Todd, M., Hulme, M., & Jones, P. (2001). Precipitation measurements and trends in the twentieth century. International Journal of Climatology: A Journal of the Royal Meteorological Society, 21(15), 1889-1922.

Nguyen, H., Bui, X. N., Bui, H. B., & Cuong, D. T. (2019). Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study. Acta Geophysica, 67(2), 477-490.

Nicodemus, K. K., &Malley, J. D. (2009). Predictor correlation impacts machine learning algorithms: implications for genomic studies. *Bioinformatics*, *25*(15), 1884-1890.

Nijssen, B., & Lettenmaier, D. P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. Journal of geophysical research: atmospheres, 109(D2).

Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Computer Methods and Programs in Biomedicine, 214, 106584.

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2011). Measuring urban rainfall using microwave links from commercial cellular communication networks. Water Resources Research, 47(12).

Oyelade, O. N., Ezugwu, A. E. S., Mohamed, T. I., & Abualigah, L. (2022). Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm. IEEE Access, 10, 16150-16177.

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accident Analysis & Prevention, 136, 105405.

Patakchi Yousefi, K., & Kollet, S. (2023). Deep learning of model-and reanalysis-based precipitation and pressure mismatches over Europe. Frontiers in Water, 5, 1178114.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

Peter I. Frazier. A Tutorial on Bayesian Optimization. 2018 on arXiv.

Prakash, Satya, Ashis K. Mitra, Amir AghaKouchak, and D. S. Pai. "Error characterization of TRMM Multisatellite Precipitation Analysis (TMPA-3B42) products over India for different seasons." Journal of Hydrology 529: 1302-1312 (2015).

Prakash, Satya, Ashis K. Mitra, D. S. Pai, and Amir AghaKouchak. "From TRMM to GPM: How well can heavy rainfall be detected from space?." Advances in Water Resources 88: 1-7 (2016).

Qiu, H., Luo, L., Su, Z., Zhou, L., Wang, L., Chen, Y., 2020. Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. BMC Med. Inform. Decis. Mak. 20, 83. https://doi.org/ 10.1186/s12911-020-1101-8.

Rabinowicz, A., & Rosset, S. (2022). Tree-based models for correlated data. Journal of Machine Learning Research, 23(258), 1-31.

Ramsauer, T., Weiß, T., & Marzahn, P. (2018). Comparison of the GPM IMERG final precipitation product to RADOLAN weather radar data over the topographically and climatically diverse Germany. Remote Sensing, 10(12), 2029.

Rasmy, M., Koike, T., & Li, X. (2014). Applicability of multi-frequency passive microwave observations and data assimilation methods for improving numerical weather forecasting in Niger, Africa. Remote Sensing, 6(6), 5306-5324.

Ren, J., Xu, G., Zhang, W., Leng, L., Xiao, Y., Wan, R., et al. (2021). Evaluation and improvement of FY-4A AGRI quantitative precipitation estimation for summer precipitation over complex topography of western China. Remote Sens. 13 (21), 4366. doi:10.3390/rs13214366.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Ridley, M. (2022). Explainable Artificial Intelligence (XAI): Adoption and Advocacy. *Information technology and libraries*, *41*(2).

Rinehart, R. E. (1983), Out-of-level instruments: Errors in hydrometeor spectra and precipitation measurements, J. Clim. Appl. Meteorol., 22, 1404–1410.

Robinson, A. C., and J. C. Rodda (1969), Rain, wind and the aerodynamic characteristics of rain-gauges, Meteorol. Mag., 98, 113–120.

Rong, G., Alu, S., Li, K., Su, Y., Zhang, J., Zhang, Y., et al. (2020). Rainfall Induced Landslide Susceptibility Mapping Based on Bayesian Optimized Random Forest and Gradient Boosting Decision Tree Models-A Case Study of Shuicheng County, China. Water. 12 (11), 3066. doi:10.3390/w12113066.

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2024). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Advanced Intelligent Systems, 2400304.

Sam, Ee. N. M., Pradhan, B., and Lee, S. (2020). Application of Convolutional Neural Networks Featuring Bayesian Optimization for Landslide Susceptibility Assessment. Catena. 186 (6), 104249.

Saouabe, T., El Khalki, E.M., Saidi, M.E.M., Najmi, A., Hadri, A., Rachidi, S., et al., 2020. Evaluation of the GPM-IMERG precipitation product for flood modeling in a semi- arid mountainous basin in Morocco. Water 12 (9), 2516.

Senocak, A. U. G., Yilmaz, M. T., Kalkan, S., Yucel, I., &Amjad, M. (2023). An explainable two-stage machine learning approach for precipitation forecast. Journal of Hydrology, 627, 130375.

Sensoy, S., 2004. The Mountains Influence On Turkey Climate Climate of Turkey 25–29.

Sensoy, S., Demircan, M., Ulupınar, U., & Balta, I. (2008). Türkiye iklimi. Turkish State Meteorological Service (DMİ), Ankara.

Serafin, R. J., & Wilson, J. W. (2000). Operational weather radar in the United States: Progress and opportunity. Bulletin of the American Meteorological Society, 81(3), 501-518.

Setiawati, M.D., Miura, F., 2016. Evaluation of GSMaP daily rainfall satellite data for f lood monitoring: case study—Kyushu Japan. J. Geosci. Environ. Protection 04 (12), 101–117.

Shapley, L. S. (1953). A value for n-person games.

Sharifi, E., Steinacker, R., & Saghafian, B. (2016). Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: Preliminary results. Remote Sensing, 8(2), 135.

Shi, J., Yuan, F., Shi, C., Zhao, C., Zhang, L., Ren, L., et al., 2020. Statistical Evaluation of the Latest GPM-Era IMERG and GSMaP Satellite Precipitation Products in the Yellow River Source Region. Water 12 (4), 1006.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017, July). Learning important features through propagating activation differences. In International conference on machine learning (pp. 3145-3153). PMlR.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Sieck, L. C., Burges, S. J., & Steiner, M. (2007). Challenges in obtaining reliable measurements of point rainfall. Water Resources Research, 43(1).

Stef, N., Başağaoğlu, H., Chakraborty, D., & Jabeur, S. B. (2023). Does institutional quality affect CO2 emissions? Evidence from explainable artificial intelligence models. Energy Economics, 124, 106822.

Stisen, S., Højberg, A. L., Troldborg, L., Refsgaard, J. C., Christensen, B. S. B., Olsen, M., & Henriksen, H. J. (2012). On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes. Hydrology and Earth System Sciences, 16(11), 4157-4176.

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems, 41, 647-665.

Stuke, A., Rinke, P., and Todorović, M. (2021). Efficient Hyperparameter Tuning for Kernel ridge Regression with Bayesian Optimization. Mach. Learn. Sci. Technol. 2, 035022. doi:10.1088/2632-2153/abee59

Su, F., Y. Hong, and D. P. Lettenmaier, 2008: Evaluation of TRMMMultisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. J. Hy drometeor., 9, 622–640, doi:10.1175/2007JHM944.1.

Sushanth, K., Mishra, A., Mukhopadhyay, P., & Singh, R. (2023). Near-real-time forecasting of reservoir inflows using explainable machine learning and short-term weather forecasts. Stochastic Environmental Research and Risk Assessment, 37(10), 3945-3965.

Sun, Q., et al., 2014. Variations in global temperature and precipitation for the period of 1948 to 2010. Environ. Monit. Assess. 186 (9), 5663–5679.

Sun, Q., et al., 2018a. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys. 56 (1), 79–107.

Tang, G., Clark, M.P., Papalexiou, S.M., Ma, Z., Hong, Y., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote Sens. Environ. 240, 111697.

Tang, G., Ma, Y., Long, D., Zhong, L., and Hong, Y. (2015). Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. J. Hydrology 533, 152–167. doi:10.1016/j.jhydrol.2015.12.008.

Tang, L., Tian, Y., Yan, F., and Habib, E., "An improved procedure for the validation of satellite-based precipitation estimates." Atmospheric Research, 163, 61-73 (2015b).

Tapiador, F. J., Navarro, A., García-Ortega, E., Merino, A., Sánchez, J. L., Marcos, C., & Kummerow, C. (2020). The contribution of rain gauges in the calibration of the IMERG product: Results from the first validation over Spain. Journal of Hydrometeorology, 21(2), 161-182.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A., ... & De Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. Atmospheric Research, 104, 70-97.

Tesfagiorgis, K., Mahani, S. E., Krakauer, N. Y., & Khanbilvardi, R. (2011). Bias correction of satellite rainfall estimates using a radar-gauge product–a case study in Oklahoma (USA). Hydrology and Earth System Sciences, 15(8), 2631-2647.

Thiemig, V., R. Rojas, M. Zambrano-Bigiarini, and A. D. Roo, 2013: Hydrological evaluation of satellite-based rainfall esti mates over the Volta and Baro-Akobo basin. J. Hydrol., 499, 324–338, doi:10.1016/j.jhydrol.2013.07.012.

Thouret, J. C., Enjolras, G., Martelli, K., Santoni, O., Luque, J. A., Nagata, M., ... & Macedo, L. (2013). Combining criteria for delineating lahar-and flash-flood-prone hazard and risk zones for the city of Arequipa, Peru. Natural Hazards and Earth System Sciences, 13(2), 339-360.

Thornes J, Bloss W, Bouzarovski S, Cai X, Chapman L, Clark J, Dessai S, Du S, van der Horst D, Kendall M, Kidd C, Randalls S. 2010. Communicating the value of atmospheric services. Meteorological Applications 17: 243–250, DOI: 10.1002/met.200

Tian, Y., & Peters-Lidard, C. D. (2010). A global map of uncertainties in satellite-based precipitation measurements. Geophysical Research Letters, 37(24).

Tian, Y., Peters-Lidard, C.D., Eylander, J.B., Joyce, R.J., Huffman, G.J., Adler, R.F., Hsu, K.L., Turk, F.J., Garcia, M. and Zeng, J., "Component analysis of errors in satellite-based precipitation estimates." Journal of Geophysical Research: Atmospheres, 114 - D24 (2009).

Toros, H., Kahraman, A., Tilev-Tanriover, S., Geertsema, G., & Cats, G. (2018). Simulating heavy precipitation with HARMONIE, HIRLAM and WRF-ARW: a flash flood case study in Istanbul, Turkey. Avrupa Bilim ve Teknoloji Dergisi, (13), 1-12.

Verma, B. P., Verma, V., & Badholia, A. (2022, July). Hyper-Tuned Ensemble Machine Learning Model for Credit Card Fraud Detection. In 2022 International Conference on Inventive Computation Technologies (ICICT) (pp. 320-327). IEEE.

Vila, D. A., De Goncalves, L. G. G., Toll, D. L., & Rozante, J. R. (2009). Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. Journal of Hydrometeorology, 10(2), 533-543.

Ward, E., Buytaert, W., Peaver, L., Wheater, H., 2011. Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. Adv. Water Resour. 34, 1222–1231

Wang, M., Li, Y., Yuan, H., Zhou, S., Wang, Y., Ikram, R. M. A., & Li, J. (2023). An XGBoost-SHAP approach to quantifying morphological impact on urban flooding susceptibility. Ecological Indicators, 156, 111137.

Wang, S., Luo, X., & Peng, Y. (2020). Spatial downscaling of MODIS land surface temperature based on geographically weighted autoregressive model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 2532-2546.

Wang, S., Zhuang, J., Zheng, J., Fan, H., Kong, J., & Zhan, J. (2021). Application of Bayesian hyperparameter optimized random forest and XGBoost model for landslide susceptibility mapping. Frontiers in Earth Science, 9, 712240.

Wang, L., Zhou, J., Qi, J., Sun, L., Yang, K., Tian, L., et al., 2017. Development of a land surface model with coupled snow and frozen soil physics. Water Resour. Res. 53 (6), 5085–5103.

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., Hou, A., 2017. Ground validation of GPM IMERG and trmm 3B42V7 rainfall products over Southern Tibetan plateau based on a high density rain gauge network. J. Geophys. Res. 122, 910–924. https://doi.org/10. 1002/2016JD025418.

Xue, X., Hong, Y., Limaye, A.S., Gourley, J.J., Huffman, G.J., Khan, S.I., et al., 2013. Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? J. Hydrol. 499, 91–99.

Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., & Verbist, K. M. (2016). Bias adjustment of satellite-based precipitation estimation using gauge observations: A case study in Chile. Journal of Geophysical Research: Atmospheres, 121(8), 3790-3806.

Yang, Y., Wang, H., Chen, F., Zheng, X., Fu, Y., and Zhou, S. (2018). TRMM-based optical and microphysical features of precipitating clouds in summer over the yangtze–huaihe river valley, China. Pure Appl. Geophys. 176, 357–370. doi:10.1007/ s00024-018-1940-8.

Yilmaz, K. K., T. S. Hogue, K. L. Hsu, S. Sorooshian, H. V. Gupta, and T. Wagener, 2005: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologicforecasting. J.Hydrometeor., 6, 497–517, doi:10.1175/ JHM431.1.

Yong, B., Ren, L., Hong, Y., Wang, J., Gourley, J.J., Jiang, S., et al., 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. Water Resour. Res. 46 (7).

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579.

Yu, C., Hu, D., Liu, M., Wang, S., Di, Y., 2020. Spatio-temporal accuracy evaluation of three high-resolution satellite precipitation products in China area. Atmos. Res. 241, 104952.

Yu, L., Zhang, Y., Yang, Y., 2020. Using high-density rain gauges to validate the accuracy of satellite precipitation products over complex terrains. Atmosphere (Basel). 11, 633.

Yucel, I. (2015). Assessment of a flash flood event using different precipitation datasets. Natural Hazards, 79, 1889-1911.

Yucel, I., Kuligowski, R. J., & Gochis, D. J. (2011). Evaluating the hydro-estimator satellite rainfall algorithm over a mountainous region. International journal of remote sensing, 32(22), 7315-7342.

Yucel, I., & Onen, A. (2014). Evaluating a mesoscale atmosphere model and a satellite-based algorithm in estimating extreme rainfall events in northwestern Turkey. Natural Hazards and Earth System Sciences, 14(3), 611-624.

Zeng, Z., Chen, H., Shi, Q., & Li, J. (2021). Spatial downscaling of IMERG considering vegetation index based on adaptive lag phase. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-15.

Zeng, Q., Chen, H., Xu, C., Jie, M., Chen, J., Guo, S., et al., 2018b. The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. J. Hydrol. 563, 106–122.

Zeng, Q., et al., 2018. Inter-comparison and evaluation of remote sensing precipitation products over China from 2005 to 2013. Remote Sens. 10 (2), 168.

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. Ieee Access, 6, 21020-21031.

Zhang, L., Chen, X., Lai, R., Zhu, Z., 2022. Performance of satellite-based and reanalysis precipitation products under multi-temporal scales and extreme weather in mainland China. J. Hydrol. 605, 127389.

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., & Ge, Y. (2021). Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. Journal of Hydrology, 594, 125969.

Zhang, S., Wang, D., Qin, Z., Zheng, Y., Guo, J., 2018. Assessment of the GPM and TRMM precipitation products using the rain gauge network over the Tibetan Plateau. J. Meteorol. Res. 32, 324–336. https://doi.org/10.1007/s13351-018-7067-0.

Zhao, Z., Anand, R., & Wang, M. (2019, October). Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In *2019 IEEE international conference on data science and advanced analytics (DSAA)* (pp. 442-452). IEEE.

Zheng, X., Yang, Y., Yuan, Y., Cao, Y. n., and Gao, J. (2021). Comparison of macro and microphysical properties in precipitating and non-precipitating clouds over central-eastern China during warm season. Remote Sens. 14, 152. doi:10.3390/ rs14010152.

Zhong, J., Sun, Y., Peng, W., Xie, M., Yang, J., & Tang, X. (2018). XGBFEMF: an XGBoost-based framework for essential protein prediction. IEEE transactions on nanobioscience, 17(3), 243-250.

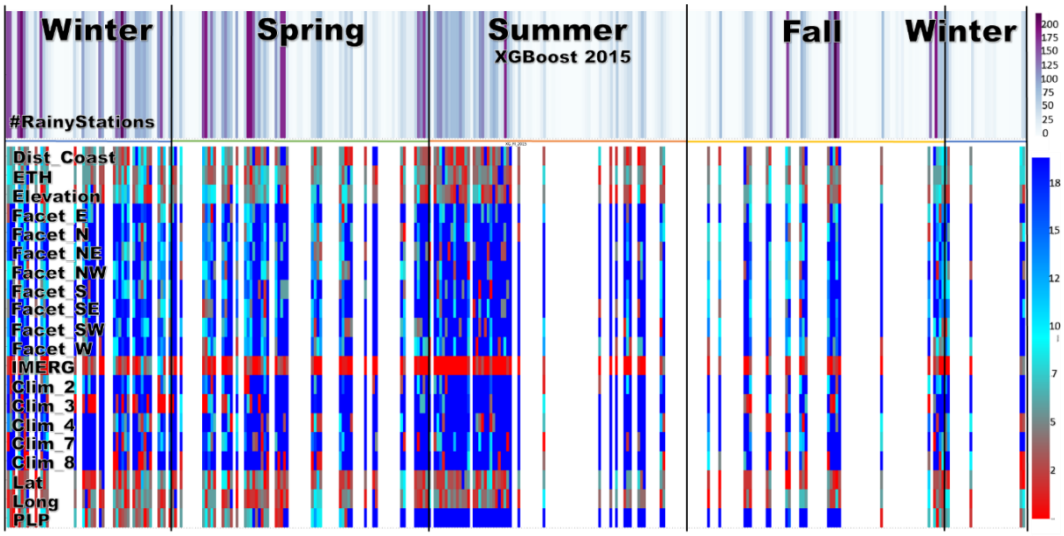
Zhou, L., Koike, T., Takeuchi, K., Rasmy, M., Onuma, K., Ito, H., ... & Ao, T. (2022). A study on availability of ground observations and its impacts on bias correction of satellite precipitation products and hydrologic simulation efficiency. Journal of Hydrology, 610, 127595.

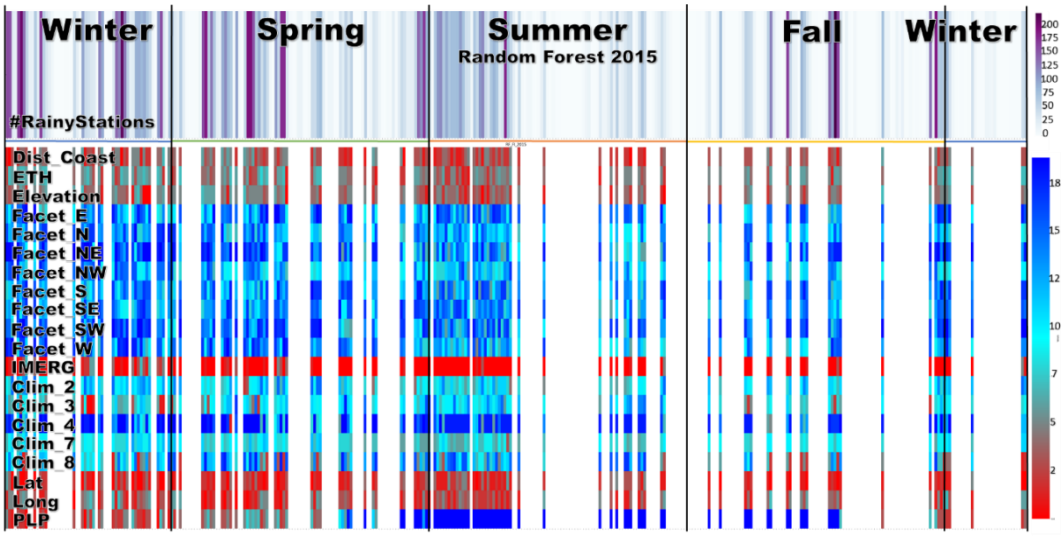
Zhu, H., Liu, H., Zhou, Q., & Cui, A. (2023). A XGBoost-based downscaling-calibration scheme for extreme precipitation events. IEEE Transactions on Geoscience and Remote Sensing.

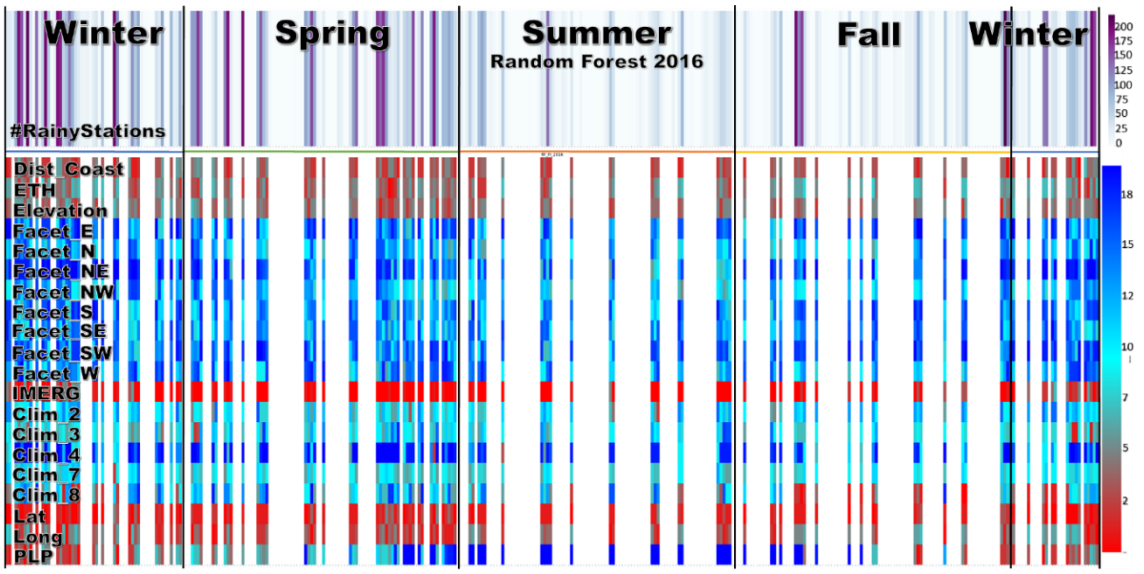
Zhu, X., et al., 2015. Comparison of monthly precipitation derived from high-resolution gridded datasets in arid Xinjiang, central Asia. Quat. Int. 358, 160–170.

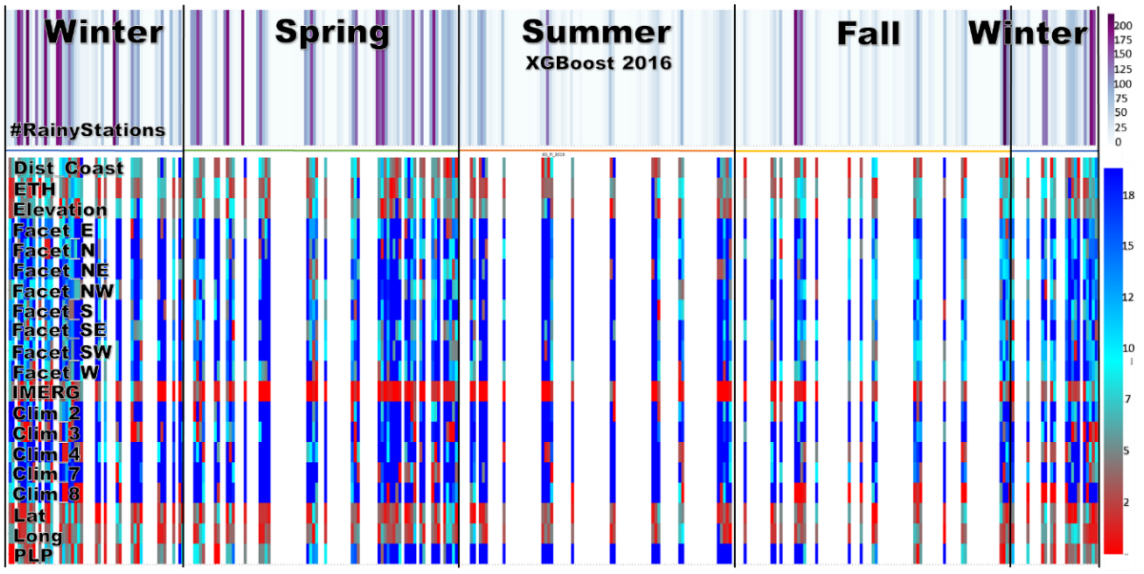
APPENDICES

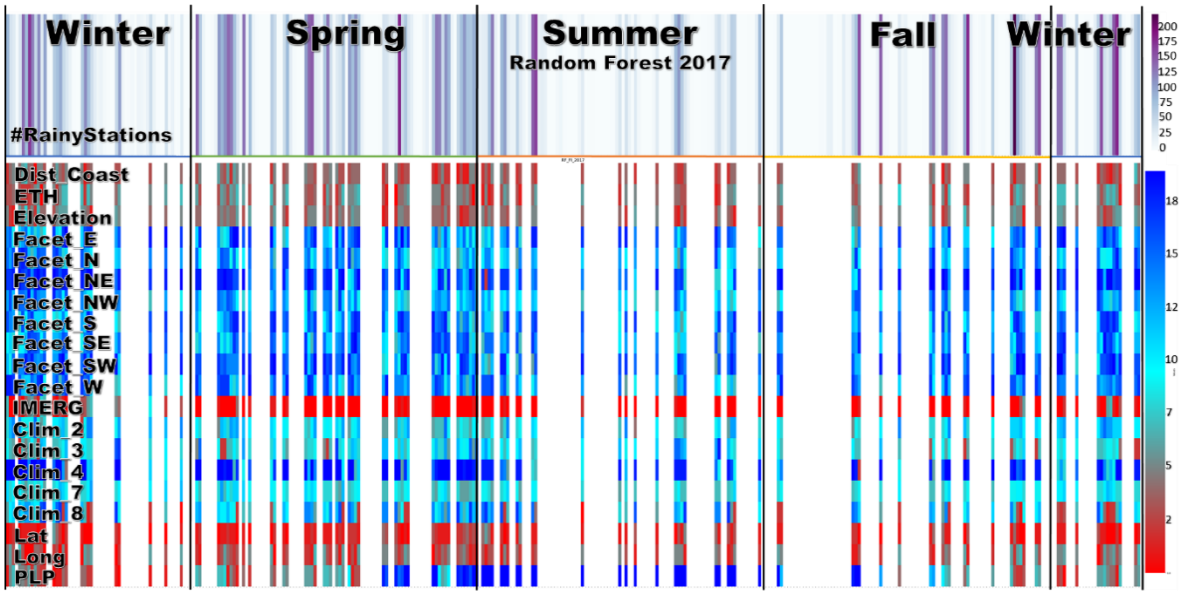
A. Appendix: SHAP feature importance scores as heatmap for Random Forest and XGBoost between 2015-2022

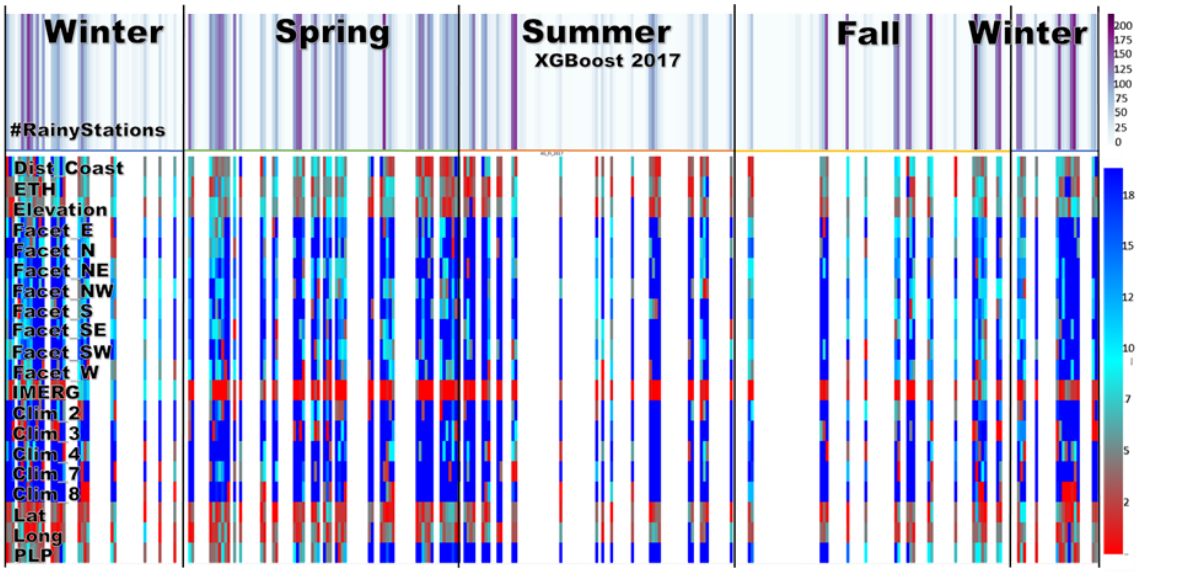


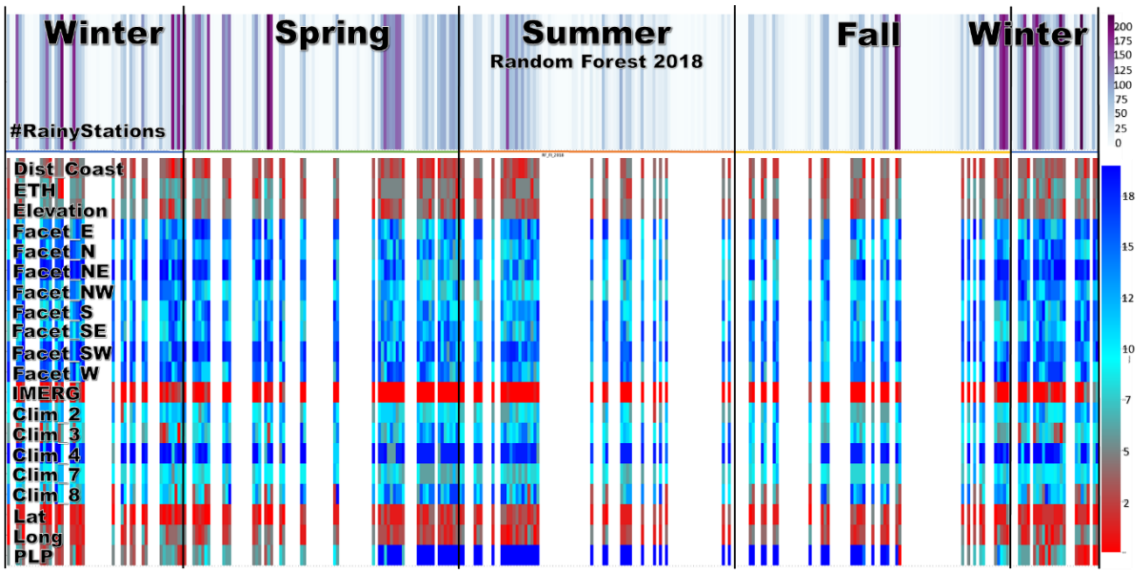


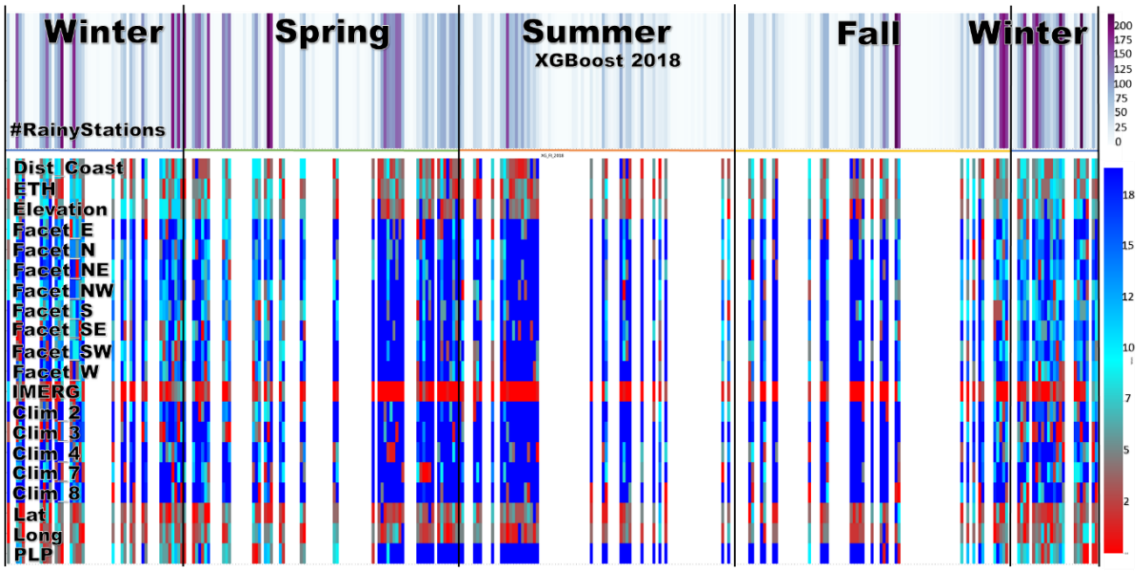

Winter
Spring
Summer
Random Forest 2016
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP

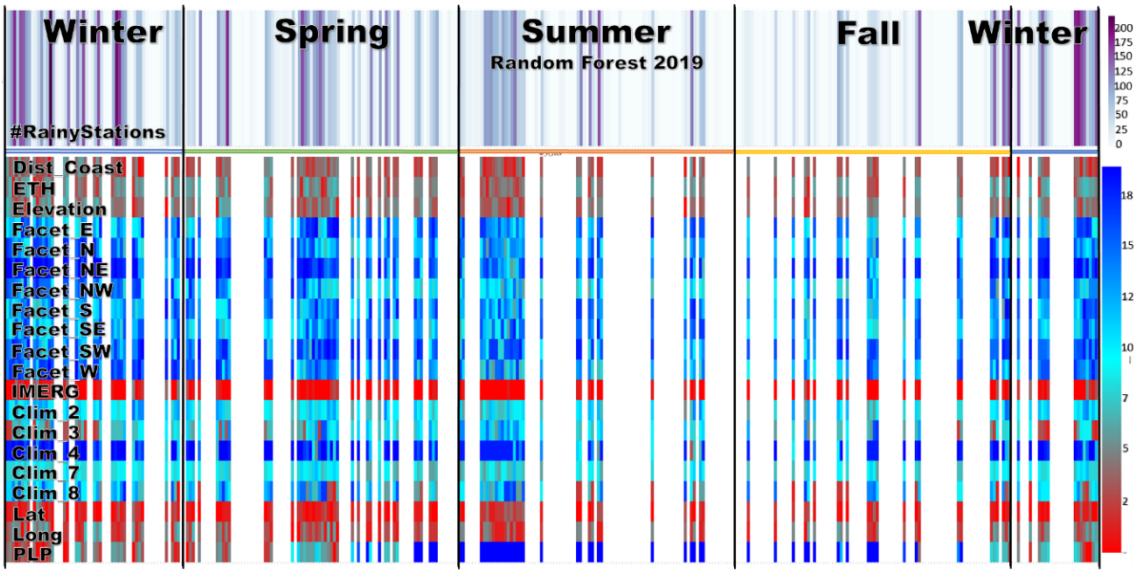

Winter
Spring
Summer
XGBoost 2016
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP

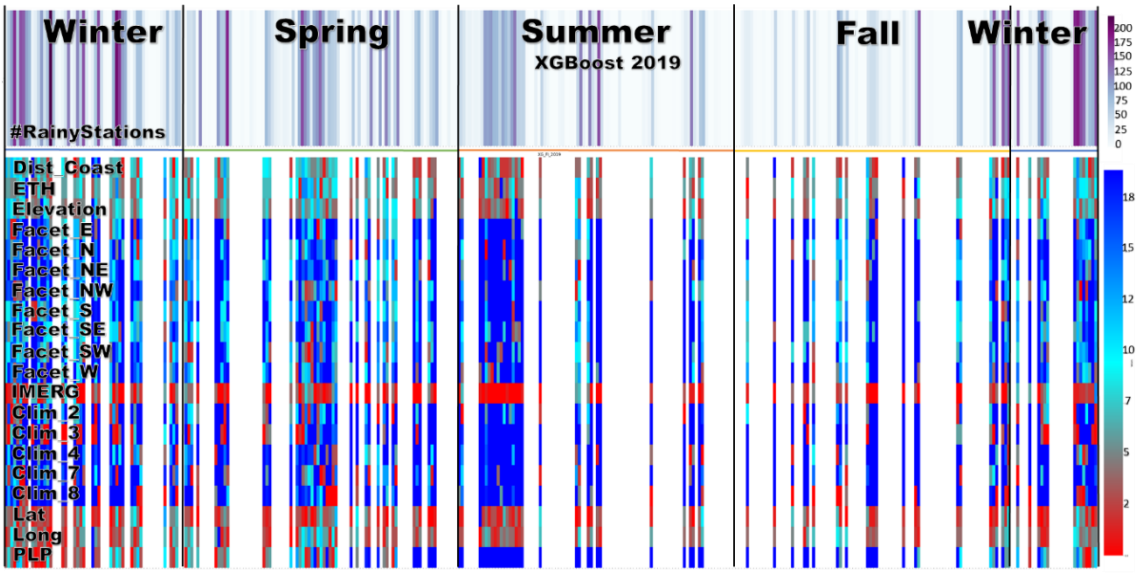

Winter
Spring
Summer
Random Forest 2017
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2

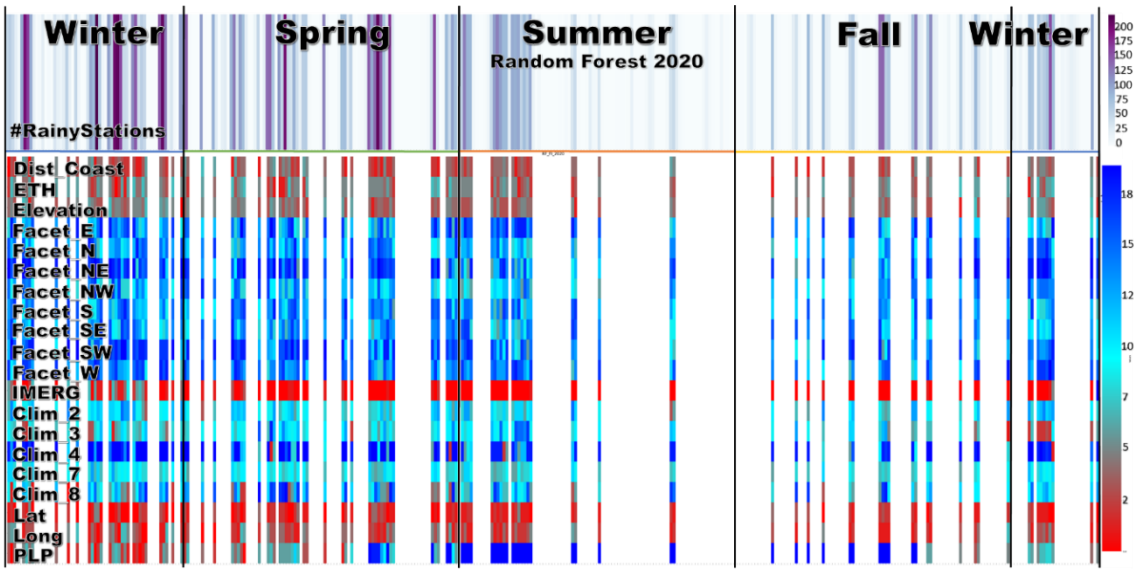

Winter
Spring
Summer
XGBoost 2017
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2

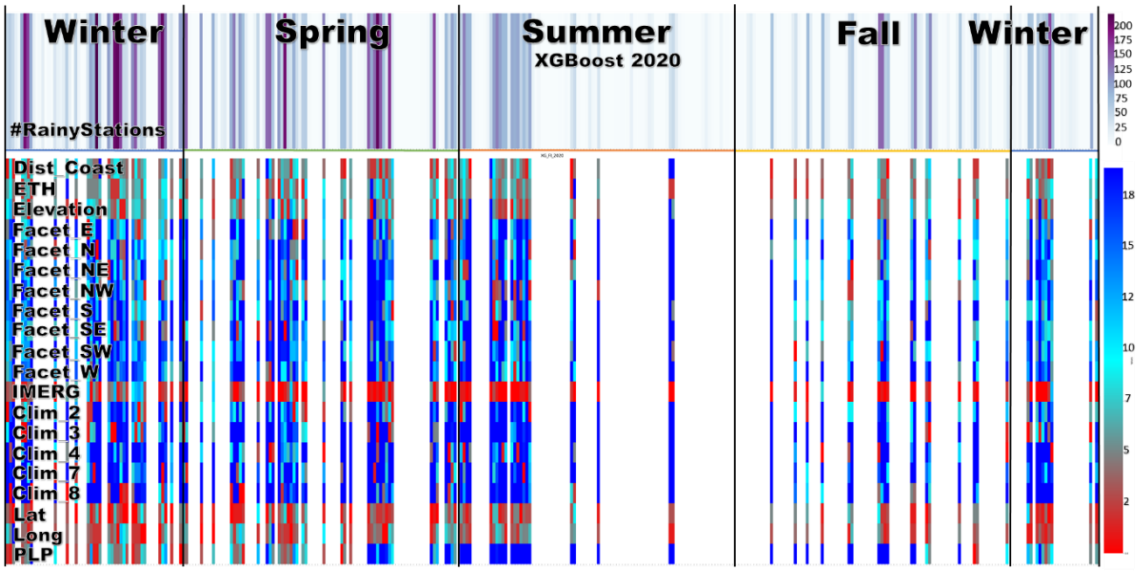

Winter
Spring
Summer
Random Forest 2018
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP

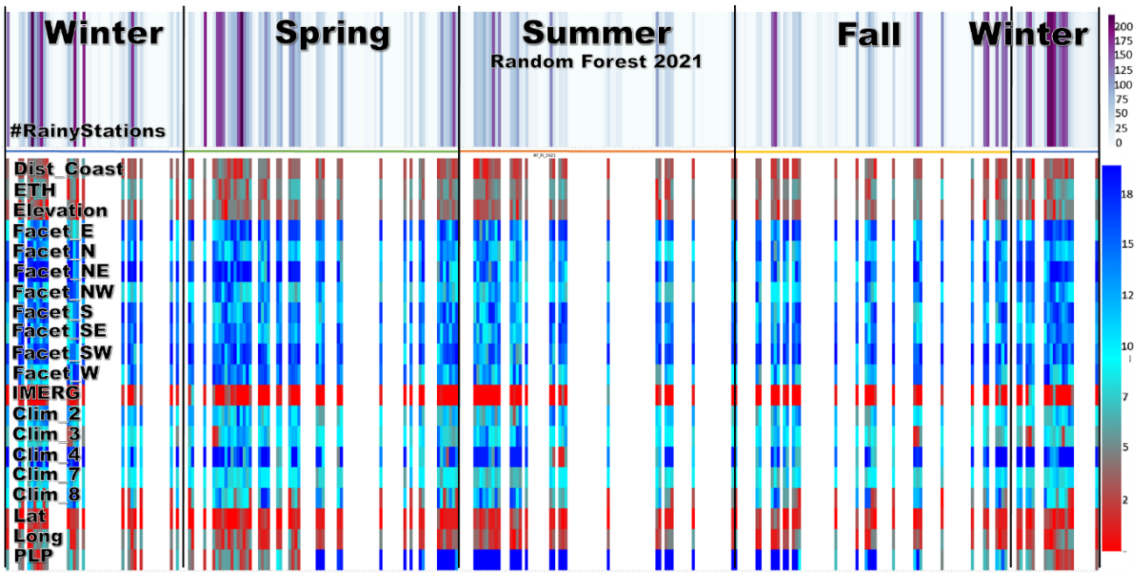

Winter
Spring
Summer
XGBoost 2018
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP

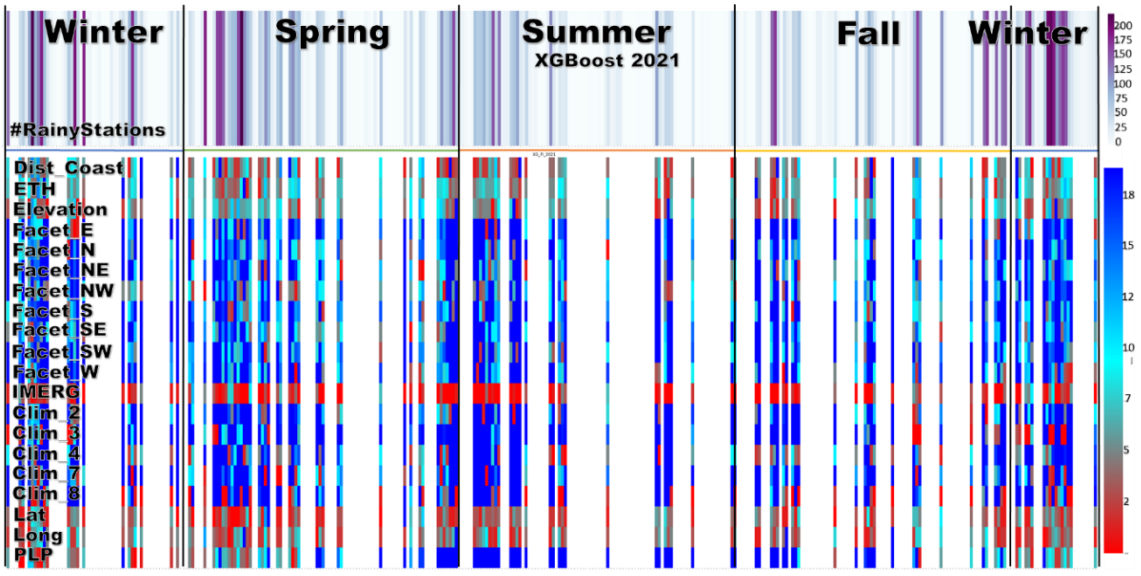

Winter
Spring
Summer
Random Forest 2019
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2

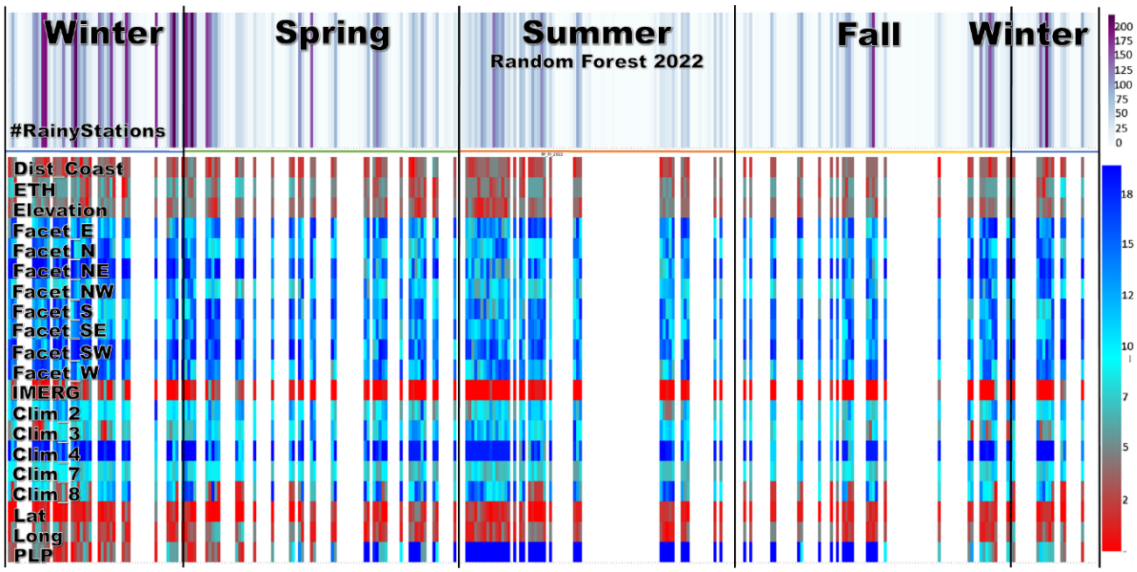

Winter
Spring
Summer
XGBoost 2019
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2

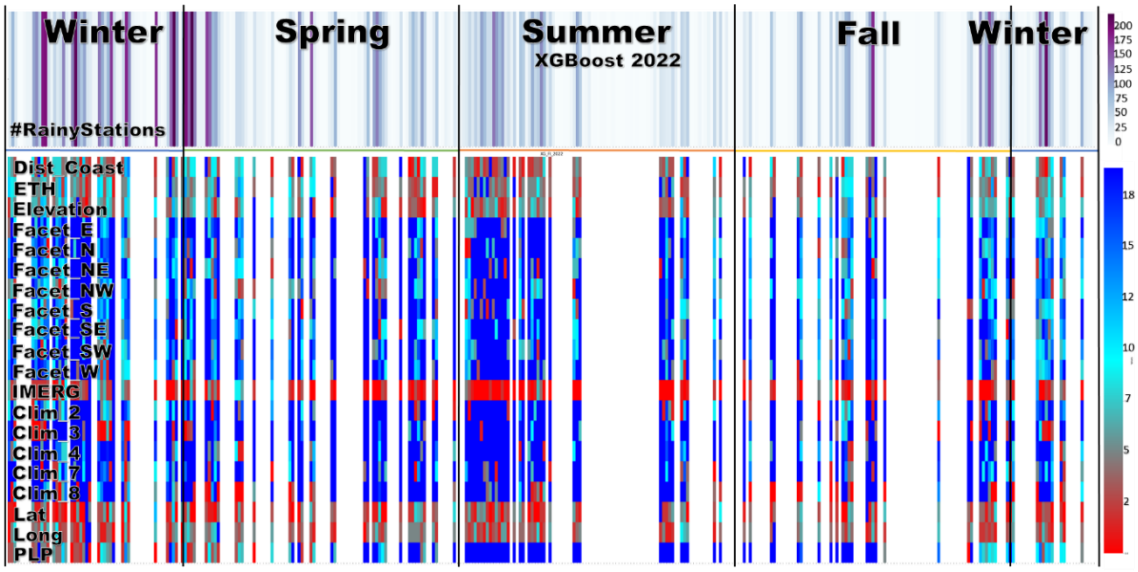

Winter
Spring
Summer
Random Forest 2020
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP


Winter
Spring
Summer
XGBoost 2020
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP


Winter
Spring
Summer
Random Forest 2021
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2


Winter
Spring
Summer
XGBoost 2021
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2


Winter
Spring
Summer
Random Forest 2022
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2


Winter
Spring
Summer
XGBoost 2022
Fall
Winter
#RainyStations
Dist_Coast
ETH
Elevation
Facet_E
Facet_N
Facet_NE
Facet_NW
Facet_S
Facet_SE
Facet_SW
Facet_W
IMERG
Clim_2
Clim_3
Clim_4
Clim_7
Clim_8
Lat
Long
PLP
200
175
150
125
100
75
50
25
0
18
15
12
10
7
5
2