Possibilities and Limitations of IMERG Datasets for Estimating Probable Maximum Precipitation

By

Kenneth Okechukwu Ekpetere

Submitted to the graduate degree program in Geography and the Graduate Faculty of the University of Kansas in Partial fulfilment of the requirements for the degree of Master of Science.

_____
Chairperson: Dr. Xingong Li

_____
Co-Chair: Dr. Jude H. Kastens

_____
Dr. David B. Mechem

Date Defended: July 26th, 2021

The Thesis Committee for Kenneth Okechukwu Ekpetere Certifies that this is the approved version of the following thesis:

Possibilities and Limitations of IMERG Datasets for Estimating Probable Maximum Precipitation

Chairperson: Dr. Xingong Li

Co-Chair: Dr. Jude H. Kastens

Dr. David B. Mechem

Date Approved: July 26th, 2021

ABSTRACT

The probable maximum precipitation (PMP), which is conventionally derived based on precipitation gauge data, is an important input for hydrological models useful for accurate storm predictions, flood modeling, and general decision making. However, precipitation data are not readily available in most regions due to limited and failing gauges. This study assessed the possibilities and limitations of deriving PMP using precipitation records from the Integrated Multi-satellite Retrievals for GPM (IMERG). PMPs for six durations were calculated with IMERG precipitation records (2000-2020) using Hershfield statistical technique built in R. The IMERG PMPs were further evaluated with NOAA-Atlas-14 station PMPs at 55 rain gauge locations in Kansas, USA, using coefficient of correlation, root mean square error, and relative bias as statistical metrics. Result of PMP evaluation for the six durations showed that, as durations increase from 30-mins to 24-hr, CC decreased from 0.956 to 0.854, RMSE increased from 4.41 mm to 10.28mm, and RB increased from -3.77% to 7.82%. RB estimates ranged between -3.77% to -7.96% at shorter durations (30-min to 12-hr) depicting underestimation in IMERG PMP. Relationship between precipitation amount and PMP errors at the six durations showed that increase durations reduces PMP error. At the 24-hr durations, the most R-squared were estimated at 51.5% and 50.3% for total accumulated, and maximum precipitation respectively. Further exploration showed that IMERG has an average percentage missing value of 74.21% with a 2-hr average length of missing value. The relationship between Missing values and PMP errors were statistically significant at the 30-mins durations with p-values of 0.00287 and 0.00127 for both percentage missing values and length of missing values respectively, while the longer intervals showed decreasing relationship between the variables with p-values of 0.665 and 0.4213 respectively. Overall assessment showed that IMERG estimate PMP better in wetter areas and longer intervals (e.g., 24-hr), while the estimated PMP errors may be influenced by missing values in IMERG datasets, with the greater influence at shorter durations (e.g., 30-min).

DEDICATION

This thesis research is dedicated to God first for His blessings and wisdom, and to my late mom Mrs. Theresa Ekpetere, whose passing mid-way into my master's program (April 21st, 2020) came as a shock. Mom, you taught me to be diligent, hardworking, and consistent, this project is a product of those virtues you imbibed in me, I love you and miss you always dear mom. Finally, I dedicate this project to the University of Kansas and graduate division (COGA) for the trust and support invested in me to do my research at the graduate level. Thank you!

AKNOWLEDEMENT

This study is very near to my heart, started after many research ideas but none came close, it will go on to forge the foundation of my research career in flood modeling and flood risk mapping. From the conception of the idea behind this research to the completion stage, it took exactly one year. Driven by data scarcity that limits many developing regions, this research was able to forge a tool that help fill those gaps, by relying on IMERG data spatial extent. In addition to filling data scarcity, flood prone states such as Kansas and other water related agencies will find this PMP tool useful. I am extremely happy to make this contribution to research and also contribute to the state of Kansas and the United State.

I would first like to thank my advisor, Dr. Xingong Li, for all his supports, guidance, and patience throughout my master's research. From far away country, you believed in my ability to succeed even before joining your research group, you have mentored, taught, showed me how to carry on as a true scholar by conducting independent research. Thank you for continuing to believe in my ability to succeed. I would also like to acknowledge Dr. Jude Kastens for his supports and teachings which helped shaped my thesis research to a better trajectory. This work would have been impossible without you and the KBS crew that have helped fund part of my research in the summer. Lastly, I would like to appreciate Dr. David Mechem, who signed my admission letter when he was the acting chair, and equally served in my thesis committee.

Specially thanks to James Coll, my senior colleague and friend in Xingong's lab for all his support and help during this one-year journey. Your assistance and support with the GEE java scripts are deeply appreciated. I thank other members of the research group particularly Chen for her inputs. What about icons such as my graduate coordinator, Dr. Alexander Diener, and others, Dr. Shi Di, Dr. K.T. Rhine, Dr. Chikanda Abel, and Cicily Corder. Specially thanks to all the staff and faculties, as wells as graduate students within and outside the department Geography for their encouragements.

Lastly, I thank my wife, Mrs. Kosi L. Ekpetere for her supports and understandings despite the 24hrs flight time away. I thank my Father, Chief Michael Ekpetere, my late mom Mrs. Theresa Ekpetere for their support. I especially thank my siblings for their distant cheers and support all through this journey. To my close pals Mr. Jacob Z. Oni, Mr. Adeniyi E. Ajobiewe, and other associates whose list of names are inexhaustible, I say thank you.

LIST OF ACRONYMS AND ABBREVIATIONS

IMERG	Integrated Multi-satellite Retrievals for GPM
TRMM	Tropical Rainfall Measuring Mission
PE	Precipitation Estimates
PMP	Probable Maximum Precipitation
PMF	Probable Maximum Flood
SPP	Satellite/Spatial Precipitation Products
GEE	Google Earth Engine
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
NOAA	National Oceanic and Atmospheric Administration
WMO	World Meteorological Organization
RMSE	Root Mean Square Error
CC	Correlation Coefficient
RB	Relative Bias

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

The challenges in flood hazard assessment caused by data insufficiency has become more pronounced in developing regions of the world. Poor data coverage problem has been linked to inadequate rain gauges, rain gauges equipment becoming obsolete and worn-out, poor maintenance of the equipment, and lack of skilled personnel to monitor and maintain the gauges. Flood disasters have been known to wreak havoc on communities around the world; ranging from property damages to deaths of hundreds to thousands of people annually (Bathrellos et al., 2016). Understanding the precipitation characteristics at a location became an important variable to model flood, hence the reliance on the Probable Maximum Precipitation (PMP). PMP is theoretically the greatest depth of precipitation for a given duration over a given size of storm area at a particular geographical location at a certain time of year (Dingman, 2014). It is used as inputs for hydrological models to estimate the largest flood that could occur in a drainage basin. This maximum flood derived from the PMPs is termed Probable Maximum Flood (PMF) (Dingman, 2014), which can be used to determine the required capacity of the emergency spillway of a dam whose failure would cause massive economic damage and loss of lives.

PMP is estimated either by examining rainfall data for the largest flood-producing storms in and near the region of interest and using meteorological reasoning to estimate the combination of conditions that could have produced the largest rainfall rates from those storms on the drainage basin of interest, or by statistically interpolating data from regions of similarities of the largest storms in a large region (Dingman, 2014). This implies that PMP can be derived in two ways, by either adopting a meteorological or statistical approach. However, the lack of reliable precipitation data for estimating PMP in many urban cities particularly in developing countries, has deterred significantly those measures to address flood vulnerabilities and adaptation (Kim et al., 2019).

Estimating PMPs requires rain gauge data on precipitation, which hinges on meteorological techniques to compute PMPs. But, when there exist insufficient/unreliable rain gauge data to compute the PMPs for those gauge, a statistical method that relies on precipitation record from remote sensing technologies are adopted (Yang et al., 2019). Hershfield statistical technique for generating PMPs is widely used in different countries such as Spain, Iran, India, Japan, Malaysia, etc., providing comparable PMP estimates to those obtained using the meteorological method (World Meteorological Organization, 2009). The World Meteorological Organization (WMO), recommends the use of the Hershfield statistical method for estimating PMP at stations with long periods of daily rainfall record, particularly those not less than 10 years (Alamri & Subyani, 2017).

Satellite precipitation products (SPPs) have emerged as an alternative means of acquiring precipitation data remotely from which PMP can be estimated (Tan & Santo, 2018). Unlike ground-based measurements such as gauges and radars, SPPs cover the precipitation system at a global scale, irrespective of difficult terrain across the globe. Ground-based radar measurements are affected by factors such as signal attenuation, surface backscattering, and reflectivity-rain-rate relationship uncertainty issues. Hence, SPPs are widely relied upon for precipitation characteristics analysis, hydrological modeling, and drought monitoring (Tan & Santo, 2018). Several SPPs have been made available and free for public accessibility, including the Global Precipitation Measurement (GPM), Tropical Rainfall Measuring Mission (TRMM), and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN). The GPM satellite launched in 2014 was intended to substitute the TRMM satellite, which ended in April 2015 after serving for 17 years. Generally, the newly released Integrated Multi-satellite Retrievals for GPM (IMERG) products have improvements in both spatial and temporal resolution (Tan & Santo, 2018).

The Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) is a new dataset that provides global precipitation measurement. The IMERG consists of an international network of satellites that provide the next-generation global observations of rain and snow (Guo et al., 2016). The IMERG Core Observatory design is an extension of TRMM which focused primarily on heavy to moderate rain over tropical and subtropical oceans. It carries the first space-borne Ku/Ka-band Dual-frequency Precipitation Radar (DPR) and a multi-channel GPM Microwave Imager (GMI). The DPR instrument provides three-dimensional measurements of precipitation structure over 78- and 152-mile (125 and 245 km) swaths, consists of a Ka-band precipitation radar (KaPR) operating at 35.5 GHz and a Ku-band precipitation radar (KuPR) operating at 13.6 GHz. The GMI instrument is a conical-scanning multi-channel microwave radiometer covering a swath of 550 miles (885 km) with thirteen channels ranging in frequency from 10 GHz to 183 GHz. With a 0.1^0 spatial grid and a temporal resolution of half-hour, the IMERG is expected to have better precipitation estimates over other SPPs when compared to data from the rain gauge (Su, Lü, Ryu, et al., 2019). IMERG applications range widely from disaster and risk management, ecological management, agricultural forecasting, energy infrastructure, development and public health, weather, and climate study.

IMERG is calibrated with the actual rain gauge estimates from the Global Precipitation Climatology Center (GPCC) gauge data (Tapiador et al., 2020) as its main inputs. The GPCC database contains precipitation data of over ~80,000 rain gauges. Different gridded precipitation products are released monthly based on this data. GPCC Monitoring Product Version 6 (GPCC-MP) is one of the GPCC products that provides monthly precipitation data at 1.0^0 x 1.0^0 in near real time used by IMERG (Schamm et al., 2014). Other data sources used to adjust the IMERG estimates to attain precision includes, data from over 2,300 rain gauges obtained from the Spanish Meteorological Agency (AEMET), data from the Global Precipitation Climatology Project

(GPCP) developed by the World Climate Research Programme (WCRP) to measure the global distribution of precipitation. The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) and ERA5 reanalysis are the additional data sources for calibrating IMERG estimates (Huffman et al., 2020).

The GPCC itself with over 80,000 rain gauges spread globally and contained in its database, receives precipitation data from numerous agencies, organizations, and 190 member countries of the WMO (George J. Huffman, 2020). In a nutshell, some of the agencies includes meteorological organizations, hydrological service agencies, and other weather monitoring units such as, Nation Oceanic Atmospheric Administration (NOAA) in Washington D.C., Deutscher Wetterdienst (DWD) Offenbach Germany, United Kingdom Meteorological Office (UKMO) Exeter UK, Japan Meteorological Agency (JMA) Tokyo-Japan. Other data sources include, Global Historical Climatology Network (GHCN), Food and Agricultural Organization of the United Nations (FAO), Climate Research Unit (CRU), Global Energy and Water Exchanges (GEWEX), among others.(George J. Huffman, 2020).

Yang et al. (2017), investigated the conventional PMP estimation approach by using both in-situ observations and mainstream satellite precipitation products in the Dadu River basin, where plenty of dams and reservoirs were built to serve the populace. Their work used several satellite precipitation products (CMORPH, PERSIAN-CDR, TRMM, and TMPA-3B42V7) with a coarse spatial resolution of 0.25^0. Their results showed that CMORPH and 3B42V7 had positive correlations of 0.68 and 0.71 respectively, compared to station data. Also, CMORPH and 3B42V7 showed better performance for their magnitude and spatial distribution of 24-h PMP in complex terrains. PERSIAN-CDR showed an overestimation in the upstream and underestimation in the downstream. However, they used SPPs with coarse resolution (0.25^0), and this might have negatively impacted the result. Similarly, Tan & Santo (2018) compared SPP's from IMERG, to

those from TMPA 3B42 and PERSIAN-CDR over Malaysia. They performed an initial comparison of three IMERG products (IMERG_E, IMERG_L, and IMERG_F) with its predecessors- TMPA 3B42 and 3B42RT products, and a long-term PERSIAN-CDR over Malaysia. Their validation used 501 precipitation gauges spread across Malaysia from 12 March 2014 to 29 February 2016. Their study adopted three widely used statistical metrics correlation coefficient (CC), root mean square error (RMSE), and relative bias (RB). Their study focused on validating precipitation amount instead of PMP.

In summary, the existing research used; (1) satellite products with coarse resolution, (2) most evaluation was carried out between satellites with no reference to station data, (3) validations were done using precipitation amount which is not ideal for flood modeling. For flood modeling and prediction to proceed with confidence, the PMP must be accounted for as an input variable (Tan & Santo, 2018). As such, this study focused on evaluating IMERG derived PMPs by extracting IMERG precipitation data and estimating PMPs from those data. The approach was seen to be particularly useful in regions lacking adequate rain gauge data. Three widely used statistical metrics (CC, RMSE, & RB) were incorporated to evaluate the IMERG estimated PMPs against station based PMPs from the NOAA-Atlas-14 precipitation product.

1.2 Research Questions and Objectives

The goal of this research is to identify the possibilities and limitations of IMERG derived PMPs by evaluating them with those from NOAA-Atlas-14 station PMPs. Salient questions asked in this study include, how well does IMERG derived PMPs compare with those from station PMPs? does precipitation variability influence satellite based PMPs? does missing values influence PMP errors at different intervals? To meet the specific questions in this study, the following objectives were pursued: (i) compute PMPs using precipitation estimates from IMERG, (ii) evaluate IMERG

estimated PMPs using rain gauge based PMPs, (iii) determine the relationship between PMP errors and precipitation amount, (iv) Determine the impact of missing values on PMP errors.

2 RESEARCH CONTEXT

This section provides more understanding on the concepts of PMP estimations, a description of the IMERG satellite and properties. The operational modality of IMERG, data fusion mechanism, and technical issues are also covered in this section.

2.1 Concepts and Definitions of PMP

The PMP originated from what used to be called Maximum Possible Precipitation (MPP) referred to as an upper maximum bound value (Salas et al., 2014). The idea was to find a maximum value of precipitation for a given storm period over a basin that physically could take place in real life but would not be exceeded. In real life, such MPP values have been exceeded as reported by several literatures, leading to the renaming of "Maximum Possible Precipitation" to "PMP" (Salas et al., 2014). Consequently, the definition also changed to "theoretically the greatest depth of precipitation for a given storm duration that is physically possible over a given size storm area at a particular geographical location at a certain time of the year" (Salas et al., 2014). Salas et al., (2014) noted that the idea behind establishing the PMP have remained the same over the years, the PMP definition used by the WMO has been slightly changed during these periods. WMO definition highlights the PMP as a physical upper limit and a quantity that cannot be exceeded. Nevertheless, WMO acknowledges the fact that the value of the PMP that is estimated for a particular location is only an approximation hinging on the physical complexity of the phenomena and limitations in data and the meteorological and hydrological sciences (Salas et al., 2014).

There are two general approaches to computing PMP endorsed by the WMO, categorized as hydrometeorological and statistical, which are also called direct and indirect PMP estimation approach respectively (Singh et al., 2018). Hydrometeorological methods include

storm transposition method, moisture maximization method, storm separation method, and generalized method. Statistical methods are the multifractal method and Hershfield's method. It's important to note that PMP estimates are dependent on the years of data availability, watershed topography, region of interest, and adopted method for estimating the PMP (Singh et al., 2018). In storm transposition method for example, the relocation of storm precipitation within a homogenous region is accounted for, as well as the relationship of the storm with terrain and meteorological features that are important to the storm rainfall.

The main idea of the storm transposition method is that a meteorological homogeneous region exists such that a major storm occurring in any particular place in the region could occur anywhere else in the region (Singh et al., 2018). The idea is that, the storm transported to a location could occur under similar meteorological conditions as the first location. The meteorological analysis of the storms to be transported, transposition limits, and the application of the appropriate adjustments factor for the change in storm location is accounted for by the storm transposition method. In the storm transposition method, it is believed that a major storm occurring somewhere in the region could occur elsewhere provided that the region is meteorologically homogenous (Singh & Singh, 2017).

Storm separation method is adapted in orographic regions where storm transposition methods are not applicable, with the assumption that relief convergence rainfall amounts can be explicitly estimated. The free-atmospheric forced precipitation (FAFP)(HMR, 57) also called convergence rainfall is notably one of the earliest reports which discussed the development of PMP in terms of relief and convergence components (National Weather Service, 1994). Convergence precipitation is a kind of precipitation with known or established terrain effect acting independently, and orographic precipitation is the example of such associated rainfall from terrain effects. As a known fact, the atmosphere is not totally free from the impacts of terrain, but cases can be found where the terrain feedback is totally

negligible especially when macro atmospheric scale is considered in explaining the storm precipitation patterns, and in these cases, precipitation is classified as pure convergence or non-orographic precipitation (National Weather Service, 1994).

The maximum recorded rainfall depths of rainstorms over extensive area are used in the generalized method. The generalized method adjusts the maximum recorded rain depths to a particular catchment (Singh & Singh, 2017). The generalized method has the benefit of using the maximum recorded rain depths in all intervals and areas combined, allowing for ease of transposition in space. Singh et al., (2018) noted how the generalized method were applied in estimating PMP values for catchments of four large dam basins in India. In the report, they assumed that an optimum combination of the available moisture in the atmosphere as well as the storm mechanism efficiency would aid in estimating PMP values using observed precipitation which were not directly measured (Singh & Singh, 2017).

In moisture maximization method, the assumption is that precipitation has a near perfect correlation with precipitable water such that the storm precipitation increases by values that are in accordance with the maximum moisture in the atmosphere for the storm location and month of occurrence. Precipitation efficiency remains constant as the storm moisture increases, and the record of high storms is adequately large to represent the most efficient storm mechanisms, but this do not inform for the optimum available moisture that would accompany a PMP event (Singh et al., 2018). In most PMP studies, important atmospheric conditions entail the efficiency with which storm converts moisture into precipitation. In addition to the amount of considered moisture content, the moisture maximization approach when adopted, approximates the highest moisture potential in the storm.

Multifractal analysis also known as multi-scaling have recently been used for PMP estimation. The multifractal analysis is generally used to describe the scaling behavior of precipitation and streamflow. This technique has also been used in the eastern United States

to calculate the physically meaningful estimates of maximum precipitation from observations (Singh & Singh, 2017). The multifractal analysis has an advantage as it provides a formal framework to infer the degree of extreme events (termed the fractal maximum precipitation (FMP)), that are suitable for orographic terrains and independent of empirical adjustments. It is worthy to note that, the multifractal method is constrained by the length of record, the spatial resolution of rain gauge network, and the lack of uncertainty estimates (Singh & Singh, 2017).

The Hershfield statistical method proposed in 1961 is the most used approach when long term data is provided. While there are no exact years of precipitation record range, a good data range will be anywhere between 15-50 years of precipitation record (World Meteorological Organization, 2009). Yang et al., (2017) used a precipitation record range of 15-years interval period (1998-2013), Gao et al., (2018) used 35 years of data range (1981-2015), Singh et al., (2018) used 30 years of precipitation record. The World Meteorological Organization (WMO) in its handbook proposed a good precipitation interval to fall between 20 or more years, whereas the NOAA rain gauge records fall anywhere between 15years or precipitation records to over 100years. The Hershfield method have been used all over the world for estimating PMP values and for comparing with other methods. Singh et al., (2018) noted that estimating PMP estimates using the Hershfield method have proven successful. Hershfield technique is based on average precipitation and standard deviation of precipitation, for computing the frequency factor (*Km*) and PMP (see eq. 1 and 2 in the study research methods). Singh et al., (2018) noted that "in 1961 Hershfield initially used 15 as the maximum value of frequency factor for computing PMP, but later in 1965 Hershfield found an upper envelope of frequency factor which had the tendency to decrease with the increasing precipitation amount. In summary, the frequency factor decreases with increasing mean annual maximum precipitation (Singh et al., 2018)". The value of *Km* varies from 5 to 20, depending upon the mean precipitation and the precipitation duration (Singh et al., 2018). This

research adopts the Hershfield technique for computing PMP using IMERG satellite precipitation record computed over a 20-years period (2000 to 2020).

2.2 Integrated Multi-Satellite Retrievals for GPM (IMERG)

The Integrated Multi-Satellite Retrievals for GPM (IMERG) emerged as a joint project between the United State (NASA) and Japan (JAXA) that began in 2014 as a replacement for TRMM. IMERG consists of a unified algorithm that provides the multi-satellite precipitation product for the GPM team. IMERG adopts the Goddard Profiling Algorithm (GPROF) for precipitation estimation from the different precipitation-relevant satellite passive microwave (PMW) sensors making up the GPM constellation. To create half-hour estimates, the GPROF is gridded, intercalibrated to the GPM Combined Radar Radiometer Analysis product (with GPCP climatology calibration) and combined into half-hourly 0.1^0 x 0.1^0 fields. This is done by using the Climate Prediction Center (CPC) Morphing-Kalman Filter (CMORPH-KF), quasi-Lagrangian time interpolation procedure, and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System (PERSIANN- CCS) infrared (IR) re-calibration procedure (Huffman, 2020). The CMORPH-KH Quasi-Lagrangian time interpolation procedure uses the MERRA2 and GEOS-FP vertically integrated vapor (TQV) fields in parallel, as well as the PMW and IR estimates (Huffman, 2020). IMERG system run twice in near-real time, Early multi-satellite product (4hr after observation time), and Late multi-satellite product (14-hr after observation time), and as soon as the monthly gauge analysis is received the final satellite-gauge product (3.5-months after the observation month) is generated (George J. Huffman, 2020).

2.2.1 IMERG Sensors and Observational Systems

The GPM Microwave Imager (GMI) (Figure 2.2) has a swath 550 miles (885 kilometers) wide, providing a wide view of extra-tropical cyclone observed off the coast of Japan on March 10, 2014 (NASA, 2021). The GMI instrument has 13 channels, each sensitive to different types of precipitation as shown in figure 2.1. Each channel is sensitive to a different frequency of microwave energy naturally emitted from or affected by precipitation. As depicted in figure 2.1, the five channels on the left are sensitive to heavy and moderate rainfall.

The four channels in the center captures precipitation mixtures of both snow and ice within the clouds. The mixed layers are often the result of ice or snow melting into rain as it drops. The four channels on the right are sensitive to light rain and snowfall (NASA, 2021). NASA, (2021) noted that "the multiple channels in each category ensures that GMI captures as full a range as possible of precipitation types. The data from GMI is used as a reference standard for an international network of partner satellites that also measure precipitation. Data from the GPM Core Observatory and the partner satellites is unified into a single global precipitation data set" (NASA, 2021).

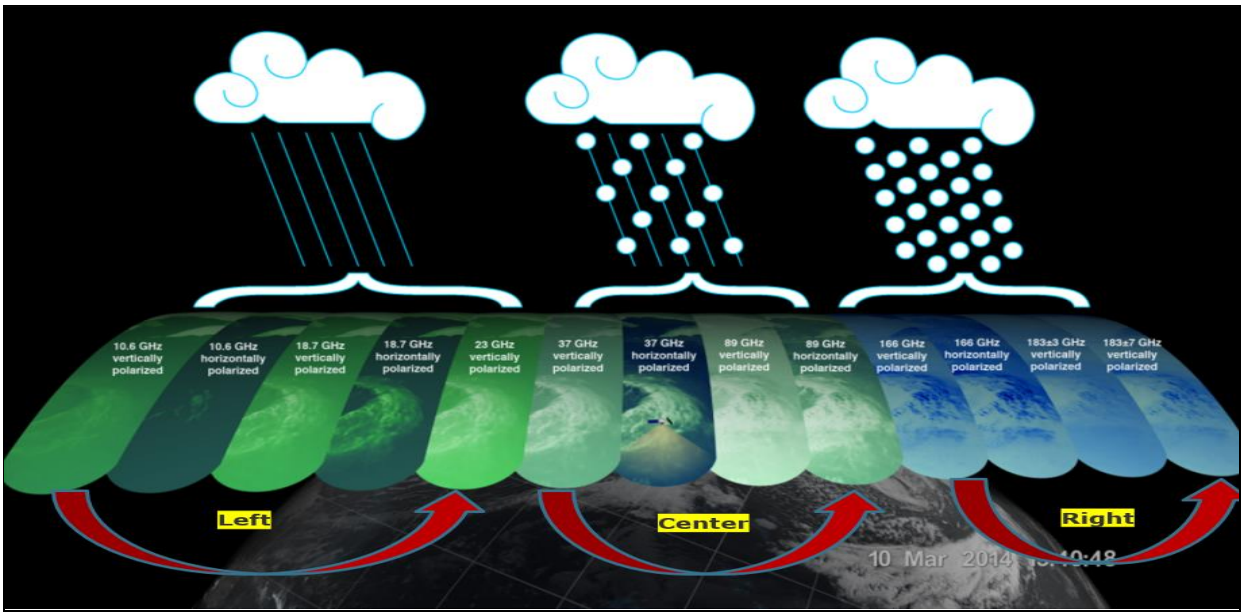

Figure 2.1. IMERG GMI showing 13 channels, each sensitive to different types of precipitation. The left channels are sensitive to heavy and moderate rains, the middle channels are sensitive to snow and ice, while the right channels are sensitive to light rain and snow fall. (Image credit: NASA).

Figure 2.1 illustrates the GMI instrument as a conically scanning multi-channel microwave radiometer, with a swath width of 550 miles (885 km) and thirteen channels ranging in frequency from 10 GHZ to 183 GHZ. NASA, (2021) noted that "the GMI uses a set of frequencies that have been optimized over the past two decades to retrieve heavy, moderate and light precipitation using the polarization difference at each channel as an indicator of the optical thickness and water content and precipitation systems".

In addition to the multi-channel GPM Microwave Imager (GMI), the GPM Core Observatory carries the first space-borne Ku/Ka band Dual-frequency Precipitation Radar (DPR) (NASA, 2021). The Ka-band precipitation radar (KaPR) of the DPR has a frequency of 35.5 GHZ, while the Ku-band precipitation radar (KuPR) operates at 13.6 GHZ. DPR provides three-dimensional estimates of precipitation structure and characteristics (fig. 2.2) (NASA, 2021). DPR originally had a swath width of 78 and 152 miles (125 and 245km) for the Ka and Ku band radars respectively prior to May 2018, but since May 2018 the swath now extends to 152 miles (245 km) for both the Ka and Ku radars (see figure 2.2) (NASA, 2021).

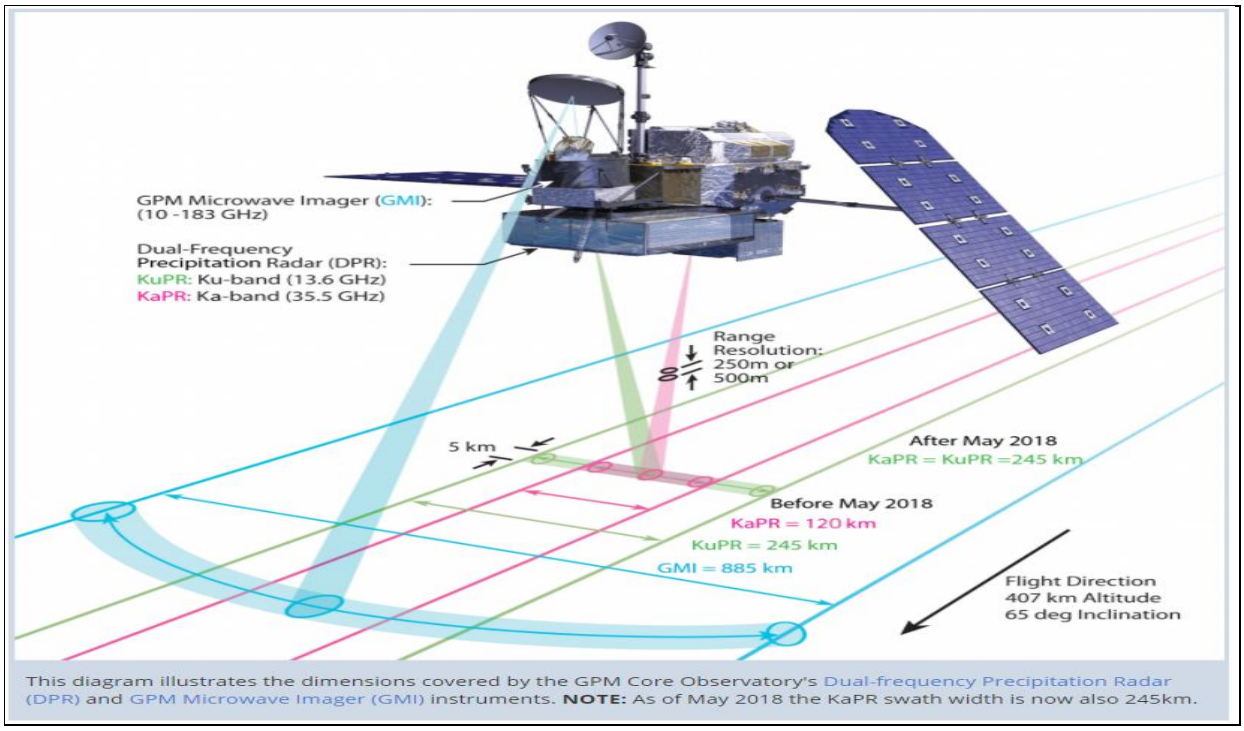

Figure 2.2. IMERG DPR showing the Ka-band and Ku-band radars and their swath width at different epoch. (Image credit: NASA).

The DPR is able to provide estimates of drop size distributions at moderate precipitation intensities (NASA, 2021). According to NASA, (2021) "the DPR is expected to provide further insights into how precipitation processes may be affected by human activities, and not limiting its capabilities in complementing microphysical measurements such as cloud and aerosol observations" (NASA, 2021).

Passive microwave (PMW) sensors provide the largest share of relatively accurate satellite-based precipitation measurements, only available from low-Earth-orbit (leo) platforms. IMERG is designed to make up for the scarce sampling available from single leo-satellites by using as many leo-satellites as possible, and then augmenting with geosynchronous-Earth-orbit (geo) infrared (IR) estimates. First, the leo-PMW data are morphed (linear interpolation following the estimated precipitation feature motion). Second, geo-IR precipitation estimates are included using a Kalman filter when the leo-PMW are infrequent. Additionally, to provide necessary regionalization and bias correction to the satellite estimates, precipitation gauge analyses are used. Not all the satellites in the IMERG constellation follows the GPM direction, therefore, IMERG uses as many satellites as possible for accurate precipitation estimation and related observations (Huffman et al., 2007).

The IMERG consists of assembled satellites that works together, such as the TRMM satellite and GPM Core Observatory which both serves as tool for calibration and evaluation of all PMW- and IR- based precipitation products combined in IMERG, providing proper comparison with all other Passive Microwave equipped leo-satellites and IR-equipped geo-satellites (Huffman et al., 2007). Multi-channel, dual-polarization PMW sensors and active scanning radars are provided in both the TRMM and GPM satellites. The GPM satellites are better than the TRMM satellites in many ways, for example, the improvements in GPM orbital inclination which was increased from 35^0 in TRMM to 65^0 (affording coverage of important additional climate zones), GPM radar was upgraded to two frequencies, adding sensitivity to

light precipitation, and "high-frequency" channels (165.5 and 183.3 GHZ) were included in the GPM Microwave Imager (GMI), which provide key information for sensing light and solid precipitation" (Huffman et al., 2007). The higher inclination for the GPM orbit reduces the radiometer and radar sampling compared to TRMM in the latitude band covered by TRMM (Huffman et al., 2007).

The GPM constellation also referred as PMW satellites consists of satellites of opportunity such as, GPM Core Observatory (owned by NASA/JAXA), Mega-Tropiques (CNES/ISRO), NOAA 18/19 (NOAA), GCOM-W1 (JAXA), DMSP F17/F18 (DoD), NOAA 20 (NOAA), MetOp A/B/C (EUMETSAT), and Suomi NPP (NASA/NOAA). The PMW satellites orbital characteristics, and operations are shown in figure 2.3. Channel selections and data policies are outside the control of NASA, apart from the GPM Core satellite (Huffman et al., 2007). For low-and mid-latitude use, the imager channels are considered best, while the sounding channels maintain same operational capability in cold and frozen-surface conditions. Three different organizations control the geo-IR satellites, by upholding the long-standing international agreements which ensured coordination of orbits and mutual aid in the event of an unexpected satellite failure. "The basic requirement is for full disk images every three hours at the major synoptic times (00, 03, …., 21 UTC)" (Huffman et al., 2007). Although piecing it together can be somewhat challenging, all satellites operators ensure this is possible by providing a great deal of imagery. These data are received as brightness temperatures (Tb) in the merged format developed at NOAA/CPC for CMORPH, and the dataset assembled at NOAA/CPC (Huffman et al., 2007).

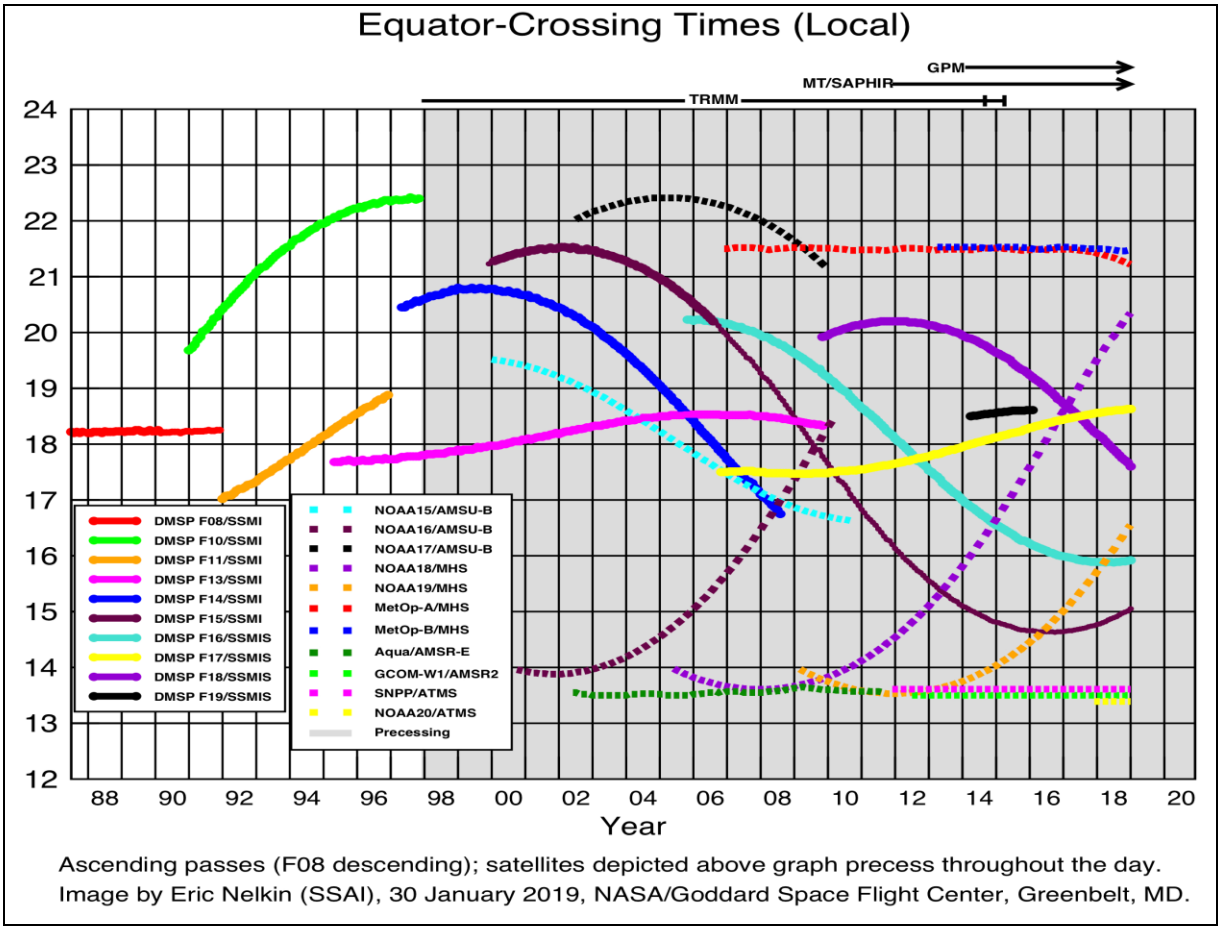

Figure 2.3. PMW sensor Equator-crossing times for 12-24 local time (LT; 00-12 LT is the same) for the modern PMW sensor era. These are all ascending passes, except F08 is descending. Shading indicates that the processing TRMM, Megha-Tropiques, and GPM cover all times of the day. Image by Eric Nelkin (SSAI; GSFC), 30 January 2019; and adapted from the NASA Algorithm Theoretical Basis Document (ATBD); *https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg holds the current version.*

Creating global precipitation products for the Global Precipitation Climatology Project (GCPC) shows that precipitation estimated from satellite soundings using the Susskind (1997) algorithm contain useful information at scales such as 1^0 daily (Huffman, 2020). However, high-frequency channels on AMSU, ATMS, MHS, GMI, and SSMIS eventually provide high-quality precipitation estimates at high latitudes, with an expectation that the cloud retrieval-based measurements may still be needed to fill gaps in the collection of high-latitude derivatives. As well, CPC has shown some improvements in using IR data from the Advanced Very High Resolution Radiometer (AVHRR) from leo- satellites for estimating precipitation at high latitudes (Huffman et al., 2007).

2.2.2 IMERG Data Fusion and Inputs from Precipitation Gauges

NASA's TRMM and GPM missions have collected rain and snow data from space for more than 22 years, and in 2019 for the first time the datasets were fused into a single dataset (NASA, 2021). The Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm combines the information from every constellation of satellites operating around earth at a given time to measure precipitation over the majority of the earth's surface (NASA, 2021). Improvement in the latest release (version 6) implies that the IMERG algorithm can now fuse the early precipitation estimates collected during past operation of the TRMM and present day GPM satellites. NASA, (2021) stated that, "the longer the record, the more valuable it is, as researchers and application developers are currently attesting. The goal is to better understand normal and extreme rain and snowfall around the world, strengthen the applications for current and feature disasters, disease monitoring, resource management, energy production, and food security". This is will help to compare and contrast past and present data (NASA, 2021).

NASA Algorithm Theoretical Basis Document (ATBD) has shown that incorporating a uniform precipitations gauges analyses is important for managing the biases that characterize satellite precipitation measurements. These observations (incorporated uniform precipitations) showed that for certain regions, monthly gauge analyses produce significant improvements (Schamm et al., 2014). The ATDB noted that, recent work at CPC proved remarkable improvements in the bias correction using daily gauge analysis for regions in which there is enough gauges (Schamm et al., 2014). The Deutscher Wetterdienst (DWD) Global Precipitation Climatology Center (GPCC) established in 1989, provides high-quality precipitation analyses over land based on conventional precipitation gauges (Schamm et al.,

2014). NASA adopts two GPCC products, the V8 Full Data Analysis for the majority of the time (1998-2016), and the V6 Monitoring Product from 2017 till date (NASA, 2021).

The monitoring product is made available two months after the time of observation and is based on surface synoptic observations (SYNOP) and monthly CLIMAT reports received in near-real time through GTS from over 80,000 stations world-wide reported through the following sources; "(1) monthly precipitation totals accumulated at GPCC from the SYNOP reports received at DWD, Offenbach, (2) monthly precipitation totals accumulated at NOAA/CPC from the SYNOP reports received at NOAA, Washington D.C., (3) monthly precipitation totals from CLIMAT reports received at DWD, Offenbach, Germany, (4) monthly precipitation totals from CLIMAT reports received at the UK Met. Office (UKMO), Exeter, UK, and (5) monthly precipitation totals from CLIMAT reports received at Japan Meteorological Agency (JMA), Tokyo, Japan" (Schamm et al., 2014). Schamm et al., (2014) noted that, the GPCC's full data analysis hinged on a data base that covers the period 1901 up to 2013 (current V7 released in late 2015). "Compared to the monitoring products, the full data analysis consists of additional data acquired from global data collections such as Global Historical Climatology Network (GHCN), Food and Agricultural Organization (FAO) of the United Nations, Climate Research Unit (CRU); datasets from the National Meteorological and/or Hydrological Services of about 190 countries of the world, and additional data from Global Energy and Water Exchanges (GEWEX)-related projects" (Schamm et al., 2014).

3 STUDY AREA AND DATA

The IMERG PMP evaluation were carried out on 55 rain gauges sites in Kansas. Figure 3.1 shows the map of Kansas with the 55 gauge sites spatially mapped, the county boundaries, and the average annual precipitation pattern unique to Kansas. Kansas is a mid-western state in the US, located at 38°30′N, 98°W, with an area coverage of 213,100 km^2. Average precipitation pattern in Kansas is such that it increases from north-west to south-east, with an hourly rate of 0.05 mm/hr in the west to 0.15 mm/hr in the eastern part of Kansas. Evaluation were carried out using IMERG estimated PMPs against those from NOAA-Atlas-14, which aligns with those 55 rain gauge sites.

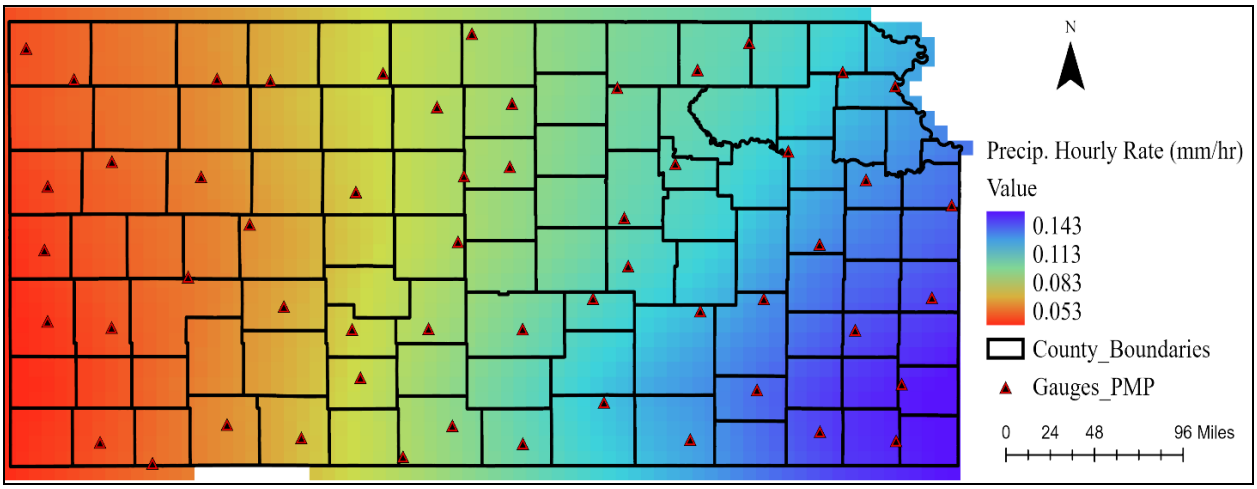

Figure 3.1. Rain gauge locations used for evaluating IMERG derived PMP, county boundaries in Kansas, and the hourly precipitation rate (mm/hr) between 2000 and 2020.

3.1 Satellite Precipitation Dataset

The precipitation dataset used in this study is the Merged Satellite-gauge precipitation estimate (Final Run- GPM_3IMERG V06), with a remarkable stretch of 65°N/S of the equator. This dataset has a spatial resolution of 0.1°, a temporal resolution of a half-hour interval, and the precipitation estimated recorded in mm. Google Earth Engine (GEE) is the platform used to extract the IMERG precipitation estimates before computing their PMPs. Extracting precipitation on IMERG can be done using either GIOVANNI or GEE. Geospatial Interactive Online

Visualization And Analysis Infrastructure (Giovanni) is a Web based application developed by the Goddard Earth Sciences Data and Information Services Center (GESDISC), for visualizing, analyzing, and accessing vast amounts of earth science remote sensing data without having to download the data (although data downloads are also supported) (NASA, 2021). GIOVANNI data visualization capability extends inter-annual time series of maximum precipitation from 2000 to present date (Fig. 3.2) for every gauge or watershed within data coverage. The challenge with GIOVANNI is the lack of flexibility for extracting the maximum precipitation with user defined time windows. It only provides the maximum precipitation for the time periods of 1 hr., 24 hr., and 30 days. As such, users must rely on interpolation to fill the durations not provided by GIOVANNI. GIOVANNI data can be accessed at https://giovanni.gsfc.nasa.gov/giovanni/.

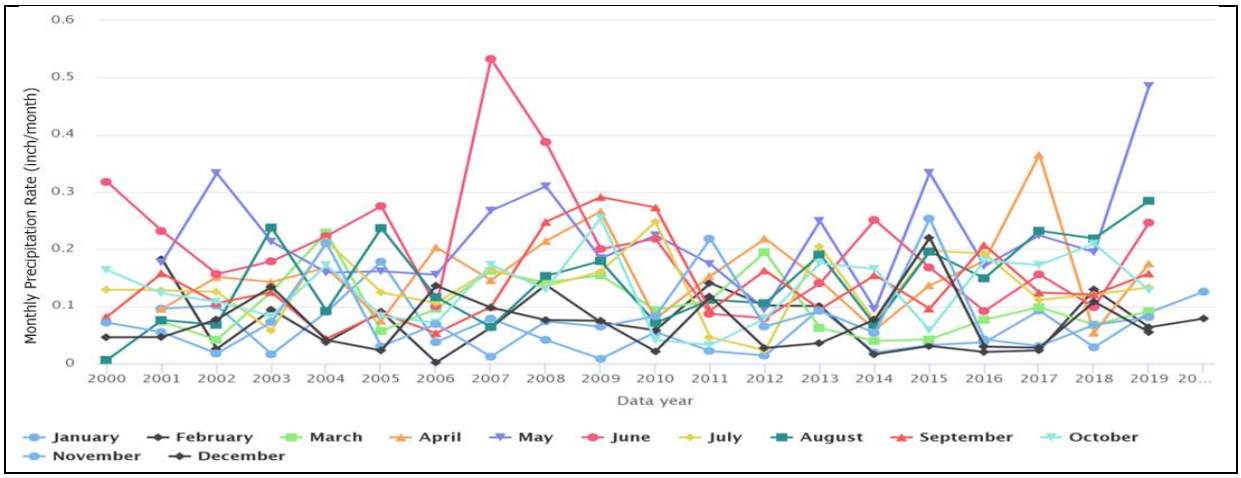

Figure 3.2. Average monthly precipitation (inch/month) for individual month in Kansas (2000-2020). Adapted from NASA GIOVANNI website: https://giovanni.gsfc.nasa.gov/giovanni/

GEE on the other hand is a cloud geospatial processing service, which can perform geospatial processing at scale, and is powered by Google Cloud Platform. GEE provides an interactive platform for geospatial algorithm development at scale, enabling high impact, data driven science (Gorelick et al., 2017). GEE uses either python or Java scripts to extract IMERG precipitation data at different durations and performs the analysis on the cloud (Mab et al., 2019). GEE's ability to analyze global data rapidly lends itself to being a useful tool for data analysis and visualization, by using JavaScript and Python Application Programming Interfaces (APIs).

This APIs allows user to develop scripts to access datasets in the cloud. This research uses the Java script built on GEE to extract IMERG precipitation estimates from GEE repository.

3.2 Rain Gauge Based PMP

The NOAA-Atlas-14 dataset is the reference data used for this research. The datasets are PMPs, which are derived using rain gauge data by the National Oceanic Atmospheric Administration (NOAA), prior to making the NOAA-Atlas-14 available for public use. The NOAA Precipitation Frequency Data Server (PFDS, https://hdsc.nws.noaa.gov/hdsc/pfds/index.html) provides the point and click interface developed to deliver NOAA-Atlas-14 precipitation frequency estimates and associated information. The PFDS provides estimates of PMPs for points, which can be extrapolated to derive PMPs for other locations within the NOAA coverage. PMP estimates from NOAA-PFDS can be obtained directly in table format or graphs.

Table 3.1 shows the differences in specifications and characteristics between the NOAA station and the IMERG satellite data. Some of the highlighted characteristics include, the spatial resolution, length of data coverage, sensors (recording devices), area coverage, calibration method, ownership, amongst others.

Table 3.1. Differences between IMERG satellite data and NOAA station data

Characteristics	**NOAA Station Data**	**IMERG Satellite Data**
Spatial Resolution	55 km (based on average distance between gauges in the study)	0.1^0 (11.1 km)
Data coverage	1857 – present (Gauge site varies)	2000- 2020
Sensor(s)	Rain gauges	GPMI (GPM Microwave Imager) DPR (Dual Precipitation Radar)
Area coverage	United States territories	65^0N/S (semi-global)
Calibration method	Observed estimates.	● TRMM & TMPA calibrated ● GPCC Monthly adjusted
Ownership	NOAA-USA	NASA & JAXA

4 RESEARCH METHOD

4.1 Procedure for Calculating PMPs

Computing PMPs from IMERG requires first extracting IMERG maximum precipitation, for different durations using GEE Java scripts at defined gauges. The output result is then imported into a python/R environment using additional script with built in statistical procedures to estimate the PMPs. The IMERG computed PMPs was then paired with NOAA-Atlas-14 using series of sorting techniques in Microsoft excel for proper evaluation purposes. Figure 4.1. below shows the workflow for extracting maximum precipitation, computing PMP, and the statistical evaluation in the study.

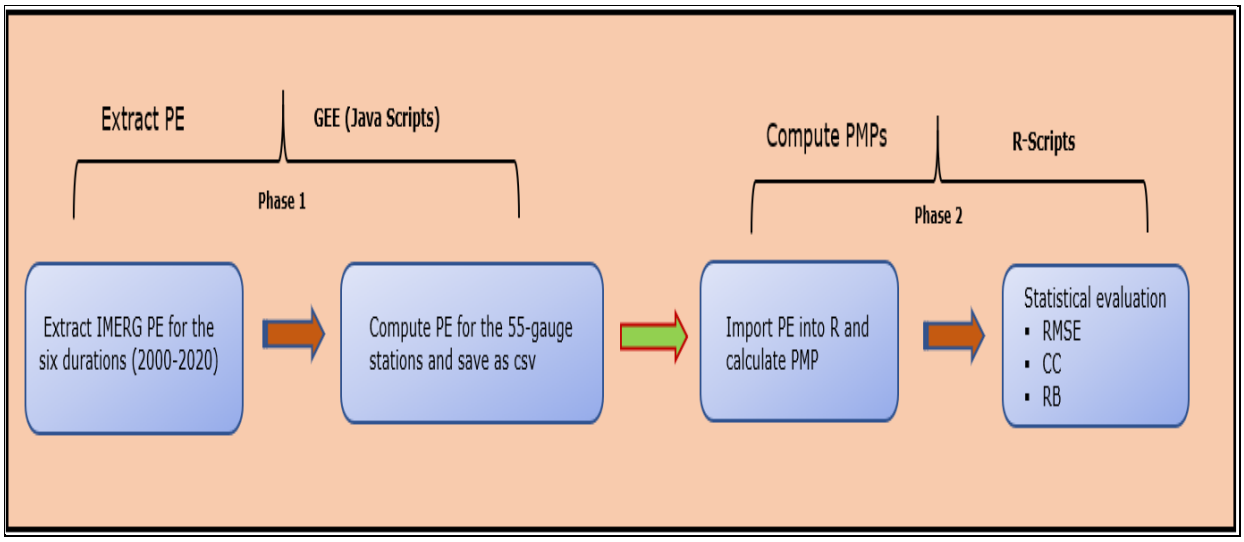

Figure 4.1. Workflow for computing PMPs from IMERG data

4.2 Calculate Maximum Precipitation Estimates

The maximum precipitation for a given time-step was calculated for each pixel from the half-hour IMERG precipitation data using Google Earth Engine (GEE). The half-hour IMERG data for a given month were summed defined by a moving temporal window, the length of which includes the PMP return periods (30-min, 1-hr, 2-hr, 6-hr, 12-hr, and 24-hr), and the maximum value from that return period is treated as the maximum depth of precipitation needed to compute the PMP for that return period. The draw back to this approach is that provisional data contains missing

values due to IMERG technical issues and must be avoided (Tan & Santo, 2018), calibration and adjustment of IMERG data takes 1-3 months before the IMERG final product is made available to the public for research purposes. The advantage of using GEE to extract precipitation estimates is that, user can define specific window intervals for extracting the precipitation, which is not the case as seen in GIOVANNI. The Java script used for extracting IMERG data is available at https://code.earthengine.google.com/70fa7105be38bb993dc9e1b765f736b5?noload=1.

4.3 Compute PMPs

A second script was written using the R-programming language to automate the computation of PMPs from the maximum precipitation for various return periods. Hershfield statistical technique is the adopted approach used for the PMP computation. This technique relies on the mean of the durations, their standard deviation, and their frequency factor concerning maximum precipitation estimates for individual durations. Hershfield has been duly endorsed by the World Meteorological Organization (WMO) and widely used by researchers. One key benefits of the Hershfield technique is that, it allows PMP to be adjustable by a factor ranging from 1.13 up to 2.9, hence a good approach for calibrating data from satellite sensors (World Meteorological Organization, 2009).

The approach involves first computing the frequency factor (K_m) and then calculating PMPs. The frequency factor is a function of the largest value in the series of duration intervals, it is also dependent on the number of observations in the series. The frequency factor was calculated using the equation 1stated below:

$$K_m = \frac{Rmax - (\frac{1}{n}\sum_{i=1}^{n}(R_{n-1}))}{S_{n-1}} \qquad (1)$$

where,

K_m is the frequency factor of the series,

R_{max} is highest value in the series of individual duration intervals,

R_{n-1} is mean excluding the largest value in the series, and

S_{n-1} is standard deviation excluding the largest value in the series.

The PMPs were derived using equation 2 outlined below:

$$X_T = \left(\frac{1}{n}\sum_{i=1}^{n} X\right) + K_m \times S_n \quad (2)$$

where,

X_T is estimated PMP value of the duration interval,

X is the Mean of the frequency,

K_m is the frequency factor which depends upon the number of observations, and

S_n is standard deviation of the series.

4.4 Statistical metrics for Evaluating IMERG derived PMPs

The three statistical metrics used for the evaluating IMERG derived PMPs include, the coefficient of correlation (CC), root mean square error (RMSE) and relative bias (RB). A third linear regression model was fitted with a line of best fit, to ascertain the predictability of IMERG derived PMP from station PMPs at different time scales. The linear regression model was also used to ascertain the degree of linear relationships between IMERG PMPs and station PMPs at varying temporal scale. Table 4.1 below shows the formula of the three statistical metrics used for the evaluation. The statistical evaluation, data comparison, and plotting were done using R-scripts. The study further applied the concepts: p-values, R-Squared, and slope to ascertain the relationship between the PMP error and mean annual total accumulation, as well as PMP error and mean maximum precipitation at different precipitation intervals. Basically, the idea was to understand the error pattern with regards to durations, and precipitation regions. Finally, we applied the concept of trend analyses, helping to interpolate our error on the surface.

Table 4.1: Statistical metrics for evaluating IMERG derived PMPs.

Metric/symbol	Formula
Correlation Coefficient (CC)	$CC = \frac{\sum_{i=1}^{n}(O_i-\bar{O})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i-\bar{O})^2}\cdot\sqrt{\sum_{i=1}^{n}(S_i-\bar{S})^2}}$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2}$
Relative Bias (RB)	$RB = \frac{\sum_{i=1}^{n}(S_i-O_i)}{\sum_{i=1}^{n}O_i}(100)$

S is IMERG PMP, O is NOAA-Atlas-14 PMP, and n is the PMP interval period.

CC is a unitless statistical metric used to determine the degree of linear agreement between IMERG estimated PMPs and NOAA PMPs, with values ranging from -1 to +1. A CC value of 0 indicates there is no correlations. On the other hand, a CC value of +1 and -1 shows perfect positive and negative correlations, respectively. RMSE measures the average absolute error magnitude of the IMERG PMP. The smaller the RMSE value, the closer the IMERG PMP to the station PMP. Statistically, RMSE assumes the unit of the variable measured (which in this case is mm for precipitation) for the evaluation. Positive values of RB indicate an overestimation of the IMERG PMP, while negative values show an underestimation (Tan & Santo, 2018).

5 EVALUATION OF IMERG PMP AND PMP ERRORS

An initial IMERG PMP assessment was conducted by using precipitation intervals as the criteria for assessment. The study aggregated PMPs from all 55 rain gauge sites and analyzed the correlation between IMERG derived PMP and NOAA station PMP at different intervals. Precipitation zones were not considered at this stage of assessment, instead the idea was to evaluate the relationship between IMERG derived PMP and NOAA station PMP, assess the PMP error at different precipitation durations, and to find any pattern between the PMP error at intervals and precipitation patterns in Kansas. In addition to the CC, RMSE, and RB for evaluating IMERG derived PMP and associated error at intervals, the linear regression model was used to confirm results from CC and justify any statistical relationship between IMERG PMP and NOAA PMP.

Result from the figure 5.1 showed that IMERG PMP recorded the highest CC at the 30-mins interval with a value of 0.95 and returned the least value of 0.85 at the 24-hr (fig.5.1a). RMSE increased with every increment in durations (fig.5.1b), an estimate of 4.41 mm of RMSE at the 30-min duration and 10.28 mm of RMSE at the 24-hr duration. In figure 5.1c, -3.77% of RB was calculated for the 30-min interval depicting underestimation, while increasing to 7.82% at the 24-hr interval to infer an overestimation. Figure 4 also showed that the non-linearity in the pattern of CC, RMSE, and RB. The early intervals (30-mins – 2-hrs) saw a sharp decline in CC compared to the rest of the intervals. Similarly, RB declined during same interval prior to peaking with consistent trajectory. On the other hand, RMSE peaked sharply at the early intervals prior to maintaining a gentler increase in direction. This unique behavior at the short intervals only confirm the unreliability of PMP estimated at shorter intervals as residual errors at individual gauges were not evaluated at this stage of analysis.

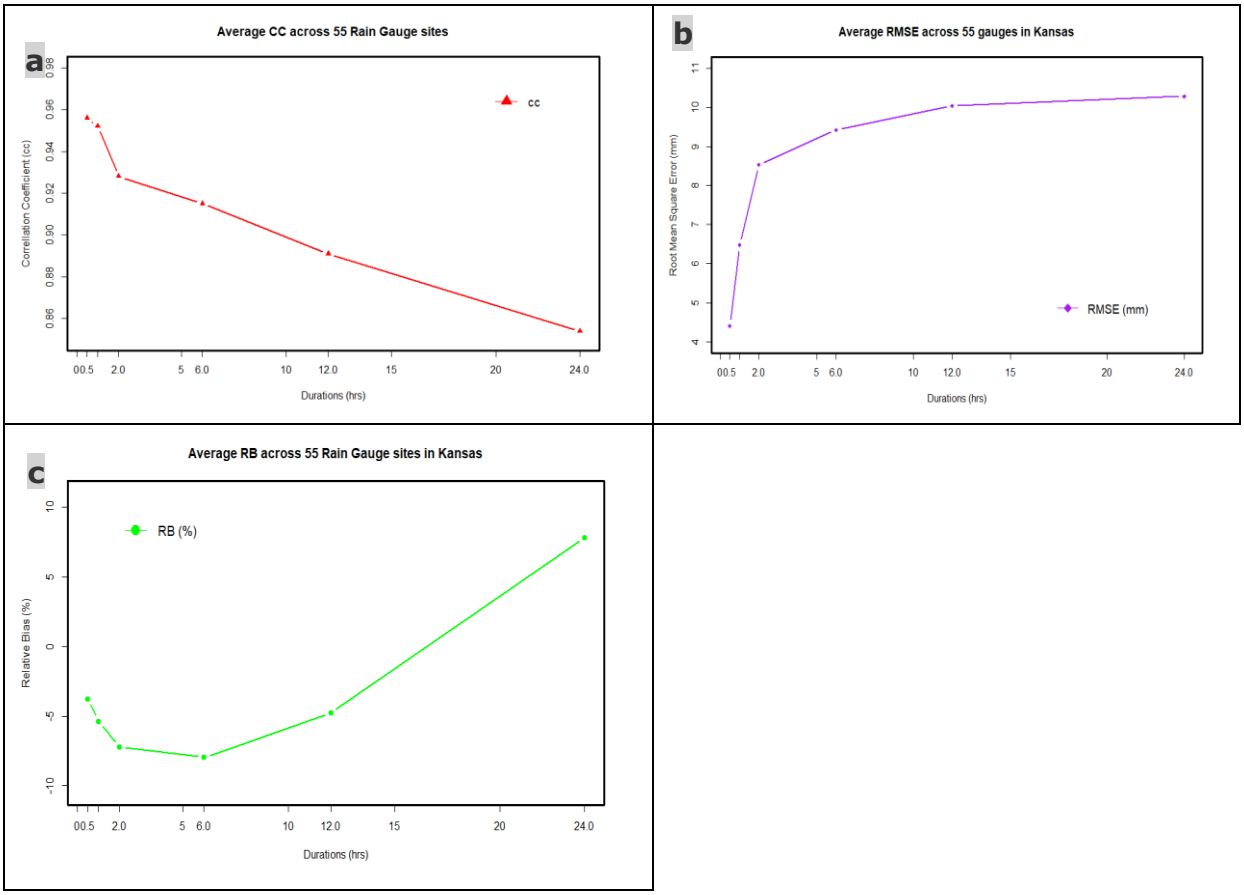

Figure 5.1. Plots showing Statistical results from the evaluation of IMERG derived PMPs at different durations. (a) coefficient of Correlation (b) the root mean square error in mm(c) the relative bias in %. The X-axes shows the durations while the Y-axes represents the statistical metric used.

Figure 5.1a implies that IMERG PMP had better correlation at the shorter durations compared to the longer periods, when assessment is focused on the durations and precipitation regions are ignored. An increase in RMSE as precipitation intervals increases (figure 5.1b) implied that there was more precipitation measured at longer intervals than the shorter intervals. Although the RMSE result were useful to measure errors at the intervals, they did not account for PMP errors with respect to precipitation pattern or zones. Findings in figure 5.1b were consistent with Tan & Santo, (2018). Result from the relative bias assessment (figure 5.1c) agrees with Yang et. al., (2017) whose RB assessment of IMERG fell between -10% and 10%, confirming the potential of IMERG to underestimates at lower intervals, while overestimating at higher intervals. Although, the statistical assessment of

PMP at different durations failed to account for spatial distribution of the gauges as well as the precipitation regions (dry, mid, and wet regions), the analysis provided an insight on how PMP at different interval could be used for future study when precipitation pattern do not matter.

Table 5.1 shows the summary of the results obtained from the statistical analysis conducted using the adopted statistical metrics. Similar to earlier assessment conducted, the precipitation regions were neglected, and the study treated the gauges as a unit precipitation zone. However, the results from the regression analyses confirm earlier observations made in the study (fig. 5.1). At the 30-min duration CC, RMSE, and RB were calculated 0.956, 4.41mm, and -7.96 respectively, while the 24-hr duration recorded 0.854, 10.28 mm, and 7.82% for CC, RMSE, and RB respectively. The fitted linear regression model (fig.5) showed a decreasing R-squared with increasing durations. The coefficient of determination was 91.4% at the 30-min interval, but then decreased to 81.3% at the 24-hr duration. Similarly, the p-value (at < 0.05 significance) confirmed a more linear relationship did exists at the 30-mins interval (0.0011) compared to the 24-hr interval (figure 5.2).

Table 5.1: Statistical results from IMERG PMP assessment at different durations

Duration (hr)	Statistical Metrics				
	CC	**RMSE (mm)**	**RB (%)**	**P-value**	**R-Squared**
0.5	0.956	4.41	-3.77	0.00111	0.914
1	0.952	6.48	-5.379	0.00147	0.904
2	0.928	8.53	-7.228	0.00126	0.859
6	0.915	9.42	-7.96	0.00103	0.838
12	0.908	10.04	-4.77	0.00757	0.821
24	0.854	10.28	7.82	0.0086	0.813

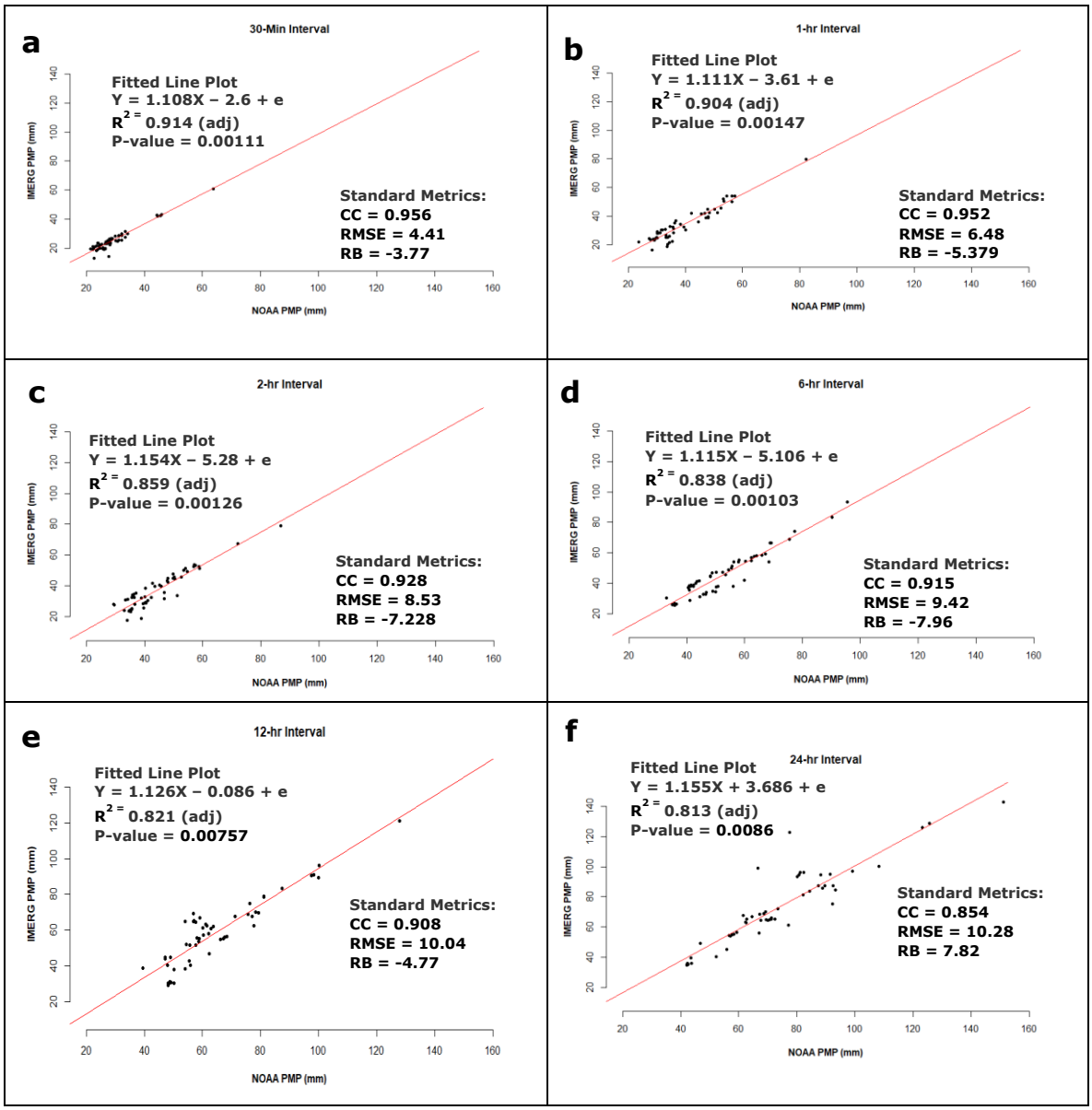

Figure 5.2. Shows the relationship between IMERG derived PMP and NOAA station PMP at different durations. Note that the y-axis represents the IMERG derived PMPs while the x-axis represents NOAA PMP in mm. (a-f) represents the duration intervals; 30-mins, 1-hr, 2-hr, 6-hr, 12-hr, and 24-hr. the red line is the fitted line (line of best fit) across the intervals. R-squared is highest in shorter durations signifying high correlation but decreases with longer durations. The p-value across a-f suggest a high relationship exists between IMERG derived PMP and NOAA PMPs. e is the residual error.

Table 5.1 and figure 5.2 further justifies observed relationship between IMERG derived PMPs and NOAA station PMPs at different intervals of duration. At shorter interval (e.g. 30min) the least precipitation was estimated, CC was highest, RMSE was lowest, whereas RB returned a negative value. However, at the longer interval (e.g. 24hr) CC was lowest,

RMSE were larger, while RB returned negative values all through. The regression model showed higher correlation between IMERG PMP and NOAA PMP at the lower interval with a coefficient of determination of 91.4%, inferring higher chances of predicting IMERG PMP using those from station. Analysis at intervals provided an insight on the characteristics of IMERG PMP when the gauge station were aggregated and analyzed as a unit, it is important to note here that, averaging PMPs from all 55 gauges has a tendency of creating a smoothening effect that concealed impacts of bias in the analyses. This part of the study analyses proved that; we may not know which exact gauge is exerting the most influence on our analyses due to the smoothening effects from the PMPs aggregation at different intervals. Further observation from the interval assessment shows that, aggregating the PMPs at all intervals resulted to underestimation of PMPs at all intervals, which is consistent with observation from Yang et. al., (2017).

5.1 Spatial Variation of PMP Errors

To better understand how the PMP errors varies spatially, a spatial interpolation technique was adopted to evaluate the magnitude of the residual error at individual gauges in relation to precipitation zones. Unlike the assessment by return periods which aggregated PMPs in all gauges by their intervals for analyses, the PMP error assessment at the gauges treated every rain gauge independently by analyzing the residual errors between IMERG computed PMPs and those from NOAA PMP. The idea was to ascertain the pattern of residual errors distribution from the PMP differences in those gauges with regards to precipitation regions (wetness). In a nutshell, the study tried to determine where the most and the least residual errors occurred with regards to precipitation pattern in Kansas. This assessment helped to address lingering concerns such as, whether the PMP error has any relationship with precipitation pattern in the study area.

The study extrapolated the residual errors from the PMPs on the surface, to determine any relationship or pattern between the residual error and the precipitation zones. The analyses were conducted using Trend (3D Analyst) tool, which explicitly uses a global polynomial interpolation that fits a smooth surface and defined by a mathematical function (a polynomial) to the inputs rain gauge sample points. As the trend technique interpolates a raster surface from points, the trend surface changes gradually and captures more coarse-scale patterns in the data. When the order of the polynomials was increased, the fitted surface becomes progressively more complex. While the order of polynomial being an integer between 1 and 12, a value of 1 fit a flat plane to the points, and higher values fits more complex surface. A polynomial number of 2 were used for the analyses as the study area fits a gentle slope criterion.

Figure 5.3(a-f) illustrates the error distribution across different precipitation durations. A close observation showed that, at the shorter intervals (e.g., 30-min, 1-hr, 2-hr, & 6-hr) (figure 5.3a-d) surface error was largely distributed between the dry and mid regions, with patches in the eastern part (wet region), showing less consistency with precipitation pattern in Kansas. However, the 12-hr and 2-4hr interval (fig. 5.3e-f), correctly represented precipitation pattern in Kansas. The most residual surface errors were recorded in the dry region (western Kansas) while the least PMP errors were found to be recorded in the eastern Kansas. These results (fig. 5.3f) simply connotes in a single view how the errors are distributed in relation to precipitation intervals and precipitation regions.

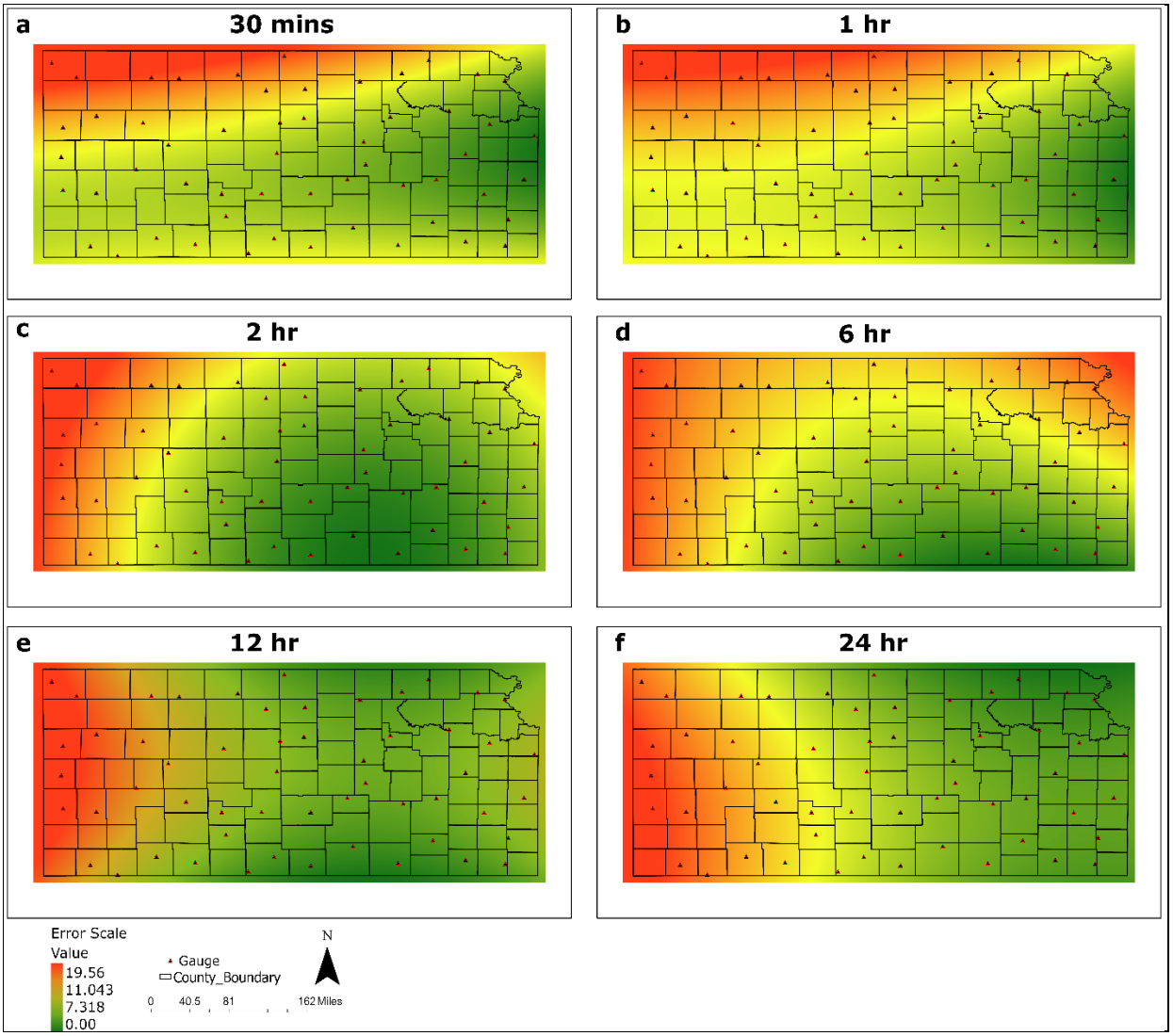

Figure 5.3. Interpolated residual errors at the surface. a-f shows the error pattern at the durations 30-mins, 1-hr, 2-hr, 6-hr, 12-hr, and 24-hr respectively.

The result from the trend analyses (figure 5.3) showed that the longer the precipitation interval (fig. 5.3f) the better the PMP estimation. Similarly, the wetter the region the lesser the error from IMERG PMP. Results from the trend assessment were consistent with observations made by Tan & Santo, (2018). This implies that, for any application of IMERG PMP for further studies (such as flood modeling), Computed PMPs at the lower intervals must be treated with caution, as the level of confidences is lower at lower interval but higher at longer interval, further suggesting preference for PMP at longer intervals (e.g. 24-hr). The result from the trend assessment further justify earlier study by Su, et al., (2019), proving that IMERG derived PMPs have higher confidence and reliance in wet regions than the dry

regions. Su, et al., (2019) found the IMERG to record less errors and higher correlation in wetter areas. The study assessment so far has shown that PMPs are both interval dependent and precipitation region dependent and great care should be taken to accommodate both variables for a better PMP estimation and assessment. The result also showed that analysis at the intervals alone can be misleading and fraught with unreliable result as gauges within same interval may be from precipitation zones with differing degree of wetness.

The study further explored the derived IMERG PMP and the station confidence interval to determine if the PMPs at the different durations were within the 90% confidence interval of NOAA station PMP (figure 5.4). Result showed that at the shorter intervals (30-min up to the 12-hr durations) IMERG derived PMPs fell outside the 90% confidence intervals of the NOAA of station PMP. However, at the 24-hr duration IMERG derived PMP fell within the upper and lower limits (77.391 mm and 67.309 mm) of NOAA 90% confidence interval. This observation further justifies those made earlier in figure 5.3, that IMERG estimates better PMP at the 24-hr durations with the least error measured.

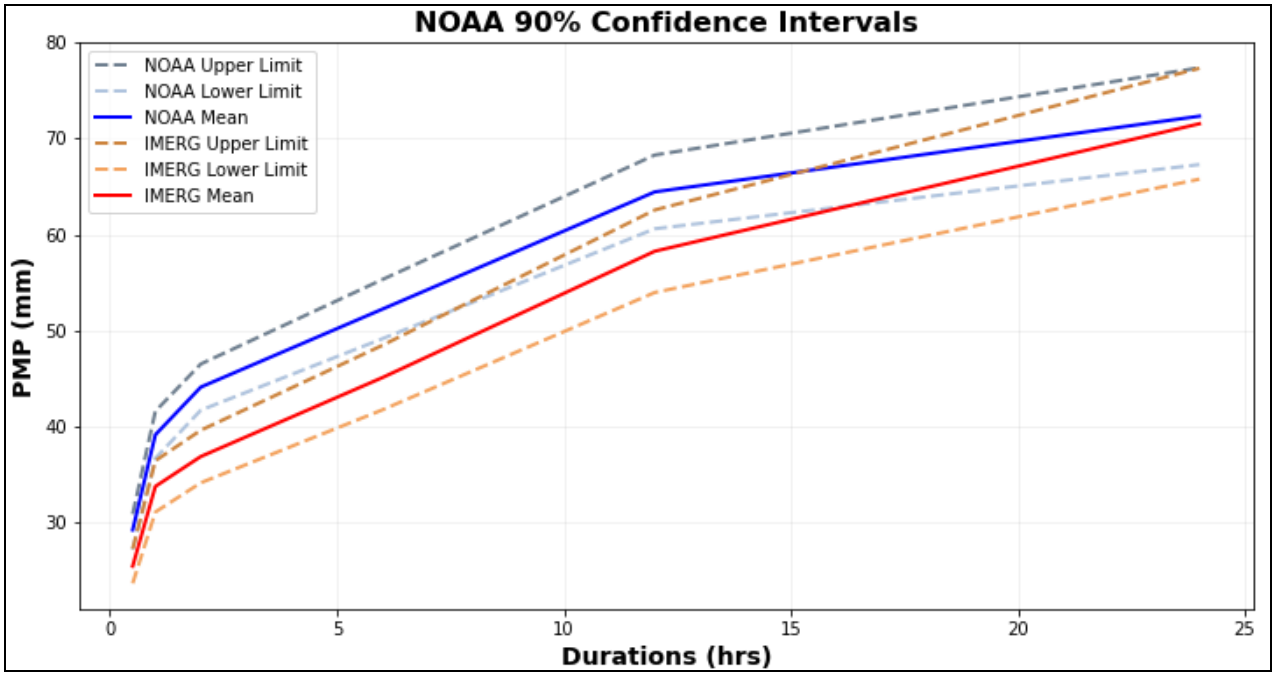

Figure 5.4. Confidence interval assessment of IMERG PMP at different durations

6 PMP ERROR AND PRECIPITATION AMOUNT

The study furthered the analysis by exploring the possible relationship between PMP error and precipitation amount within precipitation zones and at individual stations. The basic idea was to determine if the PMPs estimation were impacted by the errors (as we see in figure 5.3).

6.1 PMP Error and Precipitation Zones

Firstly, precipitation criteria were used to zone the study area into 3 regions (dry, mid, and wet regions) (see figure 6.1d). Similarly, zoning was done using the PMP at the 24-hr duration as a criterion. Both zoning criteria produced similar number of gauges in each of the regions, hence the study adopted the zoning by mean annual precipitation. The three statistical metrics (CC, RMSE, and RB) were then applied in evaluating the relationship between IMERG derived PMPs using NOAA station PMP for all three regions.

IMERG PMP at the dry region (fig. 6.1a) were consistent at the 30-mins and underestimates slightly as duration increases up to the 24-hr interval. At the Mid region, IMERG PMP were underestimated at the 30-mins interval and up to the 12-hr, although remained consistent with NOAA PMP between the 12-hr and the 24-hr with slight overestimation (fig. 6.1b). In wet regions (fig. 6.1c), IMERG underestimated PMP at the 30min duration but overestimates PMP at the 24-hr. The result showed the inconsistencies associated with IMERG PMP at the various precipitation regions. Aggregating PMP was noted to create a smoothening effect as local influences from gauges were not accounted.

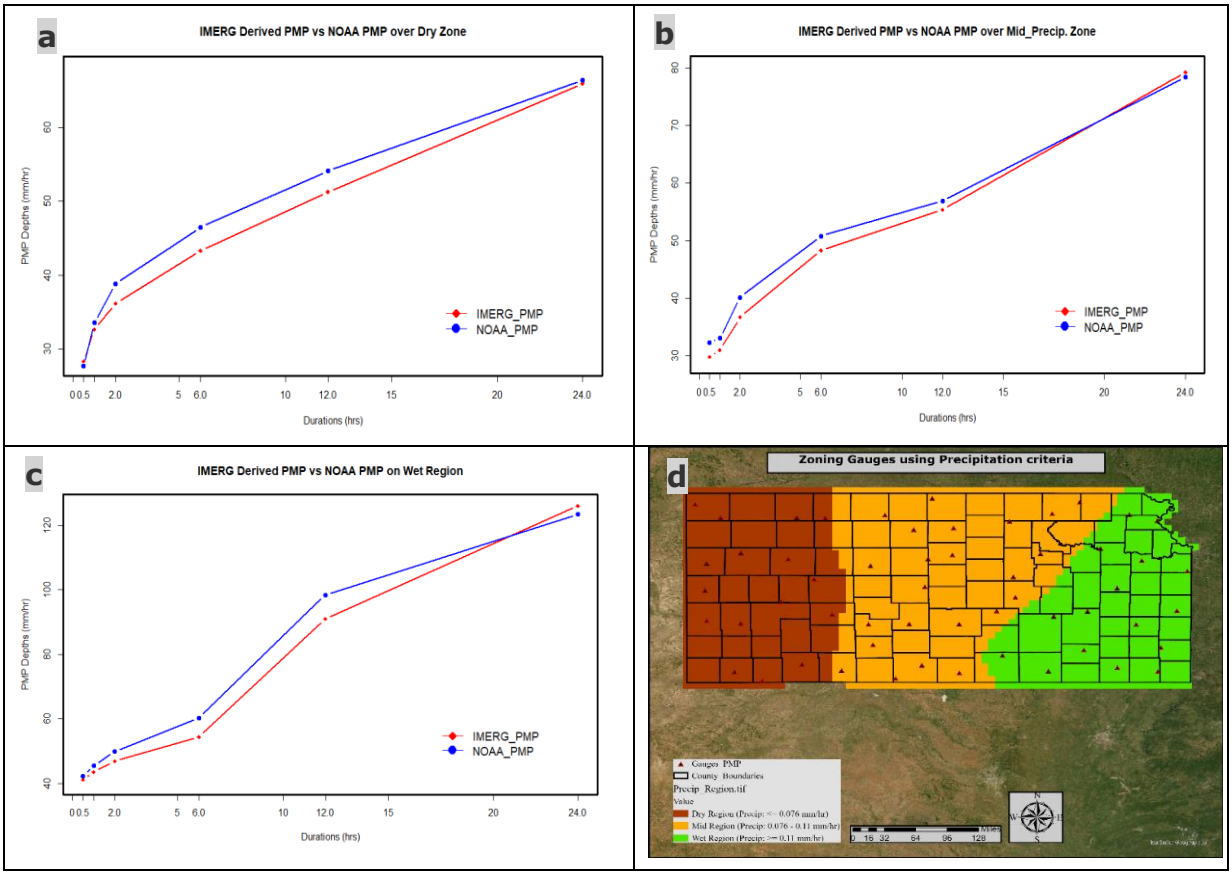

Figure 6.1 Visual assessment of IMERG and NOAA PMP by zones (a) dry region, (b) mid region, (c) wet region, (d) major regions. The vertical axes represent the PMP depths why the horizontal axes are the durations (a-c).

The influence of precipitation zones on PMP was clearly depicted in figure 6.1, as subtle differences in the level of IMERG PMP estimation were revealed. Although in comparing PMP at the three zones, the result showed a distinct pattern of correlation moving from dry to wet regions, but we may not be able to tell how much of local influences from the gauges impacted our analyses, as the study showed in figure 5.3. One thing is certain, averaging all gauge PMP at the zonal assessment created a smoothening effect that belies the residual errors from individual gauges. Result of Figure 6.1 validates findings made by Yang et.al., (2017); that IMERG has the capability of underestimating PMP at the dry regions while overestimating PMP in wet regions.

Table 6.1 showed the result from the assessment of IMERG computed PMP at the zones using the established statistical metrics. The analysis accounted for both precipitation

durations (intervals) and precipitation regions (zones). In detail, CC decreased from 0.871 at the 30-mins interval to 0.733 at the 24-hr interval for the dry region. At the mid region, CC decreased from 0.901 at the 30-min mark to 0.803 for the 24-hr interval, whereas the wet region saw highest CC value of 0.951 at the 30-min interval and 0.821 at the 24-hr interval. An assessment of the RMSE showed that highest values were at the dry region, with a value of 4.986 mm at the 30-mins interval, and 10.51 mm at the 24-hr interval, the mid region recorded 4.26 mm and 10.01 mm for the 30-min and 24-hr intervals respectively (table 6.1). The result further showed that the least RMSE were found at the wet region, with both 30-min and 24-hr intervals returning 3.75 mm and 9.45 mm respectively (table 6.1). RB was calculated to be -10.23% and -0.912% for both 30-min and 24-hr intervals respectively for the dry region. RB at both mid and wet regions were (-9.58% and 4.2%) and (-8.14% and 9.35%) respectively for the 30-min and 24-hr intervals. The negative RB that characterized the dry region connotes underestimation of PMP by the IMERG, while the Positive RB at the wet region represents overestimation in the PMP.

Table 6.1: Results of IMERG PMP evaluation at the zones

Metrics	**CC**			**RMSE (mm)**			**RB (%)**		
Durations (hr) /Regions	**Dry**	**Mid**	**Wet**	**Dry**	**Mid**	**Wet**	**Dry**	**Mid**	**Wet**
0.5	0.871	0.901	0.951	4.986	4.26	3.75	-10.23	-9.58	-8.14
1	0.845	0.881	0.93	6.21	5.85	5.58	-8.395	-7.67	-4.983
2	0.815	0.862	0.91	7.14	6.82	6.18	-5.231	-4.57	-2.13
6	0.781	0.841	0.88	8.492	8.01	7.63	-3.35	-2.89	1.256
12	0.752	0.822	0.851	9.243	8.75	8.34	-2.011	0.12	4.121
24	0.733	0.803	0.821	10.51	10.01	9.45	-0.912	4.2	9.35

The statistical result across the dry, mid, and wet regions (figure 6.2), revealed interesting pattern similar to our result in table 6.1. CC were highest in wet region, while the dry region recorded the least CC (fig. 6.2a). The most RMSE were recorded in dry areas whereas wet

region saw the least RMSE (fig. 6.2b). The effect of relative bias was mostly felt in wet areas (underestimation at shorter intervals and underestimation at longer intervals) whereas dry zone had the least impact from RB (fig. 6.2c). This observation is consistent with Yang et. al., (2017) and Tan & Santo (2018). Results discussed so far inferred that IMERG has the tendency to underestimate PMP at shorter durations and dry regions, while overestimating PMP at longer durations and wetter regions, producing consistent pattern as seen in Yang et al., (2017).

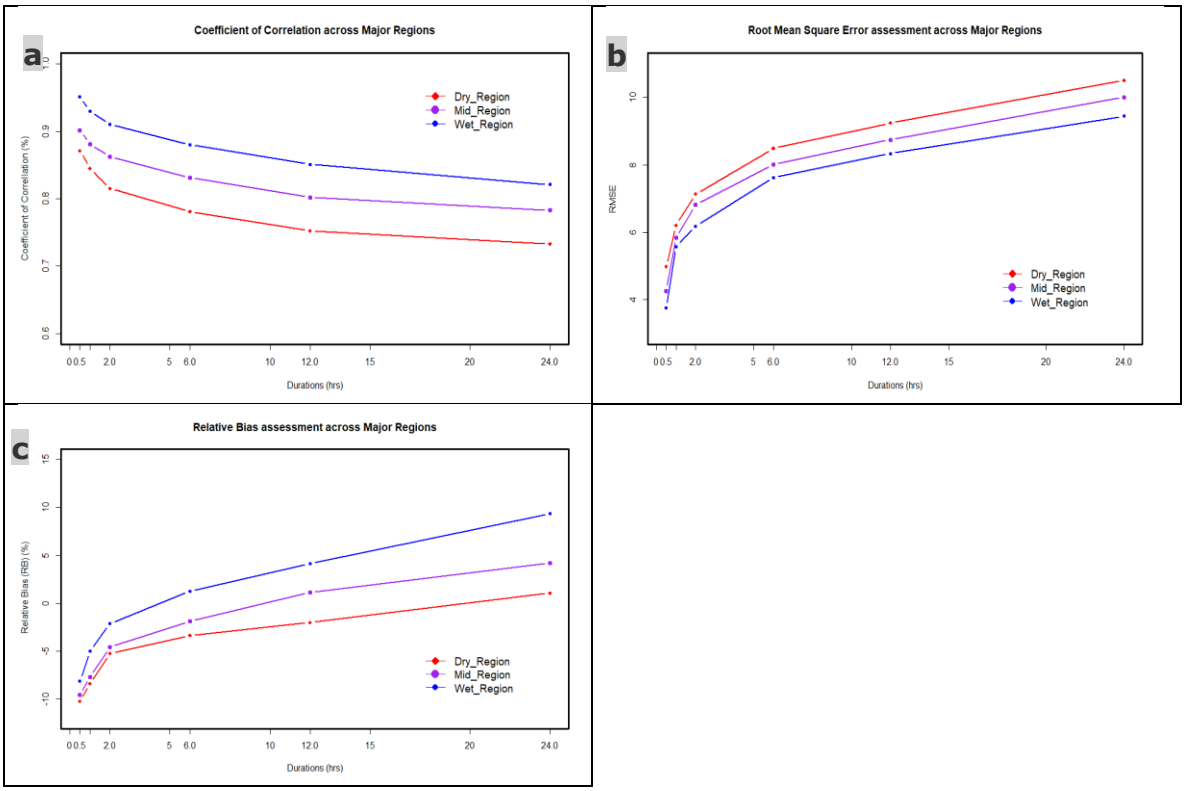

Figure 6.2. Results of IMERG PMPs evaluation at three zones (dry, mid, and wet) for different durations. (a) coefficient of correlation (b) root mean square error in mm (c) relative bias in %. The horizontal axes show the durations while the vertical axes present the statistical metrics.

Statistical evaluation of IMERG PMP at different precipitation zones (see table 6.1 and figure 6.2) showed that calculated CC ranged from 0.733 in dry region to 0.951 in wet areas, and were consistent with Tan & Santo (2018), whose CC result from earlier study stood at 0.76. Tan and Santo (2018) showed that CC range between 0.7 to 1 were statistically

significant for research, thus inferring that the relationship between IMERG computed PMP and NOAA PMP had a strong relationship in wetter regions. The study results also showed IMERG PMP had lesser error in wet region (3.75 mm at 30mins) than the dry region (4.99 mm at 30mins), however an assessment of RMSE in relation to durations proved that the highest RMSE were calculated at the longer intervals. The calculated RMSE with regards to precipitation zones agree with findings from Yang et.al., (2017) suggesting that IMERG PMP were better adapted at the longer durations with lesser errors. Similarly, an assessment of RMSE with regards to precipitation intervals were found to be consistent with previous findings in figure 5.3, suggesting that IMERG were better for estimating PMP at longer intervals.

The calculated RB (figure 6.2c) fell within the range -10% to 10%. The RB range were found to lie withing the significant threshold (-10% to 10%) as suggested by Tan & Santo (2018) and Yang et.al., (2017), inferring that the computed IMERG PMP were statistically within allowable limit of application with minimal error. The negative RB values connote an underestimation which characterized shorter precipitation intervals and the dryer regions, while the positive RB values fell within the longer precipitation intervals and wetter regions connoting overestimation. This observation further infers that IMERG derived PMPs were mostly underestimated at shorter durations and equally underestimated in dry areas, but at the longer intervals and wetter areas, IMERG tends to overestimate PMPs.

6.2 PMP Errors and Precipitation Amount at Individual Stations

The study evaluated the relationship between the PMP errors, total accumulated precipitation, and mean maximum precipitation estimates at different intervals. The annual total accumulation of precipitation (2000-2020) at different intervals were extracted and averaged to derive the mean annual total accumulation. This mean annual total accumulation were plotted against the residual error to determine their relationship PMP errors at different intervals. Figure 6.3 showed the relationship between the mean annual total accumulatio and the residual error across varying intervals. The pattern of distribution of the relative errors depicts a weak correlation with mean annual total accumulation at shorter interval (figure 6.3a) and a stronger correlation at longer interval (figure 6.3f). As PMPs increases from shorter to longer intervals, residual errors reduce, and correlation becomes more linear in the presence of negligible outliers. At the 24-hr precipitation interval, the effect of underestimation and overestimation of precipitation would cancel out producing a more reliable precipitation estimates for computing PMP, this observation agrees with findings made by Tan & Santo, (2018).

The p-value decreased from 0.311 at shorter interval (30-min) to 0.00031 at longer interval (24-hr), indicating increasing statistical significance. R-squared increased from 0.194 at the 30-mins to 0.515 reaching the 24-hr interval (see figure 6.3). This observation validates those made earlier (in figure 5.3) inferring that PMP at the longer interval showed more reliability than those at shorter intervals. Further confirming that, IMERG PMPs are better suited for wetter regions with higher performance than dry regions.

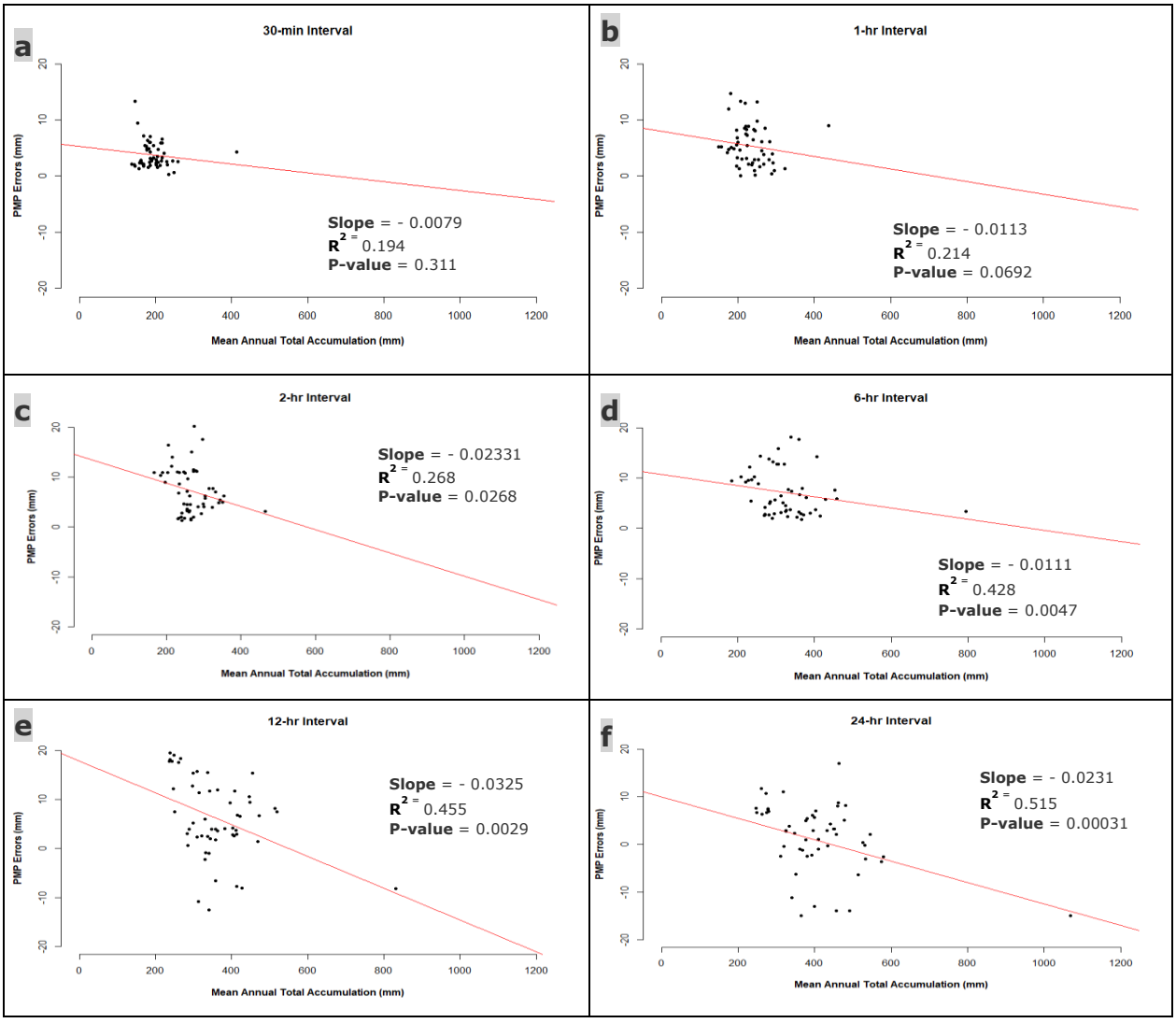

Figure 6.3. Scatterplot showing the relationship between mean annual total accumulation and PMP errors at different intervals. (a-f) shows the relationships and correlation at the durations 30-mins, 1-hr, 2-hr, 6-hr, 12-hr, and 24-hr respectively.

In addition to evaluating the relationship between the PMP errors and the mean total annual accumulation, the mean maximum precipitation at the different intervals were analyzed against the PMP error to understand the error pattern across different precipitation interval, and to find any statistical relationship between precipitation and residual error (figure 6.4). The predicted residual errors produced much outliers at shorter interval (e.g., figure 6.4a) and less outliers at longer interval (e.g., figure 6.4f). The relationship between the two numerical variables for all six intervals were non-linear and inverse. However, the

result revealed a stronger correlation at longer interval and weaker correlation at shorter interval, agreeing with those in fig.8. The p-value at shorter interval (30-min) decreased from 0.201 to 0.00043 on nearing the longer interval (24-hr mark), R-squared increased from 0.136 at the 30-mins mark to 0.503 at the 24-hr interval, indicating an increase in significance relationship between the two variables as duration increases

Like the mean annual total accumulation, the mean max precipitation estimates at various intervals showed that the 24-hr interval has measurements that are more reliable compared to shorter durations. This again can be attributed to both underestimation at shorter interval and overestimation at longer interval canceling out each other, there by producing results consistent with Tan & Santo, (2018) at the 24-hr duration. The annual total accumulation and the mean maximum precipitation both buttress the fact that IMERG derived PMP at the longer interval are more reliable than those at the shorter intervals and agrees with Tan & Santo, (2018).

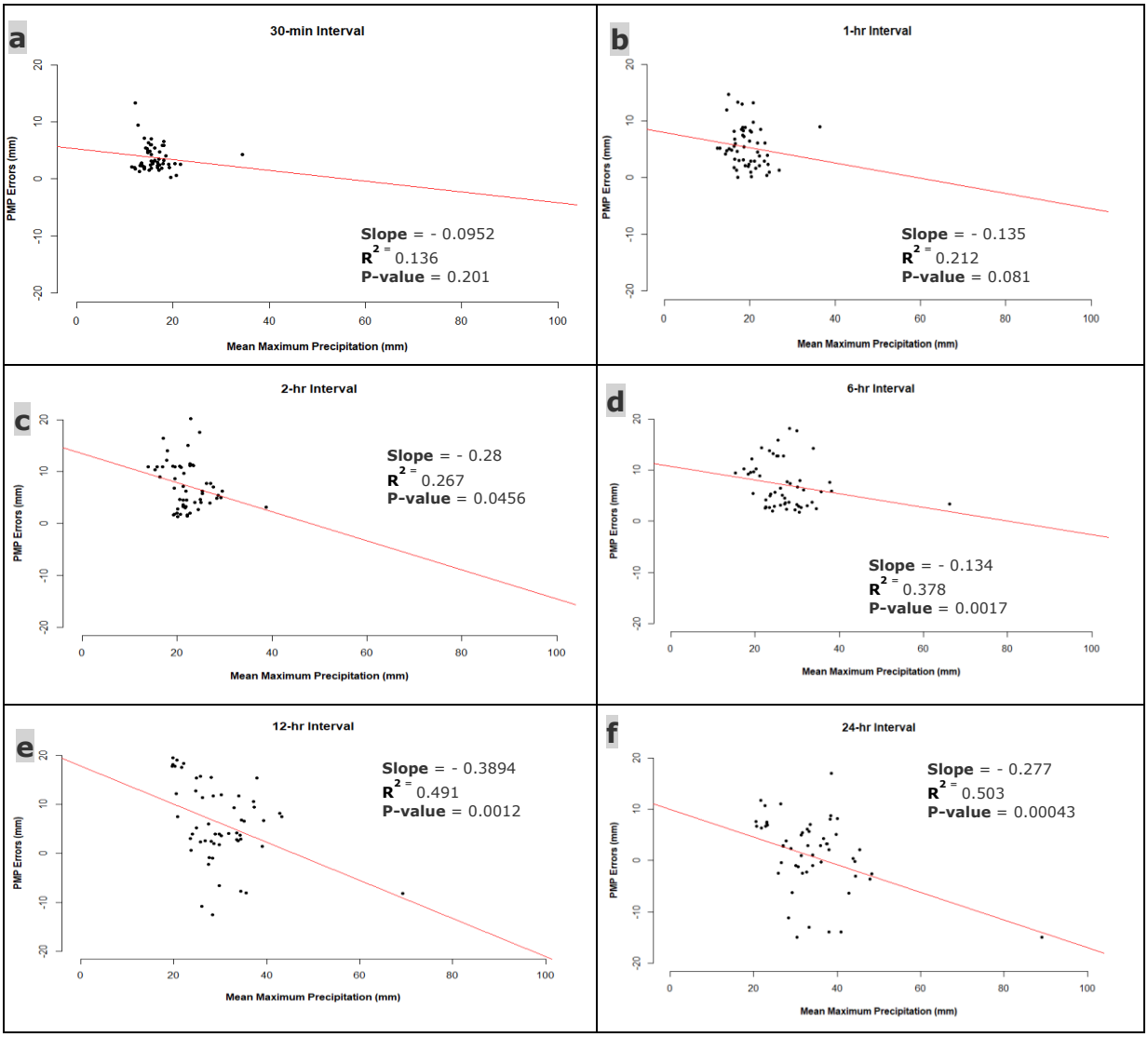

Figure 6.4. Scatterplot showing the relationship between mean maximum precipitation and relative error residual at different intervals. (a-f) shows the relationships and correlation at the durations 30-mins, 1hr, 2hr, 6hr, 12hr, and 24hr respectively. There shorter intervals (b-c) showed abundance of outliers with weak correlation, while longer interval (f) showed a strong correlation between relative residual error as the predicted variable and the mean maximum precipitation.

The statistical results from figures 6.3 and figure 6.4 further justifies the interpolated PMP error trend analysis in figure 5.3. The results showed an inverse relationship, depicting a reduction in residual error as intervals increases. Although, the result from figures 6.3 and figure 6.4 showed that the coefficient of determination (R-squared) was quiet low for the longer interval (51.5% and 50.3% for the 24-hr intervals) the general trend depicts a case of an increasing statistical significance from shorter to longer precipitation intervals. These

observations buttressed the findings made earlier in figure 5.3, confirming that IMERG PMPs are more reliable at longer intervals than shorter intervals. This further infers that longer durations characterized by higher wetness have better estimated PMP with less errors. Therefore, results here clearly shown that estimated PMPs is a function of interval and precipitation patterns in the study.

7 IMPLICATION OF MISSING VALUES ON IMERG PMP ERRORS

IMERG satellite data is fraught with missing values also referred as data holes. To understand why the IMERG holes exists, it is therefore necessary to turn attention to towards the data providers (NASA), that explained why the holes existed in the first place through its technical documentation (Huffman, 2020). Huffman, (2020) attributed the continuous missing values to sensor technical problems which were constantly been reported. Such technical issues include: a malfunction in the AMSR2 data recorder onboard the GCOM-W1 satellite resulting to loss of data, interruption of data from the Himawari-8 Geo satellite, DMSP F17 37V channel experiencing intermittent noisy values, Anomalous SAPHIR bad data being set to missing and included in all runs of version 06B. Furthermore, the Japanese GMS which is one of the IMERG satellite only provides hourly data for short periods, thus in the missing half-hours the adjacent METEOSAT and GOES-W IR values are used to the extent possible. According Huffman (2020), "all missing fields of IMERG product result from completely absent input data for the given time, if the input files are available, the product file is created, even if it lacks any valid data". One of the concerns of this research was to determine, how much of an impact the missing values has on the PMP Errors.

This study sought to analyze the relationship between IMERG missing values and the calculated PMP errors. The ideas were to determine the magnitude of impact the missing the values contributes to the errors at the various intervals over the entire 55-gauge stations. The 30mins IMERG precipitation estimates for all the gauge stations in the study were extracted for the year 2000-2020 using java script built on GEE and made publicly available at (https://code.earthengine.google.com/be56bc780193834519bdf89daeed8f5c). the study further obtained a summary statistic of each IMERG precipitation file for individual gauges using two statistical metrics written with a python script (see figure ii in appendix section).

The statistical metrics used were, the percentage of missing values (known as PMV from hereon), also defined as percentage of records with holes or no data in each file, while the second variable computed was the average length of missing values (referred to as ALMV) and calculated in hours (see table 7.1).

Table 7.1. Summary result of calculated IMERG missing value on the study gauges

Statistics	**Estimates**
Minimum missing value interval	0.5 hr
Maximum missing value interval	21.5 hr
Average length of missing values (ALMV)	2.04 hr
Standard deviation of ALMV	0.0416 hr
Average Percentage Missing value (APMV)	74.21 %
Standard deviation of APMV	1.08 %
Length of data year	20 yrs.
Interval of collection	30 mins
Total length of record	360,816

The assessment was done in two ways, first the study analyzed the relationship between the PMP errors and the PMV for each duration. The R-squared metrics was used to define the extent of correlation between the PMP error and the PMV. Similarly, the relationship between the PMP error and the ALMV were analyzed for all durations and all gauges used in the study. A higher R-Squared depicts a higher correlation and implies Missing values had higher effect on the calculated PMP errors at those intervals. Result from figure 7.1 and table 7.1 showed that the average percentage of missing value was 74.21%, while the PMP errors were at the range of 0 to 20 mm, with standard deviation of error between 2.3 mm at 30-min duration and 8.5 mm at the 24-hr interval (see table 7.2). Correlation between the PMV and PMP error were highest at the 30-mins interval with p-value of 0.00287 (at 0.05 confidence interval), while the least relationship was found at the 24-hr interval with a p-value of 0.665 (figure 7.1).

Table 7.2. Summary result of calculated PMP error and R-squared values for different durations

	Durations					
PMP Error Summary	**30 min**	**1 hr**	**2 hr**	**6 hr**	**12 hr**	**24 hr**
Minimum Error (mm)	0.25	0	1.26	1.81	-12.59	-15.35
Maximum Error (mm)	13.38	14.71	20.22	18.22	19.58	17.03
Std. Error	2.325	3.652	4.563	4.525	7.99	8.5
Mean PMP Error (mm)	3.766	5.378	7.241	7.024	6.184	1.33
PMV R-squared (%)	0.532	0.482	0.324	0.194	0.044	0.036
ALMV R-squared (%)	0.613	0.528	0.392	0.281	0.212	0.209

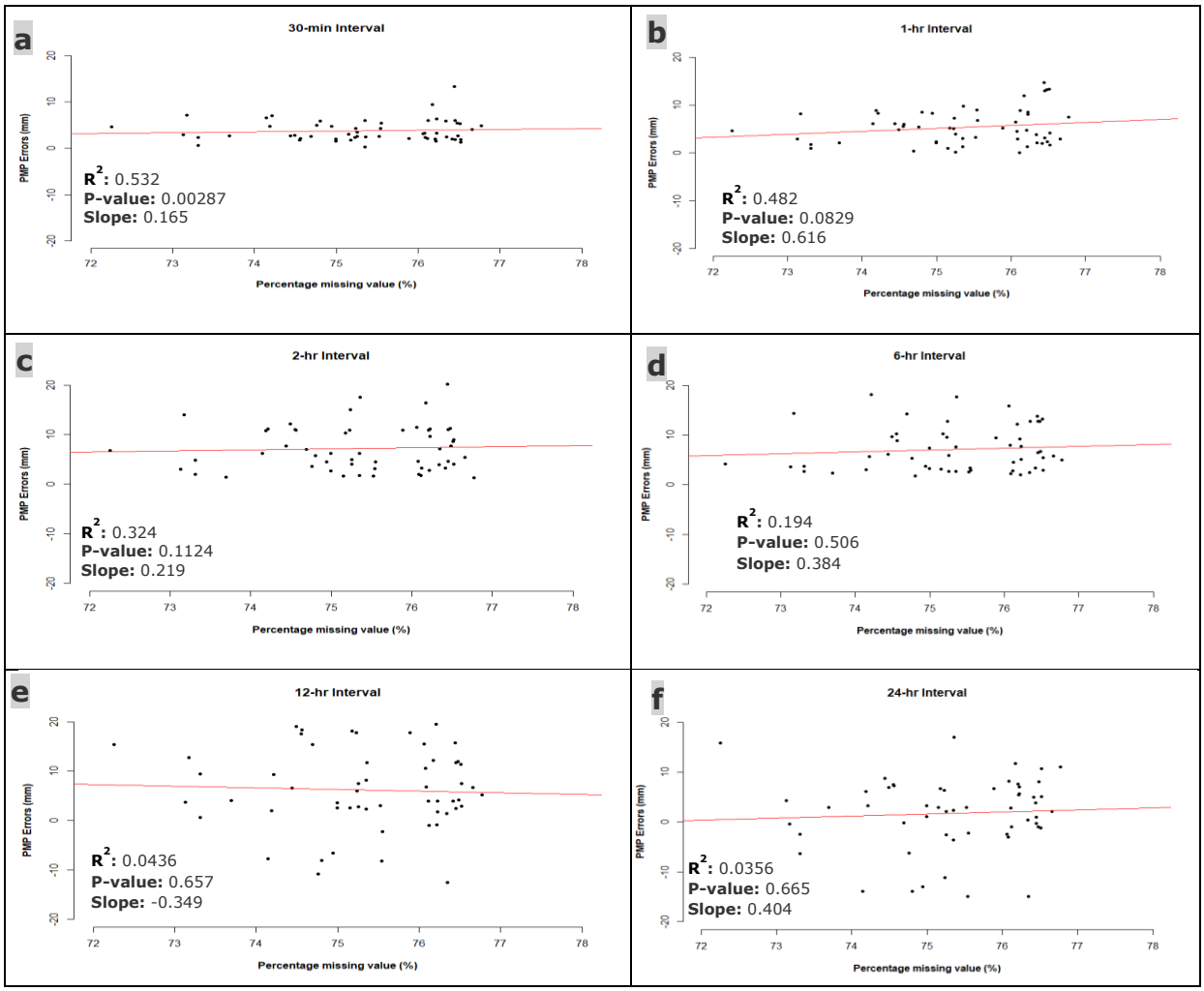

Figure 7.1. PMP error relationship with percentage of missing values for different durations. a-f represents the intervals of duration from 30-mins to 24-hrs. The analysis shows a decreasing r-squared from a-f.

Result from figure 7. 1 implies that IMERG missing values had its most impacts at the shorter duration as the PMP errors at the 30-min intervals were mostly impacted. Whereas

at the longer interval (e.g., 24-hr), the PMP errors were less influenced by the missing values problems. The result from figure 7.1 further justifies earlier findings (see figure 5.3) that PMPs at longer intervals are better suitable for further research application such as PMF mapping. Observations from results presented in figure 7.1 showed that a great deal of underestimation was caused by the missing value problem in IMERG, this agrees with Ning et al., (2017), whose IMERG dataset error assessment found that IMERG underestimates greatly at lower rain rate which is a characteristic of shorter intervals.

An assessment of relationship between the average length of missing values and the calculated PMP errors (see tables 7.1& 7.2) showed that average length of missing values was 2.04-hrs (table 7.1), with PMP errors ranged between 0 to 20mm, and standard deviation of error between 2.3 mm and 8.5 mm for the 30-min and 24-hr durations respectively (see table 7.2). The p-value increased from 0.00127 at shorter interval (30-min) to 0.4213 at longer interval (24-hr) (figure 7.2). Result from figure 7.2 agree with those in figure 7.1, implying a stronger relationship exist at the 30-min interval than the 24-hr interval. This further justifies earlier observation that IMERG missing values has stronger influence on calculated PMP errors estimated at shorter durations than longer duration. Figures 7.1 and 7.2 have been consistent with figure 5.3 showing that both Missing values, precipitation pattern, intervals of estimation, and gauge zones have a significant impact on the amount of IMERG PMP computed.

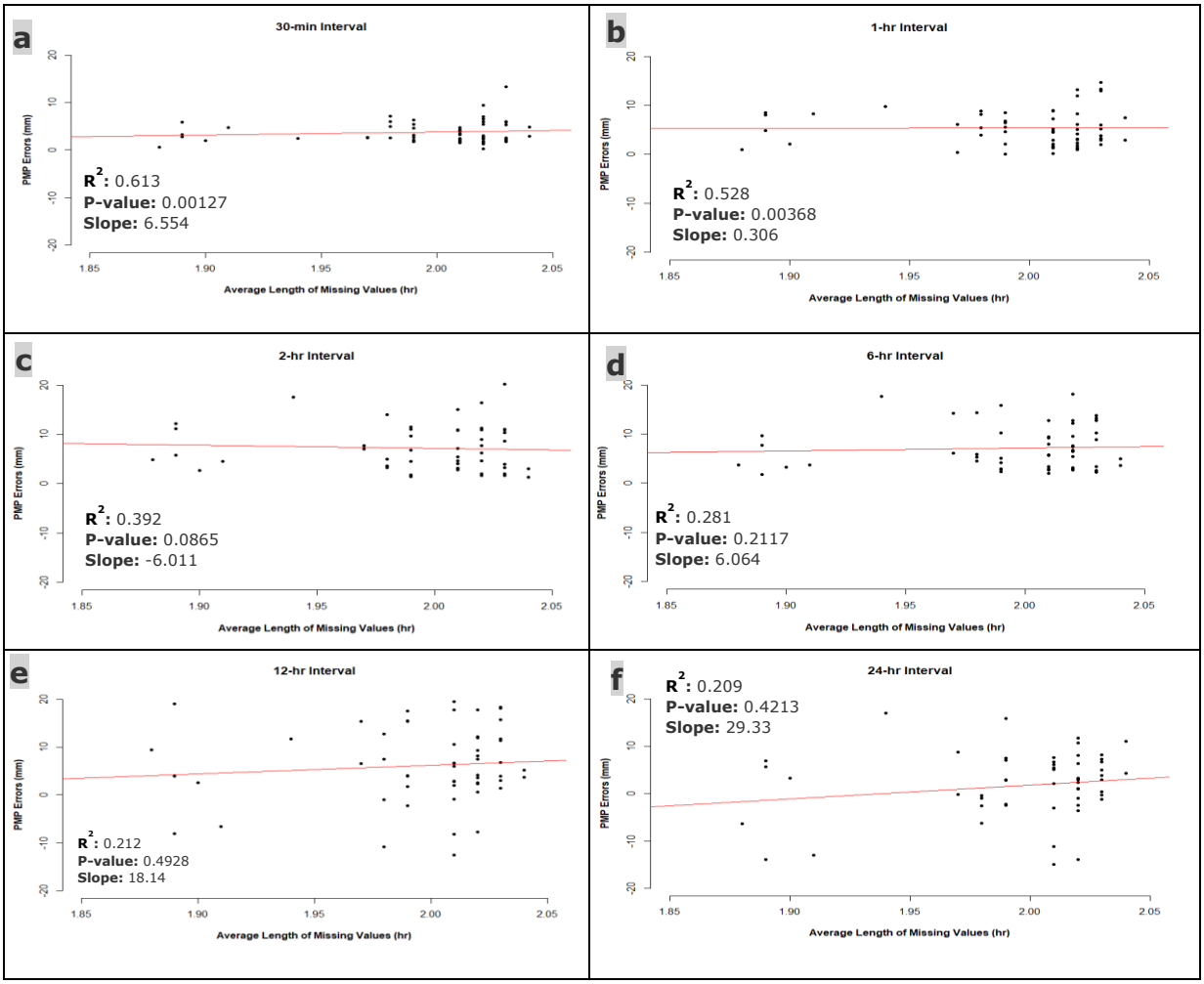

Figure 7.2. Relationship between PMP error and average length of missing values for each duration. a-f represents 30-mins to 24-hrs intervals. The analysis shows a decreasing r-squared from a-f.

A post assessment of IMERG missing values on overall PMP computation proved that the observed underestimation in IMERG derived PMP were partly influenced by missing values in IMERG dataset. To determine how these missing values, impact the derived PMP, the study utilized an independent gauge station (University of Kansas field station) data not used by NOAA (available at http://kufs.ku.edu/resources/weather-station/) to compute PMP and evaluate them against derived PMP from IMERG. Two precipitation durations (daily and monthly intervals) were collected from the station (for the period 2000-2020) and the maximum precipitation for those durations compared against those collected from IMERG at same durations (see figure 7.3). The comparison in figure 7.3 showed that IMERG

underestimates precipitation for the daily and monthly durations throughout the precipitation period (2000-2020). These underestimations recorded in IMERG maximum precipitation is directly linked to the missing values in the IMERG dataset (explained in figures 7.1 and 7.2).

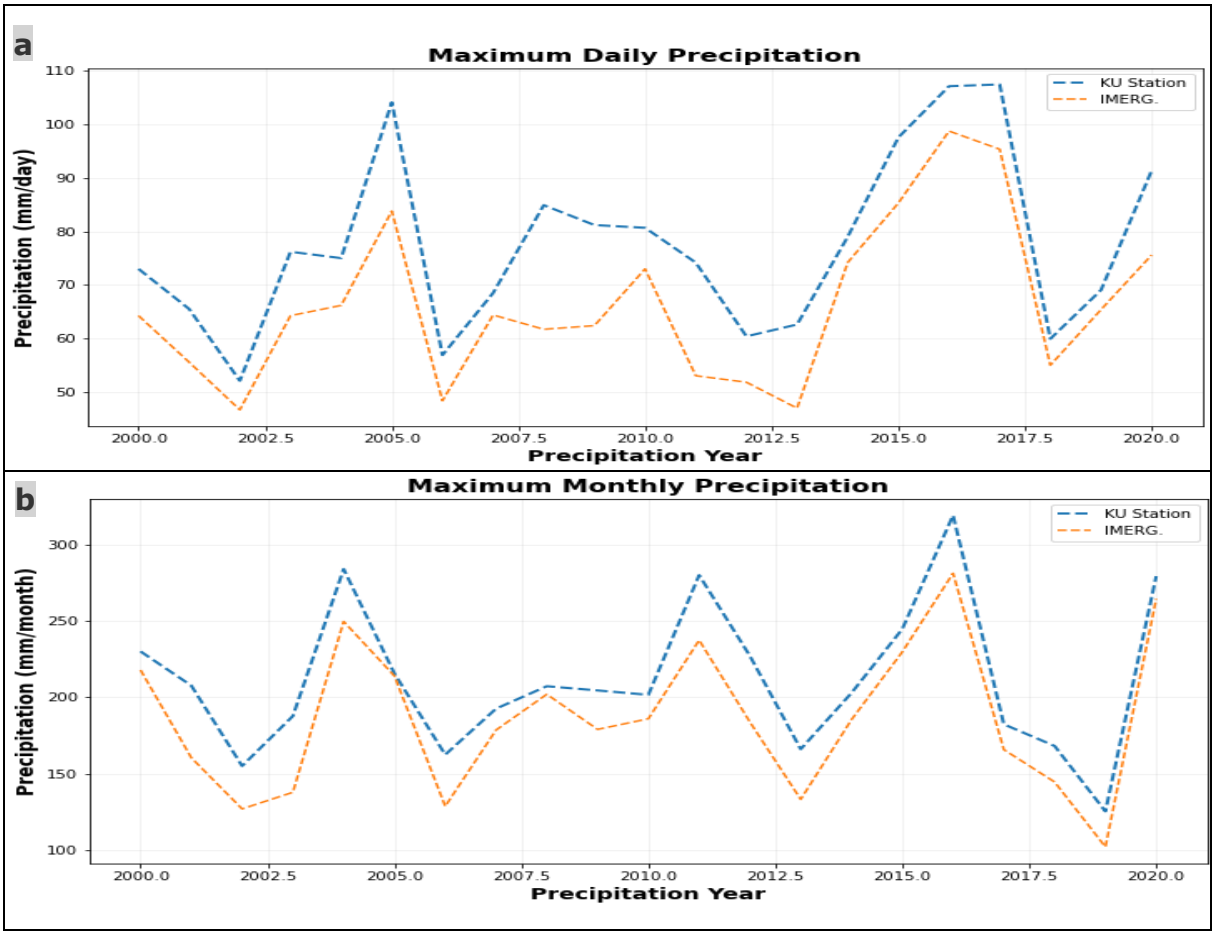

Figure 7.3. Comparing KU station data and IMERG dataset. (a) maximum daily precipitation between IMERG and KU station data (2000-2020). (b) maximum monthly precipitation assessment between IMERG and KU station data (2000-2020).

A further exploration showed that the derived PMP from those maximum precipitation at the durations results in underestimation of IMERG PMP at the six durations. Figure 7.4 showed the NOAA PMP, calculated PMP from IMERG, and field PMP derived from KU field station data, using their maximum precipitation for individual durations. The results from figure 7.4 showed that IMERG derived PMP were underestimated in all durations. The result showed that calculated PMP from KU field station were higher than NOAA

interpolated PMP. The results confirm that missing values from IMERG datasets impacts both precipitation estimates and calculated PMPs and agrees with observations made by Ning et al., (2017).

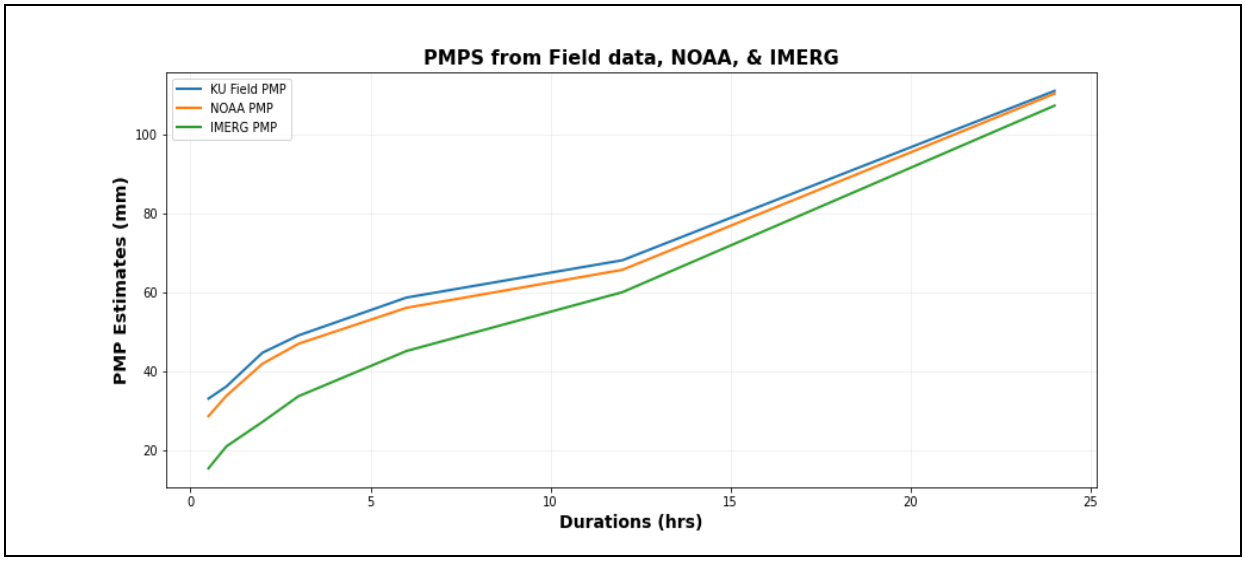

Figure 7.4. Comparing derived PMPs from IMERG, NOAA, and KU field station.

The study further determined that, IMERG PMPs between 30-mins and 12-hr fell outside the confidence interval of IMERG PMP. However, IMERG PMP at the daily durations (12-hr interval), were within the 90% confidence interval of the NOAA PMP (figure 7.5a) and KU field PMP (figure 7.5b). Results from figure 7.5 further shows that IMERG estimated PMPs are better adapted at the longer intervals. PMP at the shorter durations (30-mins up to 12-hr) proved to be unreliable for research due to high underestimation. Hence suggesting further calibration of IMERG data is necessary.

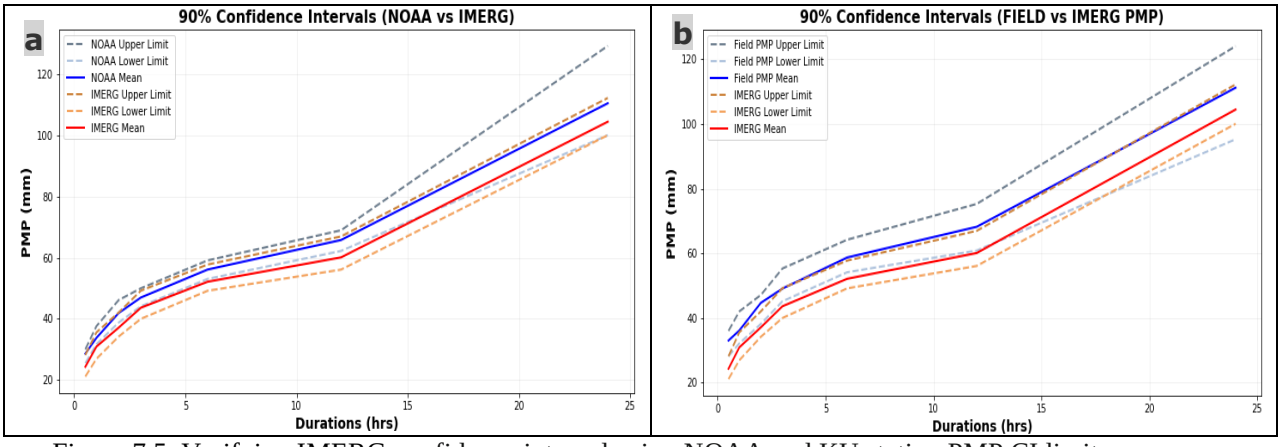

Figure 7.5. Verifying IMERG confidence interval using NOAA and KU station PMP CI limits.

There are two ways to address problem associated with missing values in precipitation records. Huffman, (2020) noted that the datasets from IMERG containing missing values were constantly corrected by data filling approach after 3 months of data collection prior to making them available to the public. This method is termed Morphing and involves interpolating from rain gauge data made available by GPCC, GPCP, UKMET, NOAA, JAXA, and other meteorological agencies from over 190 countries with over 80,000 rain gauges. The second approach relies on Hershfield adjustment technique, The WMO recommends the Hershfield adjustment factor which uses values ranging from 0.13 to 2.9 to adjusts the final PMP. Other approaches will include calibrating the derived PMP using gauge based PMP by a method of spatial interpolation.

8 CONCLUSION AND RECOMMENDATION

In this section, the summary and major findings, concluding note, research limitation and recommendation from current findings were discussed in detail.

8.1 Conclusions

As storm events are not common but sparse, hydrologist and engineers have to either rely on extreme precipitation events with longer return periods for decision making or rely on hypothetical maximum possible precipitation, recently renamed Probable Maximum Precipitation (PMP) for same decision purposes. These PMPs which are precipitation depth at different time intervals have guided engineers in siting infrastructures such as dams, bridges, culverts, levees, etc. It has since become imperative to study the biases associated with PMP, particularly those derived from recent satellites such as IMERG. To study these possibilities and limitations associated with IMERG estimated PMP, this study critically evaluated the IMERG PMP taking into consideration the PMP intervals, precipitation pattern and zones, precipitation amount, and missing value.

The study findings showed IMERG derived PMP at the 24 hr durations were within the 90% confidence interval of NOAA PMP, while the lower intervals were found to be outside the 90% confidence interval of NOAA station PMP. In addition, IMERG derived PMP at the durations showed that the 30 mins interval returned 0.95 of CC, 4.41 mm of RMSE, -3.77 % of RB, and 91.4% R-squared value between IMERG PMP and NOAA Station PMP. While the 24 hr interval recorded 0.854 of CC, 10.28 mm of RMSE, 7.82 % RB, and 81.3% of R-squared between IMERG PMP and NOAA Station PMP relationship. Relatively, the study found that IMERG underestimates PMP at shorter intervals with most errors, while the longer durations recorded the least errors. Further evaluation showed that IMERG estimates PMPs differently at varying precipitation zones (dry, mid, and wet precipitation

zone). The dry precipitation zone recorded the least CC (0.871 at 30-mins interval and 0.733 at the 24-hr), the most RMSE (4.99 mm at the 30-mins durations and 10.51 mm at the 24-hr), and least RB (-10.23% at the 30-min duration and -0.912% at the 24-hr), while the wet precipitation zone recorded the most CC (0.951 at the 30-mins duration and 0.821 at the 24-hr), the least RMSE (3.75 mm at the 30-mins duration and 9.45 mm at the 24-hr), and the most RB (-8.14% at the 30-mins duration and 9.35 mm at the 24-hr). The result from the PMP assessment at the defined precipitation zones showed that IMERG PMP had better estimation capability in wet precipitation regions with lesser errors than dry regions. Further analysis at the gauges showed that the calculated PMP errors at the gauges were consistent with the precipitation pattern of the study area (see figure 5.3), with the last error (1 mm) found in wet zones and the most error (20 mm) occurring in the dry zone (see table 7.2).

The study further determined the influence of precipitation amount and missing values on the PMP errors. It was observed that both mean maximum precipitation and the total accumulated precipitation had negatively weak correlation with PMP error. The calculated p-value decreased from 0.311 and 0.136 at the shorter duration (30-min interval) to 0.000313 and 0.00043 at the longer duration (24-hr interval) for total accumulated precipitation and mean maximum precipitation and respectively. An assessment of IMERG missing value influence on the PMP error showed that both percentage missing value and average length of missing value (74.21% and 2-hrs respectively) had impacts PMP error, with more impact on the shorter interval than the longer interval. Findings made in this research showed that missing values, total accumulated precipitation, and precipitation pattern in the study area influences the IMERG PMP, with the greater influence at the 30-mins duration. However, IMERG estimates better PMP in with areas with the least amount of errors. The findings showed a great deal of underestimation by the IMERG caused partly

by missing data and precipitation zones (dry) and were consistent with those from Ning et al., (2017), Yang et al., (2017), Tan & Santo, (2018), Tapiador et al., (2020), and Anjum et al., (2018).

8.2 Research Limitation

Contextually, this research will love to see an extension to other possible variable that could impact the estimation of IMERG PMP such as elevation. This study was solely focused on precipitation patterns, missing values impact on derived PMP across adapted gauges in Kansas, ignoring the elevation. This is partly because Kansas has more flat terrain with a polynomial order not more than 2. Question of how our results will turn out when analyses are extended to rugged terrains such as the Rockies, provides future research prospects and possible research extension. Spatially, the research was limited to Kansas, an extended evaluation across states in the US will further validate findings in this study at the macro-scale. Length of data coverage also caused concerns; one wonders what results might look like if they were extended years of data coverage (e.g., 50-years of data). The IMERG data archived 20 years of precipitation record (2000-2020) and can only bring to mind the possibility of impacts on result if more years of data were made available. High frequency of missing values also showed that an extended analysis will need some data filling using spatial interpolation techniques. Future research will extend evaluation of IMERG PMP using independent rain gauge locations not used by NOAA-Atlas-14 for PMP estimation.

8.3 Research Recommendation

Although this study advanced the method of estimating PMP for different intervals of precipitation durations using IMERG satellite dataset with fine resolutions, it is important to note here that final PMP needs calibration before applying them for flood mapping. The Hershfield technique recommends using an adjustment factor ranging between 0.13 to 2.9

depending on the precipitation interval and the purpose of usage (Kappel, 2019). This research will be beneficial to flood impacted states, regions, and counties. The study will also benefit agencies such as FEMA, Kansas Water Office, Kansas Department of Agriculture's Division of Water Resources, Bureau of Water – Kansas Department of Health and Environment, Kansas state agencies, flood agencies, other water related agencies from states, federal agencies, gauge scarce regions and states, insurance companies, engineers, hydrologists etc.

Appendix

```
var geometry = ee.Geometry.Point([-99.3000, 37.6019]);
Map.addLayer(geometry);
var dataset = ee.ImageCollection('NASA/GPM_L3/IMERG_V06')
  .select('HQprecipitation')
  .filterDate(ee.Date('2019-09-01'),ee.Date('2019-09-05'))
  .getRegion({
    geometry: geometry,
    scale: 30
  }).map(function(element){
    return ee.Feature(null, {time:ee.List(element).get(3),precip:ee.List(element).get(4)});
  });
print(ee.FeatureCollection(dataset));
Export.table.toDrive({
  collection: ee.FeatureCollection(dataset),
  description: 'Greensburg_Holes',
  folder: "",
  fileNamePrefix: 'Greensburg_Holes',
  fileFormat: "CSV"});
```

Figure i: Java script to extract files and show missing values/holes at points of interests.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('./5Greensburg_Holes.csv', index_col='Time')
```

```
## Determine/sort File Parameters
T_nans = df.isnull().sum()
real_val = df.count()
T_length = len(df)
perc_nans = (T_nans/T_length)*100

# eva = T_nans + real_val
# print(T_length, eva)
```

```
nan_vals = []
for i in range(len(df.index)):
    missval = df.iloc[i].isnull().sum()
    nan_vals.append(missval)
```

```
# Calculate holes in file
nans = []
for i, x in enumerate(nan_vals):
    if x == 0:
        nans.append(x)
    else:
        sum_value = 0
        for j in nan_vals[i::-1]:
            if j != 0:
                sum_value += j
            else:
                break
        nans.append(sum_value*0.5)
new_arr = []
idx = np.squeeze(np.where(np.array(nans[:])==0))-1
for ind, indice in enumerate(idx):
    new_arr.append(nans[indice])
final_nans = np.ma.masked_equal(new_arr, 0).compressed()
max_val = np.max(final_nans)
min_val = np.min(final_nans)
mean_val = np.mean(final_nans)
total_val = np.sum(final_nans)

#print('Maximum value = ', max_val,'\n' 'Minimum value = ', min_val, '\n' 'Mean = ', round(mean_val,2), '\n'
#      'total values = ', total_val)
```

```
## Print Statistics of Missing Values
print(" Maximum successive holes range in file:", max_val)
print(" Minimum successive hole(s) range in file:", min_val)
print(" Average hole(s) range in file:", round(mean_val,2))
print(" Total hour holes range in file:", total_val)
```

```
## Visualize / Print File Parameters / information
print(" Total Length of rows in file:", T_length)
print(" Total Length of rows with captured/real values in file:", real_val.values[0])
print(" Total Length of rows with Holes/missing values in file:", T_nans.values[0])
print(" Percentage Length of missing rows in file:", str(perc_nans.values[0]) + "%")
```

```
#df['NAN in minutes']=nans #Adding the missing values to the main dataset as a column

#Writing the data into a new file
#pd.Series(final_nans).to_csv('new_output.csv')
```

Figure ii: Python script to calculate the percentage of missing values in IMERG precipitation file

```
// Define start date for data collection
var startDate = '2020-06-01';

// Define Geometry to extract imerg data
var ExportGeometry =
    ee.Geometry.Polygon(
        [[[-170, 75],
          [-170, -56],
          [170, -56],
          [170, 75]]], null, false);
// Map.addLayer(ExportGeometry);
// Get the IMERG data from ee collection
var precipData = ee.ImageCollection("NASA/GPM_L3/IMERG_V06")
                   .filterDate(
                     ee.Date(startDate),
                     ee.Date(startDate).advance(1, 'month')
                   )
                   .select('HQprecipitation')
                   .map(function(image) {
                     return image.divide(2).set({'system:time_start':image.get('system:time_start')});            // mm/h to mm/half hour
                   })
                   .toList(17520);       // 17520 is the number of 30 minute inccrements in a year
// print(precipData)
// Map.addLayer(ee.ImageCollection(precipData),{},'Raw month');

// Define the windows to extract data
// Sum precip in each window and return the maximum sum of that window for every month of data collection
var finalimage = ee.ImageCollection(ee.List.sequence(0, (precipData.length().getInfo()-1)).map(function(pos) {
  var step30min = ee.Image(precipData.get(pos)).select(["HQprecipitation"],["HQPre_30min_max"]);
  var step1hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-1.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_1hour_max"]);
  var step2hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-2.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_2hour_max"]);
  var step3hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-3.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_3hour_max"]);
  var step6hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-6.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_6hour_max"]);
  var step12hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-12.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_12hour_max"]);
  var step24hour = ee.ImageCollection(precipData).filterDate(
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(-24.1, 'hour'),
      ee.Date(ee.Image(precipData.get(pos)).get('system:time_start')).advance(10, 'minute'))
      .sum().select(["HQprecipitation"],["HQPre_24hour_max"]);
return ee.Image(step30min).addBands([step1hour,step2hour,step3hour,step6hour,step12hour,step24hour]);
})).max();
finalimage = finalimage.addBands(ee.ImageCollection(precipData).sum().select(["HQprecipitation"],
["HQPre_month_max"])).set({'system:time_start':ee.Date(startDate)});

// print(finalimage);
// Map.addLayer(finalimage,{},'DDF data');

// Export output to Asset
Export.image.toAsset({
  image: finalimage,
  description:'PE_'+startDate.replace('-','_').replace('-','_'),
  assetId:'PE_'+startDate.replace('-','_').replace('-','_'),
  region:ExportGeometry,
  // scale:5000,
  maxPixels:1e13
});
```

Figure iii: Java script that extracts IMERG precipitation record and save them to asset.

```
var DDF_Data = ee.ImageCollection("users/crystalwhite4real2015/IMERG");
//print(DDF_Data)

var point = ee.Geometry.Point(-95.7076, 37.2266)
Map.addLayer(point);
Map.centerObject(point, 9)

// Create an image time series chart.
var chart = ui.Chart.image.series({
  imageCollection: DDF_Data,
  region: point,
  reducer: ee.Reducer.mean(),
  scale: 200
});

// Set up chart options
chart.setOptions({
  title: 'Maximum Precipitation Estimate at rain gauge cell from IMERG v6 Collections',
  vAxis: {
    title: 'mm per duration'
  }
});
chart.style().set({
  position: 'bottom-right',
  width: '500px',
  height: '300px'
});

// Add the chart to the map.
Map.add(chart);

// Create a label on the map.
var label = ui.Label('Click a point on the chart to show the image for that date.');
Map.add(label);

// When the chart is clicked, update the map and label.
chart.onClick(function(xValue, yValue, seriesName) {
  if (!xValue) return;  // Selection was cleared.

  // Show the image for the clicked date.
  var equalDate = ee.Filter.equals('system:time_start', xValue);
  var image = ee.Image(DDF_Data.filter(equalDate).first());
  var mapLayer = ui.Map.Layer(image, {
    min: 0,
    max: 60,
    palette: ['000096','0064ff', '00b4ff', '33db80', '9beb4a', 'ffeb00', 'ffb300', 'ff6400', 'eb1e00', 'af0000'],
    bands: seriesName
  });
  Map.layers().reset([mapLayer, point]);

  // Show a label with the date on the map.
  label.setValue((new Date(xValue)).toUTCString());
});
```

Figure iv: Java script that extracts precipitation records at points from asset cloud storage.

```
library(readxl)
PE <- read_excel("PE_Jim.xlsx",
                 col_names = c("Year", "P_30min", "P_1hr", "P_2hr",
                               "P_6hr", "P_12hr", "P_24hr", "P_1month"))
# Load to dataframe
(df <- data.frame(PE))


############ 1 ######################################################################
# Extract 30mins duration from df
PE_30min = df[, c(2)]
PE_30min

# Function to calculate frequency factor from individual duration
Km_30min <- function(km_30){(Rmax - Rn)/Sn}
Rmax = max(PE_30min)
Del_max = PE_30min[PE_30min!=Rmax]
# Rn <- (sum(PE_30min) - Rmax)/(length(PE_30min)-1)
Rn = mean(Del_max)
# Sn_D = PE_30min[PE_30min!=Rmax]
Sn = sd(Del_max)
Km30 = Km_30min(km_30)
Km30

# Function calculate PMP, 1.13 is harshfield adjustment factor
PMP_30min <- function(PMP_30){((Xn + (Sdn * Km30)) * 1) }
Xn = mean(PE_30min)
Sdn = sd(PE_30min)
PMP_30m = PMP_30min(PMP_30)
PMP_30m
```

Figure v: Modeled R-script to compute PMP from precipitation record.

REFERENCES

Alamri, N., & Subyani, A. (2017). Generation of Rainfall Intensity Duration Frequency (IDF) Curves for Ungauged Sites in Arid Region. *Earth Systems and Environment*, *1*, 8. https://doi.org/10.1007/s41748-017-0008-8

Anjum, M. N., Ding, Y., Shangguan, D., Ahmad, I., Ijaz, M. W., Farid, H. U., Yagoub, Y. E., Zaman, M., & Adnan, M. (2018). Performance evaluation of latest integrated multi-satellite retrievals for Global Precipitation Measurement (IMERG) over the northern highlands of Pakistan. *Atmospheric Research*, *205*, 134–146. https://doi.org/10.1016/j.atmosres.2018.02.010

Bathrellos, G. D., Karymbalis, E., Skilodimou, H. D., Gaki-Papanastassiou, K., & Baltas, E. A. (2016). Urban flood hazard assessment in the basin of Athens Metropolitan city, Greece. *Environmental Earth Sciences*, *75*(4), 319. https://doi.org/10.1007/s12665-015-5157-1

DHI (2017a) MIKE 21 flow model and MIKE 21 flood screening tool hydrodyamic module scientific documentation, published by DHI

Dingman, S. L. (2014). *Physical Hydrology, Third Edition* (3 edition). Waveland Press, Inc.

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens Environ 120:25–36. https://doi.org/10.1016/j.rse.2011.11.026 ISSN 0034-4257. http://www. sciencedirect.com/science/article/pii/S0034425712000636*

Gao, P., Carbone, G. J., Lu, J., & Guo, D. (2018). An Area-Based Approach for Estimating Extreme Precipitation Probability. *Geographical Analysis*, *50*(3), 314–333. https://doi.org/10.1111/gean.12148

George J. Huffman. (2020). *Algorithm Theoretical Basis Document (ATBD) Version 06 of NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG)*. NASA/GSFC Code 612. https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Guo, H., Chen, S., Bao, A., Behrangi, A., Hong, Y., Ndayisaba, F., Hu, J., & Stepanian, P. M. (2016). Early assessment of Integrated Multi-satellite Retrievals for Global Precipitation Measurement over China. *Atmospheric Research*, *176–177*, 121–133. https://doi.org/10.1016/j.atmosres.2016.02.020

Heidler LM (2015) Evaluation of different hydrological models in data scarce regions on the island of Ceram, Indonesia, *Technische Universitat Munchen, Germany*

Hewiston, Crane BC (1994) Precipitation Controls in Southern Mexico, in Neural Nets, Kluwer Academic Publisher

Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. *Cambridge University Press, Cambridge; New York, xiii, p 224*

Hsu K, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. *Water Resource Res 31(10):2517–2530. https://doi.org/10.1029/95WR01955*

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & 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. https://doi.org/10.1175/JHM560.1

Jose D. Salas, Germán Gavilán, Fernando R. Salas, Pierre Y. Julien, & Jazuri Abdullah. (2014). Uncertainty of the PMP and PMF. *Taylor & Francis Group, LLC*, 575–603.

Kappel, B. (2019). PMP estimation for mine tailings dams in data limited regions. In J.-P. Tournier, T. Bennett, & J. Bibeau (Eds.), *Sustainable and Safe Dams Around the World* (1st ed., pp. 3170–3182). CRC Press. https://doi.org/10.1201/9780429319778-285

Kim, D.-E., Gourbesville, P., & Liong, S.-Y. (2019). Overcoming data scarcity in flood hazard assessment using remote sensing and artificial neural network. *Smart Water*, *4*(1), 2. https://doi.org/10.1186/s40713-018-0014-5

Mab, P., Ly, S., Chompuchan, C., & Kositsakulchai, E. (2019). *Evaluation of Satellite Precipitation from Google Earth Engine in Tonle Sap Basin, Cambodia*.

Min Yang, Zhongqin Li, Muhammad Naveed Anjum, & Yayu Gao. (2019). Performance Evaluation of Version 5 (V05) of Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) over the Tianshan Mountains of China. *Water*, *11*(6), 1139. https://doi.org/10.3390/w11061139

NASA. (2021). *Giovanni | NASA Global Precipitation Measurement Mission*. https://gpm.nasa.gov/data/sources/giovanni

Ning, S., Song, F., Udmale, P., Jin, J., Thapa, B. R., & Ishidaira, H. (2017). Error Analysis and Evaluation of the Latest GSMap and IMERG Precipitation Products over Eastern China. *Advances in Meteorology*, *2017*, e1803492. https://doi.org/10.1155/2017/1803492

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., & Stender, P. (2014). Global gridded precipitation over land: A description of the new GPCC First Guess Daily product. *Earth System Science Data*, *6*(1), 49–60. https://doi.org/10.5194/essd-6-49-2014

Singh, A., Singh, V. P., & Ar, B. (2018). Computation of probable maximum precipitation and its uncertainty. *International Journal of Hydrology*, *2*(4). https://doi.org/10.15406/ijh.2018.02.00118

Singh, K. K., & Singh, A. (2017). Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier. *The Egyptian Journal of Remote Sensing and Space Science*, *20*(1), 147–155. https://doi.org/10.1016/j.ejrs.2016.04.003

Su, J., Lü, H., Ryu, D., & Zhu, Y. (2019). The Assessment and Comparison of TMPA and IMERG Products Over the Major Basins of Mainland China. *Earth and Space Science*, *6*(12), 2461–2479. https://doi.org/10.1029/2019EA000977

Su, J., Lü, H., Zhu, Y., Cui, Y., & Wang, X. (2019). Evaluating the hydrological utility of latest IMERG products over the Upper Huaihe River Basin, China. *Atmospheric Research*, *225*, 17–29. https://doi.org/10.1016/j.atmosres.2019.03.025

Tan, M. L., & Santo, H. (2018). Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia. *Atmospheric Research*, *202*, 63–76. https://doi.org/10.1016/j.atmosres.2017.11.006

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. https://doi.org/10.1175/JHM-D-19-0116.1

Warren IR, Bach HK (1992) MIKE 21: a modelling system for estuaries, coastal waters and seas. *Environ Softw 7(4):229–240, ISSN 0266-9838*. https://doi.org/10.1016/0266-9838(92)90006-P

World Meteorological Organization. (2009). *Manual on estimation of probable maximum precipitation (PMP)*. https://googledrive.com/host/0BwdvoC9AeWjUazhkNTdXRXUzOEU/wmo_1045_en.pdf

Yang, Y., Tang, G., Lei, X., Hong, Y., & Yang, N. (2017). Can Satellite Precipitation Products Estimate Probable Maximum Precipitation: A Comparative Investigation with Gauge Data in the Dadu River Basin. *Remote Sensing*, *10*(1), 41. https://doi.org/10.3390/rs10010041