Near-Real-Time Rainfall Detection and Estimation from Commercial Microwave Links and Meteosat Second Generation Data

Kwabena Kingsley Kumah

Near-Real-Time Rainfall Detection and Estimation from Commercial Microwave Links and Meteosat Second Generation Data

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr.ir. A. Veldkamp
on account of the decision of the Doctorate Board,
to be publicly defended
on Wednesday, October 26, 2022, at 14.45 hrs

by

Kwabena Kingsley Kumah

born on 15th December 1987
in Kumasi, Ghana

This thesis has been approved by

Prof. Dr Z. Su (supervisor)

Dr. Ben Maathuis (co-supervisor)

Dr. Joost Hoedjes (co-supervisor)

ITC dissertation number 420

ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-5458-9

DOI: 10.3990/1.9789036554589

Cover designed by Kumah K. Kingsley and Job Duim

Printed by CTRL-P Hengelo

Graduation committee:

Chair / secretary:	prof.dr. F.D. van der Meer
Supervisor:	prof.dr.ir. Z. Su Universiteit Twente, ITC, Department of Water Resources
Co-supervisors:	dr. B.H.P. Maathuis ITC, Department of Water Resources
	dr. J.C.B. Hoedjes Universiteit Twente, ITC, Department of Water Resources
Committee Members:	prof. dr. H. Kunstmann PhD University of Augsburg
	prof. dr. ir. R. Uijlenhoet PhD TU Delft
	prof. dr. ir. N. Van De Giesen TU Delft
	dr.ir. C.M.M. Mannaerts Universiteit Twente, ITC, Department of Water Resources
	prof.dr. M. Van Aalst PhD Universiteit Twente, ITC

Success manifests through different shades of failure....

Summary

Many operational applications, including agriculture and water resources management, local climate studies, flash flood prediction and forecasting that directly affect human livelihoods, need accurate high-resolution rainfall information. High-resolution rainfall is also vital in research applications such as evaluating satellite rainfall products to inform new satellite rainfall observation missions. However, despite its great value, rainfall's intricate characteristics, such as intermittency and spatial variation, make accurate rainfall estimation challenging.

There are existing techniques to measure rainfall, but each has an inherent limitation. For instance, rain gauges measure accurate rainfall from a point and are thus limited in their ability to measure the spatial state of rainfall. The weather radars are capable of accurate spatial rainfall information, but among other limiting factors, they are expensive to acquire, operate and maintain. Satellites are capable of global rainfall observations, but their estimates are often inaccurate. Moreover, their observations, usually averaged over squared kilometres, are not always ideal for all applications, e.g., farmers whose plots are only a few hundred squared meters.

Over a decade ago, a new technique was introduced that utilises telecom service operators' extensive mobile phone infrastructure for rainfall observations. The microwave signals beamed from one telephone tower to the other are significantly affected by rain. For this reason, researchers studied the fluctuations in the signal level recorded at the receiving tower and showed that these systems can estimate accurate rainfall; suggesting that each microwave link (MWL)–a pair of connected towers–can serve to gauge rainfall along two connecting mobile phone towers in near real-time. Albeit opportunistic and low-cost rainfall estimation technique, the problem is that the estimation accuracy is affected by many factors, including the variation of raindrop sizes along the link and the fact that the link's antenna wetting during and after rainfall introduces additional uncertainties in the retrieved rainfall estimates. Moreover, the density of the MWL network is arbitrary and often biased towards densely populated areas. This characteristic may complicate their use for retrieving spatial rainfall patterns, especially in areas with low network densities.

At the same time, geostationary (GEO) meteorological satellites, such as the Meteosat Second Generation (MSG), frequently observe the earth's atmosphere with wide spectral range radiometers capable of detecting rapidly developing storms, raining areas and estimating rainfall intensities. Several studies have

extensively studied the data with traditional rainfall measurement techniques such as rain gauges and weather radars for rainfall observation. However, combing the GEO satellite and MWL data for rainfall observation has gained very little attention, while such an alternative could benefit many places that lack traditional rainfall observation systems.

For this reason, the main objective of this study was to investigate the MWL and the MSG satellite data for high spatiotemporal resolution rainfall detection and estimation using data from Central and Western Kenya, where traditional rainfall-observing systems are often lacking or sparse. This main objective was achieved based on 4 specific objectives, which form the basis of 4 individual research studies.

Firstly, it was investigated if the MWL-MSG data combination could improve rain rate estimation and detection. We investigated the MWL's capability to estimate accurate rain rates in the study area using gauge rainfall data. Then, the MWL's rain rates were studied with the MSG satellite data using a conceptual model in which clouds with high cloud top optical thickness and particle effective radius have high rainfall probabilities and rain rates. Regarding the MWL's rainfall estimation capability, the results confirmed the robustness of the MWL rainfall estimation technique concluded by many past studies. Studying the MWL's rain rate with the MSG satellite data revealed unique spectral characteristics of daytime, night-time, raining and non-raining data. Eventually, descriptive statistics derived from the satellite spectral characteristics successfully detected rainfall on individual MWL in the study area. However, daytime rain detection, which uses reflectance data, was better than night-time detection, which uses infrared (IR) data, owing to the better rain information about cloud optical and the effective radius contained in the reflectance data.

Following the successful combination of the MWL-MSG data for rainfall observation, the subsequent investigation developed an improved rain area detection system to improve the MWL's rainfall estimates and map rainfall in the study area. This investigation evaluated multiple parametric rain detection models derived from MSG's reflectance and IR data using a conceptual model similar to the previous research. Additionally, we developed a new technique that uses rain area-specific gradient parameters to improve detected rain areas by correcting the number and sizes of the detected rain area. While comparing the rain area technique with one of the most successful satellite rainfall products, the Global Precipitation Measurement Integrated Multi-satellitE Retrievals for GPM (GPM IMERG), the results corroborate in terms of the spatial dynamics of the detected rain areas and rates.

Next, we improved the MWL's rainfall estimates for the first time in a new MSG technique that uses the rain area detection system developed in the previous study. In this investigation, we developed a new technique for wet-dry classification and baseline level estimation using the MSG-based rain area information. A new wet path length (wpl) parameter, representing the length of the raining MWL, was also developed based on the MSG-based rain area. This technique's wet-dry classification and baseline level estimation results agreed well with conventional technique results. The wpl parameter also remarkably improved the high rainfall intensities, which suggests that spatial rainfall variability over the MWL remains essential information to be considered in the MWL's rain rate retrieval.

Finally, a new technique was developed for high spatiotemporal rainfall estimation from MSG's cloud top properties and MWL rainfall intensities. This technique trains the random forest (RF) machine learning algorithm with the MWL's rainfall estimates to estimate rainfall from the MSG data. The results are convincing and promising. When compared to gauge rainfall data, the techniques' retrieval errors were comparable to errors found when comparing GPM IMERG and the European Organisation for the Exploitation of Meteorological Satellites Multi-sensor Precipitation Estimate (EUMETSAT MPE) rainfall intensities to gauge data. The spatial dynamics of the retrieved rainfall intensities also agreed well with these satellite products. The technique's advantage lies in retrieving high spatiotemporal resolution rainfall intensities regardless of the rainfall type.

Overall, this study demonstrates the great potential of using the MWL-MSG data for rainfall detection and estimation. In particular, the benefit of this rainfall observation alternative to areas with sparse or lacking conventional ground rainfall monitoring systems but growing MWL network and geostationary satellite (with capabilities like MSG) coverage may be invaluable.

Samenvatting

In veel operationele toepassingen, zoals bijvoorbeeld waterbeheer, landbouw, en het maken van hydro-meteorologische verwachtingen, zijn accurate observaties van neerslag, met hoge spatio-temporele resolutie, van groot belang. Daarnaast zijn neerslagdata met hoge resolutie ook nodig voor onderzoeksdoeleinden zoals het evalueren van neerslagproducten op basis van satellietdata, en het ontwerpen van nieuwe neerslagobservatie satellieten. Echter, de ruimtelijke en temporele variabiliteit van neerslag maken precieze bepaling van neerslag een grote uitdaging.

Nagenoeg alle bestaande technieken om regen te meten hebben inherente beperkingen. Zo zijn regenmeters in staat tot het nauwkeurig bepalen van regen op de positie van de meting, maar voor het verkrijgen van ruimtelijke informatie over de neerslag hebben deze puntmetingen duidelijke beperkingen. Deze informatie kan wel uit de data van buienradars worden gehaald, maar buienradars hebben onder andere als beperking dat de aanschaf, het onderhoud en het gebruik van de systemen grote kosten met zich meebrengen. Uit satellietdata kan informatie over regenval over grote delen van de wereld gehaald worden, maar de nauwkeurigheid van de uiteindelijke neerslagdata varieert sterk. Daarnaast is de resolutie van neerslagdata op basis van satellietdata niet altijd hoog genoeg om te voldoen aan de vereisten van bijvoorbeeld boeren, wiens percelen vaak aanzienlijk kleiner zijn (enkele honderden vierkante meters dan de hoogste resolutie van satelliet neerlsagdata (in de meeste gevallen minimaal enkele vierkante kilometers).

Aan het begin van de 21ste eeuw is een begin gemaakt met de ontwikkeling van een nieuwe techniek om regenintensiteit te bepalen. Deze techniek is gebaseerd op data van de alom aanwezige netwerkinfrastructuur van mobiele telefonie aanbieders. De signalen van microgolf straalverbindingen ondervinden signaalverlies door regen die in het microgolftraject valt. Onderzoekers hebben de ontvangen signaalsterkte bestudeerd, en hun bevindingen bewezen dat het signaalsterkte verlies gebruikt kan worden om met goede nauwkeurigheid de neerslagintensiteit op het traject te berekenen. Dit betekent dat iedere microgolfverbinding (hierna afgekort tot MWL, voor *microwave link*), het traject dat tussen de antennes van twee op elkaar gerichte microgolf zendontvangers, gebruikt kan worden ter bepaling van neerslagintensiteit op dat traject, met een zeer lage latentie van de meting. Dit opportunistisch gebruik als neerslagobservatie middel, dat bovendien zeer lage operationele en onderhoudskosten met zich meebrengt, heeft echter enkele nadelen; de nauwkeurigheid van de bepaling van de regenintensiteit is van vele factoren

afhankelijk. Voorbeelden zijn de variatie van druppelgrootte over het hele traject, en de invloed van natte antennes tijdens de regen en daarna.

Echter, er bestaan meteorologische geostationaire (GEO) observatie satellieten, zoals *Meteosat Second Generation* (MSG), die met hoge spatio-temporele resolutie en over een breed spectrum de atmosfeer van de aarde observeren, en daardoor in staat zijn om snel ontwikkelende buien en regengebieden te observeren, en informatie over de neerslag die daaruit valt te verstrekken. Er zijn veel studies uitgevoerd die satelliet regendata hebben vergeleken met traditionele observatiemethodes, zoals regenmeters en buienradars. Echter, onderzoek naar de combinatie tussen GEO satellieten en MWL data voor regenobservatie is tot op heden onderbelicht gebleven – terwijl deze combinatie zeer interessant is voor gebieden waar adequate regenobservatie systemen ontbreken.

Dit is de reden dat het doel van deze studie het onderzoek naar de mogelijkheden voor regendetectie en bepaling met hoge spatio-temporele resolutie voor midden- en West Kenia (gebieden met een ontbrekend of ontoereikend regenobservatie netwerk), op basis van MWL en MSG satellietdata. Dit hoofddoel werd gerealiseerd door middel van vier specifieke doelen, die de basis vormen van 4 individuele onderzoeken.

Ten eerste werd er onderzocht of de combinatie van MWL en MSG data gebruikt kon worden om de nauwkeurigheid van regenintensiteitsbepalingen te verbeteren. We onderzochten allereerst de mogelijkheden voor het gebruik van regenintensiteiten in het studiegebied, door deze met data van regenmeters te vergelijken. Vervolgens zijn de MWL regenintensiteiten met MSG data bestudeerd, waarbij een conceptueel model gebruikt werd. In dit conceptuele model werden hoge regenwaarschijnlijkheden en -intensiteiten toegekend aan wolken met een grote optische dikte en een hoge deeltjesdoorsnee. Het onderzoek bevestigde eerdere studies met betrekking tot de robuustheid van het gebruik van de MWL voor bepaling van regenintensiteit. Het bestuderen van de MWL regenintensiteiten met de MSG satelliet data gaf goede inzichten over de spectrale karakteristieken van dag- en nachtdata, alsmede over het verschil tussen pixels met en zonder regen. Tenslotte werden beschrijvende statistieken afgeleid van de spectrale karakteristieken van de satellietdata, op basis waarvan met succes regendetectie in het studiegebied uitgevoerd kon worden. Hieruit bleek echter wel dat de detectie gedurende de dag, die op basis reflectiedata plaatsvond, beter was dan die tijdens de nacht, welke op basis van infrarood (IR) data gebeurde. De oorzaak hiervan is dat uit de reflectiedata betere wolk optische dikte en deeltjesgrootte bepaald kon worden.

Volgend op de geslaagde combinatie van MWL-MSG data voor regenintensiteitsobservatie is er gekeken naar de ontwikkeling van gebiedsbepaling van regen, teneinde de van MWL data afgeleide regenintensiteiten te verbeteren, en vervolgens regen in het studiegebied te kunnen bepalen. In dit onderzoek zijn meerdere parametrische regendetectie modellen op basis van MSG zichtbaar licht en IR data geëvalueerd, met gebruikmaking van een conceptueel model, zoals ook in het vorige stadium gebruikt werd. Daarnaast hebben we een nieuwe techniek, op basis van gradiënt-parameters van regengebieden, ontwikkeld, teneinde regengebiedsdetectie te verbeteren en de aanvankelijk gedetecteerde regengebieden hiermee te corrigeren. Vergelijking van deze techniek met een van de meest nauwkeurige op satellietdata gebaseerde regenproducten, GPM IMERG (*Global Precipitation Measurement Integrated Multi-satellitE Retrievals for Global Precipitation Measurement*), liet erg goede overeenstemming zien in de ruimtelijke dynamiek van de gedetecteerde regengebieden en regenintensiteiten.

Vervolgens is er gekeken naar het verbeteren van regenbepaling op basis van MWL data, waarbij voor het eerst een nieuwe techniek op basis van de eerder afgeleide regengebiedsbepaling op basis van MSG data gebruikt werd. In dit onderzoek hebben we een nieuwe techniek voor de nat-droog classificatie en de bepaling van de basis signaalsterkte niveaus ontwikkeld, op basis van de uit MSG data afgeleide regengebiedsdata. Een nieuwe manier om de natte traject lengte (wpl) parameter te bepalen is ook op basis van deze data ontwikkeld. Deze parameter geeft aan over welk gedeelte van een MWL regen plaats vindt. De nat-droog classificatie en de basisniveau bepaling kwamen goed overeen met de resultaten van de conventionele techniek voor deze bepalingen. Het gebruik van de wpl parameter leidde tot een verbetering van de regenbepaling bij hoge regenintensiteiten, hetgeen aantoont dat ruimtelijke variabiliteit van regen langs het MWL traject essentiële informatie is, die bij het afleiden van regenintensiteiten uit MWL data gebruikt moet worden.

Tenslotte is er een nieuwe techniek voor hoog-resolute spatio-temporele regenbepaling op basis van MSG wolkentop karakteristieken en uit MWL-data afgeleide regenintensiteiten ontwikkeld. In deze techniek worden *random forest* (RF) machineleer algoritmes getraind met de MWL regendata om op basis hiervan met MSG data ruimtelijke regenintensiteiten te bepalen. De resultaten zijn overtuigend en veelbelovend. Een vergelijking met regenmeter data laat vergelijkbare afwijkingen zijn als wanneer GPM IMERG en EUMETSAT MPE (*European Organization fort he Exploitation of Meteorological Satellites Multisensor Preciptation Estimate*) regendata worden vergeleken met regenmeter data. De ruimtelijke dynamiek van de op onze manier berekende

regenintensiteiten komt ook goed overeen met deze satelliet regenproducten. Het voordeel van onze techniek ligt in het vermogen om, ongeacht het type neerslag, met hoge spatio-temporele resoluties goede regenintensiteiten te kunnen bepalen.

In het algemeen toont dit onderzoek het grote potentieel van de combinatie van MWL data en MSG data voor regendetectie en -bepaling aan. Vooral voor gebieden waarin conventionele *in-situ* regenobservatie netwerken ontbreken of dun gezaaid zijn, maar waar vaak zowel de dichtheid van MWL netwerken als de beschikbaarheid van bruikbare satellietdata met hoge resolutie alleen maar toeneemt, kan deze techniek zeer belangrijk zijn.

Acknowledgements

To God be the glory! This PhD journey has been unique and full of memorable experiences. The learning curve was steep, and the process was uncomfortable but relevant. In the end, I feel confident to acknowledge that I have increased in knowledge and experience. With great pleasure, I use this section to express my gratitude to all who contributed to this journey.

Firstly, I gratefully acknowledge Prof Su for giving me this PhD opportunity. Your advice, continual support and patience during this PhD journey were priceless. My gratitude extends to Dr Ben Maathuis; Ben was always available for countless meetings and detailed “pen and paper” discussions. I learned a lot from your guidance and experience. Equally important was the advice and support of Dr Joost Hoedjes. I will always remember and appreciate the time and support during the numerous field campaigns. Together, your invaluable advice, immense knowledge, and experience encouraged and guided me in all aspects of my academic research.

Besides my advisors, I am also grateful to Dr Oliver Gao from the School of Civil and Environmental Engineering at Cornel University and Dr Noam David from ATMOSCELL (an Israeli start-up) for their treasured support in numerous meetings and conversations vital to this PhD. I would also like to thank Bas Retsios for his help with programming and Drs Petra Budde for the data curation support.

This study would not have been possible without the support of Naftaly Muhoro and Maurice Miriti from Safaricom, who provided the commercial microwave link data. I also appreciate Paul Ruoya and the farmers in the Naivasha catchment who allowed us to carry out field measurements on their farm.

Many thanks to Donald Rwasoka for the fruitful discussions that shaped some of the ideas in this study. Likewise, I appreciate all my colleagues, especially Megan, Sammy and Rosa, for the happy times during this perilous journey.

Finally, yet importantly, I express gratitude to family and friends – my parents, who have been the source of enthusiasm and encouragement. Moreover, I appreciate the van Ommeren family, especially J.M.S van Ommeren for making life in Enschede homey.

Kwabena Kingsley Kumah
May, 2022
Enschede, Netherlands

Contents

List of Figures

List of tables

List of Abbreviations

ALOS	Advanced Land Observing Satellite
AMSR-E	Advanced Microwave ScanningRadiometer–EOS
AMSU	Advanced Microwave Sounding Unit
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness Temperature
BTD	Brightness Temperature Differences
CMORPH-KF	Climate Prediction Center Morphing-Kalman Filter
CPC	Climate Prediction Center
CWP	Cloud Water Path
DEM	Digital Elevation Model
DMSP	Defense Meteorological Satellite Program
DPR	Dual Precipitation Radar
ENSO	El-Niño Southern Oscillation
ERA-Interim	European Centre for Medium-range WeatherForecasts ReAnalysis Interim
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GEO	Geostationary
GMI	GPM Microwave Imager
GPCC	Global Precipitation Climatology Centre
GPM	Global Precipitation Measurement
GPM IMERG	Integrated Multi-satellitE Retrievals for GPM
GPROF	Goddard Profiling Algorithm
GSM	Global System for Mobile Communications
IDW	Inverse Distance Weighted
IOD	Indian Ocean Dipole
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
ITCZ	Intertropical Convergence Zone
LEO	Low Earth-Orbiting
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MJO	Madden-Julian Oscillation
MODIS	VIS/IR Moderate Resolution Imaging Spectroradiometer
MPE	Multisensor Precipitation Estimate
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
MWL	Commercial Microwave Link
NCEP-CFSR	National Centers for Environmental Prediction(NCEP) Climate Forecast System Reanalysis
NNET	Artificial Neural Network

NOAA	National Oceanic and Atmospheric Administration
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System
PMW	Passive Microwave
RF	Random Forest
RMSE	Root mean square error
RSL	Received Signal Level
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SSM/I	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SVM	Support Vector Machines
TAHMO	Trans-African Hydrometeorological Observatory
TMPA	TRMM Multisatellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
VIS	Visible

Chapter 1: Introduction

1.1 Background

a) Rainfall, why it is relevant and its implications

Rainfall is the amount of liquid precipitation that descends onto the Earth's surface from the atmosphere, whether on land or the oceans, expressed as the vertical water depth. It is a crucial geophysical variable of the hydrologic cycle: a solar radiation-driven process that ensures continual water movement between the atmosphere, land and oceans and regulates Earth's freshwater availability. This water cycling is linked with the energy cycling between the atmosphere, land and oceans that ultimately determines the Earth's climate and is responsible for much of its natural variability (Bolles, 2022; Duan & Duan, 2020). In its right amount, rainfall sustains aboveground primary productivity, species richness and ecosystem dynamics (Yan et al., 2015). Far too little rainfall and drought would devastate livelihood and economies, whereas too much rainfall, especially within a short time, puts life and properties at risk due to flooding.

Rainfall information about its quantity and pattern is valuable to many operational and research applications. For instance, to sustain more than 60% of the world's food through rainfed agriculture (FAO, 2020), farmers need timely rain forecasts to plan land preparations for growing seasons. Rainfall amount and variability are also used to infer discharge at a catchment's outlet, reservoir water balance, and streamflow forecasting, which informs decision-making in various water resource sectors, including hydropower production, dam protection and flood risk management (Ali & Shahbaz, 2020; Barua et al., 2013).

Additionally, rainfall is a crucial climate change variable linked to climate extremes such as drought and floods (Ojo & Ilunga, 2017). Past studies point toward an intensifying hydrologic cycle under a warming climate, with a 1 to 3% increase in global mean precipitation per degree rise in temperature (Allen & Ingram, 2002; Held & Soden, 2006). Studies show that roughly two-thirds of land on the earth will face a wetter and more variable hydroclimate on daily to multiyear timescales (Zhang et al., 2021). The recent IPCC report projects an alarming increase in temperature, frequency and intensity of heavy rainfall events over the African continent, increasing the severity and occurrence of droughts and floods with severe implications for food security, nutrition, involuntary migration and many more (IPCC, 2022).

Moreover, rainfall positively correlates with malaria, suggesting an increased risk of malaria transmission during rainy periods (Dabaro, Birhanu et al. 2021, Donkor, Kelly et al. 2021), and requires spatial rainfall pattern information to

track this vector-borne disease. Furthermore, several applications include: estimating groundwater table in hydrologic models (Kotchoni et al., 2018), estimating the amount of water in the soil capable of causing landslides (Rong et al., 2020), designing structures for runoff control, e.g. stormwater drains (Mendes et al., 2021) and monitoring water budget at the catchment level (Sieck et al., 2003) all require some rainfall information such as its amount, intensity and variability.

b) Measuring rainfall

Rainfall has high spatiotemporal variability, partly because it is a product of complex hydrometeorological processes that vary extensively in space and time (Rios Gaona et al., 2017). Spatially, rainfall variability is further complicated by factors such as orography and atmospheric processes. Temporally, rainfall is intermittent, and its distribution is highly variable. For instance, the global rainfall occurrence may be about 1 %. The modal rainfall is often zero, and while the distribution of other variables like temperature and pressure is nearly normal, rainfall's frequency distribution skews towards zero (Kidd, 2001). Due to these factors, characteristics and variability, capturing a complete and accurate picture of rainfall is challenging. Rainfall measurements broadly fall under direct and indirect techniques, albeit they continue to evolve.

i) Direct technique

The direct technique is done mainly using rain gauges and sometimes disdrometers. Rain gauges are the oldest and widely used measurement technique (Sene, 2013; Strangeways, 2010). They measure rainfall accumulation as a function of time and are broadly classified into recording or non-recording types. The latter accumulates rainfall over time using a graduated cylinder. After a rain event, the cylinder's water level directly represents the rainfall accumulation at that point. An observer manually records the measurements at a fixed time each day. The recording types use varying techniques. For instance, the tipping bucket type uses buckets placed on a fulcrum and balanced 'like a see–saw' under a rain collector. When rainwater of a calibrated volume expressed as rainfall depth accumulates in one bucket, it tips over and empty's itself while the other bucket replaces it instantaneously. This tipping action triggers a recorder that logs the time and counts based on which rain rate can be estimated. Another recording type is the weighing rain gauges. They record the mass of rainwater collected in a storage bin seated on a scale. Rain rate is measured as the difference in rainwater accumulation over time. Optical rain gauges are yet other recording types that deduce rain rate information from rain-induced attenuation (as a

function of time) of an optical signal beaming from a light source to an optical detector. Acoustic rain gauges are an even more interesting recording type; they quantitatively measure rainfall from the naturally occurring underwater sound produced by raindrops (Michaelides, 2008; Nystuen, 1998; Nystuen et al., 1996; Sene, 2013).

Instead, disdrometers provide more detailed rain hydrometeor information such as the count, drop size, shape and fall velocity. The rain rate and accumulation are then computed by summating the drops counted and sized over the sensing area per time. The common types are impact, imaging and optical disdrometers. The impact type uses a sensing head to convert the momentum of individual raindrops striking the sensor head to an electronic signal proportional to the drop size. The exact detail of the conversion is dependent on the sensor head and transducer design. Imaging disdrometers use two imaging cameras to capture the size, shape and fall velocity of raindrops passing through the camera's field of view. On the other hand, the optical type uses laser transmitters and receivers, a short distance apart, to determine the size, shape and velocity of raindrops passing through the laser beam (Michaelides, 2008; Thurai et al., 2011).

ii) Indirect technique

The indirect techniques detect and measure rainfall quite differently than the direct techniques. They remotely sense rain hydrometeors well above the ground in spatial measurement volume. Two techniques are used in rainfall remote sensing: active and passive techniques. The active rainfall remote sensing techniques use devices that beam pulses of electromagnetic radiation into the atmosphere and measure backscattered radiation or echoes from targets in clouds such as rain hydrometeors. The passive technique devices measure the incoming radiation from clouds and precipitation, more like cameras without a flash.

Weather radars are a typical example of active rainfall remote sensing devices. They measure rainfall from the radar reflectivity factor, Z (mm^6/m^3), empirically related to rainfall via a power-law relation, e.g. (Marshall et al., 1947). The Z quantifies the radar echoes proportional to the volume of all hydrometeors detected by the radar beam (Wilson & Brandes, 1979). Radars are typically characterised by the pulse frequency (or wavelength) they transmit, which has important implications on their ability to sense clouds and precipitation remotely, such as the beamwidth for a given antenna size and sensitivity to precipitation. For instance, radiation of different wavelength respond differently to clouds and precipitation hydrometeors; attenuation of the transmitted

radiation is more for small than long-wavelength radiation. Additionally, a larger antenna size leads to a smaller beamwidth for a given wavelength. Different frequency bands may be used for radar remote sensing; traditionally, categories of weather radars include precipitation (S-X bands) and cloud radars (Ku-W bands) (Hong & Gourley, 2015; Kumjian, 2018).

The passive rainfall remote sensing devices are mostly spaceborne meteorological satellites. They measure incoming radiation scattered and emitted from clouds and precipitation using visible (VIS), infrared (IR) and passive microwave (PMW) sensors on geostationary (GEO) and low earth-orbiting (LEO; including polar-orbiting) satellites. GEO satellites are located 35800 km above the equator in orbit with a singular speed the same as the Earth and therefore appear stationary relative to a location on Earth (Kidd et al., 2010a). Current GEO satellites include EUMETSAT's MSG (Schmetz et al., 2002), US's GEOS-R, S and T (https://www.goes-r.gov/mission/mission.html, accessed on April 19, 2022), and Chinese's FengYun-2 (Guo et al., 2018) satellites. Although their sensor specifications differ, these satellites share common characteristics; they carry VIS and IR sensors with spatial and temporal resolutions of about 1–4 km and 15–30 minutes (Kidd et al., 2010a). Planned GEO satellites missions such as EUMETSATS's MTG-I, S (Holmlund et al., 2021; Stuhlmann et al., 2017) and US's GEOS-U satellites (https://www.goes-r.gov/mission/mission.html, accessed on April 19, 2022) have more advanced atmospheric imaging and sounding capabilities with better measurement resolution than the existing ones.

The LEO satellites orbit the Earth such that they pass over the equator at the same local time, providing about two overpasses a day. On top of the VIS and IR sensors, LEO satellites also carry PMW imagers and sounders capable of direct precipitation measurement. The 'direct precipitation measurement' implies that the PMW radiations interact strongly with precipitation particles because their sizes are similar to the PMW wavelength. It does not directly observe the surface rainfall like e.g. rain gauges. Some currently existing polar-orbiting satellites include the NOAA series of satellites 18 and 19 (NOAA, 2009) and EUMETSAT's Metop series (Schmetz et al., 2007), orbiting the Earth every 100 minutes at an altitude of about 850 km. They carry a wide range of instrumentation, including the third-generation VIS/IR AVHRR and AMSU (Kidd et al., 2010a).

Besides operational meteorological satellites, military and research satellites are also used to estimate precipitation. These include the DMSP satellites series: SSM/I, SSMIS (Kunkee et al., 2008), AMSR-E and the MODIS onboard Terra and Aqua satellites (Kawanishi et al., 2003; Yan & Yang, 2007). Perhaps, the most dedicated satellite missions for detailed precipitation studies are the TRMM and

GPM satellites (Kummerow et al., 1998; Kummerow et al., 2007). The latter constitutes a significant advancement of the previous TRMM satellites, carrying the only spaceborne DPR and GMI. The DPR can create 3D structure and intensity estimates of precipitation, and the GMI has a greater frequency range than TRMM. Additionally, GPM has a broader global coverage (65^{o} north-south) than TRMM (35^{o} north-south) (Blumenfeld, 2015).

Owing to the plethora of satellite observations, many satellite rainfall retrieval algorithms have emerged that can be categorised based on the type of observation, namely, VIS/IR, PMW and multi-sensor algorithms. VIS observations compare closely to the human eye. Since clouds appear brighter from above, VIS observations infer rainfall with the assumption that the brightness of a cloud is related to its thickness, and thick clouds are likely to rain (Kidd et al., 2010a). The limitations of a VIS algorithm are that the observations are only available during the daytime. The data processing is also complicated because one has to account for factors such as changes in the Sun angle. Cloud top properties, including size and phase (whether ice or water hydrometeor), can also be inferred from reflected/emitted radiation in the near-IR observations for successful rainfall estimation (Rosenfeld & Gutman, 1994; Rosenfeld & Lensky, 1998). However, the observations are limited by solar illumination, restricting their usage to full daylight operations.

The thermal IR observations are related to emissions from an object and therefore have day and night usefulness. Heavier rainfall is linked to taller and larger clouds with cold tops. Thus, by observing the cloud top temperature, one can derive rainfall estimates, e.g. (Arkin & Meisner, 1987; Arking & Childs, 1985). Multichannel/multispectral techniques such as (Lazri et al., 2014; Lensky & Rosenfeld, 1997) combine observations from VIS, near-IR, water vapour and thermal IR to estimate rain area and intensities. However, several studies have focused on using multispectral channel observations for rain area delineation (Feidas & Giannakos, 2010; Lazri et al., 2013; Thies et al., 2008a, 2008c).

The PMW retrievals rely on the Earth's natural MW emissions. Two retrieval process: emission and scattering processes, which depends on the background emissivity, is used in the rainfall retrieval. The emission process is caused by raindrops and leads to increased MW radiation, whereas scattering, caused by ice particles, leads to decreased radiation. Over water, background emissivity is low and constant; thus, low frequency (< 20 GHz) PMW channels are used for rainfall retrieval from additional emissions from raindrops. Rainfall retrieval over land uses high frequency (> 35 GHz) channels to measure decreased radiation due to scattering by ice particles (Kidd et al., 2010a).

The PMW retrieval algorithms may use emission (Wilheit et al., 1991), scattering (Ferraro & Marks, 1995) or multichannel inversion (Kummerow et al., 2001) type techniques. Further, these can be split into empirical techniques, which are relatively simple and calibrated against surface data or physical techniques, which minimise the differences between modelled and observed radiance (Kidd & Levizzani, 2011). More successful physical type algorithms such as (Kummerow et al., 2001) use an a priori database of the model-generated atmospheric profiles compared with satellite observations. The physical techniques provide surface rainfall and more detailed information about precipitation (Kidd & Levizzani, 2011).

Typically, the multi-sensor algorithms employ synergistic use of the LEO PMW and the GEO VIS/IR observation to overcome their individual deficiencies when used for rainfall retrieval. Some techniques adjust the IR radiance or generate calibration curves to map IR radiance using the gauge, radar or satellite datasets. The TMPA uses PMW imaging and sounding sensors and GEO IR data, adjusted for different satellite retrievals, to estimate rainfall (Wolff et al., 2007). The GPM IMERG final run version 5 (V05) ingests datasets from the GPM core observatory, a constellation of PMW satellites, GEO IR observations and rain gauges from the GPCC to produce a single rainfall product (Hosseini-Moghari & Tang, 2020). Other techniques, such as the PERSIANN algorithm, use artificial neural networks to derive rainfall estimates from multichannel and multi-sensor observations (Hsu & Sorooshian, 2008). However, the indirect nature of the IR observation to sense rainfall limits these techniques. Nevertheless, the IR observations provide useful cloud motion information capable of morphing PMW observations between successive satellite overpasses.

Some studies suggest that lightning data may improve rainfall retrievals (Biron et al., 2012; Garcia et al., 2013; Grecu et al., 2000). For instance, Grecu et al. (2000) assessed the use of lightning information in convective rainfall estimation. They found a reduction of about 15% in RMSE of the estimated rainfall defined by convective areas. This was attributed to the fact that the errors caused by missing convective areas due to the absence of lightning were smaller than errors due to overestimated convective areas due to cirrus clouds assuming only IR data is used.

c) The limitations in the rainfall measurement techniques

The various rainfall retrieval techniques have inherent limitations. A significant drawback of the direct technique is that they observe rainfall from a single point and do not provide spatially representative rainfall information. Though spatial interpolation techniques may retrieve spatial rainfall information from their network, they easily propagate errors from the point measurements. Also, its accuracy is dependent on the density of the observing systems. Another problem common to the direct technique is the effect of environmental factors such as wind and evaporation (Michaelides et al., 2009; Villarini et al., 2008).

Weather radars are limited because the coefficient of the power-law relation that translates radar reflectivity to rain rates lacks unique information that characterises microphysical variability of different meteorological systems ranging from cold frontal to summer thunderstorms and tropical systems. Further, radars suffer from ground clutter and complex terrain effects like beam blockage, beam overshoot and range effects. Additionally, radars are expensive to acquire, operate and require technical and human resources to ensure adequate system maintenance, limiting their operations to well-funded meteorological institutions (David et al., 2013; Hoedjes et al., 2014; Michaelides et al., 2009).

The major limitation in the VIS/IR rainfall retrieval technique is that its relates precipitation to properties at the cloud top, e.g. cloud top temperature. However, precipitation of a given cloud may depend on the cloud's composition, namely cloud particle size and phase. These properties determine a cloud's precipitation properties and the difference between raining and non-raining clouds of similar top properties (Rosenfeld, 2007). Therefore, the technique often leads to retrieval uncertainties. For instance, very high cold cirrus clouds may be misinterpreted as raining, and low-level raining clouds may be missed. On top of this, there is also the issue of variation in the cloud top property-rainfall relation during the lifetime of the rainfall event, between rain systems and climatological regimes (Kidd et al., 2010a).

The limitations of PMW retrievals are that PMW sensors are on LEO platforms and therefore have a limited temporal sampling. Also, the spatial resolution of the PMW observations, 50 by 50 km (over oceans) and 10 by 10 km (over land), may not be ideal for rainfall retrieval (Kidd & Levizzani, 2011; Kidd et al., 2010a).

Furthermore, the PMW retrievals over land are challenged by the variable emissivity of soil, vegetation, and water and their variability from one location to another (Michaelides et al., 2009). Additionally, the spatial and temporal resolution difference between GEO and PMW observations and between

multispectral PMW observations require resampling and spatiotemporal collocation to combine the data. This preprocessing step may introduce uncertainties that may propagate into the retrieved rainfall based on multi-sensor techniques (Alemseged & Rientjes, 2007; Bellerby & Sun, 2005).

1.2 Commercial Microwave Link for gauging rainfall

MWL are point-to-point radio connections between antennas that commercial cellular communication service providers use for data transmission from radio, TV, internet and wireless communication between our cell phones. Rainfall physically affects the electromagnetic wave transmission between the transmitting and receiving antennas of MWL. When raindrops interact with the electromagnetic waves, a portion of the energy is scattered or absorbed, the combined effect of which reduces the energy detected at the receiving antenna (Leijnse et al., 2007a), see Figure 1. 1.

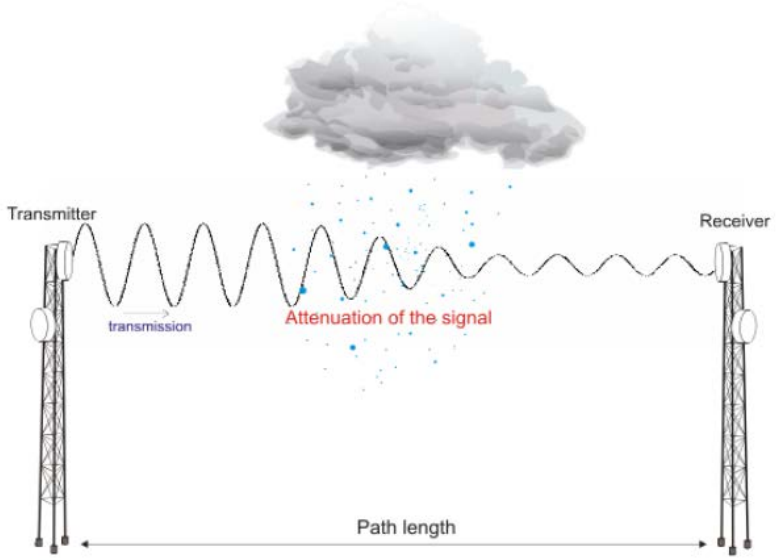

Figure 1. 1 Rain-induced attenuation of MWL signal between a transmitting and receiving antenna. (Obtained *from* https://www.epfl.ch/labs/lte/research/past-research/page-50160-en-html/)

Telecom engineers have long known and studied the relation between rainfall statistics and attenuation to design MWL networks to avoid rain-induced disruptions in signal transmission (Hogg, 1968; Olsen et al., 1978).

Studies in the recent decade (Leijnse et al., 2007a; Messer et al., 2006) have, however, shown that this relative decrease in power (dB) per unit distance, A (dB km^{-1}), can be used to retrieve average rainfall intensities along the signal transmission path. Consequently, the technique has received global research attention, and its application has been demonstrated for various hydrometeorological applications such as (David et al., 2009; Hoedjes et al., 2014; Overeem et al., 2013; Overeem et al., 2021; Pastorek et al., 2019). The global MWL network is denser than, e.g. rain gauges; estimated globally at 4 million (ERICSSON, 2015; Uijlenhoet et al., 2018) compared to about 150000 global rain gauges (Kidd et al., 2017), pointing toward an enormous and untapped potential for continental-scale rainfall monitoring.

Figure 1. 2 presents the global map of mobile service coverage across croplands (Figure 1. 2a) (Mehrabi et al., 2020) and GPCC's global rain gauge network coverage (Figure 1. 2b) (GPCC, 2020). In particular, the mobile service's wide coverage over agricultural fields in the Sub Saharan Africa compared to GPCC, which operates a few gauges per 1-degree grid box in the region, points towards a huge potential of the MWL network system for ground rainfall monitoring to benefit, e.g., agricultural management in an area where, compared to other areas, the traditional ground monitoring systems are insufficient or lacking.

Since these MWL systems already exist in the field, including widely ungauged areas and are maintained by the service providers, their operational cost for rainfall monitoring may be minimal (Kumah et al., 2021). Nonetheless, several challenges may hinder the fruition of such a goal. For instance, there are no standard procedures for acquiring the MWL data, making it challenging to access spatiotemporally continuous datasets, especially in developing countries. Besides, the MWL network is not designed for rainfall monitoring, with often an arbitrary distribution and biased towards densely populated areas that complicate accurate high spatial resolution rainfall mapping. Additionally, variation in raindrop size distribution along the MWL propagation path and issues of antenna wetting during and after a rain event limit the accuracy of the MWL rainfall estimate (Chwala & Kunstmann, 2019; Uijlenhoet et al., 2010; Zinevich et al., 2008).

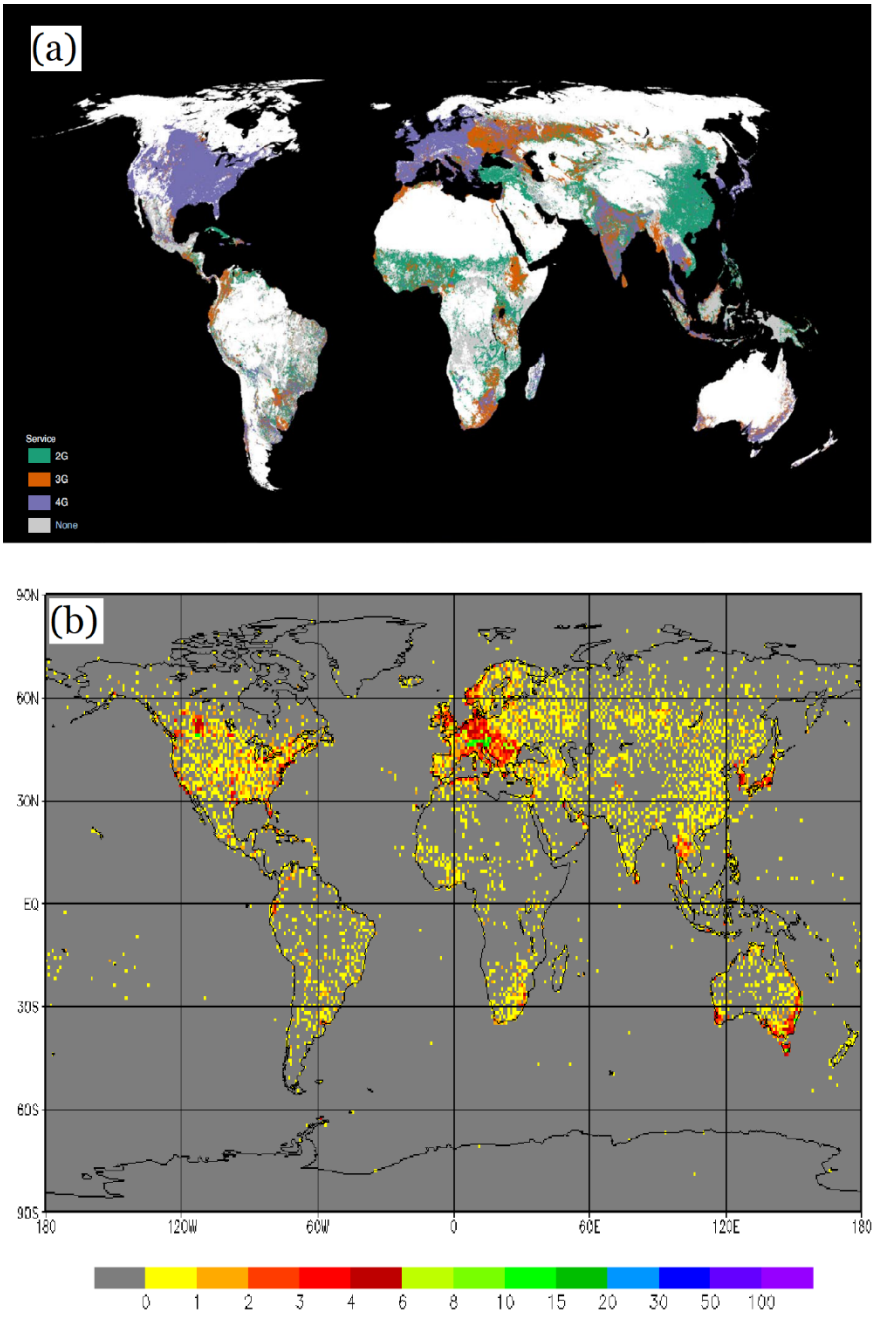

Figure 1. 2 (a) Global map of 2G, 3G and 4G mobile service coverage across croplands, acquired from (Mehrabi et al., 2020) (b) Number of rain gauges per 1-degree grid used by GPCC as of 2012, acquired from (GPCC, 2020).

1.3 Meteosat Second Generation

The MSG is a series of operational GEO meteorological satellites continuously observing the earth-atmosphere system from the equator and operating over Europe, Africa and the Indian Ocean (EUMETSAT, 2021). Its radiometric sensor, SEVIRI, has 12 spectral channels: eight thermal infrared, three solar, and one high-resolution visible (HRV) channel. SEVIRI's temporal resolution is 15 min, and its spatial resolution is 3 km for the thermal infrared and solar channels and 1 km for the HRV channel at nadir (Schmetz et al., 2002). MSG's measurement characteristics are suitable for detecting rapidly developing severe weather, weather forecasting and climate monitoring. On top of this, several studies have combined the SEVIRI spectral channels to infer cloud top properties such as optical thickness, particle size, and phase, which can successfully detect and estimate rainfall (Bendix et al., 2010; Giannakos & Feidas, 2012; Lazri et al., 2014; Thies & Nauss, 2008). MSG data is freely available in EUMETSAT's data archives (EUMETSAT, 2020).

1.4 What this thesis is about and how it is structured

a) The objective and its underlying motivation

The main objective of this dissertation is to investigate the MWL and MSG data for spatial and temporal rainfall detection and estimation. The following specific objectives investigate this main objective.

1. To investigate if the combination of MWL with MSG satellite signals could improve rainfall detection and rainfall rate estimates.
2. To develop an improved rain area detection system using MSG's high-resolution data.
3. To investigate the applicability of MSG's high-resolution rain area detection for improving the MWL rainfall intensity estimates.
4. To estimate high spatiotemporal resolution rainfall from the MWL and MSG data.

The study objective is motivated by the fact that, until now, most studies have independently used the MWL and MSG data to detect, estimate, and map rainfall. The existing MWL rainfall estimation and mapping studies are mainly based on dense MWL network data, and they have often focused on using geostatistical techniques for spatial rainfall retrieval from the MWL (Overeem et al., 2016a; Overeem et al., 2016b). However, the MWL network is arbitrarily distributed, and

the spatial density is often non-uniform. This may challenge geostatistical-based spatial rainfall retrieval from MWL in areas with sparse networks and locally intense rainfall.

On the other hand, the MSG satellite data has been widely used for high spatiotemporal resolution rainfall detection and estimation, albeit with gauge and radar data, which is not spatially dense everywhere (Bendix et al., 2010; Feidas & Giannakos, 2010; Moraux et al., 2019). Coincidentally, the MSG satellite's temporal measurement (15-minutes) and spatial coverage on land often match the MWL data, while both systems observe rainfall independently. When combined, the two systems will suit rainfall retrieval and monitoring over vast areas, including areas with sparse ground observations. Given this, new and improved techniques that improve spatiotemporal rainfall retrieval from combined MWL and MSG data may be a promising alternative that may benefit applications in vast areas, particularly areas with sparse or insufficient ground observation.

For this reason, this study investigated the main objective using data from Central and Western Kenya (longitudes 34°E, 42°E and latitudes of 5°N, 5°S), where ground data is often sparse. The area also plays a crucial role in Kenya's agricultural production and is frequently affected by drought and floods (Nicholson, 2017). It is characterised by the Kenyan Rift Valley, mountains (e.g. Mount Kenya), and forest complexes (e.g. the Mau Forest). It experiences two rainy seasons: long rains from March to June (MAMJ) and short rains from October to December (OND), which coincide with the ITCZ passage over the equator. Furthermore, ENSO, IOD, MJO, and relief features influence the area's local rainfall amounts, patterns, and characteristics (Brian et al., 2016; Nicholson, 2017; Ogwang et al., 2014).

b) The dissertation outline

This dissertation is organised into 6 chapters. Chapter 1 presents the relevance and implications of studying rainfall, the measurement techniques and their spatiotemporal caveats. The MWL and MSG are introduced in this chapter as independent rainfall observing systems. Chapter 2 investigates specific objective 1 by first assessing the MWL's ability to estimate accurate rain rates in the study area using multiple rain gauge data. Next, the satellite signals that indicate raining and non-raining cases on the MWLs were investigated for rainfall detection using a conceptual model.

Chapter 3 was dedicated to the second specific objective. The chapter investigates the MSG data for rain area detection by evaluating multiple rainfall detection models developed from solar and thermal IR channels. The chapter also developed a new gradient-based adaptive correction technique to improve the rain area retrievals from the MSG data. Chapter 4 investigated specific objective 3. Here, a new method to improve the MWL rainfall estimation using rain area detections in Chapter 3 is developed, evaluated and described.

Chapter 5 investigated specific objective 4 by applying machine learning techniques to the MWL and MSG data to estimate high spatiotemporal resolution rainfall. Finally, Chapter 6 is the conclusions and study implications chapter. Here, the significant conclusions and their research and operational implications are summarised.

Chapter 2: The MWL and MSG SEVIRI Satellite Combination for Rainfall Detection and Estimation

This chapter is originally published in:
Kumah, K.K.; Hoedjes, J.C.B.; David, N.; Maathuis, B.P.; Gao, H.O.; Su, B.Z. Combining MWL and MSG SEVIRI Satellite Signals for Rainfall Detection and Estimation. Atmosphere 2020, 11, doi:ARTN 884 10.3390/atmos11090884. 17

Abstract

Accurate rainfall detection and estimation are essential for many research and operational applications. Traditional rainfall detection and estimation techniques have achieved considerable success but with limitations. Thus in this study, the relationships between the gauge (point measurement) and the MWL rainfall (line measurement) and the MWL to the satellite observations (area-wide measurement) are investigated for (area-wide) rainfall detection and rain rate retrieval. More precisely, we investigate if combining MWL with MSG satellite signals could improve rainfall detection and rain rate estimation. The investigated procedure includes an initial evaluation of the MWL rainfall estimates using gauge measurements, followed by a joint analysis of the rainfall estimates with the satellite signals by means of a conceptual model that clouds with high cloud top optical thickness and large particle sizes have high rainfall probabilities and intensities. The analysis produced empirical thresholds that were used to test the capability of the MSG satellite data to detect rainfall on the MWL. The results from Kenya during the "long rains" of 2013, 2014 and 2018 show convincing performance and reveal the potential of MWL and MSG data for area-wide rainfall detection.

2.1 Introduction

Accurate rainfall detection and estimation are beneficial for many operational and research applications, including hydrological modelling, flash flood prediction, urban drainage planning, water resources management and many more (Hong et al., 2007). However, accurate rainfall estimation is a challenge because rainfall is intermittent, and its intensities, in some cases, vary significantly in space and time (Cristiano et al., 2017; Levizzanni et al., 2007). The state-of-the-art rainfall measurement consists of devices that can detect and quantify rainfall depending on its location (Michaelides et al., 2009).

Rain gauges measure rainfall accumulations as a function of time and generally have a high degree of accuracy, especially at low to medium intensities (Rios et al., 2012a). However, rain gauges measure at a discrete point, providing site-specific measurements with low spatial representativeness (Villarini et al., 2008). Interpolation techniques to obtain spatially continuous rainfall estimates from rain gauges are available. Nevertheless, they easily propagate errors from the point measurement and cannot adequately capture the spatial variability of rainfall (Tang et al., 2016). Also, installing and maintaining rain gauge networks can be practically challenging and expensive (Barthès & Mallet, 2013; Upton et al., 2005), resulting in sparse deployment and a rapid decline in gauge stations, especially in developing and underdeveloped countries (Dinku et al., 2008; Kidd et al., 2017).

Weather radar systems usually operate at S or C-band wavelengths, with better coverage and high spatiotemporal resolution (Lengfeld et al., 2014). The challenge is that radar-based rainfall estimation suffers from limitations, such as ground clutter, beam blockage, attenuation due to rain, mean-field and range-dependent systematic errors, see, e.g. (David et al., 2013; Dinku et al., 2002; Uijlenhoet & Berne, 2008; Villarini & Krajewski, 2009). Additionally, radars are expensive to acquire, operate and require technical and human resources to ensure adequate system maintenance. This limits their operation to well-funded national hydrometeorological institutions with the resources to ensure the valorisation of the data in the form of relevant information products (Barthès & Mallet, 2013; Hoedjes et al., 2014).

Alternatively, satellite systems can provide continuous and global-scale rainfall observation at different spatiotemporal resolutions. Satellite systems remotely sense scattered and emitted radiation from clouds, precipitation and underlying surface to estimate rainfall (Rosenfeld, 2007). This is achieved using visible and infrared sensors onboard geostationary satellites and passive and active microwave sensors onboard polar-orbiting satellites. Consequently, many

satellite rainfall estimation techniques exist e.g. Kidd et al. (2010b) provide an overview of satellite estimation techniques. Some of the most accurate satellite-based rainfall products, e.g. (Paredes Trejo et al., 2016), incorporate in-situ measurement in the retrieval process. Nonetheless, uncertainties in satellite rainfall estimates exist at varying spatiotemporal scales (Han et al., 2010). Furthermore, reliable satellite rainfall estimates require ground validation and evaluation at different spatiotemporal scales using sufficient ground data (Dezfuli et al., 2017; Doumounia et al., 2014).

Several studies in the past decade pioneered by (Leijnse et al., 2007b; Messer et al., 2006) have shown that Microwave Links (MWL) from cellular communication networks can provide near-ground average rainfall estimates. Potentially, this rainfall estimation technique could have many benefits. For instance, it is possible to estimate rainfall over large areas due to their extensive coverage (on land). Furthermore, the MWL estimates line-average rainfall, which is more representative of areal rainfall than the rain gauge's point estimates. Uijlenhoet et al. (2018) provide an overview of the history, theory, challenges and opportunities of large-scale MWL rainfall monitoring. A review of current and future challenges to MWL rainfall monitoring can also be found in (Chwala & Kunstmann, 2019).

Many studies have also combined different rainfall estimation methods to improve near-ground rainfall estimates at different spatiotemporal scales (Bianchi et al., 2013; Liberman et al., 2014; Scheidegger & Rieckermann, 2014; Sebastianelli et al., 2013; Todini, 2001). Sebastianelli et al. (2013) used rain gauge data to correct the radar rainfall estimates. By combining radar-based areal precipitation fields with point rain gauge measurements, Todini (2001) improved rainfall accuracy and spatial distribution. Furthermore, merging techniques involving MWL, radar and rain gauge data have been described by (Bianchi et al., 2013; Liberman et al., 2014; Scheidegger & Rieckermann, 2014).

The MSG is a GEO weather satellite with a wide spectral range radiometer, SEVIRI. It has eleven spectral channels of 3 km × 3 km and one high resolution visible (HRV) channel of 1 × 1 km nadir resolution that observes the Earth's atmospheric state and dynamics every 15 min (Schmetz et al., 2002). These individual SEVIRI spectral channels and combinations infer cloud top properties such as cloud top temperature for successful rainfall detection and estimation (Bendix et al., 2010; Roebeling & Holleman, 2009; Thies et al., 2008a, 2008b).

Surprisingly, the combination of MWL and MSG satellite for rainfall detection and estimation has received far too little attention, albeit it could be of great value

to area-wide rainfall monitoring. The MWL-MSG satellite combination, in which MSG observes the Earth's atmosphere at high spatial and temporal resolution and the MWL estimates accurate near-ground average rainfall, can essentially benefit large-scale rainfall detection and estimation. To date, van het Schip et al. (2017) investigated the MSG-based satellite rainfall product's potential for wet and dry classification of MWL signals in the Netherlands. They suggested that since the MWL estimates rainfall close to the ground, their combination with satellite data can provide better estimates than a satellite-only approach. Hoedjes et al. (2013) Hoedjes et al. (2014) proposed the MWL and MSG satellite data as suitable for estimating rainfall from convective systems and developing conceptual flash flood early warning systems for underdeveloped countries.

This chapter's objective is to investigate if the combination of MWL with MSG satellite signals could improve rainfall detection and rainfall rate estimates. Contrary to other MWL-based rainfall studies, for the first time, the path average rainfall is studied together with signals from MSG SEVIRI channels that provide information on cloud dynamics. Our approach includes: (i) the MWL rainfall was first evaluated using rain gauge measurements; (ii) secondly, the MWL rainfall estimates were analysed as a function of the MSG SEVIRI satellite signals. The satellite signals, in this case, were used to infer information on cloud top properties, (iii) finally, the knowledge gained from analysing the MWL rainfall with MSG SEVIRI satellites signals detected rainfall on individual rainfall MWL.

This chapter's sections are organised as follows. Section 2.2 presents the study area and the data used. Section 2.3 describes the method and performance measures used to evaluate and analyse MWL-based rainfall intensities and their relationship with MSG cloud top properties. Section 2.4 briefly presents and discusses the results, and lastly, in Section 2.5, significant findings and conclusions are summarised.

2.2 Study Area and Dataset

2.2.1 Study Area

Kericho (0.36° S, 35.28° E) and Naivasha (-0.71° S, 36.43° E), Kenya (0.02° S, 37.90° E) are the study areas for this study. The two locations were chosen because of MWL, rain gauge, and satellite data availability. Located within the Kenyan Rift Valley, both study locations are dominated by farmland. They also have similar rainfall patterns, with a long rainy season occurring in March, April, May, and June ("long rains") and a shorter rainy season in October, November, and December ("short rains") (Kimani et al., 2017). The ITCZ's seasonal passage over Kenya influences the long and short rainfall seasons (Schneider et al., 2014).

The two study locations share similar complex terrain features: high mountains, dense forests, farmlands, water bodies, and fairly populated urban areas. Kericho is characterised by hilly terrain, with elevations ranging between 1800 and 3000 m above mean sea level (a.m.s.l). Elevation in Naivasha ranges from 1980 m a.m.s.l. close to lake Naivasha to about 4000 m a.m.s.l in the Aberdare Mountains. Rainfall in both areas also varies quite noticeably with the local relief. On average, total annual rainfall varies between low and high altitudes, from 1400 to 2125 mm and 610 to 1525 mm in Kericho and Naivasha, respectively (MoAlf, 2017; Odongo et al., 2015). The temperature in Kericho ranges between 10 and 29 °C (KERICHO, 2015) and in Naivasha, between 8 and 30 °C (Odongo et al., 2015).

2.2.2 Dataset

This study focused on rainfall data for the "long rains" in May and June 2013, 2014, and 2018 (Table 2. 1). These data were obtained from MWL, rain gauges and MSG SEVIRI. Part of the data set, consisting of 2 MWL and 16 rain gauges, was used to evaluate the MWL's capability to estimate rainfall intensities. All data analysed the relationship between ground rainfall (from rain gauges and MWL rainfall estimates) and MSG satellite data for rainfall detection on MWL. The rain gauge data served as the reference data.

Table 2. 1 Characteristic of the MWL network per each study location

Study Location	Evaluation Period		Number of MWL	Frequency (GHz)	Link Length (km)
	Year	Month			
Kericho	2013	May–June	2	23	<2
			4	15	3.45–4.77
Naivasha	2014	May–June	3	23	<2
			9	15	3.47–18.95
			1	8	28.4
	2018		1	15	10

2.2.2.1 Rain Gauge Data

The gauge rainfall data were from 14 aerodynamic 'tipping bucket' rain gauges (ARG 100 rain gauges, see www.emltd.net) and two TAHMO (van de Giesen et al., 2014) rain gauges. In Kericho, five ARG rain gauges were aligned near a 15 GHz, 3.68 km MWL transect to collect the reference rainfall data during the May–June 2013 evaluation period. In Naivasha, nine ARG rain gauges aligned under a 15 GHz, 10 km MWL, and two TAHMO gauges installed near its transmitting and receiving antennas collected the reference data for the May–June 2018 evaluation period. Figure 2. 1 shows the gauges' proximity to the MWLs per each

study location. The ARG rain gauges were set to log data every minute, while the TAHMO stations recorded rainfall every 5 min. One tip of the ARG bucket equates to 0.198 to 0.202 mm of rain. Because no rain gauge data were available during the 2014 period, its evaluation did not consider ground-based rainfall information.

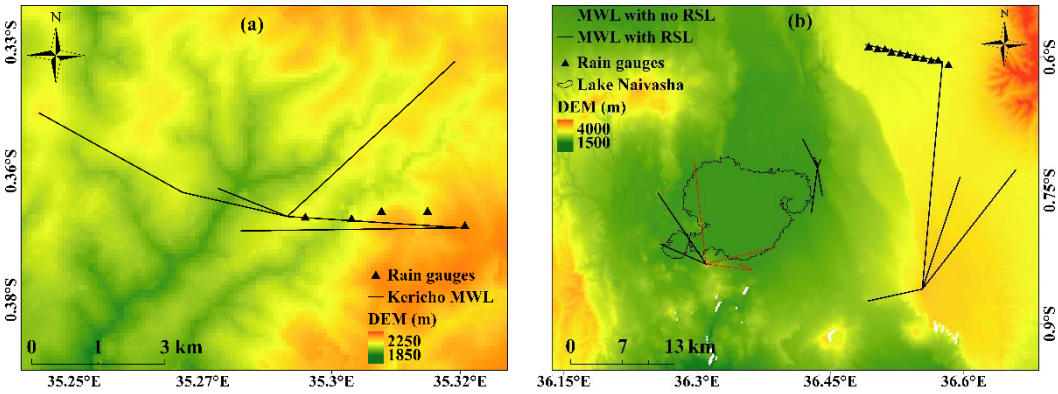

Figure 2. 1 The MWL and rain gauges in (a) Kericho and (b) Naivasha. The base map is SRTM DEM over the locations. Note: map coordinates are in decimal degrees; some MWL in Naivasha (red lines in (b)) did not have RSL data.

2.2.2.2 MWL data

Safaricom, a Kenyan telecom service provider, supplied the MWL data. Safaricom routinely collects and stores the MWL data for monitoring purposes. This study acquired received signal level (RSL) data for 19 MWLs of variable lengths and frequencies for the evaluation period. Figure 2. 1 shows the MWL network in Kericho and Naivasha on a base map using a Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) (EROS). All the MWLs used were Aviat Eclipse. Table 2. 1 gives further details on the MWL network per each study location and evaluation period.

The RSL data were characterized by minimum, maximum, and mean values at 15 min intervals and a resolution of 0.1 dBm. The MWL used were vertically polarized and had constant transmitted signal levels (TSLs). Out of the 19 MWL, one link (8 GHz, 28.4 km) was not included in this study because, at such frequencies, the attenuation rainfall relationship for estimating path average rainfall is sensitive to variation in raindrop size distribution, which can result in significant rainfall retrieval errors (Atlas & Ulbrich, 1977; Zinevich et al., 2010)

2.2.2.3 MSG Satellite Data

This study used data from the Meteosat at 0° E (2013, 2014 evaluation periods) and 41.5° E (2018 evaluation period), which corresponded to Meteosat 10 and 8 satellites (EUMETSAT, 2016), respectively, at the time this study acquired the data from (EUMETSAT, 2020). From the different kinds of SEVIRI's spectral channels, this study focused on solar and thermal infrared channels sensitive to cloud top properties: optical thickness, particle size, and phase during the day and nighttime. These correspond to visible (VIS 0.6 µm), near-infrared (NIR 1.6 µm), thermal infrared (IR 3.9 µm, IR 8.7 µm, IR 10.8 µm and IR 12.0 µm), and water vapour (WV 7.3 µm) channels.

2.3 Method

This section presents the methodology of this chapter using three broad parts. The first part, represented by subsections 2.3.1 and 2.3.2, estimates rainfall from MWL and rain gauges in Figure 2. 1a and b. Part two first briefly describes preprocessing of the SEVIRI data (i.e. 2.3.3). Next, subsection 2.3.4 describes the conceptual model that linked and analysed the satellite, MWL and gauges data using cloud top properties: cloud optical thickness, particle size and phase under daytime and nighttime conditions. SEVIRI-based cloud properties represent the cloud top information and have an indirect and nonlinear relationship with ground rainfall (Bendix et al., 2010; Rosenfeld, 2007). However, their information helps quantify rainfall due to the apparent statistical relationship between rainfall duration and amount (Bergès et al., 2005).

The spatial and temporal approach that retrieved MSG satellite, MWL and rain gauge data while considering their spatiotemporal measurement difference is also described in subsection 2.3.4. Subsequently, the retrieved data were jointly analysed to test the potential of the satellite data to detect rainfall on the individual MWL. The final part (2.3.5) presents performance measures for evaluating rainfall detection and estimation.

2.3.1 Estimating Rainfall from rain Gauges

For each ARG rain gauge, the per-minute tipping count multiplied by the tip equivalent of rain estimated the rainfall accumulation in millimetres (mm). This was then used to estimate rainfall intensities, R (mm h^{-1}), at 15 min intervals. The rainfall data from the TAHMO stations were also used to estimate rainfall intensities at 15 min to ensure temporal consistency in the gauge data. During

every 15 min, the average rainfall intensities along the MWL transmission path were calculated as the mean of the gauge rainfall estimates.

2.3.2 Estimating Rainfall from MWL Data

The average attenuation along the MWL path can be used to estimate the average rainfall, R (mm h^{-1}), using the formula (Olsen et al., 1978) in Equation 2. 1.

$$Z = aR^b \tag{2.01}$$

where Z (dB km^{-1}) is the rain-induced attenuation, a ((dB km^{-1}) (mm h^{-1})$^{-b}$) and b (−) are empirical parameters that are known from the literature (ITU, 2005a; Olsen et al., 1978) and are dependent on MWL signal frequency and polarization. Depending on the sampling strategy used for the MWL RSL data, some studies (e.g.Ostrometzky & Messer, 2014; Overeem et al., 2016b) have used different characteristics of the RSL data for rainfall estimation. The RSL data used in this study consisted of two instantaneous signal levels (i.e., the minimum and maximum RSL) and a mean signal level over a 15 min interval. We used the mean RSL to retrieve R to ensure a homogeneous comparison of gauges and MWL mean rainfall estimates. The steps below describe the estimation of R from the mean RSL data.

2.3.2.1 Wet /Dry Classification of RSL Data

MWL signal attenuation can be related to non-rainfall sources such as atmospheric water vapour content, air temperature, strong solar irradiance and multipath propagation (Chwala & Kunstmann, 2019; David et al., 2009; David & Gao, 2018). Therefore, it is essential to identify the rain periods (i.e. wet periods) in the RSL data and separate them from the no rain periods (i.e. dry periods) to estimate R from RSL data accurately. The former indicates when rain is present on the MWL path, and the latter is when rain is absent on the MWL path. Wet/dry classification in literature is based on two major concepts. One concept assumes rainfall is correlated in space so that neighbouring links will experience mutual attenuation during rain occurrence. Hence, this concept achieves wet/dry classification for a particular link by comparing its attenuation measurement with several links within its vicinity (Overeem et al., 2016b). The other concept analyses the statistical properties of the time series of link signals (e.g. (Schleiss & Berne, 2010; Wang et al., 2012).

Rainfall varies considerably in space and time over the study area (Tazalika et al., 2013; Wakachala et al., 2015). Therefore, a wet/dry classification based on a concept of mutual attenuation of nearby links might not be a practical approach in our study location. Thus, this study utilized the latter wet/dry classification

approach. For every ten mean RSL data (i.e. the equivalent of 150 min interval), we estimated the standard deviation to measure the local variability in the RSL. Since rain attenuates MWL signals, the standard deviation for a rainy interval will be high. Thus, a suitable threshold value was defined to separate the standard deviation values into wet/dry periods. Schleiss and Berne (2010) proposed that such a value can be derived from the rainfall climatology of the area as inferred from, e.g., nearby rain gauges. This study's threshold values were calibrated for each location using data from the experimental setup (i.e. Figure 2. 1, the MWL with gauges underneath their transects). The threshold value was then applied to all the links in the respective locations for wet/dry classification, assuming that the rainfall climatology is homogeneous for the small study location considered (Figure 2. 1). When a 150 min interval had less than five mean RSL data, wet/dry classification and hence rainfall was not computed for that interval.

2.3.2.2 Estimating the Reference RSL

The reference RSL or the baseline level indicates the RSL levels during dry periods. Its precise estimation depends on accurate wet/dry classification (Overeem et al., 2016b; Zinevich et al., 2010). The baseline level fluctuates even in the dry period due to signal attenuations in clear sky conditions (Chwala & Kunstmann, 2019; ITU, 2019), making it challenging to estimate accurately. This study estimated the baseline level for every wet period as the median of the mean RSL from the previous 24 h dry periods.

2.3.2.3 Estimating R from Z

After wet/dry classification and baseline level estimation, additional attenuation due to antenna wetting during and after rain was taken into account by following the dynamic model by Schleiss et al. (2013). Eventually, the path average attenuation, Z (dB km^{-1}), was estimated for every 15 min wet period by subtracting the mean RSL from the baseline level, as shown in Equation 2.2. In some cases, the method retrieved negative path average attenuation values. This is the case when a 15 min period within a 150 min interval labelled as wet is dry because rainfall was intermittent in the wet interval. In these cases, the attenuation was set to zero.

$$Z = \frac{Pref - P}{L} \tag{2.02}$$

where L is MWL length (km), while $Pref$ and P (dBm) are the reference and mean RSL (corrected from the effect of wet antenna), respectively. Finally, Equation 2.3 estimated the R (mm h^{-1}) from Z.

$$R = \left(\frac{Z}{a}\right)^{\frac{1}{b}} \quad (2.03)$$

The values of a (0.05008, 0.1284) and b (1.0440, 0.9630) for 15, and 23 GHz MWL, respectively, used in this study were from (ITU, 2005a).

2.3.3 SEVIRI Data Retrieval and Processing

The zenith viewing angle of the Meteosat at o° E, over the study area, is approximately 41°. Depending on the location and the height of clouds, this viewing angle could cause displacement of cloud tops from their actual position due to the effect of parallax, which occurs because SEVIRI observes the Earth under an oblique angle (Roebeling & Holleman, 2009). This parallax displacement can amount to about 12 km in SEVIRI pixels for very high clouds in Kenya (Hoedjes et al., 2014). Therefore, the 2013 and 2014 evaluation periods (retrieved from Meteosat at ° E) were parallax corrected using a correction algorithm from EUMETSAT (see https://cwg.eumetsat.int/parallax-corrections/, accessed on 9 June 2021). However, the zenith viewing angle of the Meteosat at 41.5° E is about 5°, and this small viewing angle does not require parallax correction.

2.3.4 The Conceptual Model, Satellite and Ground data Sampling and Analysis for Rainfall Detection

1) The SEVIRI-based Rainfall Detection Conceptual Model

This study linked MWL rainfall estimates to MSG SEVIRI data based on the assumption that clouds that rain over MWL can be detected using SEVIRI-based cloud top properties. Therefore, a conceptual model was defined that explored the relationship between spectral characteristics of different kinds of SEVIRI channels and cloud top properties (cloud top optical thickness, particle size and phase) for detecting rainfall on individual MWL. The model assumes that clouds with high optical thickness and large particle sizes (with the existence of ice or water hydrometeors at the top) have high rainfall probability and intensity, whereas clouds with low optical thickness and small particles sizes have low rainfall probability and intensity (Bendix et al., 2010; Thies et al., 2008a). The physical basis underlying this assumption is that the conditions for the development of precipitation in clouds are (i) availability of sufficient moisture, (ii) existence of an effective mechanism for converting small cloud droplets that are suspended in the air into large precipitating particles, and (iii) the existence of ice phase clouds at the cloud top to support rain generation by the Bergeron–Findeisen process (Lensky & Rosenfeld, 2003a; Thies et al., 2008b).

Past studies have inferred information about the cloud top optical thickness, particle size, and phase from SEVIRI's solar and thermal infrared satellite channels for successful rainfall detection (Bendix et al., 2010; Feidas & Giannakos, 2010; Lazri et al., 2014). Based on the numerous successful and convincing rainfall detection results, this study utilized the original reflectance and brightness temperature differences of SEVIRI channels to infer cloud top properties for rainfall detection from clouds. Because cloud top properties information differs between daytime and nighttime, detecting raining clouds is done separately for daytime and nighttime. EUMETSAT's operational cloud mask product (EUMETSAT, 2015) identified the cloudy pixels in this study to ensure that this study utilized only cloudy scenes for rainfall detection.

i. Detecting raining clouds during daytime

Reflection of solar radiation by clouds in the non-absorbing channels (between 0.4 µm and 0.8 µm visible channels) is strongly related to the cloud optical thickness. However, in the slightly absorbing 1.6 µm and 3.9 µm channels, this is related to the particle size (Baum & Spinhirne, 2000; Kawamoto et al., 2001; Kokhanovsky, 2003). Therefore, the two combined channels can provide cloud optical thickness and particle size information. Cloud's optical thickness and particle size represent a single parameter, CWP, directly related to a cloud's rainfall probability.

The CWP indicates the amount of water vertically integrated into the cloud. It depends on the diameter of particle size and the thickness of clouds formed by these particles (Lazri et al., 2013; Lazri et al., 2014; Thies & Nauss, 2008). Consequently, CWP is often implicitly inferred from VIS 0.6 µm and NIR 1.6 µm channels from SEVIRI (Thies et al., 2008b; Thies et al., 2008d). High VIS 0.6 µm reflectance indicates optically thick clouds, and low NIR 1.6 µm reflectance corresponds to large cloud particle sizes. The implication is that large CWP is observed when high VIS 0.6 µm reflectance coincides with low NIR 1.6 µm reflectance (Thies et al., 2008b).

The difference in brightness temperature between IR 8.7 µm and 10.8 µm ($\Delta T_{IR8.7-IR10.8}$) and that between IR 10.8 µm and 12.0 µm ($\Delta T_{IR10.8-IR12.0}$) can be used to infer information about the cloud phase (Thies et al., 2008b). Strabala et al. (1994) observed stronger water particle absorption between 11 µm and 12 µm than between 8 µm and 11 µm, whereas, for ice particles, the reverse is true. Following this observation, the study found that $\Delta T_{IR10.8-IR12.0}$ differences indicative of water clouds are higher than coincident $\Delta T_{IR8.7-IR10.8}$ differences. On the contrary, $\Delta T_{IR8.7-IR10.8}$ of ice clouds is higher than coincident $\Delta T_{IR10.8-IR12.0}$. Feijt

et al. (2008) have instead suggested the simultaneous use of brightness temperature IR10.8 μm ($T_{IR10.8}$) and the difference $\Delta T_{IR8.7\text{-}IR10.8}$ for identifying the cloud phase. They found that ice crystals begin to form when $T_{IR10.8} < 238$ K and $\Delta T_{IR8.7\text{-}IR10.8} > 0.25$ K. As earlier indicated, the ice phase at the cloud top supports rain generation and thus increases the likelihood of a cloud producing rain.

The $\Delta T_{IR10.8\text{-}IR12.0}$ has also been considered a good indicator of cloud optical thickness and can effectively discriminate optically thick cumulous clouds from optically thin cirrus clouds (Feidas & Giannakos, 2010, 2011; Inoue, 1987b). Optically thick cumulous clouds show small $\Delta T_{IR10.8\text{-}IR12.0}$ because of their blackbody characteristics. Optically thin cirrus clouds show larger differences because of the differential absorption between ice crystals between the two channels (Inoue et al., 2001). It is expected that optically thick cumulous-type clouds with small $\Delta T_{IR10.8\text{-}IR12.0}$ produce rain (Inoue, 1987b).

ii. Detecting raining clouds during night-time

During the night-time, the brightness temperature differences: $\Delta T_{IR3.9\text{-}IR10.8}$, $\Delta T_{IR3.9\text{-}WV7.3}$, $\Delta T_{IR8.7\text{-}IR10.8}$, and $\Delta T_{IR10.8\text{-}IR12.0}$ were used to infer information about cloud optical thickness, particle size and phase (Lensky & Rosenfeld, 2003a, 2003b; Thies et al., 2008a; Thies et al., 2008d) for rainfall detection. A sufficient empirical basis exists for using such temperature differences and combinations for night-time rainfall detection (Lensky & Rosenfeld, 2003a; Schmetz et al., 2002; Thies et al., 2008a; Thies et al., 2008d). The 3.9 μm emissions are sensitive to particle size, such that large particles have high emissions than smaller particles. The dependence on particle size is less distinct in the IR10.8 μm than in IR3.9 μm – resulting in higher ΔTIR3.9-IR10.8 differences for large particle sizes than smaller particles (Thies et al., 2008a; Thies et al., 2008d). Using $\Delta T_{IR3.7\text{-}IR11}$ of the TRMM satellite, Lensky and Rosenfeld (2003a) showed that optically thick raining clouds with large particles produced a brightness temperature difference in the interval −1 to 4 K. Concerning the $\Delta T_{IR3.9\text{-}WV7.3}$ difference, the characteristics should be similar to $\Delta T_{IR3.9\text{-}IR10.8}$ but with generally higher differences than $\Delta T_{IR3.9\text{-}IR10.8}$. This is due to the diminishing effect of the water vapour absorption and emission in the mid-to-low tropospheric levels on the brightness temperature in the WV7.3 μm channel (Schmetz et al., 2002; Thies et al., 2008a).

2) Spatial and Temporal Differences between SEVIRI and the Ground Data

The measurement characteristics of satellite and ground sensors are fundamentally different (Ha et al., 2002). The satellite measures instantaneously over a wide area, while the ground sensors continuously measure from a single location (rain gauges) or aggregated measurements over time and space (MWL). Therefore, the measurement differences suggest a possible spatial and temporal mismatch between satellite and ground measurements that must be considered when analysing the two datasets. The description below presents the spatial and temporal difference between the SEVIRI and ground sensors (MWL or rain gauges) and how they are treated in this study.

i. Spatial mismatch

The general assumption often used to compare ground rainfall, and satellite data is that the measured rainfall represents the whole satellite pixel containing the ground sensors. However, the hydrometeors from heavy rain, falling at a speed of 10 ms^{-1} from a 3 km height cloud and horizontal wind speed between 5 and 30 $ms^{-1,}$ can drift a horizontal distance between 1.5 and 90 km (Roe, 2005). This suggests that the rainfall recorded by a ground sensor might not always correlate with the satellite signal from a collocated pixel but with the signal from other adjacent pixels. Additionally, for tropical deep convective systems consisting of convective cores and anvil cloud areas having different cloud properties, dynamical regimes, and varying rainfall intensities, the retrieved cloud top properties might be biased towards the spatially dominant anvil cloud areas (Young et al., 2013).

This study adopted an approach that considered the pixel containing the ground sensor and the surrounding 3 by 3 pixels to minimise the spatial mismatches between the satellite and ground data in its analysis. This meant the gauge pixel and its surrounding 3 by 3 pixels were considered for the rain gauge. For the MWL, this corresponded to 3 by 3 pixels surrounding the centre of the MWL transect, which in this case, constituted all the pixels covering the MWL. This study used two spatial aggregation methods (i.e., summary statistics used to sample from the raw data in space) to retrieve a single satellite signal out of the 3 by 3 pixel environment comparable to the ground rainfall under daytime and night-time conditions.

The daytime spatial aggregation method previously described by (Kuhnlein et al., 2010) was used because of its simplicity and effectiveness in identifying the most effective satellite signal comparable with the ground rainfall data. The method

identifies the pixel with the most effective satellite signal as the pixel with the highest reflectance value in the VIS 0.6 µm and lowest reflectance values in the NIR 1.6 µm (indicating high optical thickness and large particle size, respectively, i.e., thick clouds). More precisely, for n ≥ 2 (where n is the number of cloudy pixels in the 3 by 3 pixels environment), the maximum and minimum reflectance values were expressed as:

$$VIS_{ref} = max_{i=1,n}(x_i) \quad (2.04)$$

where x_i is the reflectance value in the VIS 0.6 µm channel, and

$$NIR_{ref} = min_{i=1,n}(y_i) \quad (2.05)$$

where: y_i is the reflectance value in the NIR 1.6 µm channel. If the retrieved maximum VIS 0.6 µm and minimum NIR 1.6 µm values do not occur in the same pixel, the value combination that returns the highest difference between the two signals is used. The maximal difference was expressed as:

$$MaxDiff_{VISNIR} = max_{i=1,n}(x_i - y_i) \quad (2.06)$$

The theoretical assumption in this method is that by identifying the pixel with the maximum and minimum reflectance in VIS 0.6 µm and NIR 1.6 µm, respectively, the pixel with the highest optical thickness and largest particle size that indicate a large CWP pixel is detected. Once the cloudy pixel with large CWP is detected, the phase of clouds is then retrieved using $\Delta T_{IR8.7\text{-}IR10.8}$ and $\Delta T_{IR10.8\text{-}IR12.0}$.

The night-time spatial aggregation method used the mean brightness temperature differences retrieved from cloudy scenes. This mean brightness temperature difference for n ≥ 2 (where n is the number of cloudy pixels in the three by three pixels environment) was expressed as:

$$Mean = \frac{1}{n} * \sum_{i=1}^{n} \Delta T_i \quad (2.07)$$

where ΔT_i is the brightness temperature difference for the various channel combinations considered.

ii. Temporal mismatch

The potential temporal mismatch between the satellite and ground data is mainly because SEVIRI instantaneous scenes over the study area are acquired in about 6 min (depending on the latitude) into each 15 min scan interval. Therefore, the measurement may represent the cloud top conditions available during the first few minutes of SEVIRI's 15 min scan interval. In contrast, the ground measurements were continuous in time. In particular, the gauges originally recorded data at least every minute, while the MWL recorded mean RSL every 15

min. Moreover, during a 15 min interval, a raining cloud could have passed over the ground sensor during the first 6 min (when SEVIRI scenes were available) or the latter 9 min. Thus, the satellite measurement might not necessarily coincide with the ground measurement at all times. The dataset was aggregated in time to minimise the effect of this temporal mismatch. This temporal aggregation was done by computing the mean satellite signal for the satellite data, and for the ground rainfall data, the rainfall sums were computed every 30 min.

Other measurement characteristics of the satellite that may potentially present uncertainties when comparing and analysing the ground and satellite data are the effect of viewing and illumination geometries: solar zenith angle, viewing zenith, and relative azimuth angles (Kato & Marshak, 2009). These effects concern the 2013 and 2014 data acquired from solar channels of the Meteosat at 0° E. Since small study locations are considered (Figure 1) in this study, the effects of viewing and illuminating geometries should be minimal. Besides, we only used solar reflectance for daytime hours when sufficient solar illumination was available over the study area.

3) The MWL Rainfall and SEVIRI Data Analysis for Rainfall Detection

The analysis was separate for the daytime and night-time using MWL rainfall and satellite data pairs retrieved from the experimental setup in Figure 2. 1 (i.e. the MWL with rain gauges under their transect). After spatiotemporal aggregation to reduce the effect of their inherent measurement mismatch, the data pairs permitted the separation of the satellite data into raining and non-raining satellite signals. The raining satellite signals were further classified based on different rainfall categories in Table 2. 5 to investigate the satellite signals for varying rainfall intensity ranges. The rainfall classes' determination was done by analysing the frequency distribution of the gauge rainfall intensities (not shown here) and using different criteria, if possible. (1) Each rainfall class should have sufficient data to compute descriptive statistics; (2) the rainfall classes should be equal for both study locations to ensure homogenous rainfall analysis across the two areas. Subsequently, a scatter plot of rainfall intensities as a function of the satellite signals was used to investigate the rainfall in each rainfall class and the corresponding satellite signal. Then, each rainfall class' satellite signal was statistically analysed using descriptive statistics.

These descriptive statistics tested the potential of combining the information content gained from the different SEVIRI signals to detect rainfall on individual MWL during daytime and night-time. The daytime rainfall detection evaluated a four-dimensional matrix of VIS 0.6, NIR 1.6, and $\Delta T_{IR8.7\text{-}IR10.8}$, $\Delta T_{IR10.8\text{-}IR12.0}$ to

make a rain or no rain decision. In contrast, the night-time considered $\Delta T_{IR3.9\text{-}IR10.8}$, $\Delta T_{IR3.9\text{-}WV7.3}$, $\Delta T_{IR8.7\text{-}IR10.8}$, and $\Delta T_{IR10.8\text{-}IR12.0}$ for rain no rain decision. The rain detection test used independent MWL and satellite datasets.

2.3.5 Performance Measures

2.3.5.1 Evaluating MWL Rainfall Intensities

The MWL rainfall estimates (RMWL) were evaluated against the rain gauge estimates (RRG) using the relative bias (RB), coefficient of variation (CV), coefficient of determination (r2), and root mean square error (RMSE) (Table 2. 2) across the evaluation period. RB indicates whether the R_{MWL} systematically over or underestimates the R_{RG} (Walther & Moore, 2005) and ranges from −1 to + ∞, with 0 being an unbiased case. The CV indicates how the R_{MWL} varies around the mean of the R_{RG} (Rios et al., 2012b) and ranges from 0 to ∞. The r^2 shows the strength of the linear relationship between the R_{MWL} and the R_{RG}. It ranges from 0 to 1, where 1 indicates a perfect linear correlation between the R_{MWL} and the R_{RG} (Wilks, 2006). Finally, the RMSE shows how close the R_{MWL} is to R_{RG} and ranges from 0 to positive ∞, where 0 is a hypothetical case, and a larger RMSE indicates decreasing accuracies of R_{MWL} (Barnston, 1992; Wilks, 2006).

Table 2. 2 Performance measures for evaluating R_{MWL}. The full name of each measure is indicated in the text. j and n represent all timestamps for the evaluation period.

Performance Measure	Formula	Range
RB	$\frac{\frac{1}{n} * \sum_{j=1}^{n}(R_{MWL} - R_{RG})}{\frac{1}{n} * \sum_{j=1}^{n} R_{RG}}$	−1 to + ∞
CV	$\frac{\sqrt{Var(R_{MWL} - R_{RG})}}{\frac{1}{n} * \sum_{j=1}^{n} R_{RG}}$	0 to ∞
r2	$\left[\frac{Cov(R_{MWL}, R_{RG})}{S_{R_{MWL}} * S_{R_{RG}}}\right]^2$	0 to 1
RMSE	$\sqrt{\frac{1}{n} * \sum_{j=1}^{n}(R_{MWL} - R_{RG})^2}$	0 to + ∞

2.3.5.2 Evaluating the Performance of SEVIRI-Based Rain Detection on MWL

The rain detection performance test was evaluated by computing the values of the elements a, b, c and d as described in (Harold et al., 2015; Wilks, 2006). Their formulation differs for daytime and night-time because of the different SEVIRI channels and information content used for each time of the day (i.e., daytime or night-time). For daytime, a_d,b_d,c_d, and b_d were computed as:

$$a_d = R_{sat}\begin{pmatrix} Vis \geq Vis_{thres}\ AND\ Nir \leq Nir_{thres}\ AND \\ \Delta T_{\beta 1} \in [x_1, x_2]\ AND\ \Delta T_{\beta 2} \in [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} \geq 1\ mmh^{-1}) \quad (2.08)$$

$$b_d = noR_{sat}\begin{pmatrix} Vis < Vis_{thres}\ AND\ Nir > Nir_{thres}\ AND \\ \Delta T_{\beta 1} \notin [x_1, x_2]\ AND\ \Delta T_{\beta 2} \notin [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} \geq 1\ mmh^{-1}) \quad (2.09)$$

$$c_d = R_{sat}\begin{pmatrix} Vis \geq Vis_{thres}\ AND\ Nir \leq Nir_{thres}\ AND \\ \Delta T_{\beta 1} \in [x_1, x_2]\ AND\ \Delta T_{\beta 2} \in [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} < 1\ mmh^{-1}) \quad (2.10)$$

$$d_d = noR_{sat}\begin{pmatrix} Vis < Vis_{thres}\ AND\ Nir > Nir_{thres}\ AND \\ \Delta T_{\beta 1} \notin [x_1, x_2]\ AND\ \Delta T_{\beta 2} \notin [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} < 1\ mmh^{-1}) \quad (2.11)$$

where: a_d,b_d,c_d, and d_d are the hits, misses, false alarms and correct negative events, respectively. R_{sat} and noR_{sat} are raining and non-raining conditions in the satellite data. Vis,Nir are VIS 0.6 µm and NIR 1.6 µm and their respective thresholds Vis_{thres} and Nir_{thres}. $\Delta T_{\beta 1}$, $\Delta T_{\beta 2}$ are $\Delta T_{IR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$, and their brightness temperature ranges $[x_1, x_2]$ and$[y_1, y_2]$, respectively. During the nighttime, the corresponding values of a_n,b_n,c_n, and b_n were computed as:

$$a_n = R_{sat}\begin{pmatrix} \Delta T_{\gamma 1} \in [u_1, u_2]\ AND\ \Delta T_{\gamma 2} \in [v_1, v_1]\ AND \\ \Delta T_{\beta 1} \in [x_1, x_2]\ AND\ \Delta T_{\beta 2} \in [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} \geq 1\ mmh^{-1}) \quad (2.12)$$

$$b_n = noR_{sat}\begin{pmatrix} \Delta T_{\gamma 1} \notin [u_1, u_2]\ AND\ \ \Delta T_{\gamma 2} \notin [v_1, v_1]\ AND \\ \Delta T_{\beta 1} \notin [x_1, x_2]\ AND\ \ \Delta T_{\beta 2} \notin [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} \geq 1\ mmh^{-1}) \quad (2.13)$$

$$c_n = R_{sat}\begin{pmatrix} \Delta T_{\gamma 1} \in [u_1, u_2]\ AND\ \Delta T_{\gamma 2} \in [v_1, v_1]\ AND \\ \Delta T_{\beta 1} \in [x_1, x_2]\ AND\ \Delta T_{\beta 2} \in [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} < 1\ mmh^{-1}) \quad (2.14)$$

$$d_n = noR_{sat}\begin{pmatrix} \Delta T_{\gamma 1} \notin [u_1, u_2]\ AND\ \ \Delta T_{\gamma 2} \notin [v_1, v_1]\ AND \\ \Delta T_{\beta 1} \notin [x_1, x_2]\ AND\ \Delta T_{\beta 2} \notin [y_1, y_2] \end{pmatrix} AND\ (R_{MWL} < 1\ mmh^{-1}) \quad (2.15)$$

where: a_n,b_n,c_n, and d_n are the hits, misses, false alarms and correct negatives events respectively during the nighttime; $\Delta T_{\gamma 1}$and $\Delta T_{\gamma 2}$ are $\Delta T_{IR3.9\text{-}IR10.8}$, $\Delta T_{IR3.9\text{-}WV7.3}$ and their brightness temperature ranges $[u_1, u_2]$, and $[v_1, v_1]$, respectively.

Based on the computation of these elements, a set of standard verification scores (Table 2.3) were computed that evaluated an aspect of the rain detection test. The POD evaluated the fraction of the R_{MWL} correctly detected by the test. FAR answered the question, 'what fraction of the number of R_{MWL} detected by the test was incorrect?' POFD indicates the fraction of no rain (R_{MWL} < 1 mm h^{-1}) on the MWL that was incorrectly identified as rain (R_{MWL} > 1 mm h^{-1}) by the test. The ACC evaluated the overall fraction of rain and non-rain correctly detected by the test. CSI is used to show how well the rain detected by the test corresponds to R_{MWL} on individual MWL. The HSS evaluates the rain detection test's accuracy by considering the detection that was due to random chance. The N in its formula (Table 2. 3) is the sum of a, b, c and d.

Table 2. 3 Standard Verification Scores that Evaluated the Rain Detection Test

Standard verification scores	
POD	$\frac{a}{a+b}$
FAR	$\frac{a}{a+c}$
POFD	$\frac{c}{c+d}$
ACC	$\frac{a+d}{a+b+c+d}$
CSI	$\frac{a}{a+b+c}$
HSS	$\frac{(a+d)-(a_{random})}{N-a_{random}}$ $a_{random}=\frac{(a+b)\times(a+c)+(d+b)\times(d+c)}{N}$

2.4 Results and Discussion

2.4.1 Results

The first part of this section presents the results of evaluating the MWL rainfall against gauge measurements. The second part shows the results of analysing the rainfall estimates in part 1 with the SEVIRI satellite signals for rainfall detection. The final part summarizes the performance of the satellite data when used to detect rainfall on individual MWLs.

2.4.1.1 The R_{MWL} versus R_{RG}

The rainfall estimates evaluated in this section were from the experimental setup in Figure 2. 1. The evaluation period was during the long rains of 2013 and 2018 for the Kericho and Naivasha setup, respectively. The frequency of both links is 15 GHz, and their lengths are approximately 3.7 and 10 km for the Kericho and Naivasha MWL, respectively. The R_{MWL} were evaluated using R_{RG} from 16 rain gauges – 5 for the Kericho link and 11 for the Naivasha link. Here, the gauge rainfall intensities were considered as the reference rainfall measurement.

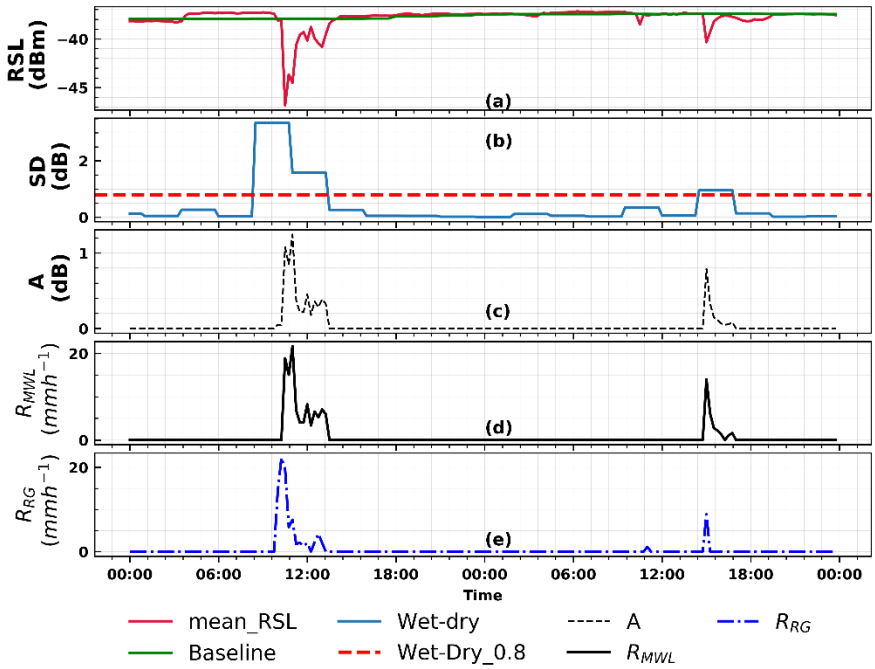

Figure 2. 2 From raw RSL to rainfall using MWL data from Kericho, 11 to 12 May 2013. (a) mean and reference RSL (b) rolling standard deviation with a threshold of 0.8 dB (red dashed line) for detecting wet/dry periods (c) attenuation (d) MWL derived rainfall intensities, and (e) rain gauge derived rainfall intensities.

Figure 2. 2 shows the transformation of MWL RSL to R_{MWL} compared with R_{RG} at 15 min intervals for Kericho, using data recorded over 48 h. It can be seen from the figure that the drop in signal levels, e.g. between 11:00 and 12:00 pm (Figure 2. 2 a), was detected as wet by the wet/dry classification (Figure 2. 2 b), indicated by the period's standard deviation above the wet/dry threshold value. This period also coincides with high attenuation (Figure 2. 2 c) and rainfall intensities averaged from the gauges under the MWL (Figure 2.2 e).

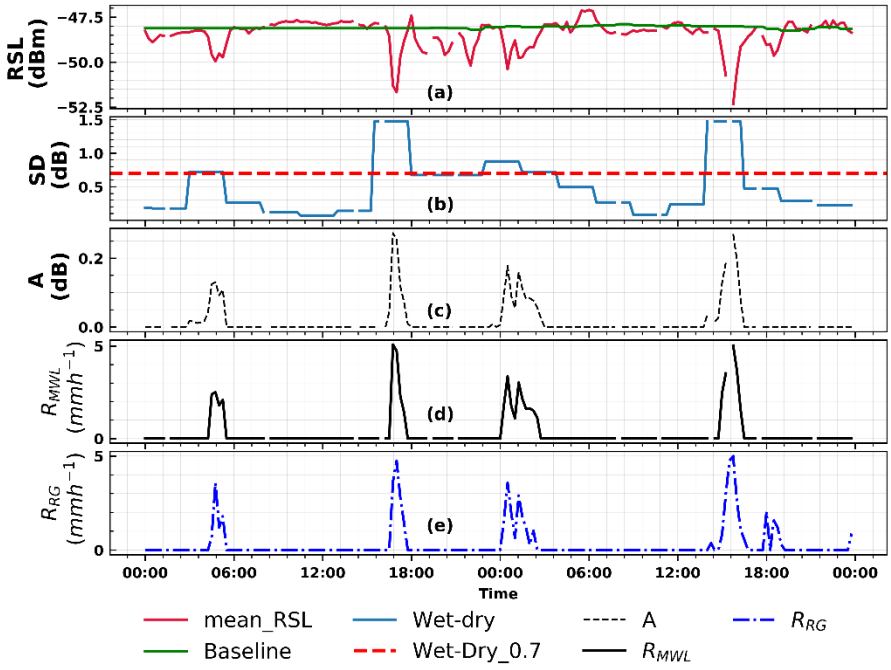

Figure 2. 3 As in Figure 2.2 but for the Naivasha MWL and rainfall events on 4 June to 5 June 2018

Figure 2. 3 is analogous to Figure 2. 2 but for the experimental MWL in Naivasha. The link had frequent intermittent periods of no data compared to the Kericho link. The data gaps were considered during the MWL rainfall estimation procedure. The estimated threshold value of the wet/dry classification was 0.7 dB, which detected most of the variable average rainfall intensities observed by the gauges under the MWL.

The R_{MWL} and R_{RG} are also compared in a scatter plot at 15 min, half-hourly and hourly evaluation timestamps for both study locations (Figure 2. 4). The half-hourly and hourly values were a summation of the 15 min rainfall intensities. Table 2. 4 summarizes the values of the performance measures for each evaluation timestamp and study location. For the scatter plot comparison and computation of the performance measures, R_{MWL} and R_{RG} pairs less than 1 mm h−1 was set to 0 mm h−1 (i.e., considered dry). However, data with the 0 mm h^{-1} were included in all analyses to evaluate the MWL's detection and estimation capabilities for both wet and dry periods. All performance measures were computed across each evaluation period.

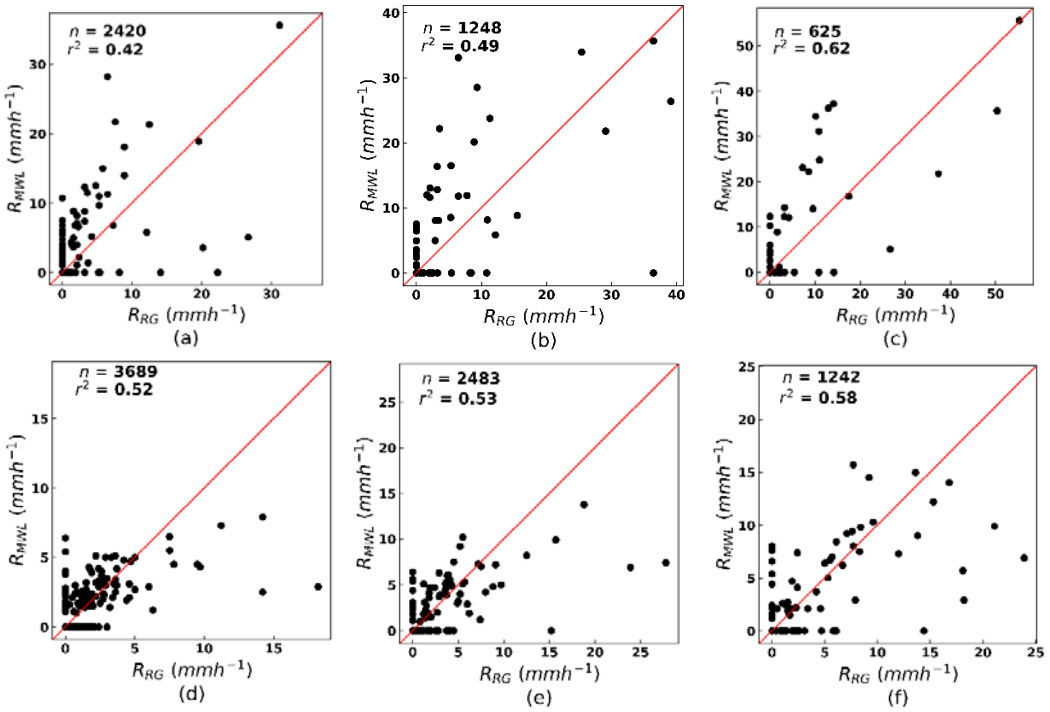

Figure 2. 4 Scatter plot comparison of R_{MWL} and the R_{RG} for Kericho (a–c) and Naivasha (d–f) at 15 min (a,d), half-hourly (b,e) and hourly (c,f) timestamp.

Table 2. 4 Performance measures calculated from R_{MWL} and R_{RG} pairs from the two study locations.

Study Location	RB			CV			r^2			RSME (mm h^{-1})		
	15 min	30 min	1 h	15 min	30 min	1 h	15 min	30 min	1 h	15 min	30 min	1 h
Kericho [1]	0.50	0.32	0.32	9.87	7.18	5.09	0.42	0.49	0.62	1.22	1.96	2.77
Naivasha [2]	-0.05	-0.14	-0.18	5.78	5.68	4.07	0.52	0.53	0.58	0.48	0.80	1.15

[1]Performance measures were computed using 26 days of R_{MWL} and R_{RG} pairs.
[2]performance measures were calculated using 52 days of R_{MWL} and R_{RG} pairs.

Table 2. 4 shows that the rainfall intensity estimation skill of both MWLs differs when comparing their R_{MWL} to the R_{RG}. The RB indicates that the Kericho MWL overestimated the observed average rainfall intensity, albeit this decreased with increasing aggregation timestamp, from 0.50 at the 15 min to 0.32 for both half-hourly and hourly timestamps. Likewise, the CV decreased from 9.87 at 15 min to 5.09 at the hourly timestamp. The strength of the relationship between R_{MWL} and R_{RG} increased for increasing timestamps, with r^2 values reaching approximately

0.6 at the hourly timestamp. The RMSE, however, increased from 1.22 mm h^{-1} at 15 min to 2.77 mm h^{-1} at the hourly timestamps.

In contrast, the Naivasha link marginally underestimated the observed rainfall. The value of RB increased minimally from −0.05 at 15 min to −0.18 at half-hourly and hourly timestamps. The CV decreased from 5.78 to 4.07 for 15 min to hourly evaluation timestamps. The strength of the relationship between R_{MWL} and R_{RG} (Figure 2. 4d–f) also increased for increasing timestamps with r^2 above 0.5 for all timestamps. Even though the RMSE increased with aggregation time, the values (Table 4) were comparatively lower than those of the Kericho link.

2.4.1.2 Joint Analysis of Rainfall and SEVIRI Satellite Data

This section presents the results of analysing rainfall estimates with SEVIRI satellite data in two parts. First, collocated ground rainfall (from MWL and rain gauge) and satellite data from the experimental setup in Figure 2. 1a,b were jointly analysed for detecting rainfall. The gauge rainfall intensities presented in this analysis were from the rain gauges close to the centre of each MWL. Note, however, that the inclusion of the gauge rainfall data gave a perspective of the rainfall satellite analysis from a reference measurement point of view. The rainfall values were grouped into different rain classes to investigate further the satellite signals for different rainfall intensity ranges (Table 2. 5). The analysis was done separately for the two study areas and for during day and night-time. Next, inferences deduced from the rainfall satellite analysis are summarised based on observations from the two study areas and separately during day and night-time.

Table 2. 5 Summary of the R_{MWL} data per each study area for the day and night-time

Study Area	R_{MWL} (mmh^{-1})	[1]Percentage of Data (%)		Accumulated R_{MWL} (mm)	
		Day	Night	Day	Night
Kericho	0	91.56 (93.81)	94.5 (97.7)	0 (0)	0 (0)
	0–5	1.95 (2.61)	2.29 (1.84)	12.22 (24.92)	13.57 (12.06)
	>5	6.49 (3.58)	3.21 (0.46)	314.25 (219.49)	98.1 (39.40)
Naivasha	0	96.68 (96.84)	95.34 (96.02)	0 (0)	0 (0)
	0–5	2.41 (2.11)	2.61 (2.88)	48.45 (34.74)	61.92 (54.14)
	>5	0.90 (1.05)	2.06 (1.1)	43.02 (88.88)	106.09 (61.41)

Note: [1]values used in calculating percentages in Kericho and Naivasha are 307, 217, 663, and 729 for day and nighttime, respectively; values in parenthesis are computed based on the rain gauge data.

(i) MSG satellite rainfall

Figure 2. 5 is a scatter plot of rainfall intensities as a function of the MSG satellite signals during the daytime in May–June 2013 (in total, 313 MSG SEVIRI scenes).

A clear observation from the figure is the difference in the scatter of data points between the raining satellite signals (0–5 mm h^{-1} and above 5 mm h^{-1} rainfall classes) and non-raining satellite signals (0 mm h^{-1} rainfall class). This feature was evident in both reflectance and brightness temperature difference combinations and the R_{RG} and R_{MWL} scatter plots.

The SEVIRI reflectance combination is in Figure 2. 5a,c. The figure shows that the 0 mm h^{-1} rainfall scatter throughout the whole range of the satellite signals, with a high concentration of the scatter occurring in the lower-left corner of the plot, where low VIS 0.6 µm reflectance is connected to low NIR 1.6 µm reflectance. On the other hand, the combination of high VIS 0.6 µm versus low NIR 1.6 µm reflectance is generally evident for rainfall occurrence. The 0–5 mm h^{-1} R_{MWL} scatter over a wide range of the satellite signal. In some cases, the value combination of the VIS 0.6 µm and NIR 1.6 µm reflectance in this rainfall class were comparable. The rainfall above 5 mm h^{-1} was generally restricted to the lower right corner of the plot, where high VIS 0.6 µm reflectance is connected with low NIR 1.6 µm reflectance.

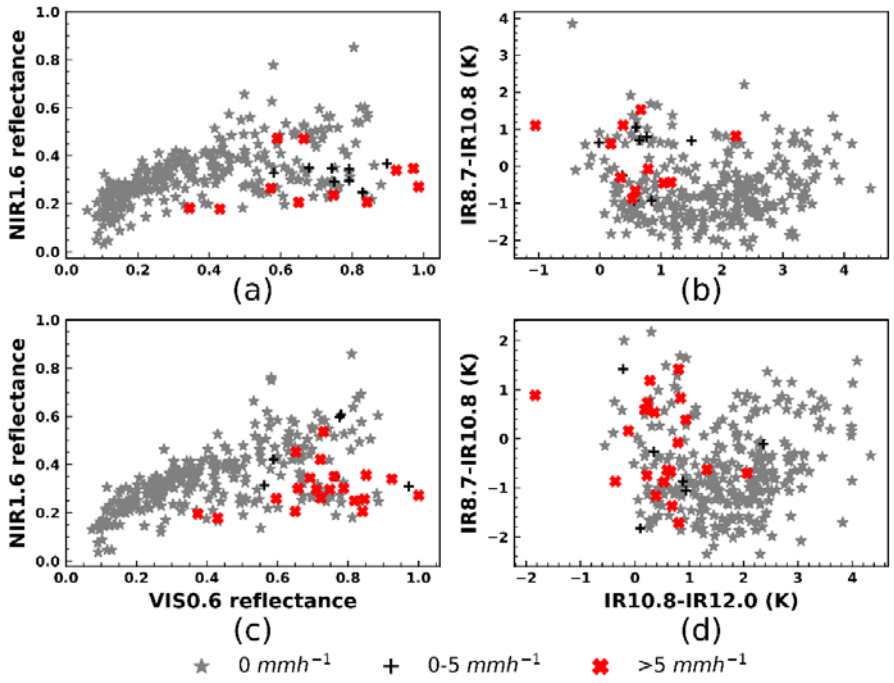

Figure 2. 5 Day time R_{RG} (a,b) and R_{MWL} (c,d) as a function of VIS 0.6 µm versus NIR 1.6 µm (a,c), and $\Delta T_{IR10.8\text{-}IR12.0}$ versus $\Delta T_{IR8.7\text{-}IR10.8}$ (b,d) for Kericho.

For the satellite brightness temperature difference (Figure 2. 5b,d) indicative of cloud phase, water ($\Delta T_{IR10.8\text{-}IR12.0}$) and ice ($\Delta T_{IR8.7\text{-}IR10.8}$), the 0 mm h^{-1} rainfall also scatter throughout the whole range of the satellite signal, albeit with higher $\Delta T_{IR10.8\text{-}IR12.0}$ differences than $\Delta T_{IR8.7\text{-}IR10.8}$. In contrast, the raining satellite signals tend to scatter in different ranges of values for $\Delta T_{IR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$. Most of the raining satellite signals of $\Delta T_{IR10.8\text{-}IR12.0}$ scatter above 0 K, with a large concentration of the scatter falling within a narrow range (approximately 0 to 1 K). However, $\Delta T_{IR8.7\text{-}IR10.8}$ scatter over a wide range of values (between -2 and 1.5 K). In addition, from the plot, some coincident values $\Delta T_{IR10.8\text{-}IR12.0}$ are larger than $\Delta T_{IR8.7\text{-}IR10.8,}$ and vice versa, which was evident for the above 0 mm h^{-1} rainfall and in both the R_{RG} and R_{MWL} plots.

Figure 2. 6 is analogous to Figure 2. 5, but for brightness temperature differences during night-time conditions in May-June 2013 (altogether 218 MSG SEVIRI scenes). The scatter of this plot's 0 mm h^{-1} rainfall satellite signal is similar to those observed during daytime conditions. However, the raining satellite signals are in different ranges for all the brightness temperatures and scatter differently for the R_{RG} and R_{MWL}.

For instance, the values of $\Delta T_{IR3.9\text{-}WV7.3}$ were larger than the values of $\Delta T_{IR3.9\text{-}IR10.8}$ (Figure 2. 6a,c). The R_{RG} plot (Figure 2. 6a) scatter below 10 K for the $\Delta T_{IR3.9\text{-}WV7.3,}$ whereas the differences shown in the R_{MWL} plot (Figure 2. 6c), particularly those between 0–5 mm h^{-1}, tend to scatter over a wide range of the satellite signals. Conversely, $\Delta T_{IR3.9\text{-}IR10.8}$ of R_{MWL} scatter above 0 K, whereas the R_{RG} plot's differences are over a wide range (between −5 and 3 K).

For the brightness temperature differences in Figure 2. 6b,d, the $\Delta T_{IR10.8\text{-}IR12.0}$ scatter above 0 K with a large concentration of the scatter between 0 and 1 K, whereas the $\Delta T_{IR8.7\text{-}IR10.8}$ scatter over a comparatively wide range of values (between −1.5 and 1 K). Similar to daytime observation, some coincident values of $\Delta T_{IR10.8\text{-}IR12.0}$ are larger than those of $\Delta T_{IR8.7\text{-}IR10.8}$ for the above 0 mm h^{-1} and vice versa and are evident from the R_{RG} and R_{MWL} plots.

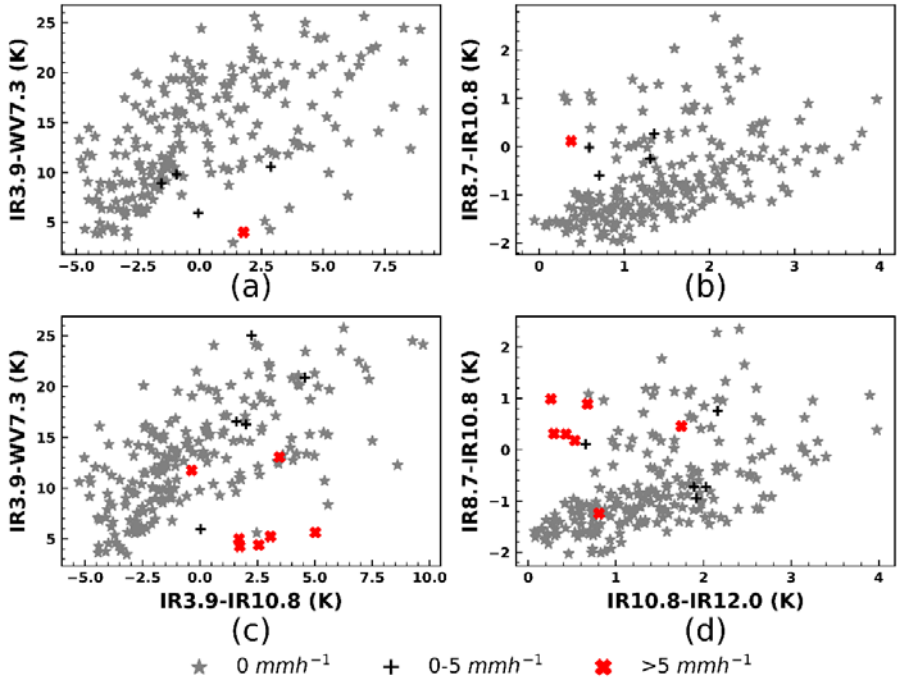

Figure 2. 6 Nighttime R_{RG} (a,b) and R_{MWL} (c,d) as a function of ΔTIR3.9-IR10.8 versus $\Delta T_{IR3.9-WV7.3}$, (a,c) and $\Delta T_{IR10.8-IR12.0}$ versus $\Delta T_{IR8.7-IR10.8}$ (b,d) for Kericho.

A scatter plot of rainfall intensity as a function of the MSG satellite signal, analogous to the daytime analysis in Kericho (Figure 2. 5), is presented for Naivasha (Figure 2. 7). Altogether, 713 MSG SEVIRI scenes were analysed during May–June 2018 period. Comparable to Kericho's analysis in Figure 2. 5, most of the raining satellite signals in R_{RG} and R_{MWL} plots did not scatter over the whole range of value combinations of satellite reflectance (Figure 2. 7a,c) and brightness temperature difference (*Figure 2. 7*b,d).

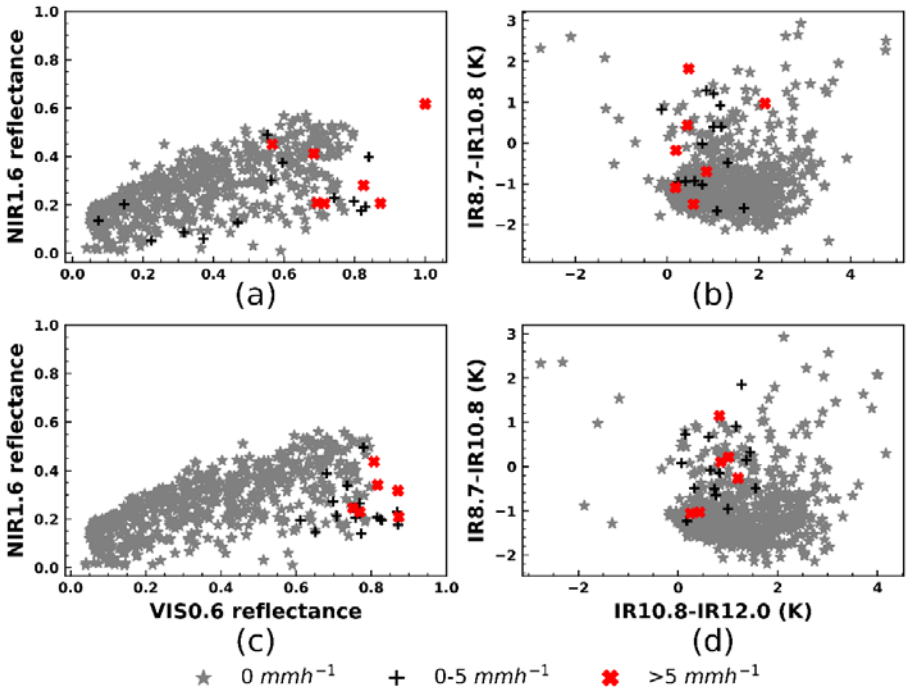

Figure 2. 7 As in Figure 2.5 but for daytime rain events in Naivasha.

The satellite reflectance combination for the 0 mm h^{-1} was comparable to Kericho's daytime observations (Figure 2. 5a,c), with a large concentration of very low VIS 0.6 µm and NIR 1.6 µm reflectance pairs. However, in contrast to the daytime analysis in Kericho, the R_{RG} between 0–5 mm h^{-1} tend to scatter over the whole range of the satellite signals (Figure 2. 7a). Also, for some of the rainfall intensities in this R_{RG} class, the value combination of the VIS 0.6 µm and NIR 1.6 µm were comparable. This plot also shows that most of the above 5 mm h^{-1} rainfall is scattered in the lower right corner with high VIS 0.6 µm and low NIR 1.6 µm reflectance pairs. For the above 5 mm h^{-1} R_{RG} that did not scatter in the lower right corner, their VIS 0.6 µm reflectance was nonetheless higher than those of the NIR 1.6 µm.

As can be seen from the $\Delta T_{IR10.8\text{-}IR12.0}$ versus $\Delta T_{IR8.7\text{-}IR10.8}$ plot (Figure 2. 7b,d), the 0 mm h^{-1} scatter over the whole range of the satellite signal, with generally larger values of $\Delta T_{IR10.8\text{-}IR12.0}$ than those of $\Delta T_{IR8.7\text{-}IR10.8}$. However, the raining satellite signals scatter over varying ranges of the two brightness temperature differences. For $\Delta T_{IR10.8\text{-}IR12.0}$, this range is above 0 K, with the majority of the signals falling between 0 and 1 K. For the $\Delta T_{IR8.7\text{-}IR10.8}$, the range is comparatively wider; in particular, the R_{RG} falls between −2 and 2 K. It can also be observed from the plot

that some coincident values of $\Delta T_{IR10.8\text{-}IR12.0}$ are larger than those of $\Delta T_{IR8.7\text{-}IR10.8}$. Likewise, values of $\Delta T_{IR8.7\text{-}IR10.8}$, in some cases, are larger than the values of $\Delta T_{IR10.8\text{-}IR12.0}$. This feature was evident for the rainfall between 0–5 mm h^{-1} and those above 5 mm h^{-1} and can be seen from the R_{RG} and R_{MWL} plots. Moreover, these observations were similar to those found in Kericho's previous analysis of daytime rainfall intensities and MSG satellite signals (Figure 5).

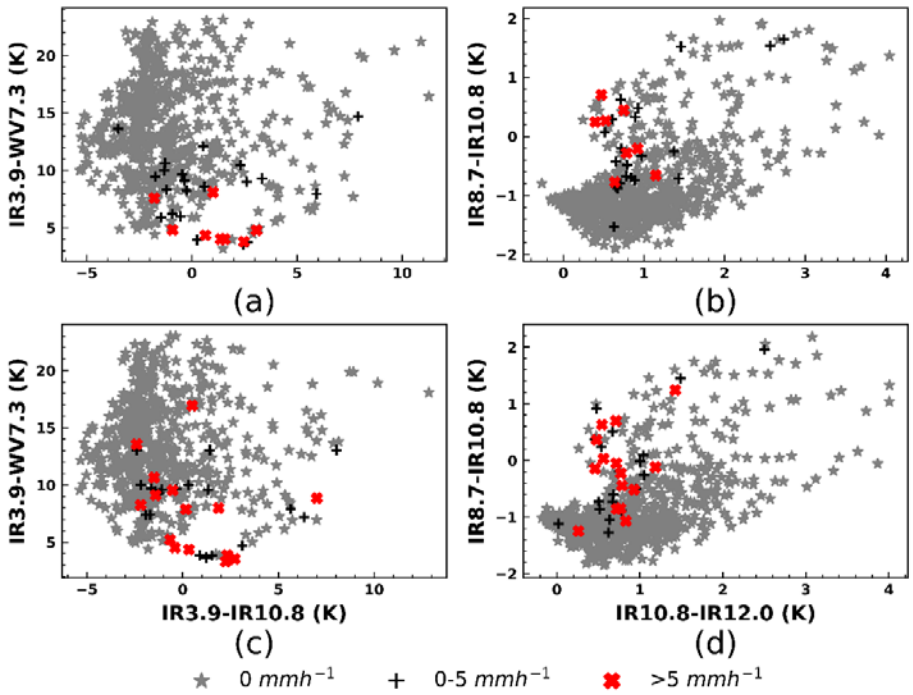

Figure 2. 8 As in Figure 2.6 but for night-time rainfall events in Naivasha.

Figure 2. 8 shows rainfall intensities as a function of MSG signals during night-time in May–June 2018 (altogether 733 MSG SEVIRI scenes) for Naivasha. A clear observation from the figure is that the satellite signals indicate 0 mm h^{-1} rainfall scattered over the whole range of values for all the brightness temperature differences.

However, the raining satellite scattered in different value ranges for the two brightness temperature differences. $\Delta T_{IR3.9\text{-}WV7.3}$ differences were below 15 K, and $\Delta T_{IR3.9\text{-}IR10.8}$ were between −5 and 5 K (Figure 2. 8a,c). Also, most R_{RG} between 0–5 mm h^{-1} scattered over a larger $\Delta T_{IR3.9\text{-}WV7.3}$ than those above 5 mm h^{-1} (Figure 2. 8a), whereas in the R_{MWL} plot (Figure 2. 8c), both rainfall intensity classes tend to scatter over a wide range.

The $\Delta T_{IR10.8\text{-}IR12.0}$ versus $\Delta T_{IR8.7\text{-}IR10.8}$ scatter plot (Figure 2. 8b,d) shows that most raining satellite signals scatter above 0 K for $\Delta T_{IR10.8\text{-}IR12.0}$ differences, and between −2 and 2 K for the $\Delta T_{IR8.7\text{-}IR10.8}$ differences. More precisely, the raining satellite signals of the 0–5 mm h^{-1} scattered over a wide range of value combinations of $\Delta T_{IR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$. By contrast, the signals of the above 5 mm h^{-1} were in a narrow range of value combinations of the brightness temperature differences (Figure 2. 8b).

(ii) Inferences from analyzing rainfall estimates with MSG satellite data

Overall, it can be stated based on the observations in Figure 2. 5 and Figure 2. 7 that most of the rain cases (i.e., 0–5 mm h^{-1} and above 5 mm h^{-1}) defined by R_{RG} and R_{MWL} during daytime were from optically thick clouds that characterized by high VIS 0.6 µm and low NIR 1.6 µm reflectance, and with different ranges of $\Delta T_{IR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$. The high VIS 0.6 µm and low NIR 1.6 µm reflectance indicate the cloud's large CWP and high rainfall probabilities and intensities (Thies et al., 2008b). The $\Delta T_{IR10.8\text{-}IR12.0}$ (between 0 and 1 K) and $\Delta T_{IR8.7\text{-}IR10.8}$ (between −2 and 2 K) ranges found in this study also indicated medium (i.e., low optical thickness with large particle sizes and high optical thickness with small particle sizes) to large (i.e., high optical thickness with large particle sizes) CWP with high rainfall probabilities and intensities according to radiative transfer calculation by Thies et al. (2008a). Further, the value ranges of $\Delta T_{IR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$ are characteristic of optically thick cumulous-type clouds with ice at the top, based on cloud classification presented by Feidas and Giannakos (2010). In contrast, most of the no rain cases (i.e., 0 mm h^{-1}) correspond to non-precipitating thin and thick cirrus clouds and N-type clouds (edges of optically thick clouds, optically thinner cumulous clouds, or low-level cumulous clouds overlaid by thin cirrus clouds) (Feidas & Giannakos, 2010; Inoue et al., 2001).

Optically thick clouds with medium $\Delta T_{IR3.9\text{-}IR10.8}$ (between -5 and 5 K) and $\Delta T_{IR3.9\text{-}WV7.3}$ (between 5 and 10 K) produced most of the night-time (Figure 2. 6 and Figure 2. 8) rain cases (i.e., 0–5 mm h^{-1} and above 5 mm h^{-1}). Thies et al. (2008a) showed that medium $\Delta T_{IR3.9\text{-}IR10.8}$ values are linked to large CWP with high rainfall probabilities and intensities. High and small $\Delta T_{IR3.9\text{-}IR10.8}$ is, however, indicative of medium CWP with low rainfall probabilities and intensities—the results for $\Delta T_{IR3.9\text{-}WV7.3}$ show comparable characteristics to $\Delta T_{IR3.9\text{-}IR10.8}$ but with generally higher $\Delta T_{IR3.9\text{-}WV7.3}$ than $\Delta T_{IR3.9\text{-}IR10.8}$ differences. The higher $\Delta T_{IR3.9\text{-}WV7.3}$ differences than $\Delta T_{IR3.9\text{-}IR10.8}$ is due to the diminishing effect of the water vapour absorption and emission in the mid-to-low tropospheric levels on the brightness temperature in WV7.3 µm channel (Schmetz et al., 2002). For the $\Delta T_{IR8.7\text{-}IR10.8}$

versus $\Delta T_{IR10.8\text{-}IR12.0}$ difference, the observations made for both rain and no rain cases were comparable to those found during the daytime analysis.

The results are consistent with the conceptual model introduced in Section 2.3.4 that clouds with high optical thickness and large particle sizes, indicative of large CWP (with the existence of ice or water hydrometeors at the top), have a high probability of producing rainfall. Based on this fact and according to the inspections of Figures 2.5–2.8, it raises the possibility of distinguishing rain and no rain cases based on a combination of parameters: VIS 0.6 μm, NIR 1.6 μm, $\Delta_{TIR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$ (for during daytime), and $\Delta T_{IR3.9\text{-}WV7.3}$, $\Delta T_{IR3.9\text{-}IR10.8}$, $\Delta_{TIR10.8\text{-}IR12.0}$ and $\Delta T_{IR8.7\text{-}IR10.8}$ (for during night-time). This is achieved using statistical machine learning or parametric (threshold) techniques, e.g., (Feidas & Giannakos, 2010; Lazri et al., 2013). Here, we empirically defined parametric thresholds based on descriptive statistics of the raining SEVIRI satellite signals for rainfall detection.

2.4.1.3 Rainfall Detection with MSG SEVIRI Data

The different combinations of MSG channels jointly analysed with rainfall data in the previous section were used to test rain detection on individual MWL. This section presents the results of detecting rain on MWL using MSG SEVIRI satellite signals. Table A2 1 shows calculated descriptive statistics of SEVIRI satellite signals per R_{MWL} class, study location, and day and night-time. A summary of the data used for calculating these statistical values is in Table 2. 5, and the data were analysed in Figures 2.5–2.8 in the previous section.

What is clear from Table A1 is the low and high standard deviation values of the reflectance and brightness temperature differences, respectively. When the standard deviation is interpreted together with the mean, mode, and median values, it becomes evident that the satellite reflectance tends to spread close to its mean. In contrast, the brightness temperature differences spread over a large range of values. Additionally, there were large differences in the statistical values between raining and non-raining satellite signals. However, the differences in statistical values between 0–5 mm h^{-1} and above 5 mm h^{-1} rainfall classes were relatively low.

Based on the descriptive statistics in Table A2 1, reflectance threshold and brightness temperature difference range (Table 2. 6) were derived that were used to compute the elements a, b, c, and d in Equations (2.8)–(2.15) for rainfall detection. Since the satellite signals of 0–5 mm h^{-1} and above 5 mm h^{-1} rainfall classes were not statistically different, the two classes' separate threshold and brightness temperature difference ranges were unnecessary. Thus, a single

threshold and brightness temperature difference range were defined separately for day and night-time and the two study locations. The reflectance thresholds: Vis_{thres} and Nir_{thres} (Table 2. 6) were derived from the median and mean statistical values, respectively, whereas the brightness temperature difference ranges were from a combination of minimum, maximum, mean and mode values depending on the channel difference. The results in Table 2. 6 show that the reflectance thresholds were comparable (during daytime). In contrast, the brightness temperature difference ranges (during day and night-time) varied across the two study locations.

Table 2. 6 Reflectance threshold and brightness temperature range used for rain detection.

Time	Study Location	Vis_{thres}	Nir_{thres}	$\Delta T_{IR8.7\text{-}IR10.8}$ K	$\Delta T_{IR10.8\text{-}IR12.0}$ K
Day	Kericho	>0.70	< 0.43	−1.0–1.42	−1.0–1.0
	Naivasha	>0.70	< 0.50	−1.10–1.15	0.0–1.2
		$\Delta T_{IR3.9\text{-}IR10.8}$ K	$\Delta T_{IR3.9\text{-}WV7.3}$ K	$\Delta T_{IR8.7\text{-}IR10.8}$ K	$\Delta T_{IR10.8\text{-}IR12.0}$ K
Night	Kericho	2.0–5.0	4.0–12.0	−0.01–1.0	0.26–1.9
	Naivasha	−3.0–1.0	3.0–15.0	−1.0–2.0	0.0–1.0

Table 2.7 summarizes the verification scores of the rain detection on individual MWL for Kericho and Naivasha during the day and night-time. The scores are based on MWL and satellite data from the two study locations independent of those used in the analysis of previous sections (Section 2.4.1.1 and 2.4.1.2). The MWL names (Table 2. 7, first column) were derived from the link identifications supplied by the telecommunication provider. The scores were computed across the evaluation period for MWLs with more than five R_{MWLs} retrieved during the day or night-time. Table 2. 7 shows the method's effectiveness in detecting no rain on individual MWL, indicated by, on average, more than 95% correct negative detections. However, the rain detection skill varied between day and night-time, as seen from the hit percentage.

The daytime detection in Kericho performed better than night-time detection. For most MWL, between 58.1% and 78.6% of the R_{MWL} were successfully detected, and between 21.4% and 42.9% were incorrectly identified as non-raining. The successful night-time detections were between 25% and 50%, whereas 50% to 75% of the R_{MWL} was incorrectly detected as non-raining. The false alarms, FAR, and POFD percentages were generally higher in the daytime than in the night-time. Nonetheless, the POD, CSI and HSS scores suggest better daytime rain detection than night-time. Altogether, the accuracy scores were high for the daytime and night-time tests, which can be attributed to the method's effectiveness in detecting no rain on individual MWL.

Table 2. 7 Computed performance verification scores of rain detection using MWL and MSG data.

MWL name	[1] R_{MWL} (mm)	Hits %	Miss %	False Alarms %	Correct Negatives %	POD	FAR	POFD	ACC	CSI	HSS
Perfect score		-	-	-	-	1	0	0	1	1	1
Kericho MWL					Day time						
13471368	311.32	58.1	41.9	29	96.4	0.58	0.33	0.04	0.92	0.45	0.58
13671368	230.87	78.6	21.4	28.6	98.4	0.79	0.27	0.02	0.97	0.61	0.75
30941368	331.15	57.1	42.9	17.9	98.4	0.57	0.24	0.02	0.94	0.49	0.62
30953094	83.53	66.7	33.3	66.7	98.1	0.67	0.50	0.03	0.96	0.40	0.55
34051368	437.98	65	35	50	96.3	0.65	0.44	0.04	0.94	0.43	0.57
					Nighttime						
13471368	124.62	50	50	30	98.5	0.50	0.38	0.02	0.96	0.39	0.5
13671368	138.43	25	75	6.2	99.5	0.25	0.2	0.01	0.94	0.24	0.4
30941368	112.65	41.7	58.3	8.3	99.5	0.42	0.18	0.01	0.96	0.39	0.5
30953094	68.72	50	50	12.5	99.5	0.50	0.2	0.01	0.98	0.44	0.6
Naivasha MWL											
					Day time						
13201328	77.0	65	35	5	99.8	0.65	0.07	0.002	0.99	0.62	0.76
34101372	203.73	61.5	38.5	15.4	99.6	0.62	0.20	0.004	0.99	0.53	0.69
13723379	201.47	57.1	42.9	28.6	99.6	0.57	0.33	0.004	0.99	0.44	0.61
1372	187.10	62.5	37.5	25	99.6	0.63	0.29	0.004	0.99	0.50	0.66
13201327	80.22	60	40	5	99.8	0.6	0.08	0.02	0.98	0.57	0.72
13071372	232.91	61.5	38.5	15.4	99.6	0.62	0.20	0.004	0.99	0.53	0.69
					Nighttime						
13201328	42.10	91.7	8.3	58.3	98.9	0.92	0.39	0.011	0.99	0.58	0.73
13723365	169.10	57.1	42.9	78.6	97.7	0.57	0.58	0.023	0.97	0.32	0.47
13201327	21.57	30	70	120	98.2	0.3	0.8	0.018	0.97	0.14	0.23
13263035	46.46	60	40	100	98.9	0.6	0.63	0.011	0.99	0.3	0.46
13263302	55.34	45.5	54.5	54.5	98.7	0.46	0.55	0.013	0.97	0.29	0.44

[1]Accumulated R_{MWL} for each MWL computed across the evaluation period.

Similarly, the daytime detection results in Naivasha were comparatively better than the night-time. For daytime, the method successfully detected between 57.1% and 62.5% of the R_{MWL}, whereas between 35% and 42.9% were erroneously detected as non-raining. However, between 30% and 91.7%, successful R_{MWL} detections were achieved during the night-time. The erroneous detections were between 8.3% and 70%.

Like the Kericho study, the percentage of false alarms, FAR, and POFD for the daytime were lower than the night-time. However, the POD, CSI, and HSS indicate a better daytime performance than the night-time. For the daytime: POD range between 0.57 and 0.65; CSI range between 0.44 and 0.62; and HSS range between 0.61 and 0.76. For the night-time, values ranging between 0.3 and 0.92; 0.14 and 0.58 and 0.23 and 0.73 were computed for POD, CSI, and HSS respectively. Regarding the overall accuracy for the day and night-time, the previous comment for the Kericho case is also valid.

2.4.2 Discussion

2.4.2.1 Accuracy of the MWL Rainfall Estimates

Two MWL of the same frequency (15 GHz) but different lengths: 3.7 and 10 km, and from different study locations: Kericho and Naivasha, were used to estimate R_{MWL}. The results were then compared, at different evaluation timestamps, with R_{RG} derived from rain gauges. Overall, the results demonstrate the potential and capacity of MWL networks to provide accurate rainfall data. The overall accuracy of the R_{MWL} derived from the two MWLs can be described as good. Despite the high CV observed, the low RB and RMSE values show that reliable rainfall information can be derived from MWL networks. The hourly r^2 values were about 0.60 for both MWLs, which can be described as convincing, considering these values were achieved at fine resolution.

The discrepancies between R_{MWL} and R_{RG} were attributed to factors such as the differences between the nature of measurement by the MWL and gauge. The R_{MWL} were retrieved from the mean RSL data derived from instantaneous RSL measurements. Additionally, the R_{MWL} represent the average rainfall over the MWL's signal transmission path. On the other hand, the R_{RG} was computed from rain gauges that recorded rainfall accumulations every minute from a single point. Moreover, uncertainties in the R_{MWL} rainfall estimation approach and spatial variability of rainfall could contribute to the differences between R_{MWL} and R_{RG} (David et al., 2019; Martin & Vojtech, 2018; Uijlenhoet et al., 2010).

2.4.2.2 The Analysis of RMWL with MSG SEVIRI Data

R_{MWL} from the two links presented above were jointly analysed with satellite signals from MSG SEVIRI solar and thermal infrared channels. Both data were temporarily aggregated, and the satellite data were spatially aggregated to reduce the effect of potential temporal and spatial mismatch between the rainfall and satellite data in our analysis. The premise for analysing the MWL rainfall with the SEVIIRI data is that raining clouds that attenuate individual MWL can be detected based on their cloud top properties using satellite data. Therefore, a conceptual model was defined for detecting rainfall based on the assumption that clouds with high cloud top optical thickness and large particle sizes (with ice or water hydrometeors) have high rainfall probabilities and intensities.

During the daytime, the model infers cloud top optical thickness and particle sizes from SEVIRI solar channels: VIS 0.6 µm and NIR 1.6 µm, respectively, whereas the brightness temperature difference between IR 8.7 µm and 10.8 µm ($\Delta T_{IR8.7\text{-}IR10.8}$) and that between IR 10.8 µm 12.0 µm ($\Delta T_{IR10.8\text{-}IR12.0}$), supplied additional information on cloud phase namely: ice and water, respectively. During the night-time, the brightness temperature differences: $\Delta T_{IR3.9\text{-}IR10.8}$, $\Delta T_{IR3.9\text{-}WV7.3}$, $\Delta T_{IR8.7\text{-}IR10.8}$, and $\Delta T_{IR10.8\text{-}IR12.0}$ inferred similar information about the cloud top properties.

When R_{MWL} was considered as a function of cloud optical thickness (VIS 0.6 µm reflectance) and particle sizes (NIR 1.6 µm reflectance) during the daytime, the raining satellite signals had high values of VIS 0.6 µm and low values of NIR 1.6 µm reflectance. On the other hand, $R_{MWL,}$ as a function of $\Delta T_{IR8.7\text{-}IR10.8}$ and $\Delta T_{IR10.8\text{-}IR12.0,}$ showed that the raining satellite signals grouped in a different range of values of the two brightness temperature differences. $\Delta T_{IR8.7\text{-}IR10.8}$, indicative of ice clouds, covered a large range of values, while $\Delta T_{IR10.8\text{-}IR12.0}$, indicative of water clouds, covered a narrow range. Night-time analysis of R_{MWL} as a function of $\Delta T_{IR3.9\text{-}WV7.3}$ and $\Delta T_{IR3.9\text{-}IR10.8}$ also showed that the raining satellite signals spread over a wide range of value combinations of $\Delta T_{IR3.9\text{-}WV7.3}$ and $\Delta T_{IR3.9\text{-}IR10.8}$. However, the ΔTIR3.9-WV7.3 differences were mostly higher than those of $\Delta T_{IR3.9\text{-}IR10.8}$. For the analysis of R_{MWL} as a function of $\Delta T_{IR8.7\text{-}IR10.8}$ and $\Delta T_{IR10.8\text{-}IR12.0}$, the brightness temperature differences in the raining satellite signals were comparable to those found in the daytime analysis.

The results corroborate the conceptual idea that clouds with high optical thickness and large particle sizes have high rainfall probability and show the possibility for daytime and night-time area-wide rainfall detection based on the different MSG satellite signals. The rain detection results indicate an appreciable capability of the satellite data to detect rainfall intensities on individual MWL.

The verification scores showed that the test performed better during the daytime and was comparable across the two study locations. This could be due to the more reliable information on CWP and rainfall available in the VIS 0.6 µm and NIR 1.6 µm reflectance pair compared to channel differences used at night (Thies et al., 2008d). The results also show a strong relationship between VIS 0.6 µm, NIR 1.6 µm reflectance and R_{MWL} during the daytime.

Moreover, descriptive statistical values from the R_{MWL} and satellite data analysis indicate that the daytime satellite reflectance was statistically different between the rain and non-raining satellite signals. Based on the daytime test, this might explain the high rain detection with good CSI and HSS scores. The statistical values derived from the night-time data were often comparable between rain and non-raining satellite signals. They could explain its high false detection with comparatively weak CSI and HSS scores.

The results suggest that the MWL and MSG data can potentially retrieve area-wide rainfall. A possible implication is that the MWL and MSG data could be potential input to optical satellite-based rainfall detection and estimation models. In this context, data fusion (Lahat et al., 2015; Safont et al., 2019) and statistical machine learning offer techniques to explore relationships (linear and nonlinear) between multiple large input variables. Further, current developments in parallel computing with machine learning have increased the training and predicting speeds of these learning algorithms and can make the automatization and improvement of the method in real-time feasible (Kühnlein et al., 2014a; Meyer et al., 2016). In addition, for MWL-based rainfall studies in areas lacking in-situ data, the results suggest that satellite-based rain information may be a valuable tool for the wet/dry classification of the MWL signal. Therefore, future research involving analysis with large datasets would also focus on investigating the MWL and satellite data for such an application.

Uncertainties in the analysis of R_{MWL} with MSG satellite data could be due to many factors, such as the measurement differences between the satellite and MWL. The satellite signal is acquired instantaneously over an area, while the R_{MWL} retrieved from RSL every 15 min represents rainfall along the link transmission path. Therefore, it is unsurprising that the satellite signal might not correspond with the rainfall intensities. Furthermore, the spatiotemporal aggregation employed to retrieve the most effective satellite signal from the satellite data is, perhaps, not enough to compensate for measurement differences inherent in the two sensors.

Additionally, horizontal wind drift within or outside raining clouds strongly influences falling hydrometeors of light rainfall. It could partly explain the high

uncertainty observed for the 0–5 mm h^{-1} rainfall intensity class. In the case of deep convective cores with anvil cloud areas, the retrieval of the satellite signal of high rainfall intensities clouds might be biased towards anvil cloud areas, which have characteristics of light rainfall intensities. This could explain why some rainfall intensities above 5 mm h^{-1} did not correspond with the satellite signals. Moreover, in the case of multi-layered clouds (Watts et al., 2011), with cloud properties differing between layers, the satellite signals may not correlate with the ground rainfall (Lensky & Rosenfeld, 1997).

To this end, distinguishing different raining cloud types, e.g., convective/stratiform, and tracking their footprints surrounding the individual MWL will provide valuable information relevant for retrieving and linking the MSG satellite signals to the MWL rainfall and ultimately improve this study's technique. Identifying different cloud types and tracking their location around the MWL should reduce the spatial aggregation problems and the resulting uncertainties between MWL rainfall and the MSG satellite signals. In this context, the promising results demonstrated by (Thies et al., 2008c; Thies et al., 2008d; Wang et al., 2019; Zipser et al., 2006) for raining cloud classification and (O'Neil et al., 2021; Turdukulov et al., 2007) for tracking raining clouds accentuate the potential of a more accurate MWL and satellite-based raining cloud detection. Future studies will consider these concepts.

2.5 Conclusion

This study investigated the combination of MWL rainfall estimates with MSG SEVIRI satellite data for rainfall detection using observations from two areas in Kenya: Kericho and Naivasha. The approach first evaluated the MWL rainfall estimates using an experimental setup consisting of two MWL and several rain gauge measurements as ground truth. Next, the MWL rainfall estimates were analysed as a function of the MSG SEVIRI satellite signals. The satellite signals were used to infer cloud top properties in a conceptual model that clouds with high cloud top optical thickness and large particle size (with ice or water hydrometeors) are linked to high rainfall probabilities and intensities. Finally, the information gained from analysing the MWL rainfall with MSG SEVIRI satellite signals successfully detected rain on individual MWL.

The results show that the MWL can estimate rainfall intensities with reliable accuracy compared with rain gauges. However, limitations, such as inherent measurement differences between the MWL and rain gauge, remain challenging and affect rainfall retrieval accuracies. Analysing the MWL rainfall estimates as a function of the MSG satellite signals revealed the raining satellite signals'

characteristics that corroborate with the conceptual model for detecting rain clouds. During the daytime, high VIS 0.6 µm and low NIR 1.6 µm reflectance (indicative of high optical thickness and large particle sizes and thus large CWP) were often consistent with MWL rainfall between 0–5 mm h^{-1} and above 5 mm h^{-1}. These raining satellite signals differed from the non-raining satellite signals (satellite signals indicative of 0 mm h^{-1}) and were comparable across the two study locations.

For night-time, ΔTIR3.9-IR10.8 and ΔTIR3.9-WV7.3 brightness temperature ranges, indicative of medium to large CWP characterized MWL rainfall between 0 and 5 mm h^{-1} and above 5 mm h^{-1}. Nonetheless, the ranges varied between the two study areas and were not different between rain and non-raining satellite signals. Additionally, daytime and night-time temperature ranges of $\Delta T_{IR8.7\text{-}IR10.8}$ and $\Delta T_{IR10.8\text{-}IR12.0}$ for the MWL rainfall between 0 and 5 mm h^{-1} and above 5 mm h^{-1} suggest that most of the rain cases were from optically thick cumulous-type clouds with icy cloud tops. Eventually, descriptive statistics: minimum, maximum, mean, mode, median and standard deviation of the satellite signal were used to determine threshold and brightness temperature difference range that tested the potential of combining information from the different satellite signals to detect rainfall on MWL. Overall, the verification scores indicate a considerable capability of the satellite data to detect rainfall on MWL, particularly during the daytime.

One should interpret the findings of this study while considering its potential limitations. The assumed relationship between MWL rainfall and MSG SEVIRI satellite signals was confirmed using a limited dataset and local origin. Thus, the descriptive statistics of the rain and non-raining satellite and the corresponding parameters (i.e., the thresholds and brightness temperature difference ranges) should be considered restricted to the study locations. Nonetheless, the approach can be applied to other areas to retrieve representative thresholds using a sufficient dataset. Furthermore, to allow for the operational and large-scale application of MWL and MSG data for rainfall studies based on this study's findings, further research involving large datasets are necessary to support the observed relation between MWL rainfall and MSG-based cloud top properties.

In this regard, the dataset should be enlarged in both space and time. A larger time series of MWL and MSG data pairs retrieved from a spatially dense MWL network should compute more representative descriptive statistics and parameter estimates while improving the retrieval technique (Bell & Kundu, 2003). Also, data fusion and machine learning techniques offer possibilities for exploring relationships between large numbers of input variables from different sensors. These learning algorithms provide efficient alternatives and may be

suitable for overcoming parametric approaches' limitations while potentially automating the retrieval approach (Kühnlein et al., 2014a; Kühnlein et al., 2014b).

Based on this study's findings, it can be stated that the combination of the MWL rainfall and MSG SEVIRI data has the potential for area-wide rainfall detection at a high temporal resolution. This is especially significant for areas lacking in-situ monitoring systems but having good MWL coverage established (and maintained) over the last decade and is still further developed. Using both in-situ and MWL data will potentially enhance the density of the observation network to monitor rainfall on the ground, and with high temporal resolution remote sensing derived information, as acquired by geostationary satellites; better temporal and spatial extrapolation will be possible.

APPENDIX A2

Table A2 1 Descriptive statistics of SEVIRI satellite signals for each rainfall intensity class and study location.

MSG signal	R_{MWL} (mm h^{-1})	Min	Max	Mean	Mode	Median	SD
Kericho							
			Day time				
	0	0.073	0.886	0.417	0.3	0.38	0.215
VIS 0.6 µm	0–5	0.562	0.972	0.74	0.8	0.77	0.15
	>5	0.374	1	0.726	0.7	0.726	0.148
	0	0.036	0.858	0.337	0.3	0.32	0.132
NIR 1.6 µm	0–5	0.31	0.608	0.433	0.3	0.384	0.137
	>5	0.178	0.536	0.304	0.3	0.296	0.09
	0	−2.357	2.177	−0.613	−1	−0.756	0.88
$\Delta T_{IR8.7\text{-}IR10.8}$	0–5	−1.826	1.417	−0.453	-	-0.57	1.10
	>5	−1.721	1.42	−0.137	−0.9	−0.354	0.915
	0	−0.55	4.336	1.711	2.3	1.736	0.98
$\Delta T_{IR10.8\text{-}IR12.0}$	0–5	−0.215	2.366	0.737	0.9	0.617	0.914
	>5	−1.836	2.064	0.469	0.8	0.561	0.749
			Nighttime				
	0	−5.252	9.709	−0.204	−1.6	−1.166	3.194
$\Delta T_{IR3.9\text{-}IR10.8}$	0–5	0.033	4.57	2.087	-	1.998	1.633
	>5	−0.366	5.026	2.454	1.7	2.563	1.686
	0	3.466	25.765	12.884	13.2	12.64	5.122
$\Delta T_{IR3.9\text{-}WV7.3}$	0–5	5.933	25.068	16.938	-	16.537	7.134
	>5	4.234	13.046	7.033	-	5.226	3.711
	0	−2.024	2.354	−0.702	−1.1	−0.959	0.894
$\Delta T_{IR8.7\text{-}IR10.8}$	0–5	−0.944	0.76	−0.306	−0.7	−0.724	0.72
	>5	−1.242	0.988	0.272	0.3	0.318	0.734
	0	0.072	3.976	1.449	1.5	1.365	0.808
$\Delta T_{IR10.8\text{-}IR12.0}$	0–5	0.655	2.162	1.732	1.9	1.918	0.611
	>5	0.258	1.748	0.678	0.3	0.532	0.511
Naivasha							
			Day time				
	0	0.04	0.795	0.338	0.1	0.314	0.221
VIS 0.6 µm	0–5	0.611	0.871	0.758	0.8	0.764	0.075
	>5	0.751	0.873	0.815	0.8	0.812	0.05
	0	0.01	0.561	0.256	0.2	0.239	0.124
NIR 1.6 µm	0–5	0.14	0.495	0.244	0.2	0.21	0.093
	>5	0.209	0.437	0.296	0.2	0.281	0.086
	0	−2.199	2.923	−0.973	−1.2	−1.184	0.775
$\Delta T_{IR8.7\text{-}IR10.8}$	0–5	−1.232	1.85	−0.066	−0.5	0.114	0.834
	>5	−1.054	1.146	−0.146	-	−0.077	0.835

$\Delta T_{IR10.8-IR12.0}$	0	−2.76	4.17	1.31	1.8	1.282	0.859
	0−5	0.073	1.556	0.786	0.1	0.748	0.497
	>5	0.249	1.202	0.763	-	0.843	0.359
				Nighttime			
$\Delta T_{IR3.9-IR10.8}$	0	−5.37	12.848	−1.162	−2.5	−1.972	2.436
	0−5	−2.404	8.041	0.844	-	0.315	3.036
	>5	−2.4	6.999	0.537	-	0.166	2.42
$\Delta T_{IR3.9-WV7.3}$	0	3.834	22.985	13.495	11.7	13.33	4.015
	0−5	3.648	13.04	8.537	13	9.44	2.967
	>5	3.317	16.932	7.845	-	7.994	3.902
$\Delta T_{IR8.7-IR10.8}$	0	−1.851	2.175	−0.84	−1.2	−1.071	0.709
	0−5	−1.277	1.953	−0.144	−0.9	−0.258	0.884
	>5	−1.246	1.239	−0.171	−0.1	−0.145	0.697
$\Delta T_{IR10.8-IR12.0}$	0	−0.181	4.008	0.945	0.9	0.851	0.715
	0−5	0.017	2.504	0.809	0.5	0.674	0.514
	>5	0.258	1.426	0.742	0.8	0.72	0.289

Chapter 3: Rain Area Detection in South-Western Kenya by Using Multi-spectral Satellite Data from Meteosat Second Generation

This chapter is originally published in:
Kingsley, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., Rwasoka, D. T., Retsios, B. V., & Su, B. Z. (2021). Rain Area Detection in South-Western Kenya by Using Multispectral Satellite Data from Meteosat Second Generation. Sensors (Basel), 21(10). doi:10.3390/s21103547

Abstract

This study presents a rain area detection scheme that uses a gradient-based adaptive technique for daytime and night-time rain area detection and correction from the MSG satellite's reflectance and IR brightness temperature data. First, multiple parametric rain detection models developed from the reflectance and IR data were calibrated and validated with rainfall data from a dense network of rain gauge stations and investigated to determine the best model parameters. The models were based on a conceptual assumption that clouds characterised by the top properties, e.g., high optical thickness and effective radius, have high rain probabilities and intensities. Next, a gradient-based adaptive correction technique that relies on rain area-specific parameters was developed to reduce the number and sizes of the detected rain areas. The daytime detection with optical (VIS0.6) and near IR (NIR1.6) reflectance data achieved the best detection skill. For night-time, detection with thermal IR brightness temperature differences of IR3.9-IR10.8, IR3.9-WV7.3 and IR10.8-WV6.2 showed the best detection skill based on general categorical statistics. Compared to GPM IMERG and the gauge station data from southwest Kenya, the model showed good agreement in the spatial dynamics of the detected rain area and rain rate.

3.1 Introduction

Accurate rainfall detection and estimation in space and time are essential for resolving scientific questions for operational purposes such as early warning, forecasting and the development of services and applications that affect and influence human welfare and agricultural production decisions at a farm-scale level (David et al., 2019; David et al., 2021; Hong et al., 2007; Macharia et al., 2020). Unfortunately, accurate detection and estimation are still open research challenges due to the high spatial and temporal variability of rainfall.

Numerous studies have focused on using geostationary weather satellites for area-wide rainfall detection and estimation because the satellite's high temporal resolution favours the study of this spatiotemporally varying phenomenon (Feidas & Giannakos, 2010; Kühnlein et al., 2014a; Kumah et al., 2020). Geostationary weather satellite-based information for retrieving rainfall is restricted to the cloud top and indirectly relates to ground rainfall observations (Rosenfeld, 2007). Nonetheless, some authors, e.g. (Bergès et al., 2005), have indicated that the information helps quantify rainfall because of the apparent relationship between rain duration and amounts.

In the past, most retrieval techniques focused on the relationship between a single cloud top property (the cloud top BT) from the IR channel, rainfall probabilities and intensities (Adler & Negri, 1988; Milford et al., 1994). This technique works best for intense cold convective clouds. Another challenge with this technique is distinguishing non-raining cold cirrus clouds from the raining convective ones. The result is an overestimation of the detected rain areas and the corresponding rain rates (Feidas & Giannakos, 2010). While some of the retrieval techniques, e.g. (Adler & Negri, 1988), consider screening out cirrus clouds, they usually require ancillary data (e.g., radar data), which is not always available everywhere.

Furthermore, the single IR retrieval technique shows drawbacks when applied to detect and quantify rainfall from stratiform clouds. These clouds have spatially homogenous (warmer) temperatures that do not differ significantly between raining and non-raining clouds. Stratiform rains are usually of low intensities but with extensive coverage. As a result, a single IR-based technique leads to uncertainties in the detected rain areas and rain rates (Kühnlein et al., 2014a). IR-only satellite rainfall products such as Milford et al. (1994) were developed for operational application in Africa. Nevertheless, ground validation over the years has shown varying accuracies over the continent see e.g., (Ross et al., 2014; Seyama et al., 2019).

To overcome these downsides, several authors suggest using multispectral satellite data (Feidas & Giannakos, 2010; Inoue & Aonashi, 2000; Kumah et al., 2020; Thies et al., 2008a, 2008b). Multispectral satellites are passive remote sensing satellites that measure reflected energy within several specific regions (called spectral bands/channels) of the electromagnetic spectrum. Inoue and Aonashi (2000) studied cloud information from visible, IR scanners and rain information from precipitating radar onboard the TRMM and GPM satellite for rain area detection. They showed that additional details from a second channel improve detection results compared to the single IR retrieval technique. Thies et al. (2008a, 2008b) proposed a method for discriminating rain from non-raining clouds during the day and night that combines visible and IR multispectral satellite data. Likewise, Feidas and Giannakos (2010) demonstrated rain detection by utilising spectral and textural cloud features inferred from several IR channels.

Multispectral satellite-based rainfall detection often applies parametric-based techniques that relate cloud top property derived from the satellite data to rainfall occurrences and rain rates see, e.g., (Feidas & Giannakos, 2010; Kumah et al., 2020). The application is straightforward and requires defining the underlying conceptual models and parametric tests. The advantage is that they directly map the conceptual knowledge of the rainfall process onto the retrieval using the satellite data as proxies (Kühnlein et al., 2014a).

In Feidas and Giannakos (2010), MSG IR satellite data was investigated for rain area delineation using a parametric technique and a conceptual model in which cloud top properties such as optical thickness and phase were used to detect precipitating clouds. Thies et al. (2008d) used visible and IR data from MSG for daytime and night-time precipitation process separation and rainfall intensity differentiation. Their conceptual design characterised convective clouds with higher rainfall intensities by their larger vertical extensions and cold top temperatures. However, the uncertainties in the detected rain areas are often significant, which calls for the need to develop new techniques for improving rain area detection.

Motivated by the need for accurate rain area detection and the need to use multispectral data to improve rain area detection, this chapter utilised reflectance and IR data from the MSG SEVIRI satellite for rainfall detection with two primary objectives. (1) To investigate the satellite data's suitability for detecting raining areas over topographically complex regions by evaluating multiple rainfall detection models. (2) To develop a rain area correction technique for improving the detected rain areas.

The primary vehicle for this investigation is a rain detection model developed based on a parametric threshold technique. Compared to state-of-the-art MSG-based parametric rainfall detection studies (Feidas and Giannakos, 2010), which evaluated the satellite data's ability to discriminate between raining and non-raining clouds, this study distinguishes itself for the following reasons. (1) We optimally searched for daytime and night-time rain area detecting parameters from both optical and thermal IR data, and (2) we improved the detected rain areas using a newly developed gradient-based adaptive technique.

The structure of this chapter is as follows. The data and method are presented in Section 3.2. The results are shown in Section 3.3 and discussed in Section 3.4. Finally, Section 3.5 presents a conclusion on the significant findings.

3.2 Study Area and Dataset

3.2.1 Study Area

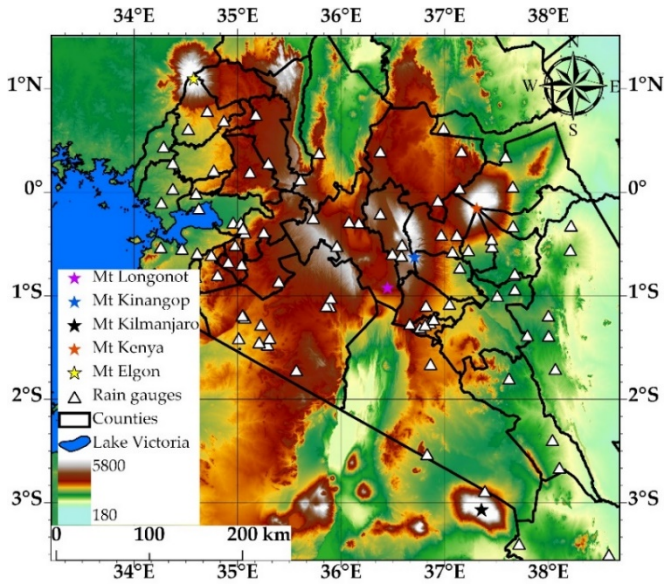

Figure 3. 1 Study area and locations of rain gauges (triangles) displayed using ALOS DEM

The area investigated (southwest Kenya) is shown in Figure 3. 1 using ALOS World 3D 30 m (AW3D30) DEM (Caglar et al., 2018). This area's rainy season occurs from March to June ("long rains"), and the second season is from October to December ("short rains"). The rainfall variability in the area is generally linked to the seasonal passage of the ITCZ over Kenya (Schneider et al., 2014), tropical pacific sea surface temperature (Maidment et al., 2015), ENSO and the IOD

(Hoell & Funk, 2013; Nicholson, 2017). Nonetheless, relief features like mountains (Mount Kenya, Rift valley) and large inland water (Lake Victoria) bodies influence the local rainfall variability (Tazalika et al., 2013).

3.2.2 Dataset

The dataset consists of MSG SEVIRI, GPM IMERG and rain gauge rainfall observations for the long rain period of 2018, 2019 and 2020. The MSG data was from the Meteosat satellite at longitude 41.5° E, including reflectance and IR channels sensitive to different cloud top properties, e.g., optical thickness, effective radius and phase during the day and night. These channels correspond to visible (VIS0.6 µm), near-IR (NIR1.6 µm), thermal IR (IR3.9 µm, IR8.7 µm, IR10.8 µm and IR12.0 µm) and water vapour (WV6.2 µm and WV7.3 µm) which were acquired at 15 min interval. The data are freely available in the EUMETSAT data archives (EUMETSAT, 2020)).

The rain gauge data was from the TAHMO. TAHMO maintains a network of ground-based weather stations across Sub-Saharan Africa. These stations (see locations in Figure 3. 1) measure standard meteorological variables such as rainfall, relative humidity, solar radiation and wind speed at 5 min intervals. (van de Giesen et al., 2014). The data ground validated the GPM IMERG satellite rainfall product over Africa (Dezfuli et al., 2017). This study used rainfall data from ninety stations distributed over the study area and had data during the evaluation period. TAHMO records rainfall accumulations every 5 min, which was used to estimate rain rates at 30 min intervals. It is also noteworthy that the TAHMO data are not part of the global network of rain gauges used by the GPCC. Therefore, this study independently compared its results with the GPM IMERG final run version 6 (V06B) satellite global precipitation product, which is calibrated with rainfall data from GPCC.

The IMERG final run version 6 (V06B) is the latest level 3 globally gridded satellite precipitation product derived from satellite radiometric observations from several GPM constellation satellites—consisting of a GPM Core Observatory satellite equipped with a dual-frequency precipitation radar and a 13-channel passive microwave (PMW) imager, and multiple partner satellites. The algorithm draws strength from previous satellite merging techniques such as the TMPA (Huffman et al., 2010). In the algorithm, rainfall estimates from the constellation satellites and based on the GPROF2017 are first gridded and intercalibrated to the estimates of the GPM Core satellite. They are then merged from their native spatial resolution to the IMERG grid at a half-hourly time step. For areas with no PMW overpass and beyond a forecast time of ±30 min from the closest PMW observation, IMERG uses the CPC CMORPH-KF Lagrangian time interpolation

scheme PERSIANN-CCS re-calibration scheme to create the half-hourly estimates.

Unlike previous versions, the motion vectors for the morphing are from hourly water vapour motion vectors from MERRA, version 2 (Joyce et al., 2004; Tan et al., 2019). Among the various IMERG precipitation products, this study utilised the MERG Calibrated precipitation estimates (precipitationCal), a gauge calibrated rainfall product using the GPCC data. The data is available freely (Huffman et al., 2014), with 10 km, 30 min resolution.

3.3 Method

3.3.1 Spectral Characteristics and Cloud Top Properties

The method for detecting rain clouds was based on investigating the relationship between the satellite data's spectral characteristics, cloud top properties (Table 3. 1) and rain. The conceptual idea used was that clouds characterised by their top properties, such as high optical thickness and effective radius (consisting of either ice or water hydrometeors), have high rainfall probability and intensities than those with low optical thickness and effective radius (Kumah et al., 2020; Thies et al., 2008a). The physical basis of this assumption is derived from the following characteristics of raining clouds: (i) the availability of adequate moisture, (ii) an effective mechanism for converting small cloud droplets that are suspended in the atmosphere into raining particles and (iii) existence of ice phase clouds to support rain generation by the Bergeron–Findeisen process (Lensky & Rosenfeld, 2003a).

This study utilised the MSG satellite's original reflectance, BT and BTD to infer cloud top properties that detected raining areas under day and night conditions. Different combinations of satellite channels were used for rain detection during the day and night periods because the MSG reflectance channels do not have night-time data. The use of IR3.9 μm during the day and twilight is also discouraged due to solar and thermal contributions and varying solar components in this channel (Kerkmann et al., 2014; Meyer et al., 2016). The daytime and night-time periods were from 04:15 AM to 3:15 PM and 4:15 PM to 03:15 AM UTC, respectively, excluding twilight periods between 03:30 AM and 04:00 AM and 3:30 PM and 4:00 PM UTC. The EUMETSAT operational cloud mask product (EUMETSAT, 2015) identified cloudy pixels in this study to ensure that only satellite data from cloudy scenes were used.

Table 3. 1 Spectral characteristics and cloud top parameters used for rain detection

Spectral Characteristics	Inferred Cloud Top Parameter	Application
VIS0.6 and NIR1.6	Optical thickness and effective radius	Daytime
IR10.8 K	Cloud top temperature	Day and night-time
IR10.8–WV6.2 K	Height	Day and night-time
IR10.8–IR12.0 K	Phase (water)/optical thickness	Day and night-time
IR8.7–IR10.8 K	Phase (ice)	Day and night-time
IR3.9–IR10.8 and IR3.9–WV7.3 K	Optical thickness and effective radius	Nighttime
WV6.2–WV7.3 K	Level (high or low)	Day and night time

Several studies have shown that the two reflectance channels: VIS0.6 and NIR1.6 µm, can be used to infer cloud top optical thickness and effective radius for successful rainfall detection (Feidas & Giannakos, 2010; Kumah et al., 2020; Lazri et al., 2013). Clouds with high optical thickness and effective radius (water or ice hydrometeors) have comparatively higher VIS0.6 µm and low NIR1.6 µm reflectance than clouds with low optical thickness and effective radius. The cloud top optical thickness and effective radius properties point towards a single parameter, cloud water path (CWP), interpreted as the amount of water vertically integrated into the cloud and directly linked to the clouds' rainfall probability. More precisely, when high VIS0.6 µm reflectance coincides with low NIR1.6 µm reflectance, it suggests clouds with high optical thickness and effective radius, and as such, large CWP is observed.

The BT at IR10.8 µm channels is a good indicator of the cloud's vertical extent because the IR10.8 µm BT of a cloud depends on its top height (Feidas & Giannakos, 2010; Giannakos & Feidas, 2012). Inoue and Aonashi (2000) observed lower BT (less than 260 K) in 11 µm for raining areas identified by a precipitation radar. The BTD between IR10.8 and WV6.2 µm was previously used for cloud classification (Lutz et al., 2003; Schmetz et al., 1997) and rainfall detection (Kidder et al., 2005). In the latter, clouds with BTD < 11 K for IR10.8–WV6.2 (an empirically determined threshold) were classified as raining.

The split window technique (i.e., BTD IR10.8–IR12.0) implemented by (Inoue, 1987a, 1987b) can be used to gain information about cloud optical thickness, which is useful for discriminating optically thick cumulus clouds from optically thin cirrus clouds (Feidas & Giannakos, 2011; Inoue, 1985). Optically thick cumulus clouds show small BTD because of their blackbody characteristics. In

contrast, optically thin cirrus clouds show a larger difference due to the differential absorption by ice crystals between the two channels (Inoue et al., 2001).

The tripsectral BTD: IR8.7–IR10.8 and IR10.8–IR12.0 have been used to infer cloud phase information (Kumah et al., 2020; Thies et al., 2008b). The basis for employing these channel differences for inferring the cloud phase lies in earlier results by Strabala et al. (1994). They showed that absorption of solar radiation by cloud hydrometeors differs (for ice and water) between the two BTDs. Water particle absorption is stronger between 11 and 12 µm (comparable to the IR10.8–IR12.0 BTD of MSG) than between 8 and 11 µm (similar to the IR8.7–IR10.8 BTD of MSG). For ice, the reverse is correct.

The BTD for IR3.9–IR10.8 and IR3.9–WV7.3 have also been used to infer cloud optical thickness and effective radius. The former is linked to the differential emissivity at IR3.9 µm and IR10.8 µm channel by large and small cloud hydrometeor particle sizes. Thus, the IR3.9–IR10.8 BTD are higher for optically thick clouds with a high effective radius than for clouds with small particle sizes (Lensky & Rosenfeld, 2003a; Thies et al., 2008d). The latter BTD are similar to that of IR3.9–IR10.8 but with generally higher differences due to the diminishing effect of the water vapour absorption and emission in the mid-to-low tropospheric levels on the BT in the WV7.3 µm channel (Schmetz et al., 2002).

Unlike WV6.2 µm, the WV7.3 µm channel is positioned at the edge of the water vapour absorption band (approximately 50 kPa). Thies et al. (2008c) performed radiative transfer simulations for the spectral ranges of SEVIRI WV and IR channels from cloud-free and variable cloud top heights. They indicated that for cloud tops below the tropopause temperature level, the WV6.2 µm BT are lower than the WV7.3 µm BT. In contrast, for cloud tops above the tropopause level, the WV6.2 µm BT is higher than that of the WV7.3 µm BT due to the stronger absorption lines of the WV6.2 µm channel. Based on these observations, it is anticipated that optically thick raining clouds with cloud tops piercing through the tropopause level will show small negative to positive WV6.2–WV7.3 BTD. Large negative differences may be observed for low-level clouds with cloud tops below the tropopause level.

3.3.2 Data Pre-Processing

This study selected rainy days from the rain gauge datasets to develop the parametric rain detection models during the evaluation period. Rainfall over the study area is highly variable in space and time (Tazalika et al., 2013; Wakachala

et al., 2015); therefore, rainy days were identified separately per gauge station as a day with accumulated rainfall above 1 mm per day.

The MSG data over the study area did not require parallax correction because of the small zenith-viewing angle of the Meteosat satellite (at 41.5° E) over the study area (Kumah et al., 2020). The satellite data were spatially aggregated. Both the satellite and rain gauge data were temporally aggregated to 30 min intervals following the method described in (Kumah et al., 2020) to reduce the effect of spatial and temporal mismatch between the satellite and gauge measurements in this study (Bendix et al., 2010; Ha et al., 2002).

The resulting dataset consisted of collocated and coincident gauge rainfall, satellite reflectance and IR BT, which inferred cloud top properties for each rainy day at a gauge station. A cloud top property was then flagged as raining if the station rainfall was equal to or above 1 mmh^{-1}; otherwise non-raining. The two sets of data, i.e., raining and non-raining cloud top properties, were analysed separately to find the optimal parametric model parameter values discussed in Section 3.3.4. Consequently, our dataset comprised gauge and satellite observations from a mixed space-time domain—implying that the dataset was derived from time-series observations sampled from different locations within the study area. As such, proper data splitting into calibration and validation sets (Moraux et al., 2019) was needed to reduce spatial and temporal bias in the amount of data per gauge station used to train and validate the model and evaluate its performance on both seen and unseen rainfall events. Table 3. 2 summarises the non-zero rain rates (i.e., above 1 mm h^{-1}) from the calibration and validation sets.

Table 3. 2 Summary of non-zero rainfall from calibration and validation sets for the daytime and night-time from the long rain period of 2018–2020.

	Calibration		Validation	
	Day	Night	Day	Night
Mean mmh^{-1}	5.59	4.39	5.26	4.00
Maximum mmh^{-1}	157.01	137.26	81.87	97.89
Standard deviation	8.97	6.32	7.42	5.47
Fraction %	13.47	16.35	13.21	14.93
N	5111	9457	1544	2957
n days	212	209	104	101

N is the total number of 30 min aggregated intervals assembled from an n (313 days) number of daytime and night-time rainy days during the evaluation period. The Fraction (%) represents the percentage of the non-zero rain rates in entire datasets (i.e., including zero and non-zero rain rates).

3.3.3 The Parametric Threshold Based Rainfall Detection Model

The rainfall detection method relies on a threshold applied to an m-dimensional space defined by the spectral characteristics that infer cloud top parameters. Here, we studied different combinations of cloud top parameters inferred from the reflectance, BT and BTD to determine the suitability of the satellite data for rain detection. It is worth mentioning that convective and stratiform rainfall is responsible for most of the area's rainfall (Houze, 1997; Kilavi et al., 2018). These two rainfall types differ in their spectral characteristics (particularly for the IR and BTD) used to infer the cloud properties, and during the day and night conditions (Feidas & Giannakos, 2011; Thies et al., 2008d). Nonetheless, we focused on the first primary objective and investigated different combinations of the BTD parameters (Table 3. 3) for detecting rain areas suggesting that the developed approach does not consider the type of rainfall.

Table 3. 3 The BTD combinations investigated for rain area detection

	BTD Combinations
BTD1	(IR10.8–IR12.0) and (IR8.7–IR10.8)
BTD2	(IR10.8–IR12.0) and (IR8.7–IR10.8) and (WV6.2–WV7.3)
BTD3	(IR10.8–IR12.0) and (IR8.7–IR10.8) and (IR10.8–WV6.2)
BTD4	(IR3.9–IR10.8) and (IR3.9–WV7.3) and (IR10.8–WV6.2)
BTD5	(IR10.8–IR12.0) and (IR3.9–WV7.3) and (IR10.8–WV6.2)
BTD6	(IR10.8–IR12.0) and (WV6.2–WV7.3) and (IR8.7–IR10.8)

For a thorough evaluation, the rain detection model was categorised into 3 groups (Table 3. 4), and the investigation was based on 2 major questions:

(1) What is the rain detection skill of the satellite data if rain detection is based on reflectance-only, IR-only and combined reflectance-IR models?
(2) Which model possesses the best rain detection skill? The model based on a single infrared (IR10.8) channel was included for comparison.

Table 3. 4 The categories of rain area detection and their parametric models for the day and night times.

Period	Reflectance-only		IR-only	Combined reflectance-IR
Daytime	Ref 1	VIS0.6 and NIR1.6	IR10.8 and BTD1	Ref1 and BTD3
	Ref 2	VIS0.6 ÷ NIR1.6	IR10.8 and BTD2	Ref2 and BTD2
	Ref 3	VIS0.6 − NIR1.6	IR10.8 and BTD3	Ref3 and BTD1
			IR 10.8	
Nighttime			BTD4	
			IR10.8 and BTD5	
			IR10.8 and BTD6	
			IR10.8	

Note: Ref is a reflectance model.

For daytime, all 3 categories of models were evaluated, and for night-time, the IR-only models were assessed due to the unavailability of the reflectance data. These models' application assumes that a cloud is more likely to rain if the cloud top parameter is above or below a defined benchmark value. More precisely, in Figure 3. 2, the model application is exemplified by dichotomous statements for raining and non-raining cases of, for instance, a reflectance-only, IR-only or combined reflectance-IR model.

A significant challenge to rainfall detection and estimation by the parametric threshold technique is non-raining thin/thick cirrus clouds because they result in erroneous estimates. The objective cloud-type classification method (Inoue, 1987a) was adopted to overcome this challenge. The basis for implementing this technique for cirrus clouds discrimination lies in the differential emissivity at 10.8 μm and IR 12.0 μm channels for cirrus clouds (Inoue, 1985), leading to larger BTD for these clouds. Therefore, this study used the empirically derived threshold of 2.5 K for IR10.8-IR12.0 BTD (Inoue, 1987a, b) to screen out cirrus clouds before calibrating and validating our parametric rain detection models.

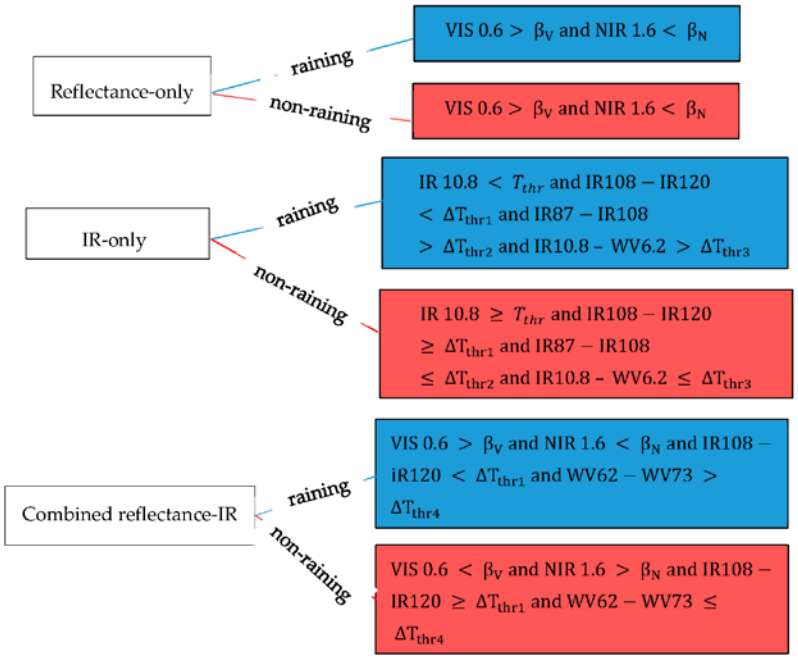

Figure 3. 2 Examples of the dichotomous statement for each category of the rain detection model. The β_V and β_V are the thresholds for VIS0.6 and NIR1.6 reflectance, respectively, T_{thr} is the threshold for IR10.8 BT, ΔT_{thr1}, ΔT_{thr2}, ΔT_{thr3}, and ΔT_{thr4} are the thresholds for IR10.8–IR12.0, IR8.7–IR10.8, IR10.8–WV6.2 and WV6.2–WV7.3 BTD, respectively.

3.3.4 Model's Calibration and Validation

1) Determining the best parametric model and parameters

The calibration of the various rain detection models (Table 3. 4) was achieved by optimally searching for the threshold value of each model parameter in a range, e.g., $x \leq \alpha \leq y$; where α is the parameter value and x, y are the upper and lower limits of the range of values, respectively. Each threshold value was used to conduct a raining/non-raining classification. The results were compared to the station data using categorical statistics to determine (1) the rain detection skill of the satellite based on the kind of satellite data used for rain detection and (2) the best rain detection model.

A 2 × 2 contingency table (Table 3. 5) was used to define the frequencies of the model (satellite-based estimates) and gauge (reference/real observations) based on the raining/non-raining observations to compute some categorical statistics. The h in Table 3. 5 are the raining observations detected in both the model and gauge observations (Hits). The m are raining observations detected by the gauge

and not the model (Misses), and f indicate those seen by the model and not by the gauge (False alarms). Finally, the frequency of non-raining observations detected by both the model and gauge observations was represented by z.

Table 3. 5 Contingency table for evaluating the raining and non-raining decision

2 × 2 Contingency Table		Gauge Observation	
		Raining	Non-raining
Model observation	Raining	h	f
	Non-raining	m	z

In Table 3. *6*, different categorical statistics were computed for each threshold value based on the contingency table elements. Each statistical parameter evaluates an aspect of the model's performance. Collectively, they were used to find the threshold value that provides optimal model performance in rainfall detection. A detailed description of each statistical parameter can be found in, e.g. (Harold et al., 2015; Wilks, 2006).

The different models were tested, and the best rain detection model was determined by optimising the value of the equitability threat score (ETS) together with the probability of detection (POD), false alarm ratio (FAR) and bias. The best model and corresponding threshold value appropriate for rain detection was defined as the model that maximises the ETS and POD values while minimising FAR and bias values.

Table 3. 6 Summary of the categorical statistics.

Statistic	Equation	Range	Optimal Value
POD	$\frac{h}{h+m}$	[0,1]	1
FAR	$\frac{f}{h+f}$	[0,1]	0
POFD	$\frac{f}{z+f}$	[0,1]	0
Bias	$\frac{h+f}{h+m}$	[0,∞]	1
ACC	$\frac{h+z}{\eta}$	[0,1]	1
CSI	$\frac{h}{h+m+f}$	[0,1]	1
ETS	$\frac{(h-h_{erandom})}{h+m+f-h_{erandom}}$	[−1/3,1]	1

where $\eta = h+m+f+z$ and $h_{erandom} = \frac{(h+m)\times(h+f)}{\eta}$

2) Validating the best parametric model and parameters

The best parametric models identified based on the daytime and night-time statistical parameters were validated using independent rainfall data. The validation approach was (1) by a point-to-pixel comparison of model rain detection with the gauge station data, (2) by comparing the model results to rain areas and rate from the GPM IMERG (Huffman et al., 2020) satellite rainfall product and (3) by comparison of both model and IMERG detected rain rates to the gauge rain rates (i.e., the ground truth).

The best rain detection model and its corresponding parameters were used for rainfall detection. The results were compared with the gauge rainfall data using all the statistical parameters (Table 3. *6*) to point validate the rain detection model. Note that the rain area correction scheme was not implemented in this validation.

For the comparison with GPM IMERG, the corrected model detected rain areas (see the method in the next section), and rain rates were compared to the results from the latest IMERG Final Run version 6 (V06b) (Huffman et al., 2019). The focus was on the precipitationCal dataset because it is a research-grade product that is climatologically adjusted using ground data from the GPCC. Although recent studies see, e.g. (Anjum et al., 2019; Maranan et al., 2020; Moazami & Najafi, 2021) have ground validated and reported this new IMERG rainfall product's performance elsewhere, its performance over areas with sparsely distributed gauge data like the study area is generally not yet reported (Jackson et al., 2020). Therefore, this comparison intends to spatially validate the newly developed parametric model's rain area and rate detection skill against the new IMERG V06b precipitationCal rainfall product.

The rain areas detected by the developed parametric model and IMERG were compared for the entire validation period. From IMERG, rain areas were identified by flagging IMERG pixels with a rain rate equal to or greater than 1 mmh^{-1} as raining, otherwise dry. IMERG has a 30 min, 10 km temporal and spatial resolution, respectively. The IMERG data were spatially resampled to 3 × 3 km using the nearest neighbourhood technique to compare IMERG with the 3 km parametric model results from the MSG data. The nearest neighbourhood resampling was used to preserve the original pixel values as much as possible.

The rain rates detected by the developed parametric model and IMERG were compared at 30 min for the entire validation period. For the developed model, the detected rain rates were all the rain rates recorded at a gauge station when

the model flagged the station pixel (i.e., the satellite pixel containing the gauge station) as raining. For IMERG, this corresponded to the rainfall retrieved from the IMERG pixel having the gauge station. The absolute difference between the means of the two detected rain rates (i.e., the parametric model and IMERG) was compared per each gauge station and separately for the day and night time to assess the model's daytime and night-time rain rate detection performances.

Finally, the probability density of the rain rates detected by the developed parametric model and IMERG, at 30 min for the entire validation period, were compared to the rain rate recorded at all the gauge stations. For this, the detected rain rates were organised into bins, and the number of rain rates in each bin was counted. The density was then computed as the count divided by the total count and the bin width. The cross-comparison purpose was to evaluate rain rate detection performance over the study area using the developed parametric model and IMERG against the ground truth and provide valuable insights that inform the ground validation wish list (Jackson et al., 2020).

3.4 Results and Discussion

3.4.1 Results

3.4.1.1 Model Calibration

i. Preliminary Analysis of the Spectral Characteristics of Cloud Top Properties

This section analyses the spectral characteristics of raining and non-raining clouds, which guided the model parameter ranges described in Section 3.3.4. The analysis is presented separately for day and night times, and descriptive statistics of the data are in APPENDIX A3.

Figure 3. 3 is a bivariate probability density distribution of the raining and non-raining spectral characteristics compared in a 2D space for the daytime observations. Each figure's contours represent the raining (blue contours) and non-raining (red contours) densities in the dichotomous dataset. A general observation from Figure 3. 3 is the clear distinction in the peak of the distribution (i.e., the area in the plot where most of the data is concentrated, indicated by high densities) for the raining and non-raining densities. This characteristic behaviour is also supported by the significant difference in descriptive statistics (APPENDIX A3, Table A3. 1 and Table A3. 2, respectively) computed from the raining and non-raining data. Although this observation is particularly noticeable in the reflectance than in the BT and BTD plots, it raises the possibility of rain and no-rain discrimination by using respective thresholds in the 2-D space.

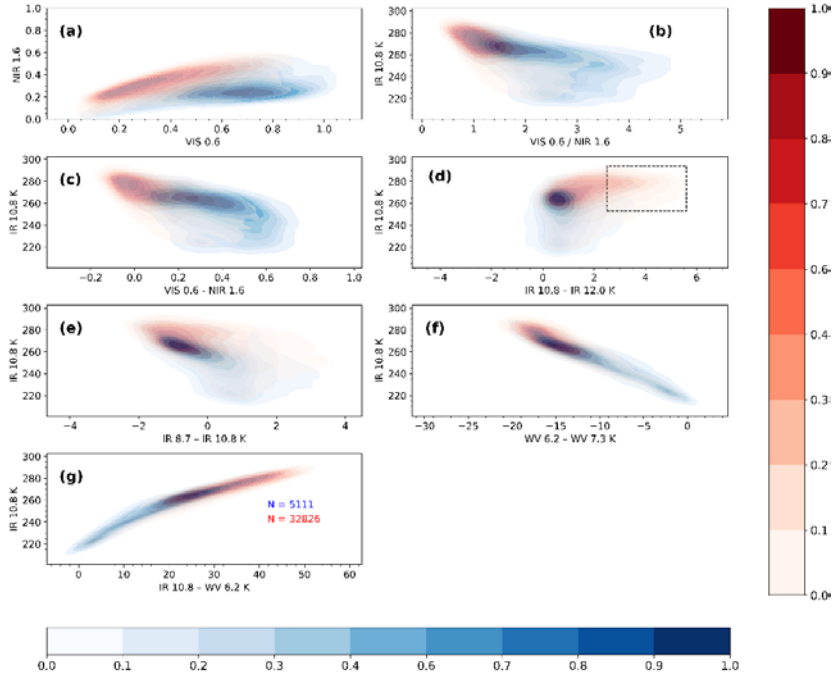

Figure 3. 3 Bivariate probability density distribution of daytime raining (blue) and non-raining (red) spectral characteristics in (**a**) VIS0.6 and NIR1.6, (**b**) IR10.8 and VIS0.6 ÷ NIR1.6, (**c**) IR10.8 and VIS0.6–NIR1.6, (**d**) IR10.8 and IR10.8–IR12.0, (**e**) IR10.8 and IR8.7–IR10.8, (**f**) IR10.8 and WV6.2–WV7.3, (**g**) IR10.8 and WV6.2–IR10.8 space. The coloured figures in (**g**) are the raining and non-raining data counts that computed the density distribution. The colour bar shows normalised densities to make the subfigures comparable.

In Figure 3. 3a, one can notice that the raining cases of the spectral characteristics peak towards the lower right corner of the plot, where large VIS0.6 (>0.6) reflectance coincides with low NIR1.6 (<0.4). Figure 3. 3b,c also show higher (above 1.5 and 0.2 respectively) ratios and differences of the VIS0.6 and NIR1.6 reflectance together with colder IR10.8 BT (less than 265 K) for the raining cases. Also, their distributions tend to be bimodal, with peaks above and below 250 K. The high VIS0.6 and low NIR1.6 reflectance and the corresponding high ratio and differences for the raining cases suggest that most of the raining cases defined by the rain gauges were from optically thick clouds with large CWP, high rainfall probabilities and intensities (Bendix et al., 2010; Kumah et al., 2020). The first peak (IR10.8 > 250 K) in Figure 3. 3b,c of the distribution for the raining cases suggests low-level optically thick clouds, whereas the second peak (IR10.8 < 250 K) is indicative of high-level optically thick cumulonimbus type clouds.

By contrast, the distribution of the non-raining cases shown in Figure 3. 3a–c all peak outside the area defined by the raining cases' distribution. In Figure 3. 3a, this corresponds to the lower left corner of the plot where low VIS0.6 (<0.6) reflectance coincide with low NIR1.6 (<0.4) reflectance. For Figure 3. 3b,c, these correspond to areas in the plot where the reflectance ratio and differences are lower than their indicated respective thresholds and with IR10.8 BT that are mostly warmer than 265 K. From the figures, one can also notice that some non-raining cases overlap the areas defined by the raining cases. Based on the IR10.8 BT in Figure 3. 3b,c, the areas defined by the non-raining cases mostly correspond to low and high-level non-precipitating thin and thick cirrus clouds and N-Type clouds (representing edges of optically thick clouds, optically thinner cumulus clouds, or low-level cumulus cloud overlaid by thin cirrus clouds) according to the cloud type classification by (Inoue, 1987a, 1989).

Figure 3. *3*d compares the raining and non-raining case distribution for the IR10.8 BT and the split window BTD. The dotted square marks the IR10.8 BT and IR10.8–IR12.0 BTD threshold ($T_{thr} > 253$ K and $\Delta T_{thr1} > 2.5$ K) used to eliminate non-precipitating cirrus clouds (Section 3.3.3) according to the cloud classification technique by (Inoue, 1987a); Inoue (1987b). As can be seen from the plot, most of the raining cases are characterised by IR10.8–IR12.0 BTD < 1.5 K and IR10.8 BT colder than 265 K. In addition, the distribution here is bimodal, having two peaks (above and below 250 K) with the IR10.8–IR12.0 BTD < 1.5 K. The first peak (IR10.8 > 250 K) comprises raining cases mostly from low-level cumulus clouds, and the second peak (IR10.8 < 250 K) contains raining cases from mainly cumulonimbus types.

However, most non-raining cases are distributed above 1.5 K (IR10.8–IR12.0 BTD) and 265 K (IR10.8 BT). Note from the plot that some of the areas defined by the non-raining cases, especially above 265 K IR10.8 BT, overlap with the raining cases. However, a large concentration of this overlap occurs in areas where the IR10.8–IR12.0 BTD > 1.5 K. Based on the cloud type classification by Inoue (1989), most of these non-raining cases are from non-precipitating thin and thick cirrus clouds and N-Type clouds.

Figure 3. 3e compares the raining and non-raining distributions of IR10.8 BT and IR8.7–IR10.8 BTD. Recall that the IR8.7–IR10.8 BTD has been used for separating water from ice clouds (Strabala et al., 1994). Larger IR8.7–IR10.8 BTD will mainly occur due to ice particles at the cloud top. It can be observed that most of the raining cases have IR8.7–IR10.8 K BTD above -2 K and are colder than 265 K (IR10.8 BT). The distribution is again bimodal, first peaking above 250 K (IR10.8 BT) and IR8.7–IR10.8 BTD between −2 and 0 K and secondly, below 250 K and above 0 K for IR10.8 BT and IR8.7–IR10.8 BTD, respectively. The first

peak consists of raining cases from mainly clouds with water droplets, while the second peak suggests clouds with ice tops–based on the threshold function by Uttal et al. (2005). The non-raining cases are warmer than 265 K IR10.8 BT and below 0 K for the IR8.7–IR10.8 K BTD. They are mostly cirrus (thin and thick) and N-Type clouds. It is also apparent from Figure. 3e that some non-raining cases overlap with the areas defined by the raining cases.

The raining and non-raining distributions of IR10.8 BT and WV6.2–WV7.3 BTD are compared in Figure 3. 3f. Most of the raining cases correspond to IR10.8 BT < 260 K and WV6.2–WV7.3 BTD > –18 K. The distribution also shows a bimodal tendency, having two peaks above and below 250 K IR10.8 BT. The first peak is between –18 and –10 K WV6.2–WV7.3 BTD and consists of raining cases from optically thick clouds below the tropopause level. The second peak (BTD mostly above –5 K) comprises raining cases from optically thick cumulonimbus-type clouds with cloud tops above the tropopause level. The non-raining cases, on the other hand, consist of N-Type clouds with WV6.2–WV7.3 BTD less than –15 K and IR10.8 BT warmer than 260 K.

Figure 3. 3g compares the raining and non-raining distributions for IR10.8 BT and IR10.8–WV6.2 K BTD. The raining cases consist of IR10.8 BT colder than 265 K and IR10.8–WV6.2 BTD, mostly less than 30 K. From the figure, the empirical threshold (<11 K) determined by Kidder et al. (2005) is consistent with raining cases from clouds with cold tops (<250 K). However, raining cases from clouds with relatively warm top temperatures (<265 K) show BTD mostly above the empirical threshold. On the other hand, the non-raining cases mainly consist of clouds with warmer top temperatures (IR10.8 BT > 265 K) and IR10.8–WV6.2 BTD > 30 K.

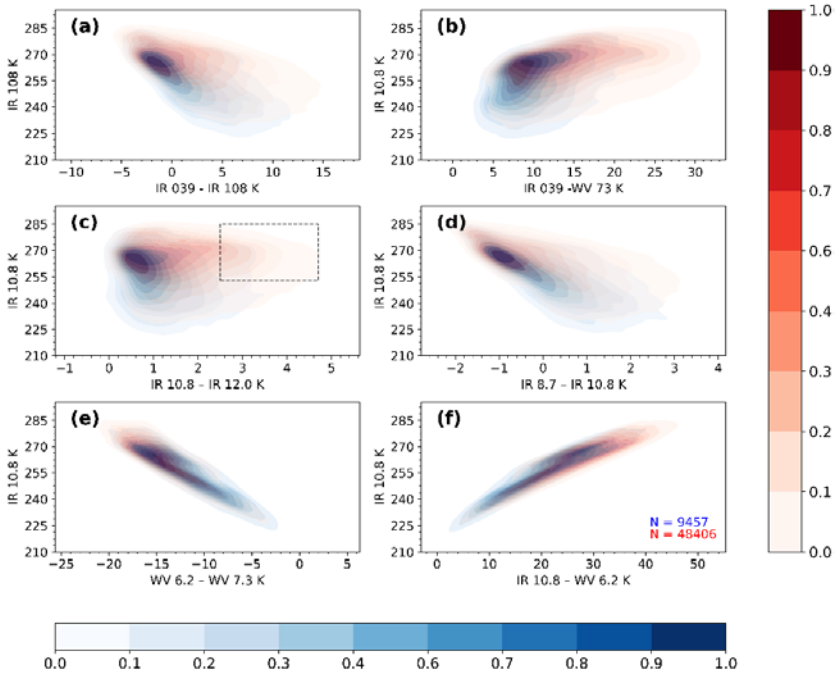

Figure 3. 4 Bivariate probability density distribution of night-time raining (blue) and non-raining (red) spectral characteristics in (**a**) IR10.8 and IR3.9–IR10.8, (**b**) IR10.8 and IR3.9–WV7.3, (**c**) IR10.8 and IR10.8–IR12.0, (**d**) IR10.8 and IR8.7–IR10.8, (**e**) IR10.8 and WV6.2–WV7.3, (**f**) IR10.8 and WV6.2–IR10.8 space. The coloured figures in (**f**) are the raining and non-raining data counts that computed the density distribution. The colour bar shows normalised densities to make the subfigures comparable.

Figure 3. 4 is analogous to Figure 3. 3 but for night-time. Unlike Figure 3. 3, one can observe a significant overlap in the raining and non-raining distributions. Their descriptive statistics (APPENDIX A3, Table A3. 3 and Table A3. 4) also paint a similar picture based on the comparatively similar statistical parameter values. Additionally, the figure shows bimodally distributed densities of IR10.8 BT above and below 255 K, slightly warmer than the temperatures observed in the daytime data.

Figure 3. 4a compares the raining and non-raining distributions for IR10.8 BT and IR3.9–IR10.8 BTD. The raining cases mainly consist of IR3.9–IR10.8 BTD between 0 and 5 K and IR10.8 BT colder than 270 K. These spectral characteristics are consistent with medium BTD found in Kumah et al. (2020); (Thies et al., 2008a) and are indicative of large CWP with high rain probabilities and intensities. The non-raining cases mostly show large positive differences in clouds with medium CWP and low rain probability and intensity.

The results comparing the raining and non-raining IR10.8 BT and IR3.9–WV7.3 BTD (Figure 3. 4b) show comparable characteristics to Figure 3. 4a but with generally higher IR3.9–WV7.3 BTD. The higher differences are due to the diminishing effect of the water vapour absorption and emission in the mid-to-low tropospheric levels on the brightness temperature in the WV7.3 μm channel.

Figure 3. 4c compares the raining and non-raining distributions for the IR10.8 BT and the split window BTD. Like in Figure 3. 3d, the dotted square indicates the IR10.8 BT and IR10.8–IR12.0 BTD threshold used to eliminate non-precipitating cirrus clouds (Section 3.3.3). The IR10.8–IR12.0 BTD defined by the raining cases is mostly between 0 and 2 K and colder than 270 K IR10.8 BT. As noted earlier, these characteristics indicate optically thick cumulonimbus clouds. On the other hand, the non-raining cases are from non-precipitating cirrus clouds and N-Type clouds. Unlike the daytime observations (Figure 3. 3d), the IR10.8 BT and the IR10.8–IR12.0 BTD are distributed over a wide range.

For the comparison of IR10.8 BT and IR8.7–IR10.8 BTD (Figure 3. 4d), the observations made for the raining cases were comparable to those found during the daytime (Figure 3. 3e). However, the non-raining cases showed a relatively wide range of IR10.8 BT values that mainly were warmer (>270 K). Figure 3. 4e,f compare the raining and non-raining distributions for IR10.8 BT and WV6.2–WV7.3 BTD and IR10.8 BT and IR10.8–WV6.2 BTD, respectively. Here, the observations made for both the raining and non-raining cases were comparable to their daytime observations (Figure 3. 3f,g respectively), except their IR10.8 BT shows a wide range of values.

ii. Determination of the Optimum Parametric Thresholds

Figure 3. 5 presents the graphical representation to determine the parametric thresholds with the best rain detection performance to answer the 2 essential questions: the satellite data's rain detection skill based on the kind of data and the best rain detection model. The categorical statistics of POD, POFD, FAR and bias were computed as a function of parameter thresholds to assess model performance and answer these two questions for the day and night-time analysis. The POD versus the POFD was analysed to infer the models' relative operational characteristic (ROC) curve—a measure of a forecasting system's relative skill and usefulness (Kharin & Zwiers, 2003). The distance of the ROC curve from the diagonal line (where POD = POFD) corresponds to a climatological skill and is often used to evaluate the quality of forecasts (Feidas & Giannakos, 2010).

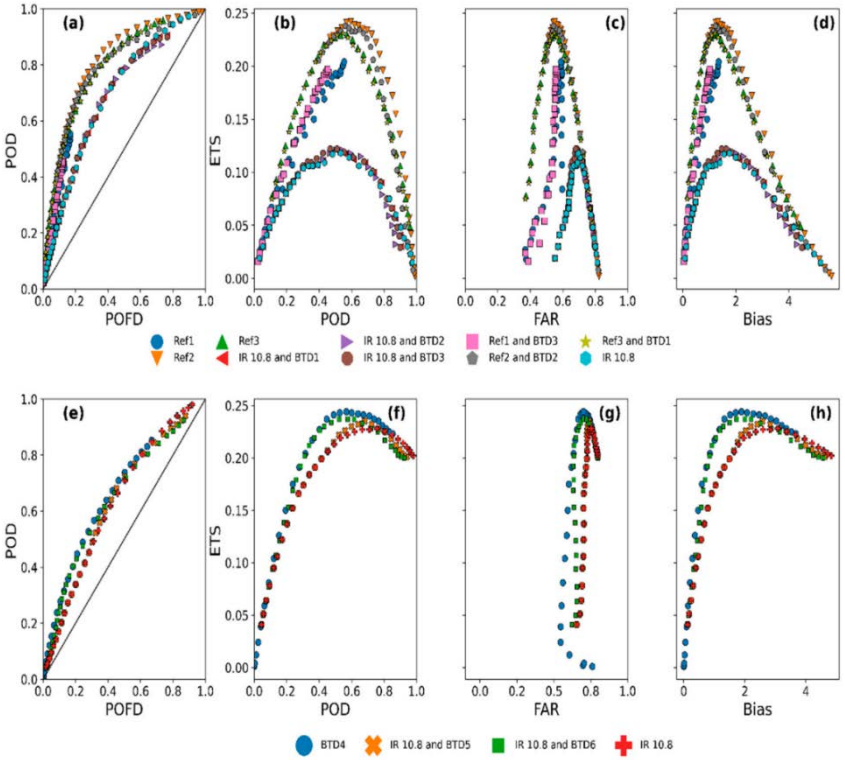

Figure 3. 5 Daytime (a–d) and night (e–h) model calibration results. (a,e) ROC curve for different model parameter thresholds. Statistical scores of ETS compared to (b,f) POD (c,g) FAR and (d,h) Bias.

From the daytime results (Figure 3. 5a–d), the ROC curves (Figure 3. 5a) for the different parametric thresholds of the rain detection models show that threshold values corresponding to models derived from the reflectance data have the largest distance from the diagonal line. Notably, the models derived from the VIS 0.6 and NIR 1.6 reflectance (Ref2 and Ref3) showed the best performance. This suggests that daytime rain detection based on reflectance measurements alone may be enough to achieve maximum detection results. The rain detection models developed from reflectance and IR showed medium performances, whereas those from the IR data alone were often poor.

The observations above are further supported by the relationship between ETS and the POD, FAR and bias values (Figure 3. 5b–d respectively). Figure 3. 5b shows that the Ref2 and Ref3 models have comparatively higher ETS and POD values. Nonetheless, all models have reasonably high FAR and Bias (Figure 3. 5c,d, respectively). It can also be seen from the figure that the rain detection

models developed from combined reflectance and IR data tend to reduce model FAR and bias values but at a cost to the POD values.

To answer the question "which is the best parametric rain detection model?" the values of all 4 categorical scores were considered. Between the Ref2 and Ref3 models, the Ref3 model was considered the best model because of its high ETS and POD and correspondingly low FAR and Bias values.

For the night-time results (Figure 3. 5e–h), the ROC curves (Figure 3. 5e) show that threshold values corresponding to BTD4 have the largest distance from the diagonal line, suggesting the best model performance. On the other hand, rain detection models developed from IR (IR10.8) BT and BTD usually resulted in medium model performance. In contrast, those based on a single IR BT often showed poor performance.

Moreover, when the ETS is compared to the POD, FAR, and Bias, the BTD4 rain detection model's superior performance is further strengthened. Its ETS and POD values are higher than those of combined IR and BTD and single IR models, with the single IR model being the lowest. Again, all models show higher FAR and Bias, although the Bias is comparatively lower than those found during daytime. Therefore, the BTD4 rain detection is considered the best model for the night-time case because of its high ETS and POD and comparatively low FAR and Bias values.

Table 3. 7 Best rain detection model and the parameter values during the day and night-time.

Application	Rain Detection Model	Parameter	Parameter Value
Daytime	Ref3	VIS0.6–NIR1.6	0.21
Nighttime	BTD4	IR3.9–IR10.8	8.18 K
		IR3.9–WV7.3	17.03 K
		IR10.8–WV6.2	33.65 K

The parameters and values for the best model are shown in Table 3. 7, and their categorical statistics from the model calibration are presented in Figure 3. 6. The daytime parameter value is comparable to the one inferred from the raining spectral characteristics (i.e., the satellite signals sampled when rainfall in the gauge was equal to or above 1 mm h^{-1}) in the bivariate distribution (Figure 3. 3c). In contrast, the night-time parameters' values are comparable to the sum of their 75th percentile and standard deviation values (APPENDIX A3 Table A3. 3) of the raining spectral characteristics' descriptive statistics.

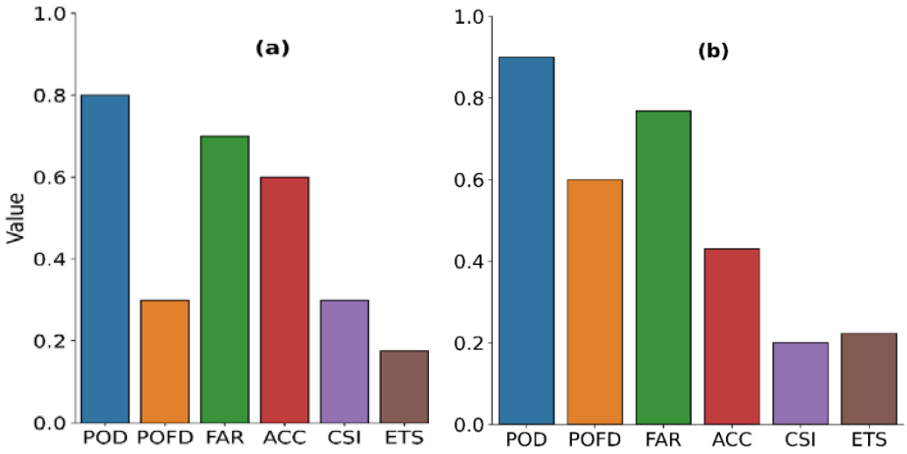

Figure 3. 6 Categorical statistics from daytime (a) and night-time (b) calibration of the best parametric rain detection model with gauge rainfall data.

Figure 3. 6 shows a clear difference in rain detection performance between the daytime (Figure 3. 6a) and night-time (Figure 3. 6b). Although their POD, FAR, and ETS are comparable, the comparatively high ACC, CSI, and low POFD and bias (2.28 compared to 3.67 for the night-time case) scores suggest that the daytime detection was better than the night-time.

3.4.1.2 Parametric Model Validation

This section validates the developed rain area detection and correction technique using the independent validation datasets for the daytime and night-time and the entire evaluation period. First, the point-to-pixel validation of the model results using the gauge station data is presented. Next, the developed rain area correction technique is described and demonstrated for two selected daytime (on 14 April 2018, 11:00 UTC) and night-time (on 6th March 2018, 17:00 UTC) periods from the validation dataset due to the variety of detected rain areas present in the scene. Finally, the model detected rain areas and rain rate are compared with the GPM IMERG satellite rainfall product results to validate the model spatially.

i. Point validation of the parametric model's rain area

Figure 3. *7* presents the categorical scores of rain area detection, indicating the best performance of the models when the model results were compared with the daytime (Figure 3. *7*a) and night-time (Figure 3. *7*b) gauge station data. The results show improved daytime detection, indicated by high ACC, low FAR and a marginal increase in bias (2.3) scores compared to the calibration scores.

However, the low ETS and high bias (4.11) scores suggest a decreased model performance for the nighttime compared to its previous calibration scores.

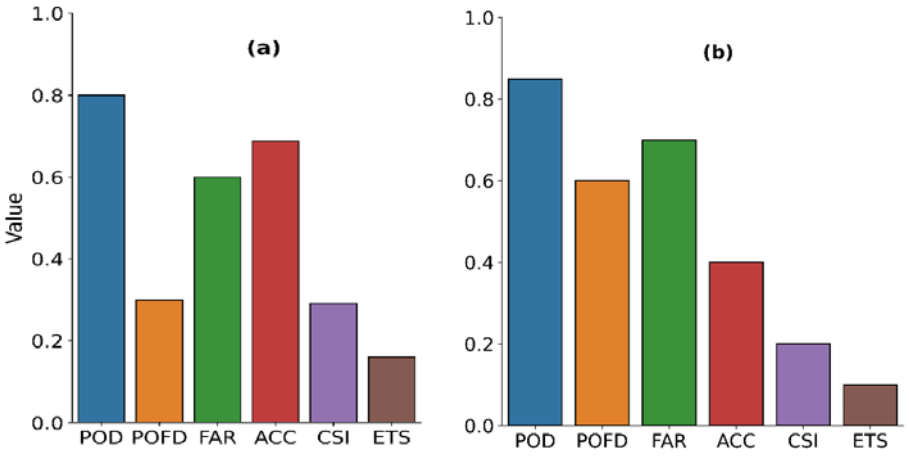

Figure 3. 7 Categorical statistics from daytime (a) and night-time (b) validation of the best parametric rain detection model with gauge rainfall data

ii. The rain area correction scheme

Corrections were applied to the detected rain areas because the indirect relationship between rain and the inferred cloud top properties from the satellite often results in retrieval uncertainties (Karaseva et al., 2011). Moreover, both the point validation of the initial rain area detection results (shown above) and its preliminary comparison with the results from EUMETSAT's MPE (Dhib et al., 2017; Heinemann et al., 2002) (not used in this study) and GPM IMERG (used in this study) showed high FAR and comparatively extensive raining areas, respectively, which required corrections. Below is a summary and results that detail the implementation of the correction technique.

This study implemented rain area correction for only the validation data because of the large number of datasets. The correction scheme relies on adaptive parametric thresholds applied to spectral characteristics from the detected rain area. This implies that the applied corrections were based on scene and rain area-specific parameters for each classification scene and detected rain area. They were more precisely derived from the gradient in the spectral characteristics computed for each rain area, i.e., the identified raining cloud object (cloud object gradient) and each pixel (pixel gradient) in a cloud object. These two kinds of gradients differ for daytime and night-time due to the different data and

information content used for rain area detection. Nonetheless, their implementations are for the same purpose during the day and night.

Cloud objects gradient aims to reduce the number of detected rain areas. It combines the gradient computed for each cloud object with the average gradient and standard deviation from all cloud objects to locate non-raining areas previously classified as raining in the initial results. This reduces the number of detected rain areas (i.e. the number of different sized raining areas) identified by the parametric model. The pixel gradient results in a reduction of the size of the cloud object because it compares the gradient computed for each pixel in a cloud object to its median and average median (from all cloud objects) to locate the non-raining high/low gradient (depending on the day/night application) pixels in the initial results.

Figure 3. 8 demonstrates the implementation of the daytime rain area correction scheme. The best daytime model was the VIS0.6 − NIR1.6 parametric model (Ref3). As was shown in Figure 3. 3c, the raining spectral characteristics were mainly above 0.2, suggesting that higher differences correspond to high rain probabilities. Figure 3. 8a is a Daytime Natural Colour RGB composite (http://www.eumetrain.org/rgb_quick_guides/, accessed on 24th March 2021) from SEVIRI NIR1.6, VIS0.8 and VIS0.6 µm channels, respectively, over the study area. High-reaching clouds with ice tops, e.g., cumulonimbus-type clouds, appear cyan in the figure. Black lines demarcate the raining areas initially detected by the developed daytime parametric model. The cloud object gradient (Figure 3. 8b) represents an area-specific gradient computed (for each detected raining area in Figure 3. 8a) as the maximum reflectance difference (max (VIS0.6–NIR1.6)) of each cloud object minus its minimum (min (VIS0.6–NIR1.6)). The cloud object gradient was combined with the average gradient (Δ) and standard deviation (σ), both indicated in Figure 3. 8b, to identify non-raining areas in the initial results and thus reduce the number of detected ted cloud objects (Figure 3. 8c).

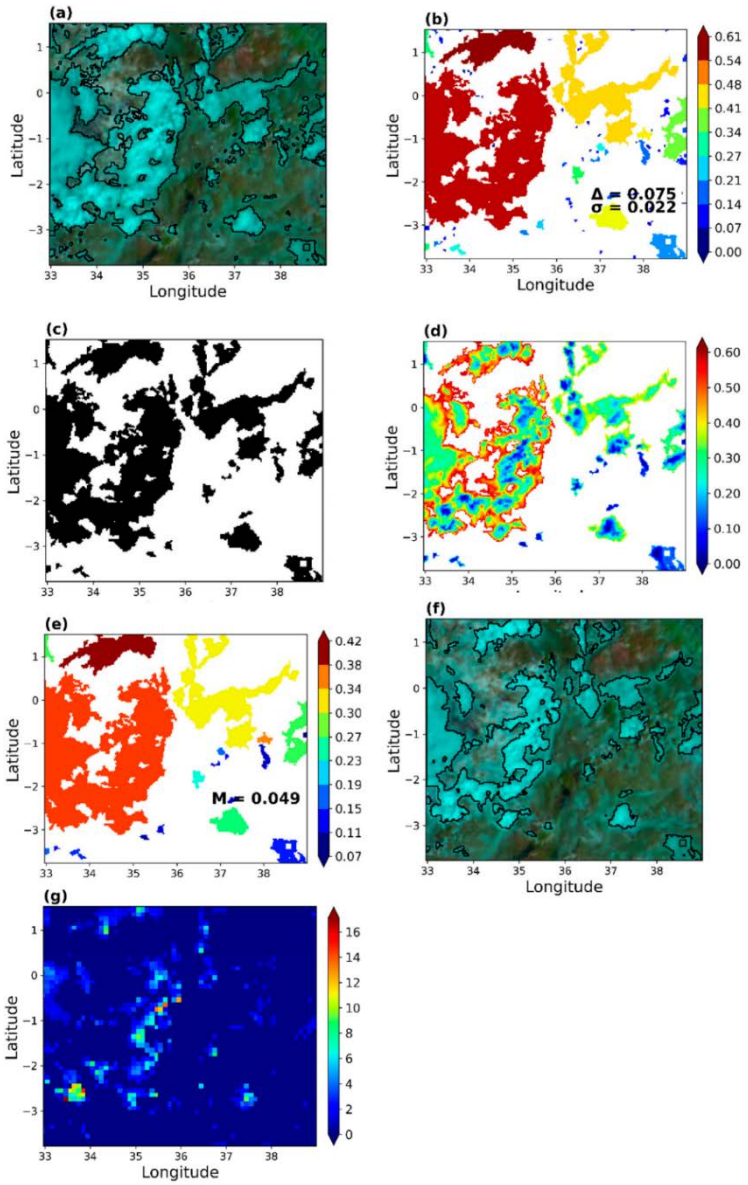

Figure 3. 8 Daytime rain area correction. (a) Initial rain area detections, (b) Cloud object gradient, (c) Rain area correction based on cloud object gradient, (d) Pixel gradient, (e) Median pixel gradient, (f) Rain area correction based on pixel gradient, (g) GPM IMERG rainfall estimate (mm/30 min) over the study area.

The pixel gradient (Figure 3. 8d) was computed for each cloud object in Figure 3. 8c as max (VIS0.6-NIR1.6) minus the pixel value. Thus, the low gradients in Figure 3. 8d correspond with high VIS0.6–NIR1.6 differences, whereas the high

gradients are the low differences. Combined with the median pixel gradient (Figure 3. 8e) for each cloud object and the average of the median (M) pixel gradient from all cloud objects, these gradients identify and reclassify high gradient pixels as non-raining. The result is a reduction in the detected rain area's size, as shown in the RGB colour composite in Figure 3. 8f. Figure 3. 8g is the GPM IMERG rainfall estimate over the study area. Comparing the initial results in Figure 3. 8a–f (the corrected version) and the rainfall estimates in Figure 3. 8g, it is evident that the correction scheme can improve the initial results to estimates comparable with the GPM IMERG satellite estimates.

Figure 3. 9 demonstrates rain area correction for the night-time. Unlike the daytime, night-time rain detection was based on a combination of IR3.9–IR10.8, IR3.9–WV7.3 and IR10.8–WV6.2 BTD (BTD4). Here, the rain area-specific parameters used in correcting the detected rain areas were from IR10.8–WV6.2 K to reduce redundancy in the data used for the correction. Figure 3. 3g and Figure 3. 4f indicate that most of the raining spectral characteristics of the IR10.8–WV6.2 are less than 30 K. In Table A3. 1 and Table A3. 3, on average, 75% are <26 K, suggesting that low differences indicate high rain probabilities.

Figure 3. 9a is a Nighttime Microphysics RGB colour composite of SEVIRI IR12.0 –IR10.8 µm and IR10.8–IR3.9 µm channel differences, and IR10.8 µm channel, respectively, over the study area. Detailed colour interpretations are in (http://www.eumetrain.org/rgb_quick_guides/, accessed 24th March 2021); of interest are the reddish-brown areas that indicate optically thick ice clouds. The initially detected rain areas by the night-time parametric model are demarcated in black. The cloud object gradient (Figure 3. 9b) and Δ were computed similarly to the daytime. The cloud object standard deviation in Figure 3. 9c and the σ (shown in Figure 3. 9c) represent an area-specific standard deviation and σ of the IR10.8–WV6.2 BTD for the detected raining area. Combined with Figure 3. 9b, they were used to reduce the number of detected cloud objects, as shown in Figure 3. 9d, similar to the daytime approach. The pixel gradient in the IR10.8–WV6.2 BTD for the areas in Figure 3. 9d is also shown in Figure 3. 9e. It was computed similar to the daytime approach. Thus, the high gradient area corresponds to the low IR10.8–WV6.2 BTD and vice versa. The median pixel gradient and the M shown in Figure 3. 9f were derived from the raining areas in Figure 3. 9e. They were used to reduce the sizes of the detected rain areas, as shown in Figure 3. 9g. Figure 3. 9h is the rainfall estimate from the GPM IMERG; its comparison with the initial and final (the corrected version) rain areas result (Figure 3. 9a,g, respectively) provided a better perspective of the effect of the developed correction scheme similar to the daytime.

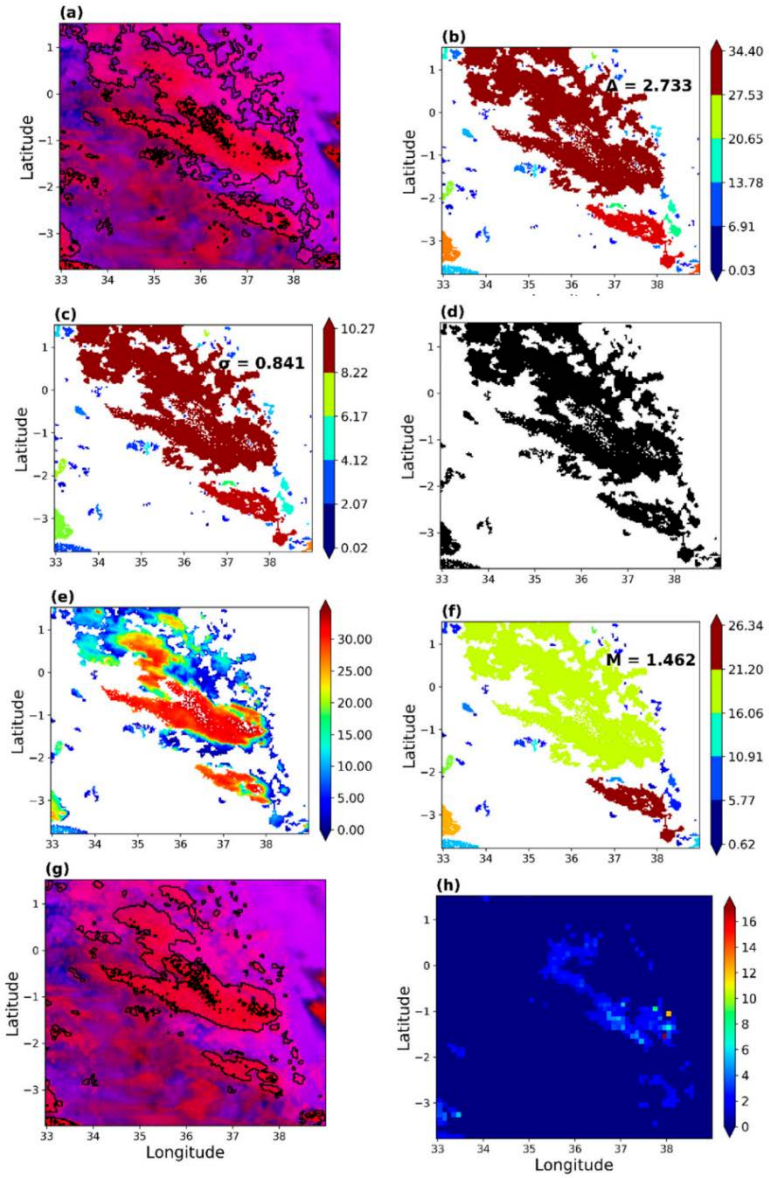

Figure 3. 9 Nighttime rain area correction. (a) Initial rain area detections, (b) Cloud object gradient, (c) Cloud object standard deviation (d) Rain area correction based on cloud object gradient, (e) Pixel gradient, (f) Median pixel gradient, (g) Rain area correction based on pixel gradient, (h) GPM IMERG rainfall estimate (mm/30 min) over the study area.

iii. Spatial validation of the parametric model's rain area and rate

Figure 3. 10 is an "eyeball" verification of the corrected rain areas (Figure 3. 8f and Figure 3. 9g) compared with the initial rain areas detections (Figure 3. 8a and Figure 3. 9a, lime green demarcations) and detections by the GPM IMERG (white demarcations). This daytime and night-time model results in an RGB colour composite compared to the IMERG results; validate the developed parametric model spatially. The daytime comparison in Figure 3. 10a shows a convincing agreement in the detected rain areas' spatial dynamics by the model and IMERG. For instance, areas between latitudes 0 to -2 and longitude 34 to 36 show a good spatial match in rain areas. These observations correlate with the high POD, ACC and CSI scores observed for the daytime in the previous validation, thus indicating high confidence in the model results.

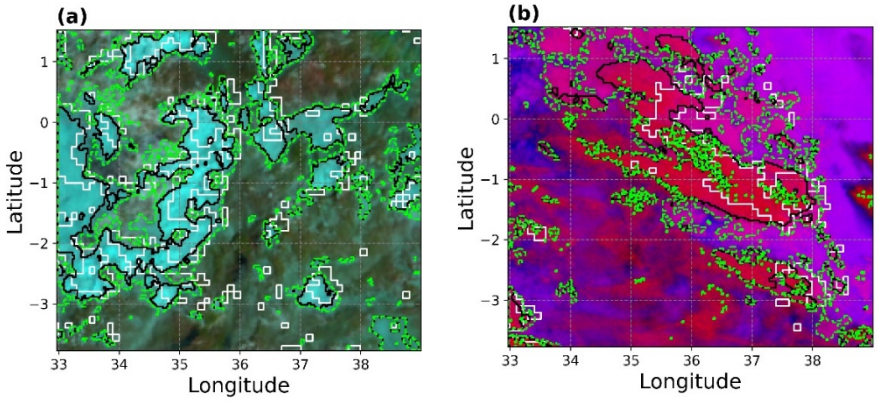

Figure 3. 10 Spatial verification of the corrected rain area detections (black extent) compared to the initial detections (lime green) and GPM IMERG (white extent) for the day (a, on 14 April 2018, 11:00 UTC) and night-time (b, on 6th March 2018, 17:00 UTC). The base maps are composite for Daytime Natural Colour (a) and Nighttime Microphysics (b).

Nevertheless, there are some differences in Figure 3. 10a. IMERG detects more rainy areas (of varying sizes) than the developed model. Instead, the raining rain areas detected by the model are mainly organised into large areas and fewer in number. Additionally, a close inspection of Figure 3. 10a reveals a slight shift in the detected rain areas by IMERG relative to the model.

Figure 3. 10b is an analogous comparison of Figure 3. 10a but for the night-time. Compared to IMERG, the corrected rain areas' spatial dynamics show good agreement, especially for the large rain areas. However, the figure also indicates that night-time detection detects more rainy areas than IMERG, which may explain the high POFD, low ACC, and high bias scores observed in the previous

validation. Moreover, the spatial shift in the detected rain areas of IMERG relative to the parametric model observed for the daytime is again noticeable.

The differences in rain areas detected by IMERG and the parametric model were mainly attributed to factors such as differences in the native resolution of the model dataset and IMERG. The spatial and temporal aggregation methods used to resolve their resolution differences could further contribute to these differences.

Table 3. 8 presents descriptive statistics of the detected rain areas' spatial properties; herein, the number and size (i.e., the area in km^2) for the uncorrected parametric model, corrected parametric model, and IMERG. The number of detected rain areas expresses the average count of all detected rain clouds, the 50% percentile, 75% percentile and standard deviation per 30 min validation time step. On the other hand, the area in Table 3. *8* expresses the average size, 50% percentile, 75% percentile and standard deviation of the clouds detected per 30 min validation timestep.

Table 3. 8 Descriptive statistics of spatial properties of the rain areas detected per 30 min time step.

Descriptive Statistics	Number Of Contiguous Cloudy Areas			Area (km2)		
	Uncorrected model	Corrected model	GPM IMERG	Uncorrected model	Corrected model	GPM IMERG
Average	186	85	19	2522	814	2275
50%	120	63	16	1732	621	1545
75%	195	103	26	3230	988	2863
Stand. dev.	191	86	15	2453	795	2547

Stand. dev. is the standard deviation, 50% and 75% are the percentile values.

Based on Table 3. *8* results, IMERG, on average, detects a comparatively smaller number of rain clouds of large sizes than the parametric model, which may be because of IMERG's larger native spatial resolution than the model.

Figure 3. 11 compares the average sizes of the parametric model's detected rain areas, before and after correction, to rain areas from IMERG for the entire validation period. The figure shows the average detected rain area for sizes ranging from 81 to 20,000 km^2 because it constitutes most of the detected rain area sizes and allows for a clear visual comparison.

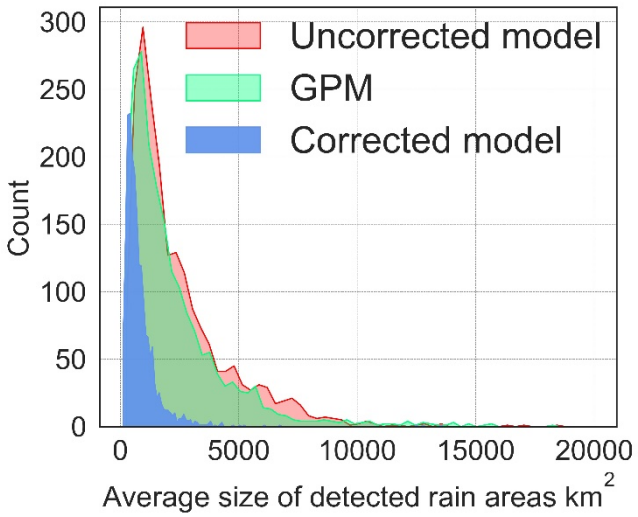

Figure 3. 11 Comparing the average sizes of detected rain areas by the parametric model and IMERG.

The initial model results generally detected extensive rain areas of varying sizes. Nonetheless, the implemented correction technique effectively reduced the number and sizes of the detected rain areas. From visual inspection of uncorrected and corrected detected rain areas, the daytime rain area correction mainly occurred at the fringes of initially detected large contiguous raining areas, whereas the small areas were mainly reclassified non-raining. On the contrary, the night-time implementation showed that large areas initially flagged as raining were reclassified as non-raining.

Figure 3. 12 compares the gauge stations' detected rain rates to those detected by the parametric model and IMERG for the entire validation period, using absolute differences and probability densities. Figure 3. 12a,b are the absolute differences between the mean rain rates detected at each gauge station by the parametric model and IMERG compared separately for the daytime and night-time, respectively. The probability densities of the rain rates from the gauge, IMERG and parametric model were also compared for rain rates below and above 20 mmh^{-1} (Figure 3. 12c,d) for the entire validation period to evaluate the models' detected rain rates against the actual observations) and IMERG.

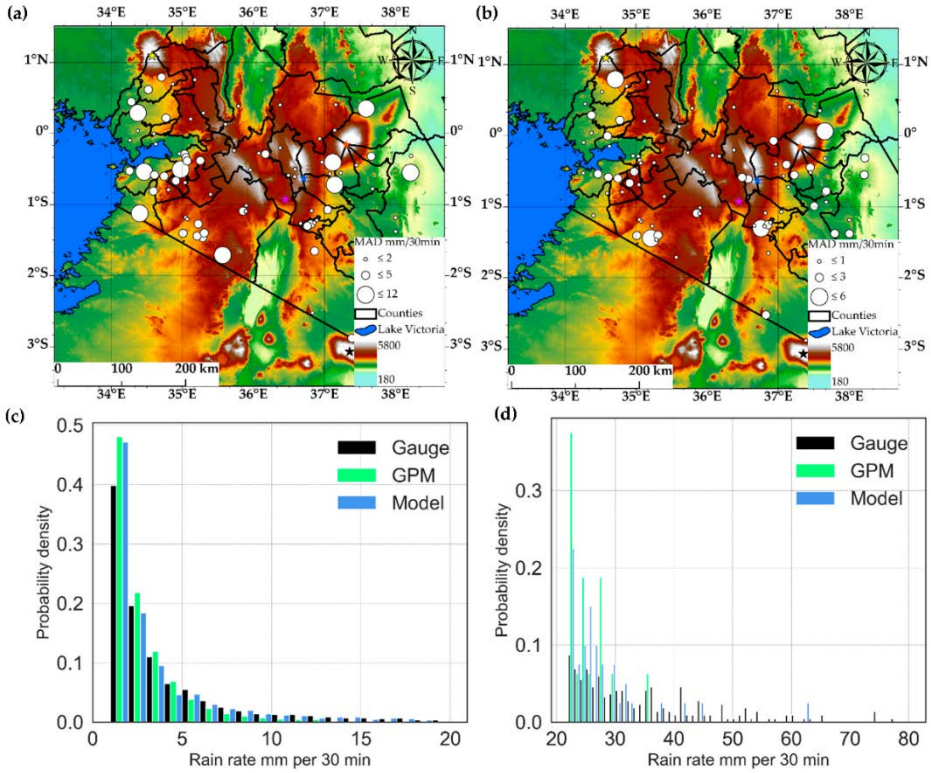

Figure 3. 12 Comparison of the detected rain rates by the gauge, parametric model and IMERG. (a) Daytime absolute differences between the means of the parametric model and IMERG, (b) Nighttime absolute differences between the means of the parametric model and IMERG, (c) Probability density of below 20 mm per 30 min rain rates, and (d) Probability density of above 20 mm per 30 min rain rates detected by the gauge, IMERG and model.

Figure 3. 12c shows comparable probability densities of the detected rain rates by the gauge, model and IMERG for rain rates below 20 mm per 30 min interval. By contrast, the densities for the above 20 mm rain rates in Figure 3. 12d show that IMERG's maximum detected rain rate was below 40 mm. On the other hand, the model detections were mainly comparable to the ground truth and above 50 mm per 30 min, suggesting that IMERG has a low tendency to detect very high rain rates. The model's comparable detected rain rates with the ground truth is because it was calibrated using a dense rain gauge network. IMERG's low tendency to detect very high rain rates may be due to the spatial and temporal averaging technique used to merge and intercalibrate rainfall estimates from several PMW sources.

Based on the results, it can be stated that the developed rain area detection and correction technique using multispectral data from the MSG SEVIRI was successful in detecting rain areas at 2 scales. (1) At the point and pixel scale when the detected rain areas were compared with gauge station data, and (2) at a large scale when compared to the spatial dynamics of rain areas detected by IMERG. Furthermore, retrieval and comparison of the model, IMERG, and gauge rain rates reveal the model's capability to detect rain rates comparable to rain gauge data and with a better tendency for detecting higher intensities than IMERG.

3.4.2 Discussion

3.4.2.1 The Day and Nighttime Rain Detection Technique

Rain area detection using multispectral satellite data from MSG SEVIRI was demonstrated for the daytime and night-time over topographically complex terrain in south-western Kenya. The technique relies on developing and calibrating multiple parametric rain detection models using rainfall data from a dense network of rain gauges to determine the best model parameters and parametric values for successful rain area detection. The models were rooted in the conceptual assumption that clouds characterised by their top properties, such as high optical thickness and effective radius (comprising ice and water hydrometeors), have high rainfall probabilities and intensities. Several models in 3 categories: reflectance, IR and combined reflectance-IR models, were developed to answer 2 primary questions: (1) what the rain detection skill of the proposed spectral models is? (2) Which model possesses the best detection skill?

The results, determined using standard categorical statistics and ROC curves, show that daytime rain detection based on models using reflectance alone data over the study area outperformed those using the IR and combined reflectance-IR models. Combining reflectance and IR data showed medium performance in rain area detection and reduced model FAR and Bias scores. However, the IR-only-based models, particularly the model based on a single IR BT, often showed poor performances. The best daytime model was determined from the reflectance models to be VIS0.6–NIR1.6, and the parameter value above which the best detection performance was achieved was 0.21. Based on the premise that the VIS0.6 and NIR1.6 reflectance indicates cloud optical thickness and effective radius, respectively, and point towards CWP, the results suggest that the reflectance differences above 0.21 detect clouds with high CWP indicating high rainfall probabilities.

The night-time models consisted of a single IR BT, BTD and combined IR BT-BTD models developed from IR-only spectral data. The results indicate that the

BTD model showed the best performance in rain area detection over the study area. In contrast, the combined IR BT-BTD and the single IR BT models showed medium and poor performances, respectively. The best model was determined to be a combination of IR3.9–IR10.8, IR3.9–WV7.3 and IR10.8–WV6.2 BTD. The corresponding parameter values, below which the best detection performance was achieved, were 8.18 K, 17.03 K and 33.65 K, respectively. The IR3.9–IR10.8 and IR3.9–WV7.3 parameters indicate cloud optical thickness and the IR10.8–WV6.2 BTD indicates the height of the cloud top. Unlike daytime detection, the results suggest that night-time rain area detection may be best achieved with cloud optical thickness and height information.

Overall, the day and night-time models demonstrated high FAR scores due to many factors, such as the non-linear relationship between rainfall and the cloud top properties (Karaseva et al., 2011). The daytime rain area detection performed better than the night-time, which could be attributed to the relevant information content on CWP and rainfall available in the VIS 0.6 and NIR 1.6 reflectance pair (Kumah et al., 2020). Furthermore, the bivariate analysis of the raining and non-raining spectral characteristics (Section 3.4.1.1) and their descriptive statistics (APPENDIX A3) reveal uniqueness in reflectance data that supports its high detection skill. The night-time model's comparatively weak performance, indicated by its high POFD and low CSI and ETS values, was also observed by Kumah et al. (2020) for a relatively small study area. It could be partly explained by the comparable distribution observed for their raining and non-raining spectral characteristics (Section 3.4.1.1). Nonetheless, the results support the conceptual model that raining clouds characterised by their top properties, such as high optical thickness and effective radius, have high rainfall probabilities.

3.4.2.2 The Rain Area Correction and Validation Technique

The detected rain areas were corrected for their number (i.e., the number of areas detected as raining) and sizes (i.e., the sizes of the detected rain areas) using rain area-specific parameters and adaptive parametric thresholds. Specifically, two kinds of gradient correction: the cloud object and pixel gradient, were defined for the daytime and night-time rain area correction. Although their determination differed for the day and night-time, the implications were the same for both periods. The cloud object gradient reduced the number of areas detected as raining by comparing a rain area-specific cloud gradient with the average gradient and standard deviation from all detected rain areas. On the other hand, the pixel gradient reduced the detected rain areas' size by comparing the gradient computed for each pixel in a raining area to its median and average median from all detected rain areas.

The developed rain detection parametric model was validated using independent validation sets and comparing the model results to data from the gauge stations and IMERG. The comparison between model rain detection results and the gauge station's rainfall data showed improvement in daytime detection, indicated by high ACC and low FAR scores compared to its previous calibration results. In contrast, low ETS scores (compared to previous calibration results) observed for the night-time detection suggested reduced night-time detection skills for the validation sets.

The detected rain areas by the parametric model and IMERG were compared to validate the model's rain area detection skill spatially. Their detected rain rates were compared to the gauge station rainfall data to evaluate the model's rain rate detection against ground truth and IMERG.

By visual inspection, the model's agreement in spatial dynamics of detected rain areas with IMERG's detected rain areas was generally convincing, particularly for large contiguous raining areas and better during the day than nighttime. However, there were some noticeable differences between the detected rain areas. For instance, a slight shift between the detected rain areas by the model and IMERG could be discerned from both the day and night-time results. Further, IMERG, particularly during the daytime, detected more rain areas of varying sizes, whereas the model's detection was mainly organised into large contiguous areas. Also, the absolute differences in mean rain rate detected by the model and IMERG at the gauge stations showed similar spatial dynamics for the day and night. However, the mean differences were comparatively higher for the day than the night-time.

3.4.2.3 Uncertainties and Implications of the Rain Area Detection and Correction Technique

The uncertainties in the model results may be related to multiple factors. For instance, the spatial and temporal mismatch between the MSG satellite and gauge observations, although spatiotemporally aggregated following the method by Bendix et al. (2010); Kumah et al. (2020), impacts the satellite's data sampled to calibrate and validate the model. Additionally, for multi-layered clouds with cloud properties differing between layers (Watts et al., 2011), the satellite's information may not agree with the ground observation (Lensky & Rosenfeld, 1997). This could explain uncertainties in the bivariate comparison of rain and non-raining spectral characteristics.

Many factors contribute to the differences in rain area and rates detected by the parametric model and IMERG. For instance, the difference in spatial resolution of the MSG and IMERG datasets (approximately 3 × 3 and 10 × 10 km,

respectively) suggests that the IMERG data is averaged at a comparatively large spatial resolution than the parametric model; however, in this study, the IMERG pixel was compared to a single gauge. Such a spatial disparity could affect the MAD and explain for, e.g., the comparatively less number but large contiguous (raining) cloud areas detected by IMERG than the model. Moreover, the developed model was based on reflectance and IR data from a single geostationary satellite, MSG, calibrated using rainfall data from a dense gauge station network.

On top of this, the parametric model's underlying conceptual assumption is the relationship between cloud top information such as high optical thickness and effective radius and rain probability and intensities. IMERG, on the other hand, is a multi-satellite algorithm that combines microwave observations from multiple satellite sensors to estimate half-hourly globally gridded precipitation. Central to IMERG is the morphing technique, which uses motion vectors to fill gaps in passive microwave precipitation estimates using a quasi-Lagrangian interpolation. This latest version of the IMERG rainfall product (V06b) uses motion vectors derived from total precipitable water vapour retrieved from numerical models, unlike geostationary IR BT in the previous versions (Tan et al., 2019). This could explain the differences in the number and sizes of the detected rain areas by the model and IMERG.

Furthermore, the shift/dislocation between IMERG relative to the model detected rain areas observed in this study should be due to the procedures undertaken in developing IMERG, such as the morphing passive microwave measurements using the motion vectors (Jackson et al., 2020). It is worth pointing out that spatial displacement or dislocation errors have not received much attention in satellite rainfall validation. Nevertheless, where the rainfall field is located has significant implications on operational applications' effectiveness and efficacy, such as flood and flash flood forecasts. Recent work has suggested some spatial displacement error metrics, e.g. (Acharya et al., 2020; Carlberg et al., 2020), to address this gap, albeit applied to reanalysis and rainfall forecast data. However, satellite rainfall evaluations that address this possible spatial displacement are still lacking. The results presented herein imply the possibility of a spatial displacement in satellite rain fields, which needs to be investigated, quantified and corrected. Therefore, future studies will investigate and address this potential spatial displacement error in rain areas from satellites.

The rainfall and satellite datasets used in this study represent a significantly larger space and time domain than the previous study (Kumah et al., 2020). Thus, the developed model and parameters may be regionally applicable in the tropical

(Eastern) African region to detect and correct rain areas using multispectral geostationary satellite data. Nevertheless, Milford et al. (1994) suggest that the rain area detecting parameters are regional and climate-dependent. Therefore, more research on the technique's applicability over a comparatively more extensive scale, e.g., continental, must be completed until a final rain area detection scheme is available.

In this regard, the near continental scale coverage of TAHMO's stations in Africa (van de Giesen et al., 2014) and the existence of opportunistic rain sensors such as MWL (Uijlenhoet et al., 2018) are, e.g., two valuable sources of rainfall information that heightens the potential of developing such a spatiotemporally high detection scheme. Despite the challenges, e.g. (Chwala & Kunstmann, 2019), in the last decade, it has been well established that accurate near-ground rainfall monitoring using CML data is possible and could benefit the sparsely gauged regions or complement conventional monitoring techniques. The study will be extended to cover sub-Saharan Africa and match TAHMO's spatial coverage in a future step. In addition, cloud-top information from Cloud Property Dataset Using SEVIRI, edition 2 (CLAAS-2) (Stengel et al., 2014) will replace the indirect cloud-top property information inferred from the SEVIRI optical, near and thermal IR data in this study. This could potentially reduce uncertainties due to analysing ground rainfall with indirect cloud top information.

3.5 Conclusions

A gradient based adaptive technique capable of day and night-time rain area detection and correction is presented using reflectance and IR data from the MSG SEVIRI satellite observations from south-western Kenya. In this investigation, we first developed, calibrated and validated multiple parametric rain detection models using rainfall data from a dense gauge station network to determine the best model parameters for the day and night-time. We then developed a new technique to correct the detected rain areas—the method uses rain area-specific parameters to reduce the number and sizes of the detected rain areas.

Compared to the GPM IMERG and gauge station data, the developed model shows convincing agreement in both the detected rain area and rain rates, suggesting that the new technique could provide valuable insights to satellite rainfall retrievals to benefit many operational applications.

The technique's limitation is related to the fact that it calibrates the satellite data using gauge rainfall data which may not be available everywhere. Additionally, the rain detecting parameters identified from the satellite data may be regionally and climate-dependent—implying the technique should be calibrated per study

region to obtain suitable parameter values for a successful rain area detection. In that sense, the study by Kumah et al. (2020) is for, e.g., a proof of concept that near-ground rainfall from microwave attenuations on CML can be used in place of the gauge data to overcome these limitations.

APPENDIX A3

The tables in this section are the descriptive statistics of the spectral characteristics during the daytime and night-time. The notations: Stan. Dev, 25%, 50%, and 75% are standard deviation, percentile values respectively.

Table A3. 1 Descriptive statistics of the spectral characteristics of daytime raining cloud top properties.

Descriptive Statistics	VIS0.6	NIR1.6	VIS0.6 / NIR1.6	VIS0.6 – NIR1.6	IR 10.8	IR10.8 – WV6.2	IR8.7 – IR10.8	IR10.8 –IR12.0	WV6.2 – WV7.3
Minimum	0.06	0.01	0.46	−0.11	213.02	0.38	−3.61	−0.95	−27.82
Maximum	1.00	0.81	17.47	0.90	288.02	50.01	5.49	6.56	0.87
Mean	0.63	0.27	2.51	0.35	253.03	18.89	0.16	1.01	−10.58
Mode	0.70	0.20	1.60	0.40	262.70	23.80	−0.80	0.50	−14.80
Median	0.65	0.26	2.40	0.36	256.24	19.09	0.05	0.81	−11.42
Stan. Dev	0.21	0.11	1.02	0.18	16.51	10.30	1.11	0.89	4.93
25%	0.50	0.20	1.78	0.22	241.42	10.86	−0.68	0.43	−14.56
50%	0.65	0.26	2.40	0.36	256.24	19.09	0.05	0.81	−11.42
75%	0.77	0.34	3.10	0.49	264.96	25.74	0.83	1.38	−6.78

Table A3. 2 Descriptive statistics of the spectral characteristics of daytime non-raining cloud top properties.

Descriptive Statistics	VIS0.6	NIR1.6	VIS0.6/ NIR1.6	VIS0.6 – NIR1.6	IR 10.8	IR10.8 – WV6.2	IR8.7 – IR10.8	IR10.8 – IR12.0	WV6.2 – WV7.3
Minimum	0.03	0.01	0.24	−0.31	212.49	0.15	−4.04	−9.31	−23.84
Maximum	1.00	0.89	10.67	0.78	294.91	57.02	8.27	15.10	2.49
Mean	0.43	0.32	1.39	0.11	266.62	29.86	−0.10	1.96	−14.34
Mode	0.40	0.30	1.00	0.00	268.90	30.10	−1.00	0.70	−16.60
Median	0.42	0.32	1.22	0.07	268.93	30.42	−0.40	1.70	−15.24
Stan. Dev	0.20	0.12	0.69	0.16	14.46	10.75	1.28	1.34	3.97
25%	0.27	0.24	0.95	−0.02	260.23	22.98	−1.02	0.92	−17.03
50%	0.42	0.32	1.22	0.07	268.93	30.42	−0.40	1.70	−15.24
75%	0.58	0.41	1.61	0.20	277.23	38.10	0.54	2.75	−12.67

Table A3. 3 Descriptive statistics of the spectral characteristics of night-time raining cloud top properties.

Descriptive Statistics	IR3.9 – IR10.8	IR3.9 – WV7.3	IR 10.8	IR10.8 – WV6.2	IR8.7 – IR10.8	IR10.8 – IR12.0	WV6.2 – WV7.3
Minimum	−7.57	−1.26	220.10	2.19	−1.89	−0.72	−22.84
Maximum	21.13	33.06	285.66	48.16	7.15	8.47	−0.70
Mean	1.91	10.54	254.93	20.41	0.18	1.21	−11.78
Mode	−1.50	8.50	258.40	17.10	0.20	0.70	−14.70
Median	1.16	9.61	256.04	20.30	0.04	0.99	−12.12
Stan. Dev	3.72	4.37	12.52	7.97	1.07	0.85	3.72
25%	−0.96	7.45	245.62	14.43	−0.67	0.63	−14.63
50%	1.16	9.61	256.04	20.30	0.04	0.99	−12.12
75%	4.17	12.71	264.73	26.00	0.82	1.56	−9.09

Table A3. 4 Descriptive statistics of the spectral characteristics of night-time non-raining cloud top properties.

Descriptive Statistics	IR3.9 – IR10.8	IR3.9 – WV7.3	IR 10.8	IR10.8 – WV6.2	IR8.7 – IR10.8	IR10.8 – IR12.0	WV6.2 – WV7.3
Minimum	−9.38	1.22	220.10	2.12	−2.20	−0.45	−23.89
Maximum	28.00	37.97	289.38	51.20	6.29	39.32	4.57
Mean	2.68	15.35	261.92	26.69	0.09	1.80	−14.01
Mode	−1.90	11.80	267.10	28.50	−1.00	0.80	−15.40
Median	1.81	14.37	264.02	27.22	−0.13	1.58	−14.65
Stan. Dev	4.68	6.21	12.08	8.50	1.22	1.13	3.49
25%	−0.97	10.49	254.78	20.96	−0.90	0.94	−16.54
50%	1.81	14.37	264.02	27.22	−0.13	1.58	−14.65
75%	5.41	19.48	270.42	32.78	0.86	2.44	−11.98

Chapter 4: Using Rain Area Detection from Meteosat Second Generation to Improve Commercial Microwave Link Rainfall Intensity

This chapter is originally published in:
Kumah, K. K., Hoedjes, J. C. B., David, N., Maathuis, B. H. P., Gao, H. O., & Su, B. Z. (2021). The MSG Technique: Improving Commercial Microwave Link Rainfall Intensity by Using Rain Area Detection from Meteosat Second Generation. Remote Sensing, 13(16). doi:10.3390/rs13163274

Abstract

According to studies in the past decade, MWL used by mobile telecom operators for data transmission can provide hydro-meteorologically valid rainfall estimates. For the first time, this study investigated a new method, the MSG technique, that uses MSG satellite data to improve MWL rainfall estimates. The investigation, conducted during the daytime, used MSG optical (VIS0.6) and near IR (NIR1.6) data to estimate rain areas along a 15 GHz, 9.88 km MWL for classifying the MWL signal into wet-dry periods and estimate the baseline level. Additionally, the MSG technique estimated a new parameter, wet path length, representing the length of the MWL that was wet during wet periods. Finally, MWL rainfall intensity estimates from this new MSG and conventional techniques were compared to rain gauge estimates. The results show that the MSG technique is robust and can estimate gauge comparable rainfall estimates. The evaluation scores every three hours of RMSD, relative bias, and r^2 based on the entire evaluation period results of the MSG technique were 2.61 mm h^{-1}, 0.47, and 0.81, compared to 2.09 mm h^{-1}, 0.04, and 0.84 of the conventional technique, respectively. For convective rain events with high intensity spatially varying rainfall, the results show that the MSG technique may approximate the actual mean rainfall estimates better than the conventional technique.

4.1 Introduction

An MWL is a communication between two antennas (i.e., transmitter and receiver antennas) usually installed on telephone towers or roofs of buildings by mobile telecom service providers for data transmission from radio, TV, internet, and wireless communication between our cell phones (Edstam et al., 2018; 2015; Patrick, 2020). MWL uses 10GHz–80GHz frequency ranges for data transmission, which are attenuated mainly by rainfall such that the more intense the rainfall, the stronger the MWL experiences attenuation. For this reason, previous studies pioneered by (Leijnse et al., 2007a; Messer et al., 2006) have investigated and converted the MWL signal to hydro-meteorological valid rainfall estimates.

Indeed, the MWL signal data have been studied for estimating rainfall for many applications (e.g. David et al., 2013; David et al., 2019; David et al., 2021; Doumounia et al., 2014; Kumah et al., 2020; Overeem et al., 2011). For example, Overeem et al. (2011) used the MWL data for measuring urban rainfall, and (David et al., 2013) demonstrated the data's potential application for monitoring rainfall in dry climatic regions. In Africa, David et al. (2019); David et al. (2021) investigated the MWL data's potential for providing valuable rainfall information for agricultural needs and (Doumounia et al., 2014; Kumah et al., 2020) tested its application for rainfall monitoring. Other studies have applied the data for country-wide rainfall monitoring (Overeem et al., 2013, 2016a) and complimenting gauge and radar rainfall estimates (Rahimi et al., 2004).

Such extensive research of the MWL data for rainfall estimation is due to their specific advantages for rainfall monitoring comparable to prevailing techniques. For instance, their network on land is relatively dense and can estimate rainfall over vast areas comparable to weather radars. Additionally, this naturally dense MWL network allows for spatially redundant rainfall observing systems with potentially no single point of failure (i.e., unlike radars, when one MWL fails, several other links are usually active). Moreover, line-average rainfall estimated from the MWL is spatially representative of areal rainfall than point estimates from rain gauges. Further, the potential costs of running and maintaining the MWL network for rainfall estimation and monitoring are minimal since the telecom service providers have already established and maintained the infrastructure.

Nonetheless, there are limitations to MWL rainfall estimation and monitoring. Access to MWL data can be a challenge. Usually, there are no standard procedures (Chwala & Kunstmann, 2019), so some studies (e.g. Leijnse et al., 2007a; Upton et al., 2005) utilised self-made MWL data in their MWL rainfall

estimation. Additionally, the MWL network is designed for a purpose other than rainfall monitoring, often arbitrary in space and mostly biased towards densely populated areas (Zinevich et al., 2008). This complicates rainfall mapping from the MWL and can affect the retrieval accuracies for low MWL network density areas. Furthermore, the low sampling frequency (usually 15 minutes), precision (often 1 dB), and the noisy nature of the MWL data present additional challenges to accurate rainfall estimation from MWL data (Chwala & Kunstmann, 2019; David et al., 2015; Leijnse et al., 2008; Uijlenhoet et al., 2018).

Currently, three primary steps are used to estimate rainfall from the MWL data. Firstly, the MWL received signal levels (RSL) are classified into wet and dry periods, describing periods when rain is present or absent on the MWL. This is essential because MWL rainfall estimation uses data from only the wet periods. The prevailing methods for this classification are centred on two concepts. One concept assumes rainfall is naturally correlated in space and relies on mutual attenuation on neighbouring for the wet-dry classification (Overeem et al., 2016b). This concept favours high MWL network density areas but may significantly challenge areas with low network density and high spatially varying rainfall. The other concept classifies the MWL RSL data by analysing the statistical properties of the time series of the individual MWL (Schleiss & Berne, 2010) and thus may not be affected by the network's density. However, gaps in the MWL RSL and low sampling frequency data may affect the wet and dry classification accuracy.

Step two estimates the baseline level to represent the RSL behaviour during the dry period. The accuracy of this baseline level estimate is affected by the classification accuracy in the previous step and the fact that the MWL RSL fluctuate during the dry period due to attenuation caused by other non-rainfall-related sources (Chwala & Kunstmann, 2019). Previous studies (e.g. Kumah et al., 2020; Rios Gaona et al., 2015) estimated the baseline level as the median signal of all dry periods in the previous 24 hrs. The final step computes rain-induced specific attenuation (i.e., the relative loss of signal attributed to the MWL length) by subtracting the signal level from the baseline level and dividing it by the length of the MWL. Before estimating the path average rainfall from, e.g., the power-law model (Olsen et al., 1978), attenuations due to antenna wetting are often corrected using varying techniques, see, e.g. (Minda & Nakamura, 2013; Schleiss et al., 2013)). The wet antenna attenuation describes the additional attenuation caused by wetting the MWL antenna surfaces during and after rainfall; this needs to be estimated and corrected to prevent the overestimation of the MWL rainfall.

The conventional MWL rainfall estimation technique described above implies that the rain-induced attenuation and rainfall retrieved from the MWL represent average attenuation and rainfall over the entire MWL propagation path (Rios Gaona et al., 2015). Nevertheless, rainfall, in some cases, can vary spatiotemporally along the MWL propagation path. This suggests that the conventional technique may not, at all times, accurately represent the actual average rainfall intensity, e.g., as identified by rain gauges (Villarini et al., 2008).

Due to the existing challenges in MWL rainfall estimation, this study advocates incorporating high-resolution information on clouds and rainfall from meteorological satellites, including MSG, in the rainfall retrieval procedure. MSG represents a significant advancement in observation capabilities from previous geostationary meteorological satellites. MSG's radiometric sensor, SEVIRI, has a wide spectral range and a frequent repeat cycle (Schmetz et al., 2002). These measurement characteristics permit a quasi-continuous observation of rainfall distribution, making it possible to study spatiotemporally varying rainfall in near-real-time (Bendix et al., 2010).

Surprisingly, only a few studies have combined the MWL and satellite data for rainfall detection and estimation (Kumah et al., 2020; van het Schip et al., 2017). To our knowledge, no study has used the MSG satellite data to improve MWL rainfall estimations. Both data already exist at a comparable spatial coverage (on land) and temporal resolution while being used independently for rainfall detection, estimation, and monitoring. Nevertheless, the MWL and MSG data synergy could be valuable for areal rainfall estimation from the MWL. More precisely, a combination of the MWL and MSG, where the satellite estimates high spatiotemporal resolution raining area information, could be valuable to the MWL rainfall estimation procedure. For example, this study shows how MSG-based rain area information could benefit the MWL's wet-dry periods and baseline level estimation during the daytime. Additionally, it is shown in this study that the spatially distributed raining areas identified by the MSG could be critical information for identifying the approximate wet path of the MWL, particularly during spatially varying raining conditions such as convective rainfall (Hoedjes et al., 2014), and improve the MWL rainfall estimates thereof during the daytime.

Therefore, this study benefits from MSG's high measurement (temporal, spatial, and spectral) resolution for improving MWL rainfall intensity estimation during the daytime. More precisely, it is investigated whether MSG's high-resolution rain area detection could achieve wet-dry and baseline level estimation for a successful MWL rainfall estimation. Further, the rain area information from the satellite is

used to investigate a new parameter, wet path length (wpl in km), representing the approximate length of the wet MWL (i.e., length of the MWL covered by the rainfall) during each wet interval. This is particularly important under spatially varying raining conditions for improving the retrieved MWL rainfall estimates. Ultimately, this new technique—the MSG technique—and the conventional technique are compared to the actual mean rainfall intensities from rain gauges to evaluate the MSG-based rain areas' effect on improving the MWL rainfall estimates.

4.2 Study Area and Dataset

The data used in this study were MWL, rain gauges, and MSG SEVIRI from a topographically complex area (-0.61° S, 36.6° E) close to the Aberdare mountain in Kenya (0.02° S, 37.90° E). The evaluation period was between May and June 2018. Previous research in this region during the evaluation period demonstrated the data's capability for rainfall detection and estimation (Kumah et al., 2020).

Rainfall data from nine aerodynamic 'tipping buckets' (ARG TB) rain gauges and two rain gauges from TAHMO (van de Giesen et al., 2014) served as ground truth in this study. The ARG TB were aligned under the MWL transect. In contrast, the TAHMO gauges were placed close to its transmitting and receiving antennas. The ARG TB logged data every minute using a Gemini Tinytag data logger, while the TAHMO gauges recorded rainfall data every 5 minutes. One tip of the ARG TB equates to 0.198 to 0.202 mm of rain.

A Kenyan telecom service provider, Safaricom, supplied the received signal level (RSL) data for a 15 GHz, 9.88 km MWL. The RSL data were characterised by minimum, maximum, and mean values at 15 minute intervals and a resolution of 0.1 dBm. It is an Aviat Eclipse MWL, vertically polarised, and has a constant transmitted signal level (TSL). The data was accessed through Safaricom's head office in Nairobi, Kenya.

The MSG SEVIRI data was obtained from the Meteosat at 41.5o E, which corresponded to Meteosat 8 (EUMETSAT, 2016) when the data was retrieved. The SEVIRI channels used were visible (VIS0.6 μm) and near-infrared (NIR1.6 μm), provided by EUMETSAT (2020) at 3km and 15 minutes spatiotemporal resolution.

4.3 Method

4.3.1. Rainfall Intensities Estimated from Rain Gauges

The rainfall for all gauges was used to estimate rainfall intensities (R mm h^{-1}) at 15 minutes. For the ARG TB, this was computed from the per-minute rain rate in millimetres estimated from the tipping count and the gauge-specific tip equivalent of rain in mm provided by the manufacturer. On the other hand, R (mm h^{-1}) from the TAHMO data were computed from their 5 minutes accumulated rain rates. In this study, a gauge was considered raining if the R (mm h^{-1}) was above 0.5 mm h^{-1}; otherwise, it was non-raining. Table 1 summarises the non-zero rainfall data for all the gauges. The differences in the gauge rainfall data records are mainly due to spatial variability of rainfall in the area due to topography and gauge malfunctions during the field campaign. For instance, gauge G1 was often non-operational during the field campaign, thus having the least days with data records. Nonetheless, these gauge rainfall records depict high spatial rainfall variability over the MWL propagation path.

Table 4. 1 Summary of non-zero rainfall intensities from all rain gauges.

	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
Mean	7.12	7.65	8.74	8.25	6.79	4.41	2.88	4.36	6.7	3.82	3.98
Maximum	43.46	23.64	72.36	49.53	27.72	43.63	5.54	27.47	19.3	12.86	28.02
Standard dev	9.8	7.65	15.01	12.79	8.16	8.56	1.43	5.61	7.33	3.98	5.71
Fraction %	3.29	1.72	2.4	1.98	1.46	2.82	1.14	2.47	0.62	0.82	1.88
N	70	7	23	19	14	44	11	46	6	8	40
n days	53	12	26	26	26	39	26	46	26	26	53

N is the total number of 15 minutes of rainfall data assembled from n number of days during the daytime in the evaluation period. The fraction (%) represents the fraction of raining periods. Standard dev is the standard deviation

4.3.2. Rainfall Intensities Estimated from MWL

The rain-induced specific attenuation A(dB km-1) can be used to estimate R (mm h^{-1}) from, e.g., the power-law model in (Olsen et al., 1978):

$$A = aR^b \tag{4.01}$$

where the a ((dB km^{-1}) (mm h^{-1})-b) coefficient and b (-) exponent depend on the MWL frequency, polarisation, and local rainfall climatology (Olsen et al., 1978; Uijlenhoet et al., 2018), which can be acquired from the literature, such as (ITU, 2005b).

4.3.2.1. The Conventional Technique

The conventional technique used to estimate MWL rainfall intensities in this study is described in detail by Kumah et al. (2020) and in the methodology of Chapter 2. Here, a summary of the steps used to estimate the path average rainfall from the mean RSL data is provided. (1) The wet-dry classification of the MWL signal was by the rolling window approach, which uses the time series of the individual MWL signals separately. (2) Next, a reference or baseline level was estimated as the median of the mean RSL of the previous 24 hours labelled as dry periods by the preceding step. Before computing A from Equation 4.02, a dynamic wet antenna correction model (Schleiss et al., 2013) was used to correct the mean RSL from attenuation due to antenna wetting.

$$A_{fp} = \frac{B - P}{L_{fp}} \tag{4.01}$$

where:

A_{fp} (dB/km)—is the rain-induced specific attenuation averaged over the entire MWL

L_{fp}—is the MWL length, and B and P are the baseline and the mean RSL, corrected for the effect of antenna wetting according to the dynamic model by Schleiss et al. (2013).

Finally, Equation 4.03 estimated the R (mm h^{-1}) from A_{fp}.

$$R_{fp} = \left(\frac{A_{fp}}{a}\right)^{\frac{1}{b}} \tag{4.02}$$

where R_{fp} is the path average rainfall computed based on the entire MWL length, and a (0.05008) and b (1.0440) were from (ITU, 2005b).

4.3.2.2. The New MSG Technique

The MSG technique incorporates an MSG-based rain area detection and correction method recently developed by Kingsley et al. (2021) into the MWL rainfall estimation procedure. The method detects rain areas at 3 km and 30 minutes spatiotemporal resolution. However, in this study, it was implemented at MSG's 15 minutes temporal resolution to match the temporal resolution of the MWL RSL data. Additionally, the method is capable of daytime and night-time rain area detection, but this study focused on daytime detection due to its high success rate of rain detection. Further details of the method and its accuracy can be found in (Kingsley et al., 2021); here, a summary and its application for MWL rainfall estimation are provided.

This rain area detection method is instantaneous, which means that, for each individual MSG scene, the method detects rain areas independent of the previous and subsequent scenes. It employs a parametric threshold model developed from a conceptual framework in which clouds characterised by top properties such as high top optical thickness and large effective radius have high rainfall probabilities and intensities. The daytime model is from MSG SEVIRI optical (VIS0.6) and near-infrared (NIR1.6) reflectance differences. Specifically, the model application assumes that a cloud is raining if the reflectance difference is above 0.21; otherwise, it is non-raining. The method subsequently corrects the detected rain areas by employing a gradient-based adaptive correction technique that uses rain area-specific parameters to reduce the number and sizes.

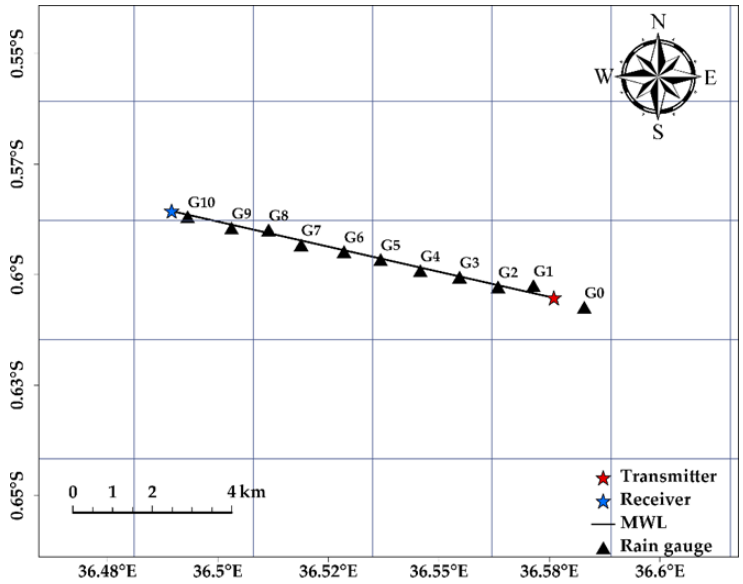

Figure 4. 1 The MWL and rain gauges displayed in MSG pixels.

The following steps describe how the MSG-based rain area information is incorporated into the MSG technique for estimating rainfall intensities from the mean RSL data. (1) During each 15 minute interval in the MWL mean RSL data, the rain area detection method classified the pixels over the link as raining or non-raining. When a pixel over the MWL was classified as raining, the new parameter, wpl, was retrieved from the length (km) of the MWL touching the raining pixel. Figure 4. 1 displays the MSG pixels over and around the neighbourhood of the MWL.

A 15 minute interval in the mean RSL data was classified as wet if the wpl was larger than 15% of the MWL length; otherwise, the interval was classified as dry. This was to ensure the retrieval of realistic path averaged specific attenuation values in the subsequent step. Simultaneously, the rainfall intensities measured by the rain gauges situated in the wpl were retrieved and averaged as the actual mean rainfall intensities. (2) After identifying the wet and dry periods in the RSL data, the baseline level was estimated from the mean RSL. The latter is the median of the previous 24 hours classified as dry by the MSG-based rain area detection technique. (3) Finally, the wet antenna correction method by Schleiss et al. (2013) was implemented before estimating the rain-induced specific attenuation from:

$$A_{wp} = \frac{B_{MSG} - P}{L_{wp}} \tag{4.03}$$

where:

A_{wp} (dB km^{-1})—the rain-induced specific attenuation averaged over wpl
L_{wp}—the wpl (km) retrieved from the MSG-based rain area information
B_{MSG}—the baseline, retrieved from dry periods identified by the MSG-based rain area information.

The average R (mm h^{-1}) was estimated from A_{wp} using:

$$R_{wp} = \left(\frac{A_{wp}}{a}\right)^{\frac{1}{b}} \tag{4.04}$$

where R_{wp} (mm h^{-1}) represents the average rainfall intensity over wpl.

4.3.2.3. Conditions and Uncertainties in Estimating the R_{wp}

Some conditions under which the R_{wp} may be uncertain, and the approach to retrieving more accurate estimates is described. The first is when the wpl is between 1.5 (the threshold MWL length used to make a wet-dry decision in the RSL data) and 3 km (i.e., the width of the MSG pixel). Figure 4. *2*a shows rain area detection over the MWL in binary classification (1 is rain and 0 is no rain), where wpl is approximately 2 km. The R_{wp} for such wpl are high and do not correlate with the gauge rainfall and MSG pixel intensity, such as those from convective raining pixels, because the MWL attenuation is computed over a short MWL length. For these cases, the R_{wp} from Equation 4.04 was multiplied by $\alpha = \frac{wpl}{L}$, where $\alpha < 1$ and L is the MWL length to retrieve more accurate estimates.

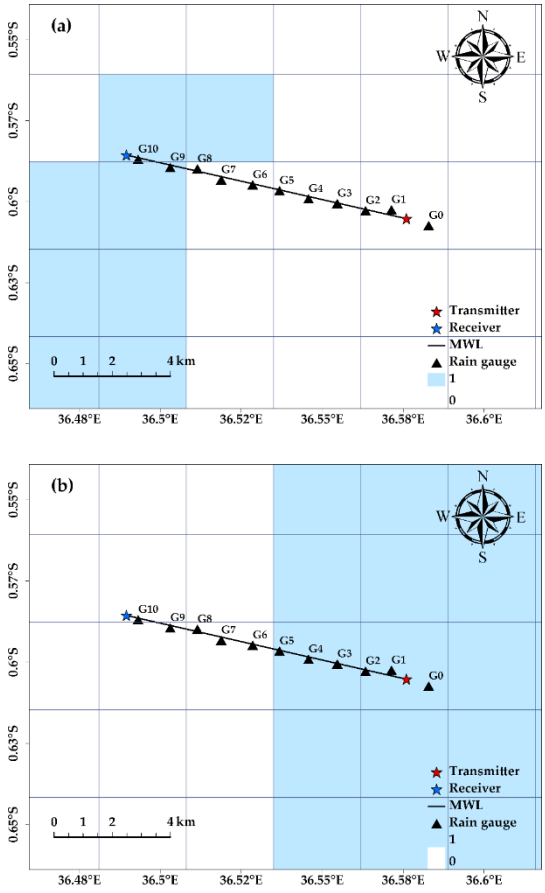

Figure 4. 2 MSG-based rain (blue), no rain (white) area detections over the MWL and wpl for different conditions. (a) Rain areas cover about 20% of the MWL's length (b) Rain areas cover about 50% of the MWL's length.

The second is when wpl is estimated from rain areas defined by mixed pixels (i.e., raining MSG pixels of varying reflectance difference intensities) from a convective rain cloud. A convective rain cloud covered a few rain gauges (see Figure 2b), had high gauge rain intensities, and mostly lasted for less than an hour. The intensities of the raining area MSG pixels also varied from a maximum reflectance difference to a relatively low difference. Capturing the high rain intensities for such cases requires determining the centre of the convective rainstorm. This was determined as the raining pixels with reflectance difference greater than the mean reflectance difference of all raining pixels covering the MWL, and wpl was estimated from the length of the MWL covered by these pixels.

4.3.3. Error Metrics

The average rainfall intensities by the conventional and MSG technique were compared to the rain gauges using the root mean squared deviation (RMSD), relative bias (RB), and coefficient of determination (r^2) to evaluate their performances against the actual mean rainfall intensities. The actual mean rainfall intensities for the conventional technique were computed from all gauges under the MWL; for the MSG technique, this was computed from gauges under wpl. Detailed descriptions for these metrics are in (Barnston, 1992; Walther & Moore, 2005; Wilks, 2006) and Equations 4.05 and 4.06, respectively.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(R_{Mi} - R_{RGi})^2}{N}} \tag{4.05}$$

$$RB = \frac{\frac{1}{N} * \sum_{i=1}^{N}(R_{Mi} - R_{RGi})}{\frac{1}{N} * \sum_{i=1}^{N} R_{RGi}} \tag{4.06}$$

where:

R_{Mi} —represents all possible MWL rainfall intensity estimates by the conventional and MSG technique

R_{RGi} —represents all possible gauge mean rainfall estimates, and N is the number of samples.

4.4 Results and Discussion

4.4.1 Results

This section compares the MWL rainfall intensity estimates by the conventional and the new MSG technique to actual mean rainfall intensities from gauges to evaluate the new technique's accuracy for improving MWL rainfall intensity estimates. From the raw mean RSL to the MWL rainfall intensity estimates, line plots for specific rainy periods were used to compare MWL rainfall intensity estimates to rain gauge estimates. Due to the variable rain intensities, the selected periods allowed for an effective visual comparison of the new MSG and conventional techniques. Finally, performance metrics computed based on all the 15 minute intervals in the MWL RSL data evaluate the accuracy of the new technique for MWL rainfall intensity estimation.

4.4.1.1 From Raw RSL to Rainfall Intensity Estimates: a Comparison of the Conventional and MSG Technique

Figure 4. 3 demonstrates the transformation of the mean RSL to rainfall intensities and its comparison with the actual mean intensities from rain gauges

according to the conventional and MSG techniques. Figure 4. 3a shows comparable baseline levels estimates by the conventional and MSG techniques. Figure 4. 3b compares the conventional and MSG technique's wet-dry classification. The dashed pink line indicates a standard deviation threshold value of 0.7 (dB), an empirically determined value by Kumah et al. (2020), that separates the wet (above 0.7 dB) and dry (below 0.7 dB) in the conventional technique. Instead, the MSG technique uses a binary classification of 0 and 1 to indicate when the MWL is wet (1) and dry (0).

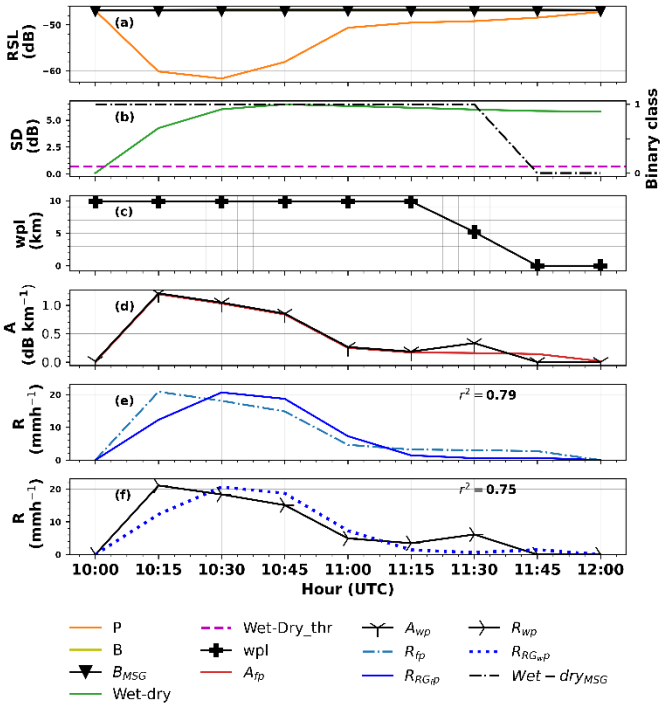

Figure 4. 3 From MWL mean RSL to rainfall for 15 minute interval rain events of 8 May 2018. (a) compares the conventional (B) and MSG technique (B_{MSG}) baseline level to the mean RSL (P), (b) compares wet–dry classification by the conventional (wet–dry) and MSG technique (wet–dryMSG)—the standard deviation threshold (wet–dry_thr, dashed pink line) value (0.7 dB) separates the wet and dry periods in the conventional technique; a binary class showing raining (1) and non-raining (0) periods separates the wet and dry periods in the MSG technique—(c) is the wpl over which the MSG technique computed attenuation (A_{wp}), (d) compares the conventional (A_{fp}) and MSG technique (A_{wp}) attenuation, (e) compares the conventional MWL rainfall intensity (R_{fp}) and the actual mean rainfall intensity (R_{RGfp}), and (f) compares the MSG technique MWL rainfall intensity (R_{wp}) and actual mean rainfall intensity (R_{RGwp}).

According to the wpl in Figure 4. 3c, the MWL was fully wet from the onset of the rain event until 11:15 and then partially wet at 11:30. Thus, this suggests that the rain events during the former periods occurred over the entire MWL length. In contrast, the event occurred over approximately half the MWL length for the latter period. Following these observations, one can observe comparable rain-induced specific attenuation estimates (Figure 4. 3d) by the conventional and MSG techniques from the rain event onset until 11:15 since their attenuation estimates were computed over the entire MWL length. However, the attenuation estimate at 11:30 was comparatively higher for the MSG because, unlike the conventional technique, it was estimated over a shorter MWL length (approximately 5 km).

The MWL rainfall intensities by the conventional and MSG techniques in Figure 4. 3e,f, respectively, were comparable most of the time for this rain event. Overall, they had a convincing agreement with the actual mean rainfall intensities from the rain gauges according to the r^2 values computed based on the entire rain event period. One noticeable feature in Figure 4. 3e,f is that both the conventional and MSG techniques overestimated the actual mean intensities at the onset of the rainfall event.

Figure 4. 4, like Figure 4. 3, demonstrates the transformation of the mean RSL to rainfall intensities by the conventional and MSG techniques compared to the actual mean rainfall intensities from rain gauges. The baseline level by both techniques in Figure 4. 4a again shows comparable estimates. Their wet–dry classifications (Figure 4. 4b) are also primarily comparable, especially for wet periods between 11:00 and 12:00. As was observed in Figure 4. 3, these observations coincide with a decrease in mean RSL (dB) attributed to rainfall as observed by the rain gauges.

However, there are some differences in Figure 4. 4b. The wet-dry classification by the MSG technique seems to detect wet periods before the onset of the wet period, which does not agree with the mean RSL and the no rainfall occurrence in the rain gauges, e.g., between 10:30 and 10:45. On the other hand, the conventional technique missed the onset of the wet periods between 10:45 and 11:00. For the MSG technique, this may be due to the raining cloud top properties available before the onset of the rain event, which resulted in false rain detection by the rain detection method used by the MSG wet–dry classification technique. By contrast, the conventional wet-dry technique missing the beginning of the wet period may be due to low rainfall intensities that resulted in a mean RSL that is not entirely different from the RSL in the preceding dry periods. Again, the

conventional technique continues to report wet periods even after it has ceased, likely due to the wet antenna effect.

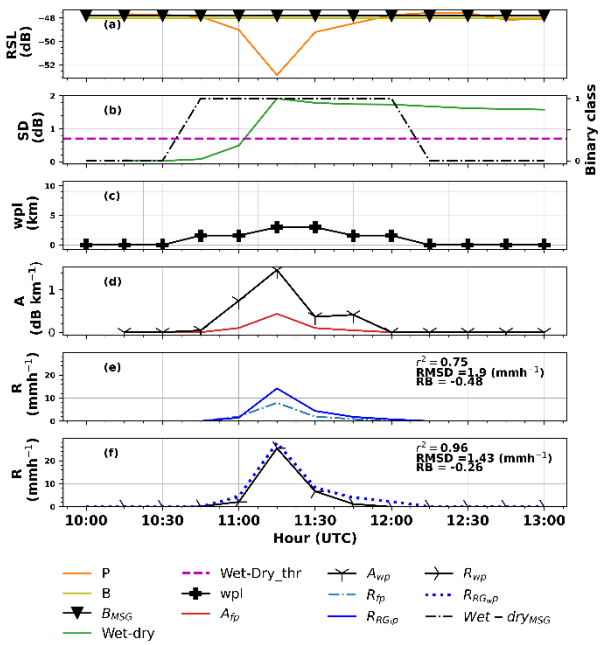

Figure 4. 4 (a-f) As in Figure 4. 3 but for 15 minutes interval rain events of 2 June 2018.

Figure 4. 4c suggests spatiotemporally distributed rainfall events occurred along the MWL length, indicated by time-varying wpl, for the entire raining period. Such rainfall events are often attributed to convective rainfall and are associated with sub-hourly high rain intensities. In Figure 4. 4d, the MSG technique estimated comparatively high attenuations than the conventional technique because it was estimated over a portion of the MWL length rather than the entire length (i.e., according to the conventional technique).

As shown in Figure 4. 4e,f, the rainfall intensities by the MSG technique are also comparatively higher than the conventional technique due to its higher attenuation estimates. Its rainfall intensity estimates better reflect the high rainfall intensities observed for that rainfall event and compare better to its actual mean rainfall than the conventional technique according to the error metrics shown in the figures. This can be attributed to determining the rainstorm's centre for a more accurate estimation of wpl to capture the high rainfall intensities.

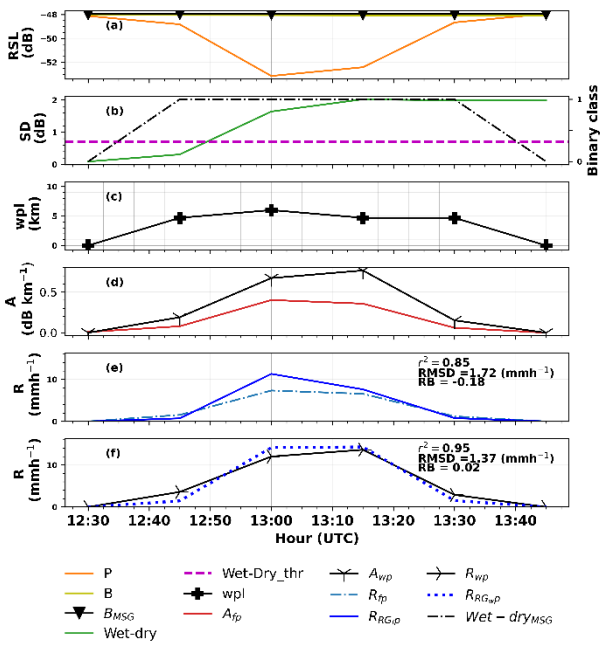

Figure 4. 5 (a–f) As in Figure 4. 3 but for 15 minutes interval rain events of 9 June 2018.

Figure 4. 5 is an analogous comparison of Figure 4. 3 and Figure 4. 4, but for different date-time periods. Again, Figure 4. 5a shows comparable baseline levels estimates by the MSG and conventional techniques. The wet-dry classification in Figure 4. 5b shows that the MSG technique captures the dynamics in the mean RSL and rainfall observation in the rain gauges better than the conventional technique. It is clear from the figure that the technique coincides nicely with the onset and end of the wet period and the dry periods preceding and after the wet periods. On the other hand, the conventional technique missed the onset of the wet period and continued to detect wet periods even after they had ceased.

The wpl shown in Figure 4. 5c suggests spatiotemporally varying rainfall events occurred along the MWL length, as in the previous analysis (Figure 4. 4c). Correspondingly, Figure 4. 5d shows comparatively higher attenuation estimates by the MSG than the conventional technique. This is because it was estimated over varying portions of the MWL rather than the entire length based on the conventional technique. According to their figures' error metrics, their rainfall estimates in Figure 4. 5e and Figure 4. 5f show good agreement with the actual

mean rainfall estimates from the gauges, albeit better in the MSG than the conventional technique.

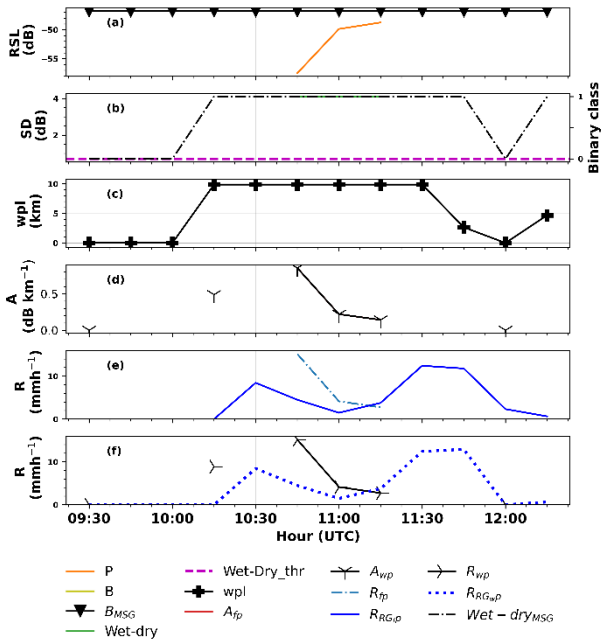

Figure 4. 6 (a–f) As in Figure 4. 3 but for 15 minutes interval rain events of 3 May 2018

Albeit analogous to previous figures, Figure 4. 6 is a unique comparison of the MSG and conventional techniques regarding frequent gaps in the MWL RSL data. The mean RSL in Figure 4. 6a shows that the MWL data was available for only a few minutes of this rainfall event. For this reason, wet-dry classification, attenuation, and rainfall (Figure 4. 6b,d,e, respectively) were unsuccessful in the conventional technique. However, the MSG technique instantaneously determines the wet-dry periods and wpl from the rain areas information (Figure 4. 6b and c). Therefore, the technique could estimate attenuation and rainfall intensities (Figure 4. 6d,f, respectively) for the period when the MWL data was available. Note that the r^2 values are not computed in Figure 4. 6e,f because the MWL data gaps did not allow for a fair comparison of MWL–gauge rainfall intensities.

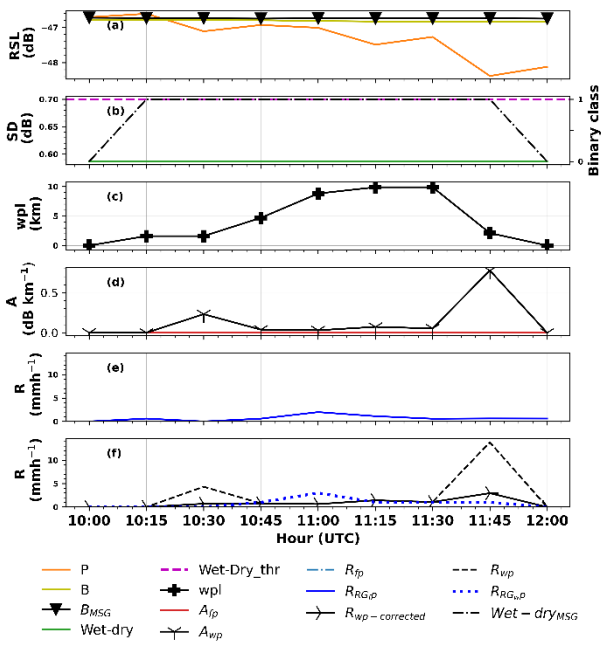

Figure 4. 7 (a–f) As in Figure 4. 3 but for 15 minutes interval rain events of 15 May 2018.

Figure 4. 7 demonstrates the effect of α in correcting the R_{wp} described in Section 4.3.2.3. As shown in Figure 4. 7b,d,e, the conventional technique missed this period's rain event, possibly due to the low rain rates (< 5 mm h-1) observed based on the gauge rainfall data. The MSG technique's wet-dry classification detected the event (Figure 4. 7b), which may be attributed to the MSG-based rain area detection's ability to detect low rain rates. The wpl in Figure 4. 7c shows the time-varying MWL length covered by the rain. For instance, at the beginning and end of the event, the estimated wpl was < 3km. Subsequently, the attenuation and rainfall retrieved based on this MWL length were high and did not agree with the gauge rainfall intensities, thus requiring correction. After correction, using the α, the MSG technique's rainfall intensities $R_{wp-corrected}$ visually shows a better agreement with the observed rainfall intensities by the rain gauges than its previous estimate based on the wpl.

Overall, the results show that rainfall estimation from MWL is robust in approximating the actual mean rainfall intensities over the MWL propagation path. Additionally, the MSG technique successfully estimated wet-dry and

baseline-level MSG techniques. This can be attributed to the MSG-based rain area detection, which uses relevant information content on cloud-top properties and rainfall available in the VIS 0.6 and NIR 1.6 reflectance pair.

In particular, the successful baseline level estimates by the MSG technique can be attributed to the robustness of the MSG data in detecting dry areas (Kumah et al., 2020). Furthermore, the new parameter, wpl, derived from the MSG-based rain area information, effectively estimated attenuation and rainfall intensities comparable to the conventional and actual gauge estimates, especially when the entire MWL length was wet. Based on the results, it can also be stated that when the rainfall is not spatially covering the entire length of the MWL, the MSG technique provides a better estimate of the actual mean rainfall as retrieved from, e.g., rain gauges.

4.4.1.2 Appraisal of the MSG and Conventional Technique for MWL Rainfall Intensity Estimation

Table 4. 2 presents the error metrics computed based on MWL and gauge rainfall intensity pairs for evaluation timestamps at 15 minutes, 30 minutes, hourly, and every three hours for both the MSG and conventional techniques. The RB suggest an overestimated rainfall intensities by the MWL relative to the actual gauge estimates, albeit comparatively higher in the MSG than the conventional technique. This is reflected in the RMSD, which is comparable for the two techniques, except for the 3-hourly scores. Nonetheless, their r^2 values were above 0.5 at 15 min, which increased due to aggregation (Rowe, 1976) above 0.8 at hourly and 3-hourly evaluation timestamps, indicating that both techniques can estimate gauge comparable rainfall intensities.

Table 4. 2 Error metrics computed for varying evaluation timestamps.

Estimation technique	RMSD mm h^{-1}				RB				r^2			
	15 min	30 min	1 h	3 h	15 min	30 min	1 h	3 h	15 min	30 min	1 h	3 h
MSG	0.63	0.84	1.32	2.61	0.47	0.47	0.47	0.47	0.70	0.78	0.83	0.81
Conventional	0.60	0.80	1.23	2.09	0.02	0.02	0.03	0.04	0.63	0.73	0.80	0.84

The total number of MWL–gauge rainfall intensity pairs (including raining > 0.5 mm h-1 and non-raining < 0.5 mm h-1) that computed the error metrics were 2088, 1380, 660, and 240 for the evaluation timestamps at 15, 30 minutes, one hour, and three hours, respectively, covering the evaluation period. These data and periods corresponded to when the MWL, MSG satellite, and rain gauges coincided.

Various factors may account for the overestimation of rainfall intensities by the MWL relative to the gauge estimates. The MWL rainfall intensities represent

areal average estimates derived from 15 minutes instantaneous mean RSL data. In contrast, the gauge computed average rainfall estimates from point measurements recorded per minute and 5-minute intervals. Additionally, spatial variability of rainfall in the study area and uncertainties in the MWL rainfall estimation may contribute to the discrepancies between the MWL and gauge estimates (Kumah et al., 2020; Uijlenhoet et al., 2010).

In particular, compared to the conventional technique, the comparatively high overestimation by the MSG is mainly attributed to uncertainties in the MSG-based rain area detection (Kingsley et al., 2021). For instance, false alarms in the rain area detection method could significantly affect the wet-dry classification and baseline level estimation. A dry interval in the MWL data, incorrectly identified by the MSG-based information as wet, could compute rainfall for a dry period and affect the RMSD and RB scores. Furthermore, the MSG technique estimated high attenuation and rainfall intensities for wpl less than the MWL length (indicating spatial variability of rainfall over the MWL). Additionally, this wpl is tied to the MSG rain area information's accuracy and may further impact the computed error metrics that evaluate the MSG technique's accuracy.

4.4.2 Discussion

A new technique for MWL rainfall estimation is investigated and described using 15 GHz, 9.88 km MWL, and MSG SEVIRI VIS0.6 μm and NIR1.6 μm satellite data. The investigation, conducted during the daytime, used the MSG data for detecting rain areas, according to the method by Kingsley et al. (2021), over the MWL propagation path. This spatial information on rain areas provided by the MSG data estimated three significant parameters: wet-dry periods, baseline level, and wpl for estimating MWL rainfall intensities.

All three parameters were estimated instantaneously from the MSG satellite data. The wet-dry periods indicate when the MWL was wet or dry, whereas the baseline level represents the MWL's behaviour in the dry period. The wpl is a property of the MWL's length, indicating the approximate length of the MWL during a wet period based on which the rain-induced attenuation was estimated. Eventually, rainfall intensity was estimated using all parameters. The results were compared to intensities estimated by a conventional method and rain gauges (where the gauge served as the actual estimates).

The results demonstrate an effective skill of the new MSG technique. The wet-dry periods and baseline level estimates were comparable to the conventional technique. In addition, the wpl effectively estimated the MWL signal attenuations

for wet periods. Subsequently, the rainfall estimates agreed well with the conventional and rain gauge estimates when the rainfall was spatially covering the entire length of the MWL (i.e., when the wpl was equal to the MWL length). However, when the rainfall is convective and spatially covered a portion of the MWL, determining the centre of the rainstorm was required to estimate a more accurate wpl and capture the high rainfall intensities that reflect the actual mean intensities better than the conventional technique. Moreover, unlike the conventional technique, the MSG technique's rainfall detection and estimation were unaffected by periods with no MWL RSL measurement.

Furthermore, the current study's results were better than the previous study's results (Kumah et al., 2020) for the same link, albeit with overestimation. The hourly RB and r^2 values were −0.18 and 0.58, respectively, estimated from combined daytime and night-time rainfall. The differences in performance may be due to the daytime rainfall and its high rainfall intensities measured by this current study, for which the influence of wet antenna and wet-dry classification errors may be minimal (Overeem et al., 2021)

The MSG technique showed higher RB and RMSD scores than the conventional technique. Generally, the differences can be attributed to differences in their measurement techniques and uncertainties in the MWL rainfall estimation procedure. The differences resulting from the uncertainties in the rain area detection provided by the MSG, such as false rain detections, could be interpreted in the MWL data as wet periods and be used to compute rainfall during dry periods, impacting the metric scores. It is noteworthy that this study's results were based on daytime MSG satellite reflectance data. Additionally, convective clouds with cold cloud top temperatures are responsible for most of the rainfall in the area. Therefore, our analysis did not consider the overall effect of warm rainfall, i.e., rain from clouds with top temperatures warmer than 273 K (Feidas & Giannakos, 2010; Thies et al., 2008d). However, for night-time applications and applications with rain areas derived from thermal infrared satellite data, e.g., (Feidas & Giannakos, 2010), warm rains may further impact the error metrics. For instance, the MSG technique may underestimate the actual mean rainfall estimates for warm rains that are not detected by the satellite-based rain area information.

The study's results may have many implications. For instance, the successful wet−dry classification and baseline level estimation by this MSG technique indicates that it may be applied when the conventional technique is limited. In particular, since their estimation is independent of the MWL RSL data and instantaneous from the satellite data, it may not be affected by the MWL network

density or sampling frequency of the MWL RSL data, which are common to the conventional technique. Additionally, wpl could estimate mean rainfall intensities more reflective of the actual mean rainfall intensities. This fact indicates that spatial variability of rainfall along the MWL (as shown in Section 4.3.1) may be essential information to consider in the MWL rainfall estimation. More detailed information on the rainfall process and type (such as convective, stratiform) from the satellite, e.g., (Thies et al., 2008c; Thies et al., 2008d), may also inform the a and b parameters in Section 4.3.2, because these parameters may differ according to the rainfall type, e.g., convective rainfall (Olsen et al., 1978), and may improve the MWL rainfall estimates. An overall implication of the success of this MSG technique based on this study's results is that the MWL MSG synergy may be beneficial for large-scale rainfall estimation and monitoring or complement existing techniques.

Despite the new MSG technique's robustness and accurate rainfall estimates, further studies are needed, for instance, using multiple MWL. Additionally, wpl needs further investigation, especially for wpl less than the width of the MSG pixel (i.e., ~3 km). Furthermore, from a scientific research perspective, it is necessary to estimate and validate wpl using other remote sensing systems such as weather radars. In fact, radars may provide better estimates since they can provide rainfall locations more accurate than those derived using cloud-top information from geostationary satellites such as MSG. These questions constitute our future research.

4.5 Conclusions

A new MSG technique for MWL rainfall estimation was presented, which incorporates rain area detections by MSG satellite for the MWL rainfall retrievals. Based on the presented results, the technique is robust and is capable of wet-dry, baseline, and gauge comparable rainfall intensity estimates to benefit many operational and research applications. Specifically, the results confirm the capability of the MWL to estimate accurate mean rainfall estimates that had occurred over the MWL propagation path, especially when the rainfall occurred over the entire MWL path. However, when rainfall is spatially varying over the MWL (often the case for high-intensity convective rainfall), the results suggest that the MSG technique may approximate the actual mean rainfall better than the conventional technique.

Nevertheless, the technique is limited because its accuracy is linked to the spatial information on rain areas provided by the MSG satellite data. Nonetheless, this limitation may be somewhat prevented using more accurate information from

remote sensors such as weather radars. The study's results are from a single 15 GHz, 9.88 km MWL with two months of gauge, MWL, and satellite data. Despite this limitation, the results improve the MWL rainfall estimation, specifically from a spatially varying rainfall occurrence perspective. Additionally, it is shown that the satellite information is capable of wet-dry and baseline level estimation, which may benefit the large-scale application of the MWL and satellite for rainfall retrievals.

Overall, the new MSG technique may largely contribute to rainfall estimation and monitoring in many ungauged areas where the MWL and satellite data are readily available. In particular, its application for estimating high rainfall intensities from convective systems may benefit many applications in flash flood warnings and the nowcasting of hazardous storms.

Chapter 5: Near-Real-Time Estimation of High Spatial Resolution Rainfall from Cloud Top Properties of the MSG satellite and Commercial Microwave Link Rainfall Intensities

This chapter is originally published in:
Kumah, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., & Su, Z. (2022). Near real-time estimation of high spatiotemporal resolution rainfall from cloud top properties of the MSG satellite and commercial microwave link rainfall intensities. Atmospheric Research, 279. doi:10.1016/j.atmosres.2022.106357

Abstract

High spatiotemporal resolution rainfall is needed in predicting flash floods, local climate impact studies and agriculture management. Rainfall estimation techniques like satellites and commercial microwave links (MWL) rainfall estimation have independently made significant advancements in high spatiotemporal resolution rainfall estimation. However, their combination for rainfall estimation has received little attention, while it could benefit many applications in ungauged areas. This study investigated the usability of the random forest (RF) algorithm trained with MWL rainfall and Meteosat Second Generation (MSG) based cloud top properties for estimating high spatiotemporal resolution rainfall in the sparsely gauged Kenyan Rift Valley. Our approach retrieved cloud top properties for use as predictor variables from rain areas estimated from the MSG data and estimated path average rainfall intensities from the MWL to serve as the target variable. We trained and validated the RF algorithm using parameters derived through optimal parameter tuning. The RF rainfall intensity estimates were compared with gauge, MWL, GPM IMERG and EUMETSAT MPE to evaluate its rainfall intensities from point and spatial perspectives. The results can be described as good, considering they were achieved in near real-time, pointing towards a promising rainfall estimation alternative based on the RF algorithm applied to MWL and MSG data. The applicative benefits of this technique could be huge, considering that many ungauged areas have a growing MWL network and MSG and, in the future, Meteosat Third Generation (MTG) coverage.

5.1 Introduction

Understanding the hydrologic and energy cycles to enhance our meteorological and hydrological monitoring capabilities, predict flash floods, manage water resources and make agricultural decisions at a farm-scale level require high spatiotemporal resolution rainfall information, including its distribution and quantity. However, rainfall's intricate characteristics, such as high spatiotemporal variability, hinder accurate spatial rainfall retrieval from prevailing techniques (Hu et al., 2019).

Spatial interpolation techniques such as deterministic, geostatistical and multiple regression have been widely used to retrieve the spatial state of rainfall from gauge rainfall data (Hu et al., 2019; Ly et al., 2013). However, rain gauges are often sparsely distributed, and the accuracy of these methods depends on the density and spacing of rain gauges. Even if one could install a spatially dense gauge network with extensive coverage that can accurately capture the spatial characteristics of rainfall, such a task will be expensive to install and maintain. Besides, the gauge provides point rainfall information that may not spatially represent the entire rainfall field (Gyasi-Agyei, 2020; Yan et al., 2021).

MWLs used by commercial telecom service providers for data transmission can estimate rainfall (Leijnse et al., 2007a; Messer et al., 2006). Following a successful demonstration of such a unique rainfall retrieval technique, some studies have utilised the MWL for spatial rainfall retrieval and demonstrated the potential of using the globally spread MWL system for rainfall mapping (Messer et al., 2008; Overeem et al., 2016b; Silver et al., 2021). Nonetheless, various factors may limit accurate spatial rainfall estimation from the MWL. The accuracy of the MWL's rainfall estimates is affected by the variation of raindrop size distribution along the MWL path and the fact that the MWL antenna wetting during and after rainfall introduces additional uncertainties to the MWL signal. Furthermore, the MWL's network is arbitrary, and the density is often biased towards more developed countries and urban areas, affecting retrieval accuracies in underdeveloped countries and rural areas (Kumah et al., 2021; Zinevich et al., 2008).

Additionally, remote sensing systems such as weather radars and satellites provide spatially continuous rainfall information and have been a valuable source of spatial rainfall information for operational and research applications. The weather radars estimate spatial rainfall from backscattered radar power from precipitation particles, typically using low frequency (S or C band) high power

radar systems (Michaelides et al., 2009). Nonetheless, radars cannot be installed everywhere, e.g. over oceans and topographically complex regions. Also, various error sources, including uncertainties in the backscattering-rainfall (Z-R) relationship, beam overshoot and range effects, and vertical profile reflectivity, limit the radar estimates' accuracy (Uijlenhoet & Berne, 2008; Yan et al., 2021).

Satellites are spaceborne in low earth or GEO orbit, and their rainfall estimates have extensive coverage that fills the spatial rainfall information gap. In particular, the GEO satellite-based spatial rainfall information retrieval has been the focus of many studies due to its high spatial and temporal resolution that permits the study of sudden and intense rainfall with thunderstorms from convective systems. Notably, retrieval from the MSG satellite has received significant attention because of its high temporal resolution and wide spectral range consisting of different channels that infer cloud top properties and rainfall. Most MSG-based retrievals use multispectral data to infer optical and microphysical cloud top properties such as cloud top optical thickness and effective radius for rainfall detection and estimation (Bendix et al., 2010; Roebeling & Holleman, 2009; Thies et al., 2008d). Other retrieval techniques relate the MSG's spectral features to cloud top properties and rainfall (Feidas & Giannakos, 2010; Kingsley et al., 2021).

A parametric approach that relates the cloud top properties to rainfall is at the core of these retrieval techniques. Typically, their application requires a definition of parametric tests and underlying conceptual models. The advantage is that their application is straightforward, requiring few input variables, and they directly map the conceptual knowledge of the rain generation process onto the retrieval using the satellite data as proxies (Kingsley et al., 2021). In contrast, the non-linear and complex relation between cloud top property and rainfall may be beyond the skill of parametric tests and conceptual models (Kühnlein et al., 2014b).

In this regard, machine learning algorithms that rely on data-driven analysis to explore the relationship between variables and have strong capabilities in dealing with non-linear relations may be suitable for retrieving rainfall from the multivariate satellite data to overcome the limitations of the parametric techniques (Hu et al., 2019; Kühnlein et al., 2014a; Kühnlein et al., 2014b). Several studies have successfully used machine learning algorithms such as RF, NNET, SVM and deep-learning models for spatial rainfall estimation (Kühnlein et al., 2014a; Lazri et al., 2014; Meyer et al., 2016; Moraux et al., 2019). In particular, the RF machine (Breiman, 2001) learning algorithm has gained significant attention. It is an ensemble classification and regression algorithm

that assumes that a whole set of trees can make more accurate predictions than a single tree or network. The RF algorithm has many features that suit its application for rainfall retrievals. For instance, it efficiently handles large datasets and can capture non-linear relations between predictor and target variables (Kühnlein et al., 2014a). However, most of its applications to MSG data, such as (Kühnlein et al., 2014a; Kühnlein et al., 2014b; Meyer et al., 2016), used gauge-adjusted radar data as the training target, which may be sparsely distributed or non-existent depending on the study area. To the best of our knowledge, no study has applied the RF algorithm to MSG data and used MWL-based rainfall as the training target, while the application could be beneficial to areas with insufficient ground data but with a growing MWL network.

Therefore, the objective of this chapter is to evaluate the usefulness of the RF algorithm trained with MWL-based rainfall intensities for estimating high spatiotemporal resolution rainfall from cloud top properties of the MSG satellite. Compared to existing studies, this study's uniqueness is due to the following reasons:

1) This study applied the RF algorithm for rainfall estimation in a topographically complex area in the Kenya Rift Valley, where gauge data is sparsely distributed
2) This study trained the RF algorithm using MWL rainfall as the target variable for the first time.

5.2 Study Area and Dataset

Figure 5. 1 shows the study area using ALOS World 3D 30 m (AW3D30) DEM (Caglar et al., 2018) to visualize the area's location within the Kenyan Rift Valley. The area's temperature ranges between 8 and 30 °C. It experiences a bimodal rainfall pattern influenced by the passage of the ITCZ over Kenya. There is a long rainy season from March to June and a shorter rainy season from October to December. Additionally, rainfall varies noticeably with relief features, with the total annual rainfall of the low and high altitudes varying between 610 to 1525 mm (Odongo et al., 2015), respectively.

This study's evaluation period was during the long rain period of 2014, 2018 and 2019, which constitute the periods we had consistent and collocated ground and satellite data. For the 2018 and 2019 periods, gauge rainfall data from TAHMO were available as 5 min rainfall accumulations. These computed the 15-minutes rainfall intensities that served as the ground truth in this study. The TAHMO gauges are shown as white triangles (labelled by the station codes provided by

TAHMO) in Figure 5. 1 and illustrate a sparse distribution of ground data in the study area.

Safaricom provided the RSL data for the set of MWL with arbitrary geometry, frequency and length in the study area, as shown in Figure 5. 1 For the 2014 and 2019 periods, data from multiple MWL were available. In contrast, for the 2018 period, only a single 15 GHz MWL data was available. These MWL are Aviat Eclipse MWL, vertically polarised, and has a constant transmitted signal level (TSL). Their RSL was characterised by minimum, maximum, and mean values at 15 minute intervals and a 0.1 dBm resolution.

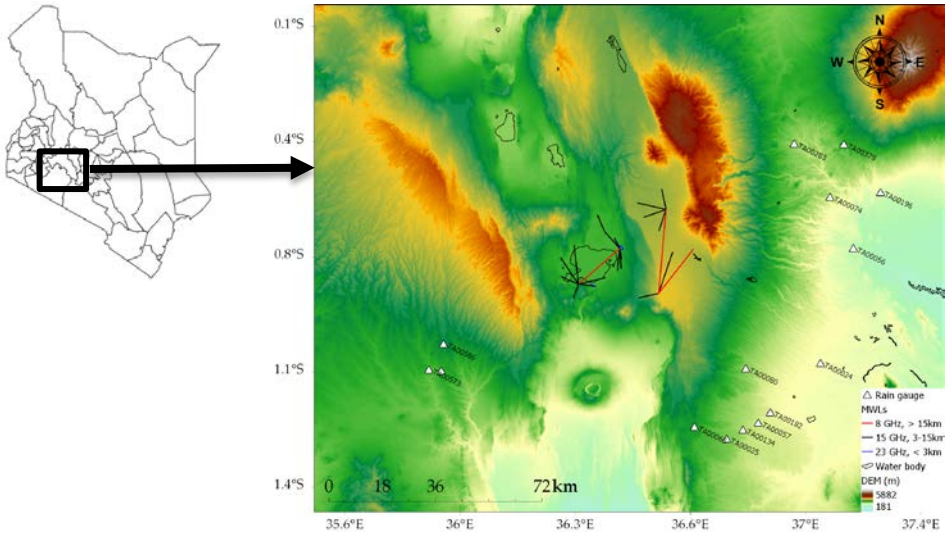

Figure 5. 1 Study area in Kenya, locations of rain gauges and topology of MWL shown using ALOS DEM as a base map

The IR (IR10.8 µm and IR12.0 µm) and water vapour (WV6.2 µm and WV7.3 µm) channels used in this study were from the SEVIRI radiometer onboard the Meteosat at 0^0 (2014 period) and 41.5^o E (2018 2019 period). This corresponded to Meteosat 10 and 8 satellites, respectively (EUMETSAT, 2016), when the data was acquired from (EUMETSAT, 2020) at 3×3 km and 15 min spatial and temporal resolution. This spatial resolution is preserved over the study area. These channels are sensitive to cloud top properties such as cloud top temperature and height. The data from the Meteosat at 0^o were parallax corrected because of the satellite viewing angle, which could cause displacement in the actual position of cloud tops depending on their location and height (Kumah et al., 2020; Roebeling & Holleman, 2009).

The IMERG final run version 6 (V06B) and EUMETSAT MPE rainfall products verified this study's retrieved rainfall spatially. The MPE is a near real-time rainfall product derived for each repeat cycle of Meteosat 7, 8 and 9 satellites from the thermal IR channel. The MPE algorithm relies on a weather-dependent monotonic function that relates the IR brightness temperatures to the PMW SSM/I rain rates. For this reason, MPE continuously adjusts the retrieval function geographically and temporarily, using the PMW rain rates as calibration values. The retrieval function is based on the histogram matching technique derived from collocated IR images and PMW data accumulated over up to 12 hours and in 5^o x 5^o geographical boxes to account for the poor spatial coverage of the PMW measurement. The MPE rainfall product is most suitable for convective rainfall because the monotonic function assumes that colder clouds produce more rain than warm clouds (Heinemann & Kerényi, 2003). This study retrieved MPE data from EUMETSAT (2020) at 15 minutes and 3×3 km resolution for the evaluation period.

The description of the IMERG precipitationCal data in this study is given in Chapter 3. This study used the precipitationCal data because it is a research-grade product that is climatologically adjusted using ground data from the GPCC. Moreover, previous studies in the study area (Kingsley et al., 2021) had found good agreement when they compared the data with MSG-based rain areas and ground data. IMERG data can be retrieved from https://gpm.nasa.gov/data/imerg (accessed on 16th February 2022) at approximately $0.1^0 \times 0.1^0$ and 30 minutes resolution.

5.3 Method

5.3.1. General methodology of the rainfall retrieval

This study retrieved high spatiotemporal resolution rainfall intensities from MSG satellite data using the RF algorithm trained with MWL rainfall intensity estimates. Figure 5. 2 shows a flow chart of the retrieval procedure, comprising the three steps:

(1) initial detection of raining areas
(2) estimating MWL rainfall intensities
(3) estimating the rainfall intensity of the detected raining areas step 1.

In this study, steps 1 and 2 were based on techniques described in previous studies (Kingsley et al., 2021; Kumah et al., 2020, 2021) and in chapters 2, 3 and 4 of this dissertation. Step 3 was realised by using the RF algorithm. These steps

agree with the optical rainfall retrieval approach that separates rain area detection from the rain rate estimation (Bendix et al., 2010).

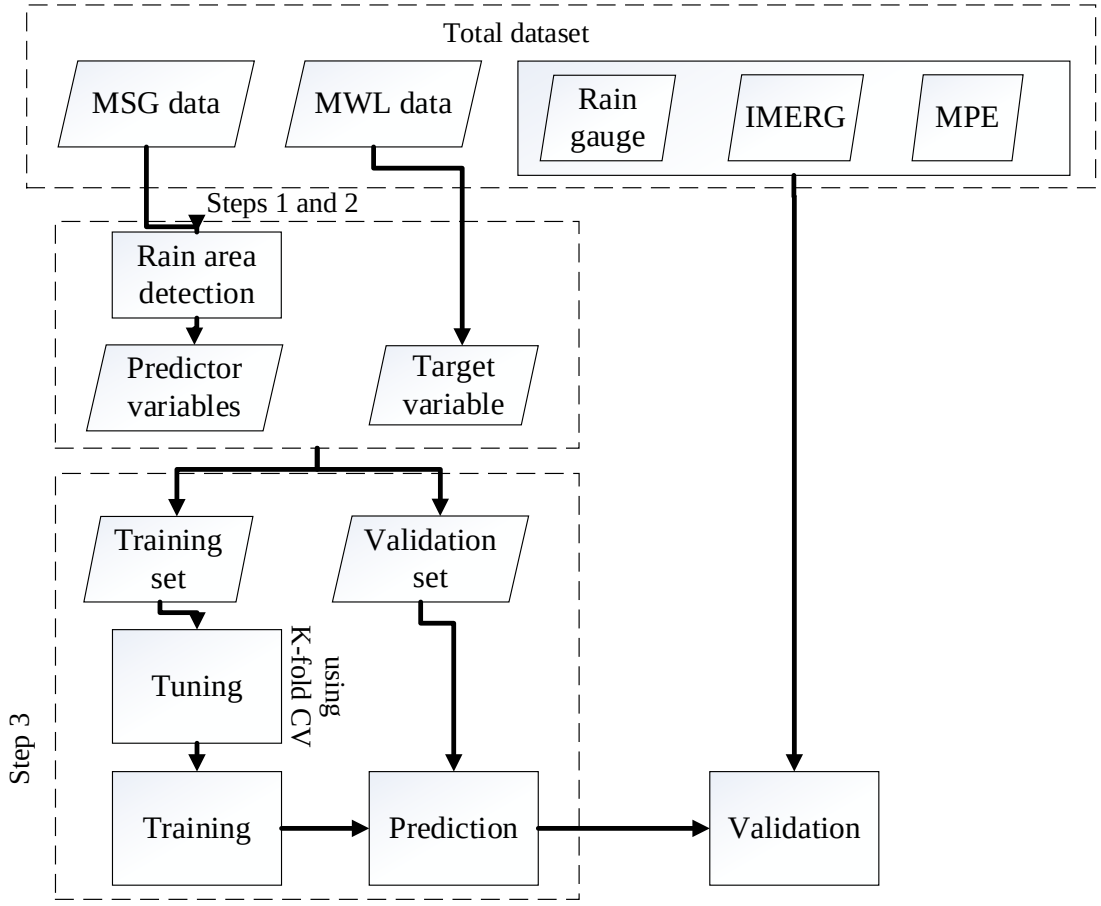

Figure 5. 2 Flow chart of this study's rainfall retrieval procedure

5.3.1.1 Detecting rain areas

The rain area identification technique was based on the approach described in Chapter 3. It relies on a parametric threshold model based on the concept that clouds with high cloud top optical thickness and height have high rain probabilities and intensities and vice versa. The basis of this conceptual model is rooted in the characteristics of raining clouds provided by Lensky and Rosenfeld (2003a). The rain detection model uses differences in brightness temperature of the thermal IR and water vapour channels such as IR10.8–IR12.0 K and IR10.8–WV6.2 K BTD from IR10.8 µm, IR12.0 µm and WV6.2 µm SEVIRI channels to

infer the cloud top optical thickness and height properties. It applies a threshold to a 2-D space defined by these BTDs, assuming that a cloud is more likely to rain if the parameter (i.e. the BTD) is below the threshold value. The threshold values were determined by calibrating and validating the detection model using gauge rainfall and satellite data. Subsequently, a gradient-based adaptive correction technique reduces the number and sizes of the detected rain areas by using rain area-specific parameters.

5.3.1.2 Estimating rainfall intensities from the MWL data

This section presents this study's MWL rainfall estimation method, which summarises the conventional approach previously described in Chapters 2 and 4 and in (Kingsley et al., 2021; Kumah et al., 2020). The approach estimates rainfall intensities from the mean RSL data by first classifying the data into wet and dry periods using a rolling window statistical technique. Next, a baseline level is estimated as the median of the mean RSL of the previous 24 hours labelled as dry periods by the wet and dry classification step. Finally, the mean RSL data is corrected for the effect of the wet antenna (Schleiss et al., 2013) before retrieving attenuation and rainfall from equations 5.01 and 5.02, respectively.

$$A = \frac{B-P}{L} \qquad 5.01$$

$$R = \left(\frac{\mathrm{A}}{a}\right)^{\frac{1}{b}} \qquad 5.02$$

where:

A (dB/km)—is the rain-induced specific attenuation averaged over the entire MWL

L—is the MWL length, and B, P are the baseline and the mean RSL, corrected for the effect of antenna wetting by using a dynamic model by Schleiss et al. (2013).

R—is the MWL rainfall intensity, a(0.05008,0.1284) and b(1.0440,0.9630) values were from (ITU, 2005b) for the 15,23 GHz MWL respectively.

5.3.1.3. Estimating spatial rainfall intensities using RF

i. The predictor variables

Based on conceptual ideas used by optical rainfall retrieval models in the last decades, optical cloud properties most relevant to rain areas and rain rates are

cloud top temperature, height, and cloud water path (represented by the cloud optical thickness and particle effective radius). Retrieval techniques such as those that use only the cloud top temperature often consider the cloud top temperature to indicate the cloud top height and assume that cold clouds produce (more) rainfall (Arkin & Meisner, 1987). Though this worked for convective clouds, the technique considered cold non-raining cirrus clouds as raining or missed rainfall from the relatively lower warm clouds. The cloud water path retrievals, e.g. (Bendix et al., 2010; Thies et al., 2008d), assume that rain clouds have high optical thickness and effective radius with extended tops.

This study utilised two kinds of information as predictor variables: (1) spectral features and (2) gradient features, summarised in Table 5. 1. The spectral features were derived from SEVIRI channels and differences. They are consistent with those used by previous studies (Kingsley et al., 2021; Kühnlein et al., 2014a; Kühnlein et al., 2014b) to infer cloud top properties such as cloud top temperature, height, optical thickness and particle effective radius, for rain area and rain rate retrievals.

Table 5. 1 The predictor variables used for rainfall retrieval

Spectral features		Pixel gradient features
Cloud top properties	Channels and channels differences	
Cloud top temperature	IR10.8 K	ΔIR10.8 K
Cloud top height	IR10.8−WV6.2 K	ΔIR10.8−WV6.2 K
	IR12.0−WV7.3 K	ΔIR10.8−IR12.0 K
Cloud height	WV6.2−WV7.3 K	ΔWV6.2−WV7.3 K
Cloud optical thickness	IR10.8−IR12.0 K	ΔIR12.0−WV7.3 K

The gradient features indicate pixel gradients in the cloud top properties. This was computed based on past study's (Kingsley et al., 2021) method described in Chapter 3. Previous studies used gradient features in satellite rainfall retrievals (Gao et al., 2004; Li et al., 2021). The reason for including gradient features as a predictor for retrieving rainfall is that different raining cloud types, such as convective and stratiform clouds, have distinguishable characteristics, such as temperature gradient and local pixel temperature variations with corresponding rain rates. For instance, fully grown convective clouds have overshooting tops with high temperature gradients indicating the convective core, characterized by high rainfall intensities. By contrast, stratiform clouds exhibit gradual temperature gradients and low pixel temperature variations with relatively low rainfall intensities. The gradient feature measures the cloud patch average pixel gradient to determine these distinct characteristics to improve the retrieved rainfall estimate.

ii. Compiling training and validation datasets

This study utilised common machine learning techniques consisting of training, validation and testing datasets to develop and test the rainfall retrieval method. The training set was used to train the model by optimising its learning parameters, whereas the validation and testing set assessed the model's ability to generalize well to unseen data.

The training dataset consisted of target and predictor variables sampled from mixed space-time observations from the study area during the 2014 and 2019 periods. More precisely, they were retrieved from multiple MWLs and the corresponding MSG pixels covering the MWL, shown in Figure 5. 3 for the raining ($R > 1$ mmh^{-1}) and non-raining ($R < 1$ mmh^{-1}) periods. For the MWL with transmission paths covered by multiple MSG pixels, the mean of the satellite data estimated from these pixels was retrieved for estimating the average rainfall of a pixel to allow a fair comparison with other satellite rainfall estimates used by this study. Besides, unlike, e.g., the minimum or median values, the mean value of the satellite data considers neighbouring pixel information.

The validation dataset was from the 2018 and 2019 periods. It consisted of all MSG pixels in the study area, assuming that the rainfall and MSG-based cloud top properties would not change much for the small study area we considered (see Figure 5. 1). For the 2019 period, this excludes data from those pixels covering individual MWL since they were used to train the RF model. The data from the 2018 and 2019 periods validated the RF model because they coincided with the periods when gauge, independent MWL, MPE and IMERG data were available in the study area, thereby allowing for a thorough validation of the RF model against different rainfall estimation techniques.

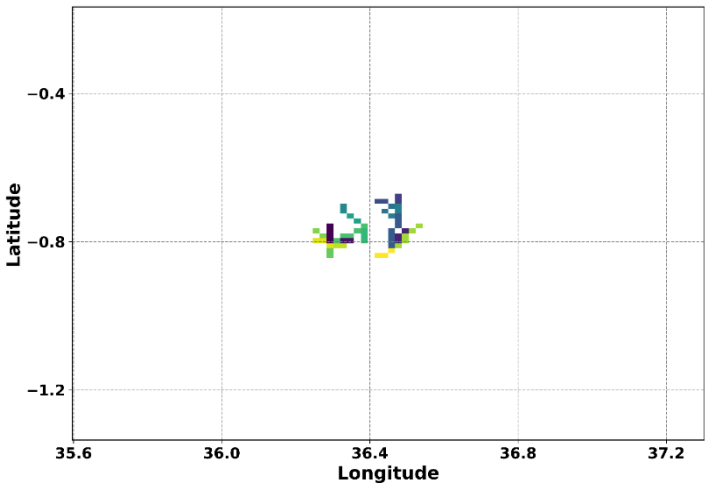

Figure 5. 3 The MSG pixels covering multiple MWL that trained the RF model

iii. The RF regression model and parameter tuning

This study applied the RF for rainfall retrieval based on its advantages and good performance for rainfall estimation (Kühnlein et al., 2014a; Kühnlein et al., 2014b; Wolfensberger et al., 2021). Besides, Meyer et al. (2016) investigated the performance of several machine learning algorithms, including the RF, for rainfall retrieval and found no single algorithm performed considerably better than the other. They concluded that finding more suitable satellite-based predictor variables is more necessary than optimization through the choice of the machine learning algorithm.

The RF is an ensemble approach used for classification and regression purposes. It is based on the idea that the outcome of a group of weak learners (i.e. decision trees) when combined with a voting scheme, can yield an improved estimate with better performance (Breiman, 2001). RF uses bootstrap sampling and random feature selection to ensure the heterogeneity of these weak learners. Assuming an input dataset with $N \times M$ dimensions (where N and M are the numbers of samples and input features, respectively), RF grows each tree in the forest using bootstrap samples (randomly selected, with replacement, samples from N). About two-thirds of the sample is used to grow the decision tree for each bootstrap, while the remaining one-third is not included in the learning sample. This out-of-bag (OOB) sample is later used to get an unbiased estimate of the generalised error and to estimate the importance of the variables used in constructing the tree.

When growing trees, only a number of m features (where m < M) are used in deciding the best split at each node of a tree, and features with the lowest residual sum of squares are chosen for the split. The process is repeated through parallel processing until several trees are grown. For RF regression, the final estimate is the average of all outcomes of all trees in the forest (Wolfensberger et al., 2021). This study implemented the RF regression model in Python 3.7.3 using the scikit-learn package (Pedregosa et al., 2011). There are over a dozen parameters to adjust in this package to achieve a robust RF performance. However, this study focused on the number of decision trees ($n\_estimators$) and the number of input features to consider when looking for the best split ($max\_features$), following previous study's account (Turini et al., 2021).

Since RF may perform poorly for the highly imbalanced dataset (Liu et al., 2006), the imbalance between the raining (representing 8% of the dataset – the minority class) and the non-raining (representing 92% of the dataset – the majority class) dataset was considered before assessing the optimal values of the RF parameters. Balancing class distribution may be achieved by oversampling the minority class or downsampling the majority class. The latter was a better strategy for our dataset because of the comparatively low percentage of the raining class. Besides, Liu et al. (2006) showed that downsampling the majority class is a better class balancing strategy. Therefore, this study addressed the imbalance in the dataset by keeping all the data from the minority class and randomly sampling (without replacement) several observations (less than the original) from the majority class.

This study searched for optimal parameter values by performing a stratified 5-fold-cross-validation on several tuning values. Stratified 5-fold-cross-validation randomly splits the training samples into 5 equal-sized folds regarding the distribution of the target variable. In effect, each (1/5) fold has a similar target variable distribution as the training sample. Then, models were fitted while repeatedly leaving one fold out to evaluate the model's performance using the mean squared error (MSE) metric in equation 5.01. The model performance for the respective tuning values is the average of the MSEs from the hold-one-out iterations.

$$MSE = \frac{1}{N} \sum_{i=1}^{N-1} \left(R_{rfi} - R_{ti}\right)^2 \qquad 5.01$$

where:

R_{rfi} —represents all possible RF rainfall intensity estimates

R_{ti} —represents all possible target variable observations, and N is the number of samples.

An important parameter is the total number of trees, *n_estimators*, to grow in the forest. According to Breiman (2001), the generalization error converges as the number of trees increases. Increasing the number of trees in the forest does not result in over-adjustment, except this increases the computational time. In essence, *n_estimators* should be optimized to obtained a computationally feasible value. To determine the optimal value of the n_estimators parameter, many RF models were created using the training data for all possible values of n_estimators and max_features. The maximum n_estimators were 2000 trees, whereas the max_features values ranged from 3 to 9 representing 30% to 90% of the total number of input features.

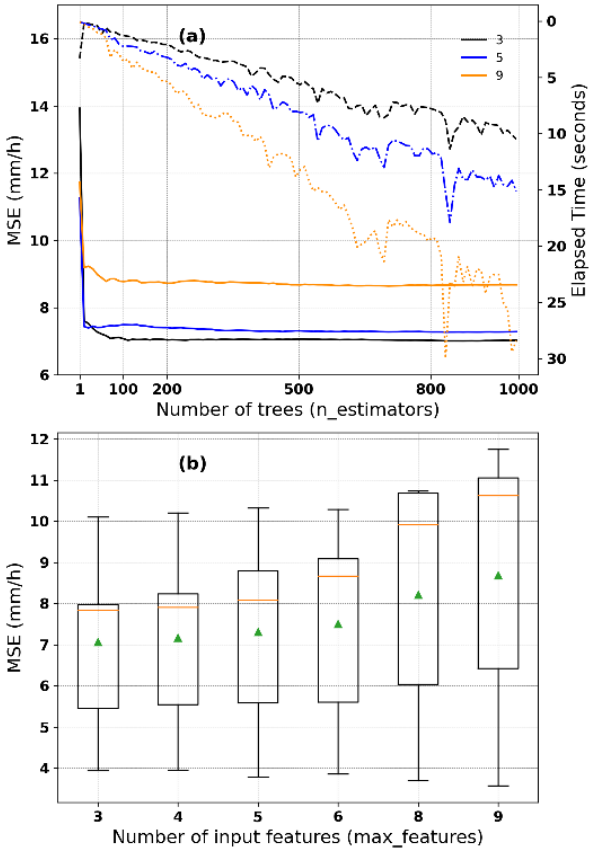

Figure 5. 4 The RF parameter tuning. Effect of (a) number of decision trees with 3, 5, and 9 input features on rainfall retrieval errors and computational time and (b) the number of input features on rainfall intensity retrieval errors. In (b), boxes show the first quartile, median (orange lines), and third quartile; whiskers (lines outside the box) extend from the minimum to the first quartile and from the third quartile to the maximum, and the average MSE is shown as green triangles.

Figure 5. 4a exemplifies how the number of trees with 3, 5, and 9 max_feature values affects rainfall intensity retrieval errors and computational time. Based on the dataset, the figure shows that increasing the number of decision trees and input features increases the computational time. Nonetheless, regardless of the number of input features, the rainfall intensity retrieval errors decrease rapidly with an increase in the number of decision trees until approximately 100 trees, where the error rate stabilises. This suggests that more than 100 trees in the forest can be considered sufficient for a robust RF model performance. Thus, this study set the *n_estimators* to 100, with a reasonable computational time of about 2 seconds.

Breiman (2001) shows that the RF error rate largely depends on the correlation between any two trees and the strength of individual trees in the forest. Increasing the correlation increases the RF error rate, whereas increasing the strength of individual trees decreases the RF error rate. The *max_features* parameter affects these two aspects such that reducing it reduces both the correlation and strength, whereas increasing it increases both. In practice, the *max_features* value is often treated as a tuning parameter (Kühnlein et al., 2014a; Kühnlein et al., 2014b). To determine the optimal max_features value, many models were created using the training data for different possible max_features values ranging from 3 to 9, representing 30% to 90% of the number of input features while setting the n_estimators parameter to 100 in each scenario. Figure 5. 4b presents the descriptive statistics of the MSE of rainfall intensities based on the different *max_features* values. Based on these results, *max_features* = 3 was used because this leads to low rainfall intensity errors.

iv. RF model prediction and validation

The tuned model parameter values were used to train the RF regression model, and the trained model predicted rainfall intensities of the validating MSG pixels. The mean absolute error (MAE) (Wilks, 2006) described in equation 5.02 evaluated the RF model performance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|R_{rfi} - R_{oi}| \qquad 5.02$$

$$RB = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(|R_{rfi} - R_{oi}\right)}{\frac{1}{N}\sum_{i=1}^{N}R_{oi}} \qquad 5.02$$

where:

R_{rfi} —represents all possible RF rainfall intensity estimates

R_{oi} —represents all possible gauge and MWL rainfall intensity estimates, and N is the number of samples.

The validation approach was by:

1) comparing the rainfall intensity estimates by the RF to gauge, IMERG, and MPE pixel to validate the RF model and evaluate its capability to estimate rainfall comparable to already existing rainfall estimation techniques. For this, we used the point-to-pixel approach to compare the gauge's rainfall estimate to the RF, IMERG and MPE estimates. The approach assumes that the gauge's estimate is representative of the RF, IMERG and MPE pixel being compared to;
2) comparing averaged RF rainfall estimates from pixels covering the MWL to the MWL's rainfall estimates to assess the RF's capability of path average rainfall intensities estimation
3) spatially comparing the RF model estimates to those of the MPE and IMERG rainfall products to evaluate the RF model against existing satellite rainfall products

Since this study focused on evaluating the RF model's usability for high spatiotemporal resolution rainfall retrieval, the validation was done at 30 minutes and 3×3 km resolution. Also, to ensure a comparison of collocated rainfall intensity estimates, the spatial and temporal mismatch in the dataset was considered. For this, the IMERG estimates were spatially resampled, using the nearest neighbourhood technique that preserves the pixel values, to the spatial resolution of the RF and MPE. On the other hand, the gauge, RF and MPE estimates were temporally aggregated to IMERG's 30 minutes temporal resolution by summing their respective rainfall intensity estimates.

5.4 Results and Discussion

5.4.1 Results

5.4.1.1 Comparing rainfall intensity estimates at a pixel by the RF, MPE, IMERG and gauge

This section evaluates the RF rainfall intensity estimates at a pixel using MWL, gauge, MPE, and IMERG estimates. Firstly, a point evaluation is presented through visual and statistical analysis of the RF estimates compared to gauge, MPE, and IMERG for rainfall events observed from two different locations in the study area. Secondly, a performance evaluation of the RF estimates against gauge estimates compared to MPE, and IMERG estimates are presented. Thirdly, the

probability density of all rainfall intensities observed by the gauge, RF, IMERG, and MPE from the gauge pixel is presented.

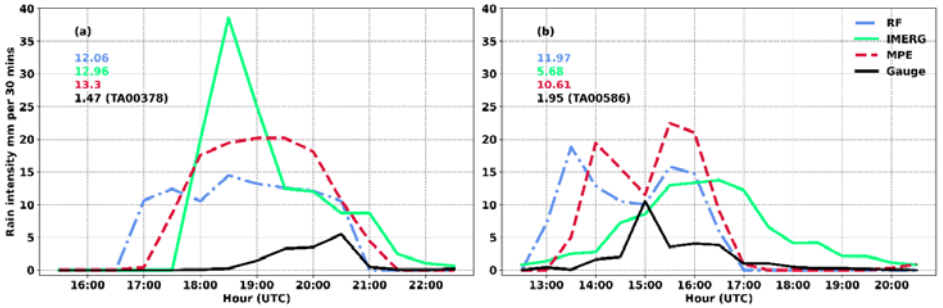

Figure 5. 5 Comparing rainfall intensity estimates by the RF, MPE, IMERG, and rain gauge. Coloured figures are the mean intensity estimates (excluding R = 0 mmh^{-1}) of RF, IMERG, MPE and gauge rainfall events, respectively. The names in brackets are the gauges that computed the mean values.

Figure 5. 5 presents rainfall intensities of two rainfall events captured by the RF, MPE, IMERG and rain gauges. The gauge estimates are from gauges TA00378 (Figure 5. 5a) and TA00586 (Figure 5. 5b), shown in brackets in the figure, situated at different locations within the study area. The RF, MPE and IMERG are estimates retrieved from the pixels containing the two gauges. Figure 5. 5a shows rainfall events that occurred on 8 May 2018. It is clear from the figure that, although all the rainfall retrieval techniques captured the rainfall events that occurred between the hours of 16:00 to 21:00 UTC, the characteristics of their rainfall events differ. For instance, the RF, MPE and IMERG captured more rainfall than the gauge, which is also evident from the mean rainfall computed for the event. Moreover, the peak rainfall intensity captured by IMERG was above 30 mm per 30-minutes intervals. In contrast, the RF and MPE were comparable and below 30 mm per 30-minutes, compared to the gauge's peak rainfall intensity below 10 mm per 30-minutes interval.

The rainfall events in Figure 5. 5b occurred between 13:00 to 18:00 UTC (based on the gauge observation) on 4 April 2019. The figure shows that the gauge, RF, MPE, and IMERG captured the rainfall event with fairly differing characteristics. On average, the RF observed the most rainfall, followed by MPE, IMERG and gauge, as shown by the mean rainfall intensity of the rainfall event. In addition, the RF and MPE captured two comparable peaks below 25 mm per 30-minutes. However, IMERG's event extends beyond 18:00 UTC and its peak rain intensity, like the gauge, was below 15 mm per 30-minutes.

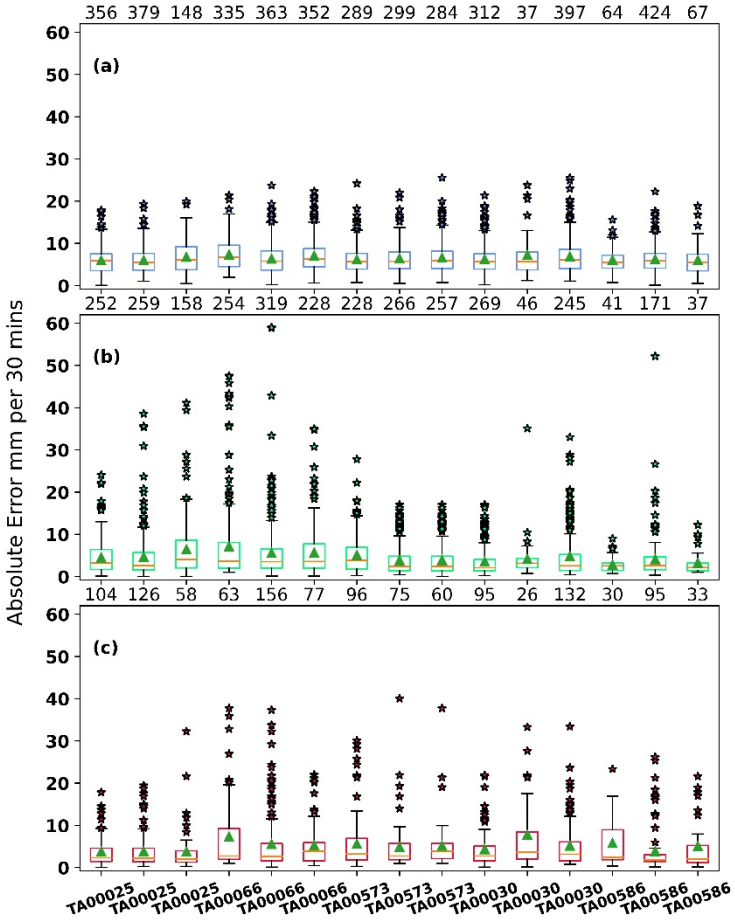

Figure 5. 6 Performance evaluation of the RF compared with MPE and IMERG. Box and whisker plot showing descriptive statistics of the absolute error of (a) RF versus gauge, (b) IMERG versus gauge, and (c) MPE versus gauge. Boxes show the first quartile, median (orange lines), and third quartile; whiskers (lines outside the box) extend from the minimum to the first quartile and from the third quartile to the maximum; stars indicate outliers; the MAE is shown as green triangles. Each plot shows the total number of 30-minutes data points at the top of the plot (excluding 0 mm) that computed the descriptive statistics. The x-axis shows the station codes provided by TAHMO

Figure 5. 6 shows the absolute error of RF versus gauge rainfall intensity estimates compared to MPE and IMERG. The data used in computing the absolute error in this figure were collocated observations by the gauge, RF, MPE

and IMERG, excluding the 0 mm estimates, during the validation period. On average, the absolute errors in RF versus gauge (Figure 5. 6a) estimates were about 5 mm per 30-minutes, comparable to those of the IMERG and MPE vs gauge estimates. Based on the average errors, the RF's rainfall estimation performance can be considered as good as IMERG (Figure 5. 6b) and MPE (Figure 5. 6c). Nonetheless, its outliers mostly below 30 mm per 30-minutes (Figure 5. 6a) compared to those of the IMERG and MPE, which mainly were below 50 mm per 30-minutes, may point to differences in their high rainfall intensity estimates.

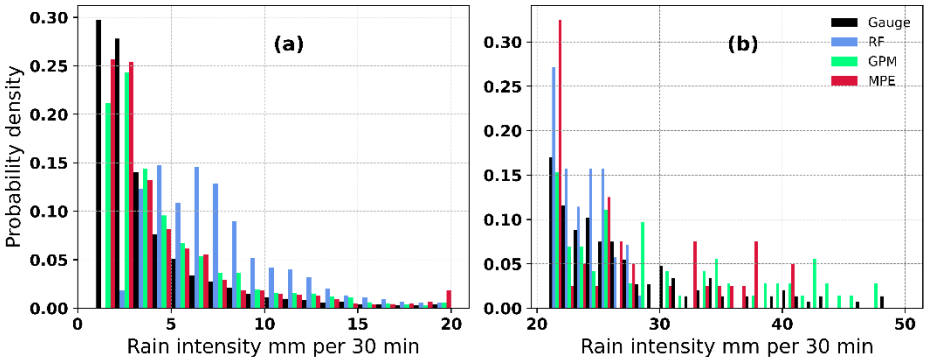

Figure 5. 7 Probability density of rainfall intensity estimates by the gauge, RF, IMERG, and MPE for (a) less than 20 mm and (b) above 20 mm

Figure 5. 7 shows the density distribution of collocated rainfall intensity estimates from the gauge, RF, IMERG and MPE. The distribution of rainfall intensities in Figure 5. 7a suggests that compared to IMERG and MPE, the RF mostly overestimates the gauge rainfall intensities below 15 mm per 30-minutes. When this distribution is compared with that in Figure 5. 7b, it is evident that the RF underestimates the gauge's rainfall intensities, judging by its estimates largely below 30 mm per 30-minutes. Nonetheless, these estimates were from sparse gauge pixels in the study area, which may not be a fair representation of the full range of the area's rainfall estimates.

The discrepancies in rainfall intensity estimates by the measurement techniques may be due to various factors. Their spatial resolution differences may explain some of these discrepancies. To be precise, the gauge observes rainfall from a single point, making it easy to miss or underestimate a high-intensity local rainfall event, depending on its proximity to a rainstorm. The RF, IMERG and MPE all estimate the average rainfall intensity of a pixel, which is spatially more extensive than the gauge's point observation and, therefore, may more likely

capture a rainfall event, albeit with intensity differences that depend on the measurement technique. Additionally, the RF's rainfall intensity estimate represents an average of predictions from all trees, which may explain why it overestimates the low (and underestimates the high) intensities (Kühnlein et al., 2014b; Wolfensberger et al., 2021).

5.4.1.2 Comparing the RF and MWL rainfall intensity estimates

We next compared the RF rainfall intensity estimates with independent MWL RSL data estimates to assess the RF's capability of path average rainfall estimates. Here, the RF's mean, median and maximum rainfall intensity over the MWL is included in the comparison to provide an idea of the range of rainfall intensities estimated by the RF and how it compares with the MWL's estimates. Table 5. *2* presents descriptive statistics of the absolute errors when comparing the RF's mean, median and maximum rainfall intensities to the MWL's estimates.

Table 5. 2 Descriptive statistics of the absolute errors of the RF and MWL rainfall intensity estimates. 25%, 50% and 75% indicate percentile levels

	Absolute errors of RF versus MWL rainfall (mmh^{-1})		
Descriptive statistics	Mean	Median	Maximum
Mean	4.1	4.0	6.8
Minimum	0.0	0.0	0.1
Maximum	18.0	21.4	23.4
25%	1.5	0.0	4.0
50%	3.9	4.0	6.5
75%	6.3	6.5	8.9

These statistics are from a total of 920 15-minutes rainfall intensity data

On average, absolute errors of the mean and median versus MWL rainfall intensities were around 4 mmh^{-1} compared to about 7 mmh^{-1} when comparing the RF's maximum to the MWL's estimates. This suggests a better agreement between the RF's mean and median and the MWL rainfall intensity values. Nonetheless, the maximum error and the 75% percentile value of the mean comparison suggest the RF's mean rainfall estimates may better agree with the MWL rainfall than the median.

The better agreement between the RF's mean and MWL rainfall intensities is because both represent average intensities over the MWL's path. However, comparatively high absolute errors of the RF's maximum versus MWL rainfall intensities are because the maximum rainfall intensities represent the highest intensities observed over the MWL's path. Furthermore, discrepancies in RF and MWL estimates that contribute to errors in Table 5. *2* may be attributed to other

factors, including differences in their rainfall retrievals. The RF's estimates are based on nonlinear relationships between MWL rainfall and cloud top properties aloft, whereas the MWL derives rainfall intensities from average rain-induced attenuation over its path.

5.4.1.3 Comparing spatial rainfall estimates by the RF model, MPE and IMERG

We finally validated the RF rainfall intensity estimate spatially by comparing it with the IMERG and MPE rainfall products on a scene-by-scene basis. First, an exemplary scene is shown from 4 April 2019 at 13:00 to visually analyse the RF, IMERG, and MPE estimate. Next, the MAE and RB were computed based on all above 0 mmh^{-1} rainfall intensity estimates by the RF, IMERG, and MPE during the validation.

Figure 5. 8 compares spatial rainfall intensity estimates by the RF to IMERG and MPE to validate the RF estimates. The white pixels in the centre of the figure are the MSG pixels over the MWL that trained the RF model. There is a good agreement in the spatial distribution of rain areas by IMERG and RF, whereas MPE shows fewer rain areas that are more localised than RF and IMERG. There are also some differences in their rainfall intensity estimates. For instance, MPE captured high rainfall intensities around latitude $0^{o,}$ which the RF and IMERG underestimated. Overall, based on a visual inspection of Figure 5. 8, it can be stated that the rain areas in the RF are comparable to IMERG but with intensities that compare better with the MPE.

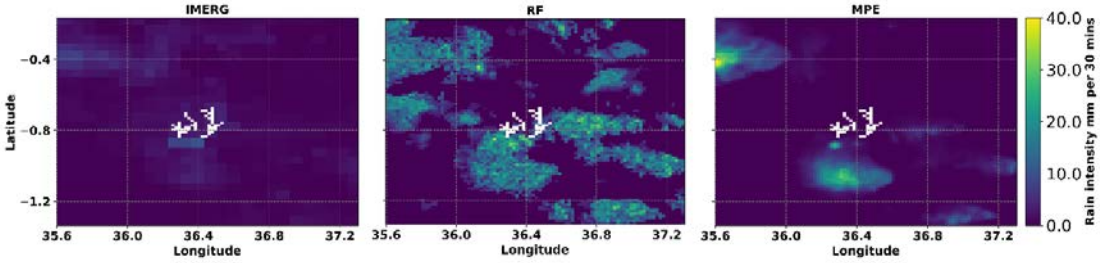

Figure 5. 8 Spatial rainfall estimates by the RF compared to IMERG and MPE

These discrepancies in rain areas and intensities in Figure 5. 8 may be attributed to measurement differences in the retrieval techniques. For instance, the MPE algorithm's design captures convective rainfall of local origin and high intensities. By contrast, the rain area detection system used by the RF is not dependent on the rain cloud type (Kingsley et al., 2021), and its rainfall intensity estimates were based on a nonlinear relationship between IR-based cloud properties aloft and ground-level rainfall. Moreover, the fact that the RF estimates represent an

average of predictions by all trees (Kühnlein et al., 2014b; Wolfensberger et al., 2021) may contribute to some of the rain area and intensity differences between the RF, IMERG and MPE. On the other hand, IMERG uses spatiotemporal average rainfall from multiple microwave rainfall estimates, which may explain its low rainfall intensities in Figure 5. 8 compared to the RF and MPE.

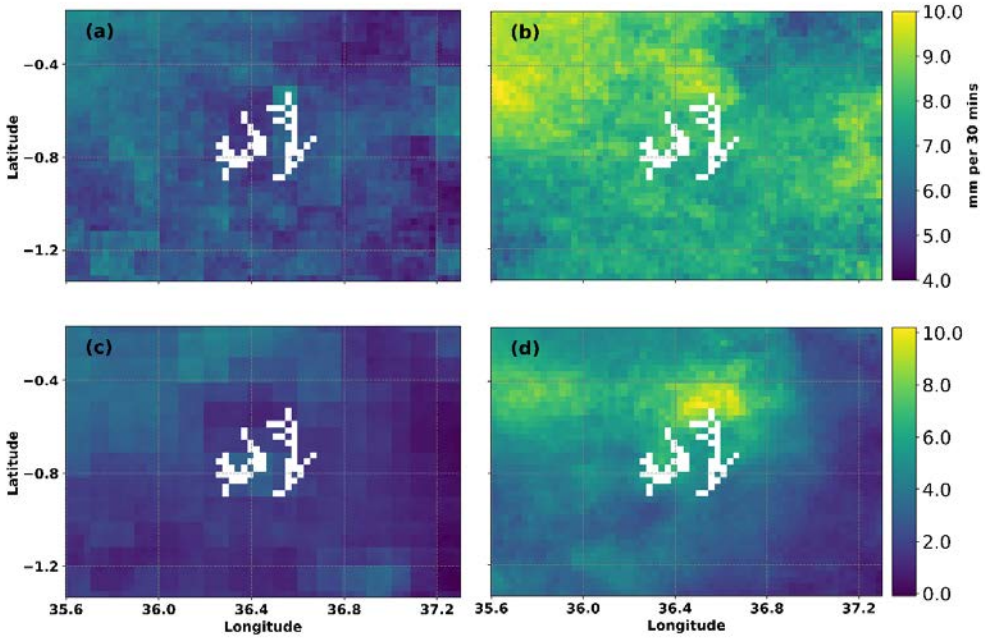

Figure 5. 9 Spatial variabilities of MAE (a,b) and RB (c,d) computed from RF and IMERG (a,c) and RF and MPE (b,d) rainfall pairs.

Figure 5. 9 shows the spatial variability of MAE (Figure 5. 9a,b) and RB (Figure 5. 9b,d) computed from RF versus IMERG (Figure 5. 9a,c) and RF versus MPE (Figure 5. 9b,d) pairs during the validation period over the study area. It is clear from the figure that the RF estimates agree better with IMERG than MPE estimates. On average, the MAE and RB computed from RF versus IMERG values were below 6 mm per 30-minutes and 3, compared to the values of RF versus MPE, which were below 8 mm per 30-minutes and 5, respectively.

Nonetheless, both IMERG and MPE show high differences compared to the RF, indicated by their respective high MAE and RB values, particularly between latitude -0.2 and -0.6, attributed to probably the high rainfall intensities observed in these areas with complex topographic features (see Figure 5. 1). On average, the RF-IMERG MAE values are below 6 mm per 30-minutes compared to the RF-MPE estimates below 8 mm per 30-minutes.

5.4.2 Discussion

The usability of the RF machine learning algorithm trained with MSG-based cloud top properties and MWL rainfall intensities for estimating high spatial and temporal resolution rainfall intensities in a topographically complex area in the Kenyan Rift Valley is investigated and evaluated. The investigation followed three major steps: (1) rain area detection based on the method described by (Kingsley et al., 2021) in Chapter 3 and retrieval of MSG-based cloud top properties that served as predictor variables, (2) rainfall estimation from MWL RSL data to serve as target variables and (3) rainfall intensity estimation using the RF algorithm. We compared the RF estimates with gauge, MWL, IMERG, and MPE estimates to evaluate the RF's rainfall intensity estimation performance.

The results based on the study area can be described as good, considering that they were achieved at a high spatial and temporal resolution of 3×3 km and 30 minutes, pointing towards a convincing skill of the RF algorithm for rainfall estimation. An analysis of rainfall events from different locations in the study area revealed the capability of the RF to estimate rainfall events in the study area with mean rainfall characteristics comparable to IMERG and MPE. Comparing rainfall intensity estimates by the RF, IMERG, and MPE retrieved from all gauge pixels in the study area to the gauge estimates reveals the RF's overestimation of low intensities (mostly below 15 mm per 30-minutes), whereas the high intensities (above 30 mm per 30-minutes) are underestimated. On average, when compared to gauge estimates, the absolute errors were about 5 mm per 30-minutes, comparable to the IMERG and MPE versus gauge estimates, suggesting an RF rainfall estimation performance in the study area that may be as good as the IMERG and MPE technique. However, the fact that the RF's estimation, unlike MPE, is not dependent on the cloud type and its estimates are at high spatial and temporal resolution than IMERG suggests an effective skill that needs future investigation.

This study also compared the RF's rainfall intensity estimates over the MWL transmission path to estimates derived from independent MWL RSL data to determine the RF's ability to estimate average rainfall over the MWL path. Overall, the RF's mean, median and maximum rainfall intensities indicate that the RF can quantify rainfall over the MWL transmission path. However, the RF's mean intensities compare better with the MWL estimates, which was attributed to the fact that both represent the average rainfall intensity along the MWL transmission path. The differences in the RF and MWL rainfall estimates were rather due to differences in the retrieval techniques.

When comparing the spatial distribution of the RF rainfall intensities to IMERG and MPE over the study area using an exemplary scene, the MPE showed fewer rain areas of local origin but with intensities that agree with the RF. However, the RF and IMERG raining areas were extensive and comparable, though the IMERG's intensities were comparatively lower. Overall, MAE and RB values computed using all scenes during the validation period reveal that the RF's spatial rainfall estimates agree better with IMERG than MPE. Nevertheless, some areas showed noticeably high MAE, and RB values which may be due to the high rainfall intensities observed related to complex topographic features.

The discrepancies found when comparing the RF estimates to the gauge, IMERG and MPE are somewhat expected when comparing rainfall estimates from different techniques and may be due to many factors. The spatial resolution is a contributing factor; particularly, the gauge observes rainfall from a single point with low spatial representativeness compared to the RF, IMERG and MPE estimates. For this reason, the gauge may easily miss or underestimate a local rainfall event, depending on its proximity to the storm. In contrast, the RF, IMERG and MPE are more likely to capture a rainfall event, though their intensity estimates may differ based on the measurement technique.

Additionally, the RF and MPE estimates represent average estimates of 3 × 3 km, whereas IMERG's estimates represent approximately 10 × 10 km area. Moreover, differences in the measurement techniques used by the RF, gauge, IMERG and MPE may explain the discrepancies in their rainfall estimates. For instance, the rain gauge records rainfall accumulations continuous in time from a single point. The accuracy of the MWL rainfall data that trained (using tuned parameters) the RF model is affected by various factors such as the wet antenna effect and variation of raindrop sizes along the MWL path. Moreover, the MSG-based cloud properties that estimated the RF's rainfall estimates represent instantaneous properties at the cloud top. Besides, the RF rainfall intensity estimates represent an average of predictions by all trees, which may explain its overestimation (underestimation) of low (high) intensities.

On the other hand, the MPE algorithm relates IR brightness temperatures to the SSM/I rain rates to target convective rainfall that is mostly of high intensities and localized. Therefore, the MPE is likely to miss non-convective rainfall events. IMERG is a multi-sensor technique; its estimates represent a spatiotemporally averaged rainfall from multiple microwave estimates. Also, the accuracy of PMW rainfall retrievals over mountainous areas is affected by the orographic effect on rainfall area (Adhikari & Behrangi, 2022; Kingsley et al., 2021; Petković & Kummerow, 2017), which may explain the high differences observed between the

RF versus IMERG and RF versus MPE estimates for areas with complex topography. Furthermore, the high spatiotemporal rainfall variability in the study area may also contribute to the differences in the rainfall estimates by the gauge, RF, IMERG and MPE (Wakachala et al., 2015).

This study's results may have implications for rainfall retrievals, benefiting various operational and research applications such as agriculture and water resources management and evaluating satellite products, particularly in the many ungauged areas. The reason is that our rainfall intensity retrievals rely on MWL and MSG satellite data, already existing in vast areas, including areas lacking conventional ground rainfall monitoring systems.

5.5 Conclusion

A new technique to estimate high spatiotemporal resolution rainfall from MSG-based cloud top properties using the RF algorithm trained with MWL rainfall intensities is investigated and evaluated for a topographically complex area in the Kenyan Rift Valley. The technique uses MSG spectral IR data not affected by solar illumination, making it applicable under daytime and nighttime conditions.

In general, the presented results show a promising technique. When comparing the technique's rainfall intensities to gauge data, the average retrieval errors were about 5 mm per 30-minutes, comparable to errors found when comparing IMERG and MPE to gauge data. Additionally, the spatial distribution of rainfall intensities retrieved agreed well with the IMERG and MPE satellite products. On top of this, the technique's advantage is that the rainfall intensities are retrieved at high spatiotemporal resolution and are not limited by the rainfall type. Besides, it employs a machine learning technique that may potentially allow for rainfall retrievals in an automated manner.

The study's evaluation was based on a small area and limited MWL network data. On top of this, central to this study's retrieval procedure is a rain area detection step requiring site-specific threshold and gradient parameters that may limit the direct transferability of this study's model parameters to other areas. However, the method for retrieving site-specific rain area detection parameters (Kingsley et al., 2021) is transferable to other study areas. It can be used to replicate this study's rainfall retrieval technique elsewhere. In spite of the limitations, the promising results suggest that with the inclusion of data from a spatially extensive MWL network and by considering site-specific rain area detection parameters, better retrieval accuracies over vast areas are possible.

Chapter 6: Concluding remarks and the study implications

6.1 Concluding remarks

The significance of accurate spatial and temporal rainfall information for many applications such as climates studies, water resources management and agriculture cannot be overemphasized. Despite their great value, traditional techniques are mainly challenged in accurate rainfall detection and estimation by rainfall's high spatiotemporal variability.

The limitations of traditional techniques to estimate accurate rainfall information may be primarily technique-dependent. For instance, the gauge's measurement provides point rainfall information and challenges its usage for spatial rainfall retrieval, especially when retrieving from sparsely distributed gauge observations. Though powerful for retrieving spatiotemporal rainfall, radars cannot be installed everywhere, and they are limited by error sources such as beam blockage, beam overshoot and range effects. Satellites provide area-wide rainfall that fills spatial rainfall gaps. Yet, in some cases, their estimates are uncertain, and they need ground evaluation to advance new satellite rainfall measurement missions and increase the reliability of their estimates.

The MWL-rainfall estimation is an opportunistic technique that potentially offers a low-cost global rainfall observation. Nonetheless, the issue of raindrop size distribution and the MWL's antenna wetting during and after rain are some factors that may challenge its accurate rainfall retrieval. Moreover, the low density of MWL in rural and underdeveloped areas may challenge accurate spatial rainfall estimation. With its wide range and high temporal resolution radiometer, SEVIRI, the MSG satellite permits quasi-continuous rainfall observation in near-real-time over large areas, including ungauged areas or areas with sparse monitoring systems. Consequently, its combination with the MWL data could improve rainfall observation in ungauged areas or areas with sparse data.

Hence, this study's objective was to investigate the application of the MWL and MSG data for high spatial rainfall detection and estimation using data from Central and Western Kenya, where ground data is often sparse. Four specific objectives investigated using individual research studies achieved this main objective.

Chapter 2's study achieved the first specific objective of investigating if the MWL-MSG data combination could improve rain rate estimation and detection. This investigation first evaluated the MWL's ability to estimate gauge-comparable rain rates. Then, the MWL's rain rates were used to study MSG satellite signals of the day and night-time by utilizing a conceptual model in which clouds with high

optical thickness and effective radius have high rainfall probabilities and rain rates. Eventually, descriptive statistical information from the analysis successfully detected rainfall occurrence on individual MWL.

The evaluation of the MWL rainfall estimation technique in the study area confirmed the robustness of the technique's rainfall retrieval skill pointed out by numerous studies elsewhere. Studying the MWL's rainfall estimates with MSG data for rain detection on individual MWL was also successful, with a better daytime performance than nighttime. This is due to its better rain information content about cloud optical depth and effective radius that lies in daytime reflectance than the night-time IR data. Overall, the study highlighted the unique potential of the MWL-MSG data for area-wide rainfall detection and estimation over the study area.

Following the successful combination of the MWL-MSG data for area-wide rainfall observation, the study in Chapter 3 used the high-resolution MSG data to develop an improved rain area detection system that achieved specific objective 2. This rain area detection system later improved the MWL rainfall and estimated high spatiotemporal rainfall over the study area. The investigated procedure involved developing and evaluating multiple parametric models derived from cloud top properties inferred from reflectance and IR data and using a conceptual idea similar to Chapter 2's study. This study also introduced a new technique to overcome uncertainties in MSG cloud top property-based rain estimates using rain area-specific gradient parameters to improve the detected rain areas.

Over the study area, this study found that the reflectance-based model VIS0.6–NIR1.6 with a threshold value of 0.21 was consistent with daytime rainfall occurrences in rain gauges. In contrast, the night-time IR-based model from IR3.9–IR10.8, IR3.9–WV7.3 and IR10.8–WV6.2 brightness temperature difference (BTD) with parametric thresholds 8.18 K, 17.03 K, 33.65 K, respectively, achieved the best performance. It is noteworthy that the daytime detection model performed better than the night-time, which confirms the previous observation and conclusion in Chapter 2.

In general, the detected rain areas also agreed well with estimates from the IMERG satellite rainfall product, albeit with a better agreement in the daytime than night-time estimates. Nonetheless, the study found differences in rain area locations by MSG and IMERG, suggesting spatial displacement errors in rainfall exist and may affect operational applications in floods and flash flood forecasts.

For the first time, the MSG data improved the MWL rainfall estimate in Chapter 4, a study that achieved specific objective 3. This investigation utilized rain areas detected in Chapter 3 to develop a new method for the MWL's wet-dry classification and baseline level estimation. Additionally, a new parameter, the wpl, was also derived from the satellite data representing the length of the MWL's path that is raining.

The MSG-based rain area estimates successfully detected wet and dry periods and estimated a baseline level from the MWL data. Also, the wpl parameter notably improved the high rain intensity estimates, suggesting that spatial rainfall variability along the MWL transmission path is essential to be considered in the MWL rainfall retrieval.

Finally, Chapter 5's study focused on achieving specific objective 4. The study used the RF machine learning algorithm to investigate a new technique for high spatiotemporal resolution rainfall estimation from the MWL-MSG data. The investigation trained the RF algorithm with MWL rainfall estimates to estimate rainfall from MSG-based cloud top properties. The validation results demonstrated a promising rainfall retrieval technique with good accuracy. The retrieval errors were comparable to the IMERG and MPE satellite rainfall errors when evaluated based on gauge estimates. The technique's spatial rainfall estimates also agreed well with these satellite rainfall products. The advantage is that the technique retrieves rainfall at a high spatiotemporal resolution regardless of the rainfall type, e.g. convective or stratiform.

6.2 The study implications

Ground rainfall information is often lacking, particularly in rural and underdeveloped parts of the world, where it directly impacts many applications such as agriculture and water resource management. Additionally, this limits satellite rainfall products' ground evaluation and reliability in these areas. It is expected that the techniques and models developed in this dissertation will generally contribute to area-wide rainfall detection and estimation, particularly by improving rainfall observation in areas where ground monitoring systems are sparsely distributed but with geostationary satellites (like MSG) data coverage and growing MWL network.

The MWL and MSG data can independently observe rainfall. This dissertation confirms the MWL's accurate rainfall retrieval in areas where ground monitoring systems are lacking or insufficient, suggesting the MWL is a vital source of rainfall information for many applications, particularly in these areas. Furthermore, this

dissertation shows that rain area-specific gradient parameters can reduce uncertainties in MSG cloud top property-based rain area estimates that may improve the satellite's rainfall estimates.

Additionally, the dissertation highlights the great potential of combining the MWL and MSG data to improve rainfall observation to benefit areas with insufficient monitoring systems. For instance, the accuracy of the MWL's high rainfall intensities estimates can be remarkably improved when considering the spatial variability of rainfall along the MWL signal transmission path. This may directly affect the study of convective rain and flash floods, especially in areas where monitoring systems are lacking or insufficient.

Moreover, high spatial and temporal rainfall retrieval from the MWL and MSG data is demonstrated in this study. The implications of this realisation for area-wide rainfall observation are enormous, considering that both the MWL and MSG data (together with other geostationary satellites) have global coverage on land. Notably, its application for rainfall estimation in areas with sparse ground data may invaluably impact applications such as satellite data ground evaluation, local climate studies and water resource management.

Bibliography

Acharya, Suwash Chandra, Nathan, Rory, Wang, Quan J., Su, Chun-Hsu, & Eizenberg, Nathan. (2020). Ability of an Australian Reanalysis Dataset to Characterise Sub-Daily Precipitation. *Hydrology and Earth System Sciences, 24*(6), 2951-2962. doi:10.5194/hess-24-2951-2020

Adhikari, Abishek, & Behrangi, Ali. (2022). Assessment of Satellite Precipitation Products in Relation with Orographic Enhancement over the Western United States. *Earth and Space Science, 9*(2). doi:10.1029/2021ea001906

Adler, Robert F., & Negri, Andrew J. (1988). A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall. *Journal of Applied Meteorology, 27*(1), 30-51. doi:10.1175/1520-0450(1988)027<0030:asitte>2.0.co;2

Alemseged, T. H., & Rientjes, T. H. M. (2007, 2007). *Spatio-Temporal Rainfall Mapping from Space: Setbacks and Strengths.* Paper presented at the 5th International Symposium Spatial Data Quality 2007

Enschede, Netherlands.

Ali, Shahzad, & Shahbaz, Muhammad. (2020). Streamflow Forecasting by Modeling the Rainfall–Streamflow Relationship Using Artificial Neural Networks. *Modeling Earth Systems and Environment, 6*(3), 1645-1656. doi:10.1007/s40808-020-00780-3

Allen, M. R., & Ingram, W. J. (2002). Constraints on Future Changes in Climate and the Hydrologic Cycle. *Nature, 419*(6903), 224-232. doi:10.1038/nature01092

Anjum, Ahmad, Ding, Shangguan, Zaman, Ijaz, . . . Yang. (2019). Assessment of Imerg-V06 Precipitation Product over Different Hydro-Climatic Regimes in the Tianshan Mountains, North-Western China. *Remote Sensing, 11*(19). doi:10.3390/rs11192314

Arkin, Phillip A., & Meisner, Bernard N. (1987). The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western Hemisphere During 1982-84. *Monthly Weather Review, 115*(1), 51-74. doi:10.1175/1520-0493(1987)115<0051:trblsc>2.0.co;2

Arking, Albert, & Childs, Jeffrey D. (1985). Retrieval of Cloud Cover Parameters from Multispectral Satellite Images. *Journal of Climate and Applied Meteorology, 24*(4), 322-333. doi:10.1175/1520-0450(1985)024<0322:Roccpf>2.0.Co;2

Atlas, David, & Ulbrich, Carlton W. (1977). Path- and Area-Integrated Rainfall Measurement by Microwave Attenuation in the 1–3 Cm Band. *Journal of Applied Meteorology, 16*(12), 1322-1331. doi:10.1175/1520-0450(1977)016<1322:Paairm>2.0.Co;2

Barnston, Anthony G. (1992). Correspondence among the Correlation, Rmse, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. *Weather and Forecasting, 7*(4), 699-709. doi:10.1175/1520-0434(1992)007<0699:Catcra>2.0.Co;2

Barthès, L., & Mallet, C. (2013). Rainfall Measurement from the Opportunistic Use of an Earth–Space Link in the Ku Band. *Atmospheric Measurement Techniques, 6*(8), 2181-2193. doi:10.5194/amt-6-2181-2013

Barua, Shishutosh, Muttil, Nitin, Ng, A. W. M., & Perera, B. J. C. (2013). Rainfall Trend and Its Implications for Water Resource Management within the Yarra River Catchment, Australia. *Hydrological Processes, 27*(12), 1727-1738. doi:10.1002/hyp.9311

Baum, Bryan A., & Spinhirne, James D. (2000). Remote Sensing of Cloud Properties Using Modis Airborne Simulator Imagery During Success: 3. Cloud Overlap. *Journal of Geophysical Research: Atmospheres, 105*(D9), 11793-11804. doi:10.1029/1999jd901091

Bell, Thomas L., & Kundu, Prasun K. (2003). Comparing Satellite Rainfall Estimates with Rain Gauge Data: Optimal Strategies Suggested by a Spectral Model. *Journal of Geophysical Research: Atmospheres, 108*(D3). doi:10.1029/2002jd002641

Bellerby, Timothy J., & Sun, Jizhong. (2005). Probabilistic and Ensemble Representations of the Uncertainty in an Ir/Microwave Satellite Precipitation Product. *Journal of Hydrometeorology, 6*(6), 1032-1044. doi:10.1175/jhm454.1

Bendix, Jörg, Nauß, Thomas, Thies, Boris, & Kühnlein, Meike. (2010). Rainfall-Rate Assignment Using Msg Seviri Data—a Promising Approach to Spaceborne Rainfall-Rate Retrieval for Midlatitudes. *Journal of Applied Meteorology and Climatology, 49*(7), 1477-1495. doi:10.1175/2010jamc2284.1

Bergès, Jean Claude, Chopin, Franck, Bessat, Frédéric, & Based, Satellite. (2005). *Satellite Based Downscaling Algorithm for Rainfall Estimation Hal* Paper presented at the IV Colóquio Brasileiro de Ciências Geodésicas - IV CBCG, Curitiba, Brazil. https://hal.archives-ouvertes.fr/hal-00378462

Bianchi, Blandine, Jan van Leeuwen, Peter, Hogan, Robin J., & Berne, Alexis. (2013). A Variational Approach to Retrieve Rain Rate by Combining Information from Rain Gauges, Radars, and Microwave Links. *Journal of Hydrometeorology*, *14*(6), 1897-1909. doi:10.1175/jhm-d-12-094.1

Biron, Daniele, SIST, Massimiliano, MELFI, Davide, ZAULI, Francesco, VOCINO, Antonio, FACCIORUSSO, Leonardo, & LEONIBUS, Luigi DE. (2012). *Exploitation of Lightning Data for Rainfall Estimation by Multi-Sensor and Cross-Platform Approach*. Paper presented at the 22nd International Lightning Detection Conference, Broomfield, Colorado, USA.

Blumenfeld, Josh. (2015). From Trmm to Gpm: The Evolution of Nasa Precipitation Data. Retrieved on 16th-12-2021 from https://earthdata.nasa.gov/learn/articles/tools-and-technology-articles/trmm-to-gpm

Bolles, Dana. (2022). Water Cycle. Retrieved on 10-01-2022 from https://science.nasa.gov/earth-science/oceanography/ocean-earth-system/ocean-water-cycle

Breiman, Leo. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32. doi:10.1023/a:1010933404324

Brian, Odhiambo Ayugi, Wang, Wen, & Daisy, Chepkemoi. (2016). Analysis of Spatial and Temporal Patterns of Rainfall Variations over Kenya. *Journal of Environment and Earth Science*, *6*(11).

Caglar, Bayik, Becek, K., Mekik, C., & Ozendi, M. (2018). On the Vertical Accuracy of the Alos World 3d-30m Digital Elevation Model. *Remote Sensing Letters*, *9*(6), 607-615. doi:10.1080/2150704x.2018.1453174

Carlberg, Bradley, Franz, Kristie, & Gallus, William. (2020). A Method to Account for Qpf Spatial Displacement Errors in Short-Term Ensemble Streamflow Forecasting. *Water*, *12*(12). doi:10.3390/w12123505

Chwala, Christian, & Kunstmann, Harald. (2019). Commercial Microwave Link Networks for Rainfall Observation: Assessment of the Current Status and Future Challenges. *Wiley Interdisciplinary Reviews-Water*, *6*(2), e1337-e1337. doi:10.1002/wat2.1337

Cristiano, Elena, ten Veldhuis, Marie-Claire, & van de Giesen, Nick. (2017). Spatial and Temporal Variability of Rainfall and Their Effects on Hydrological Response in Urban Areas – a Review. *Hydrology and Earth System Sciences*, *21*(7), 3859-3878. doi:10.5194/hess-21-3859-2017

David, N., Alpert, P., & Messer, H. (2009). Technical Note: Novel Method for Water Vapour Monitoring Using Wireless Communication Networks Measurements. *Atmospheric Chemistry and Physics, 9*(7), 2413-2418. doi:10.5194/acp-9-2413-2009

David, N., Alpert, P., & Messer, H. (2013). The Potential of Cellular Network Infrastructures for Sudden Rainfall Monitoring in Dry Climate Regions. *Atmospheric Research, 131*, 13-21. doi:10.1016/j.atmosres.2013.01.004

David, N., & Gao, H. O. (2018). Using Cell-Phone Tower Signals for Detecting the Precursors of Fog. *Journal of Geophysical Research: Atmospheres, 123*(2), 1325-1338. doi:10.1002/2017jd027360

David, Noam, Gao H Oliver, Kumah K.K, Hoedjes C.B. Joost, Su Z., & Liu, Yanyan. (2019). *Microwave Communication Networks as a Sustainable Tool of Rainfall Monitoring for Agriculture Needs in Africa*. Paper presented at the 16th International Conference on Environmental Science and Technology, Rhodes, Greece.

David, Noam, Liu, Yanyan, Kumah, Kingsley K., Hoedjes, Joost C. B., Su, Bob Z., & Gao, H. Oliver. (2021). On the Power of Microwave Communication Data to Monitor Rain for Agricultural Needs in Africa. *Water, 13*(5), 730. doi:10.3390/w13050730

David, Noam, Sendik, Omry, Messer, Hagit, & Alpert, Pinhas. (2015). Cellular Network Infrastructure: The Future of Fog Monitoring? *Bulletin of the American Meteorological Society, 96*(10), 1687-1698. doi:10.1175/bams-d-13-00292.1

Dezfuli, A. K., Ichoku, C. M., Huffman, G. J., Mohr, K. I., Selker, J. S., van de Giesen, N., . . . Annor, F. O. (2017). Validation of Imerg Precipitation in Africa. *J Hydrometeorol, 18*(10), 2817-2825. doi:10.1175/jhm-d-17-0139.1

Dhib, Saoussen, Mannaerts, Chris M., Bargaoui, Zoubeida, Retsios, Vasilios, & Maathuis, Ben H. P. (2017). Evaluating the Msg Satellite Multi-Sensor Precipitation Estimate for Extreme Rainfall Monitoring over Northern Tunisia. *Weather and Climate Extremes, 16*, 14-22. doi:10.1016/j.wace.2017.03.002

Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S. J., & Ropelewski, C. F. (2008). Validation of High-Resolution Satellite Rainfall Products over Complex Terrain. *International Journal of Remote Sensing, 29*(14), 4097-4110. doi:10.1080/01431160701772526

Dinku, Tufa, Anagnostou, Emmanouil N., & Borga, Marco. (2002). Improving Radar-Based Estimation of Rainfall over Complex Terrain. *Journal of Applied Meteorology, 41*(12), 1163-1178. doi:10.1175/1520-0450(2002)041<1163:Irbeor>2.0.Co;2

Doumounia, Ali, Gosset, Marielle, Cazenave, Frederic, Kacou, Modeste, & Zougmore, François. (2014). Rainfall Monitoring Based on Microwave Links from Cellular Telecommunication Networks: First Results from a West African Test Bed. *Geophysical Research Letters, 41*(16), 6016-6022. doi:10.1002/2014gl060724

Duan, Q., & Duan, A. (2020). The Energy and Water Cycles under Climate Change. *Natl Sci Rev, 7*(3), 553-557. doi:10.1093/nsr/nwaa003

Edstam, Jonas, Olsson, Andreas, Flodin, Jonas, Öhberg, Mikael, Henriksson, Anders, Hansryd, Jonas, & Ahlberg, Jonas. (2018). *Ericsson Microwave Outlook*. Retrieved from Göteborg, Sweden: https://www.ericsson.com/4a312c/assets/local/reports-papers/microwave-outlook/2018/ericsson-microwave-outlook-report-2018.pdf

ERICSSON. (2015). *Ericsson Microwave Towards 2020: Delivering High-Capacity and Cost-Efficient Backhaul for Broad Band Networks Today and in the Future*. Retrieved from Stockholm, Sweden: http://cdn.everythingrf.com/live/microwave-2020-report.pdf

EROS, USGS. Usgs Eros Archive - Digital Elevation - Shuttle Radar Topography Mission (Srtm) 1 Arc-Second Global. Retrieved on 28 April 2020 from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects

EUMETSAT. (2015). Cloud Mask Product: Product Guide. In. Darmstadt, Germany: EUMETSAT.

EUMETSAT. (2016). Meteosat-8 Satellite's New Position of 41.5e Provides Weather and Climate View over the Indian Ocean. Retrieved on 14 July 2020 from https://phys.org/news/2016-09-meteosat-satellite-position-415e-weather.html

EUMETSAT. (2020). Archived Meteosat Data. Retrieved on 19 March 2020 from https://www.eumetsat.int/archived-meteosat-data

EUMETSAT. (2021). Meteosat Series. Retrieved on 22 December 2021 from https://www.eumetsat.int/our-satellites/meteosat-series

FAO. (2020). *The State of Food and Agriculture 2020*. Rome, Italy: FAO.

Feidas, Haralambos, & Giannakos, Apostolos. (2010). Identifying Precipitating Clouds in Greece Using Multispectral Infrared Meteosat Second Generation Satellite Data. *Theoretical and Applied Climatology*, *104*(1-2), 25-42. doi:10.1007/s00704-010-0316-5

Feidas, Haralambos, & Giannakos, Apostolos. (2011). Classifying Convective and Stratiform Rain Using Multispectral Infrared Meteosat Second Generation Satellite Data. *Theoretical and Applied Climatology*, *108*(3-4), 613-630. doi:10.1007/s00704-011-0557-y

Feijt, Arnout J., Roebeling, Robert A., & Wolters, Erwin L. A. (2008). Evaluation of Cloud-Phase Retrieval Methods for Seviri on Meteosat-8 Using Ground-Based Lidar and Cloud Radar Data. *Journal of Applied Meteorology and Climatology*, *47*(6), 1723-1738. doi:10.1175/2007jamc1591.1

Ferraro, Ralph R., & Marks, Gerard F. (1995). The Development of Ssm/I Rain-Rate Retrieval Algorithms Using Ground-Based Radar Measurements. *Journal of Atmospheric and Oceanic Technology*, *12*(4), 755-770. doi:10.1175/1520-0426(1995)012<0755:Tdosrr>2.0.Co;2

Gao, Xiaogang, Sorooshian, Soroosh, Hsu, Kuo-Lin, & Hong, Yang. (2004). Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology*, *43*(12), 1834-1853. doi:10.1175/jam2173.1

Garcia, João V. C., Stephany, Stephan, & d'Oliveira, Augusto B. (2013). Estimation of Convective Precipitation Mass from Lightning Data Using a Temporal Sliding-Window for a Series of Thunderstorms in Southeastern Brazil. *Atmospheric Science Letters*, *14*(4), 281-286. doi:10.1002/asl2.453

Giannakos, Apostolos, & Feidas, Haralambos. (2012). Classification of Convective and Stratiform Rain Based on the Spectral and Textural Features of Meteosat Second Generation Infrared Data. *Theoretical and Applied Climatology*, *113*(3-4), 495-510. doi:10.1007/s00704-012-0802-z

GPCC. (2020). National Center for Atmospheric Research Staff (Eds). Last Modified 27 Feb 2020. "The Climate Data Guide: Gpcc: Global Precipitation Climatology Centre. Retrieved on April, 11 2022 from https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre.

Grecu, Mircea, Anagnostou, Emmanouil N., & Adler, Robert F. (2000). Assessment of the Use of Lightning Information in Satellite Infrared Rainfall Estimation. *Journal of Hydrometeorology, 1*(3), 211-221. doi:10.1175/1525-7541(2000)001<0211:Aotuol>2.0.Co;2

Guo, H., Ma, J., & Fu, W. (2018). Chinese Programs. In S. Liang (Ed.), *Comprehensive Remote Sensing* (pp. 220-245). Oxford: Elsevier.

Gyasi-Agyei, Yeboah. (2020). Identification of the Optimum Rain Gauge Network Density for Hydrological Modelling Based on Radar Rainfall Analysis. *Water, 12*(7). doi:10.3390/w12071906

Ha, Eunho, North, Gerald R., Yoo, Chulsang, & Ha, Kyung-Ja. (2002). Evaluation of Some Ground Truth Designs for Satellite Estimates of Rain Rate. *Journal of Atmospheric and Oceanic Technology, 19*(1), 65-73. doi:10.1175/1520-0426(2002)019<0065:Eosgtd>2.0.Co;2

Han, Hyo-Jin, Sohn, B. J., & Seo, Eun-Kyoung. (2010). Validation of Satellite-Based High-Resolution Rainfall Products over the Korean Peninsula Using Data from a Dense Rain Gauge Network. *Journal of Applied Meteorology and Climatology, 49*(4), 701-714. doi:10.1175/2009jamc2266.1

Harold, Brooks, Barb, Brown, Beth, Ebert, Chris, Ferro, Johannes, Jenkner, Ian, Jolliffe, . . . David, Stephenson. (2015). Wwrp/Wgne Joint Working Group on Forecast Verification Research. Retrieved on 2020 October 10 from https://www.cawcr.gov.au/projects/verification/

Heinemann, T, Latanzio, A, & Roveda, F. (2002, September 2002). *The Eumetsat Multi-Sensor Precipitation Estimate (Mpe).* Paper presented at the Proceedings of the second International Precipitation Working Group (IPWG) meeting, Madrid, Spain,.

Heinemann, Thomas, & Kerényi, Judit. (2003). *The Eumetsat Multi Sensor Precipitation Estimate (Mpe) : Concept and Validation*.

Held, Isaac M., & Soden, Brian J. (2006). Robust Responses of the Hydrological Cycle to Global Warming. *Journal of Climate, 19*(21), 5686-5699. doi:10.1175/jcli3990.1

Hoedjes, Joost, Kooiman, André, Maathuis, Ben, Said, Mohammed, Becht, Robert, Limo, Agnes, . . . Su, Bob. (2014). A Conceptual Flash Flood Early Warning System for Africa, Based on Terrestrial Microwave Links and Flash Flood Guidance. *ISPRS International Journal of Geo-Information, 3*(2), 584-598. doi:10.3390/ijgi3020584

Hoedjes, Joost, Said, Mohammed, Becht, Robert, Kifugo, Shem, Kooiman, André, Limo, Agnes, . . . Wright, Iain. (2013, 2013/04//). *Towards near Real-Time Convective Rainfall Observations over Kenya*.

Hoell, Andrew, & Funk, Chris. (2013). Indo-Pacific Sea Surface Temperature Influences on Failed Consecutive Rainy Seasons over Eastern Africa. *Climate Dynamics, 43*(5-6), 1645-1660. doi:10.1007/s00382-013-1991-6

Hogg, D. C. (1968). Millimeter-Wave Communication through the Atmosphere: The Known and Unknown Features in Propagation of Short Radio Waves Are Discussed. *Science, 159*(3810), 39-46. doi:10.1126/science.159.3810.39

Holmlund, K., Grandell, J., Schmetz, J., Stuhlmann, R., Bojkov, B., Munro, R., . . . Blythe, P. (2021). Meteosat Third Generation (Mtg): Continuation and Innovation of Observations from Geostationary Orbit. *Bulletin of the American Meteorological Society, 102*(5), E990-E1015. doi:10.1175/bams-d-19-0304.1

Hong, Yang, Adler, Robert F., Negri, Andrew, & Huffman, George J. (2007). Flood and Landslide Applications of near Real-Time Satellite Rainfall Products. *Natural Hazards, 43*(2), 285-294. doi:10.1007/s11069-006-9106-x

Hong, Yang, & Gourley, Jonathan J. (2015). Radar Hydrology: Principles, Models, and Applications. In (1st ed.). Boca Raton: Taylor & Francis Group.

Hosseini-Moghari, Seyed-Mohammad, & Tang, Qiuhong. (2020). Validation of Gpm Imerg V05 and V06 Precipitation Products over Iran. *Journal of Hydrometeorology, 21*(5), 1011-1037. doi:10.1175/jhm-d-19-0269.1

Houze, Robert A. (1997). Stratiform Precipitation in Regions of Convection: A Meteorological Paradox? *Bulletin of the American Meteorological Society, 78*(10), 2179-2196. doi:10.1175/1520-0477(1997)078<2179:Spiroc>2.0.Co;2

Hsu, Kuo-Lin, & Sorooshian, Soroosh. (2008). Satellite-Based Precipitation Measurement Using Persiann System. In S. Sorooshian, K.-L. Hsu, E. Coppola, B. Tomassetti, M. Verdecchia, & G. Visconti (Eds.), *Hydrological Modelling and the Water Cycle* (pp. 27-48). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hu, Qingfang, Li, Zhe, Wang, Leizhi, Huang, Yong, Wang, Yintang, & Li, Lingjie. (2019). Rainfall Spatial Estimations: A Review from Spatial Interpolation to Multi-Source Data Merging. *Water, 11*(3). doi:10.3390/w11030579

Huffman, George, Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., & Xie, P. (2014). Integrated Multi-Satellite Retrievals for Gpm (Imerg). version 4.4. Retrieved on 19 March 2021 from https://gpm.nasa.gov/data/directory

Huffman, George, Bolvin, David T., Nelkin, Eric J., & Tan, Jackson. (2020). Integrated Multi-Satellite Retrievals for Gpm (Imerg) Technical Documentation. Retrieved on 17 January 2021 from https://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc_180207.pdf

Huffman, George J., Adler, Robert F., Bolvin, David T., & Nelkin, Eric J. (2010). The Trmm Multi-Satellite Precipitation Analysis (Tmpa). In M. Gebremichael & F. Hossain (Eds.), *Satellite Rainfall Applications for Surface Hydrology* (pp. 3-22). Dordrecht: Springer Netherlands.

Huffman, George, Stocker, E.F., Bolvin, D.T., Nelkin, E.J., & Tan, Jackson. (2019). Gpm Imerg Final Precipitation L3 Half Hourly 0.1 Degree X 0.1 Degree V06. 06. Retrieved on 02 January 2021 from https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary

Inoue, Toshiro. (1985). On the Temperature and Effective Emissivity Determination of Semi-Transparent Cirrus Clouds by Bi-Spectral Measurements in the 10ΜM Window Region. *Journal of the Meteorological Society of Japan. Ser. II, 63*(1), 88-99. doi:10.2151/jmsj1965.63.1_88

Inoue, Toshiro. (1987a). A Cloud Type Classification with Noaa 7 Split-Window Measurements. *Journal of Geophysical Research, 92*(D4), 3991-4000. doi:10.1029/JD092iD04p03991

Inoue, Toshiro. (1987b). An Instantaneous Delineation of Convective Rainfall Areas Using Split Window Data of Noaa-7 Avhrr. *Journal of the Meteorological Society of Japan. Ser. II, 65*(3), 469-481. doi:10.2151/jmsj1965.65.3_469

Inoue, Toshiro. (1989). Features of Clouds over the Tropical Pacific During Northern Hemispheric Winter Derived from Split Window Measurements. *Journal of the Meteorological Society of Japan. Ser. II, 67*(4), 621-637. doi:10.2151/jmsj1965.67.4_621

Inoue, Toshiro, & Aonashi, Kazumasa. (2000). A Comparison of Cloud and Rainfall Information from Instantaneous Visible and Infrared Scanner

and Precipitation Radar Observations over a Frontal Zone in East Asia During June 1998. *Journal of Applied Meteorology, 39*(12), 2292-2301. doi:10.1175/1520-0450(2001)040<2292:Acocar>2.0.Co;2

Inoue, Toshiro, Wu, Xiangqiuan, & Bessho, Kotaro. (2001, 16 October 2001). *Life Cycle of Convective Activity in Terms of Cloud Type Observed by Split Window*. Paper presented at the Proceedings of 11th Conference on Satellite Meteorology and Oceanography, Madison, WI.

IPCC. (2022). *Summary for Policymakers [H.-O. Pörtner, D.C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem (Eds.)]. In: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group Ii to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (Eds.)]*. Retrieved from https://www.ipcc.ch/report/ar6/wg2/

ITU. (2005a). *Recommendation Itu-R P.838-2 Specific Attenuation Model for Rain for Use in Prediction Methods* (1992199920032). Retrieved from Geneva, Switzerland: https://www.itu.int/rec/R-REC-P.838-3-200503-I/en

ITU. (2005b). *Recommendation Itu-R P.838-3: Specific Attenuation Model for Rain for Use in Prediction Methods*. Retrieved from Geneva, Switzerland: https://www.itu.int/rec/R-REC-P.838-3-200503-I/en

ITU. (2019). *Itu-R Attenuation by Atmospheric Gases and Related Effects*. Retrieved from Geneva: https://www.itu.int/rec/R-REC-P.676

Jackson, Tan, George, Huffman, David, Bolvin, & Eric, Nelkin. (2020). Imerg V06 Ground Validation Wishlist. Retrieved on 02 March 2021 from https://gpm.nasa.gov/sites/default/files/document_files/IMERG-Validation-Wishlist.pdf

Joyce, Robert J., Janowiak, John E., Arkin, Phillip A., & Xie, Pingping. (2004). Cmorph: A Method That Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology, 5*(3), 487-503. doi:10.1175/1525-7541(2004)005<0487:Camtpg>2.0.Co;2

Karaseva, Marina O., Prakash, Satya, & Gairola, R. M. (2011). Validation of High-Resolution Trmm-3b43 Precipitation Product Using Rain Gauge Measurements over Kyrgyzstan. *Theoretical and Applied Climatology, 108*(1-2), 147-157. doi:10.1007/s00704-011-0509-6

Kato, Seiji, & Marshak, Alexander. (2009). Solar Zenith and Viewing Geometry-Dependent Errors in Satellite Retrieved Cloud Optical Thickness: Marine Stratocumulus Case. *Journal of Geophysical Research: Atmospheres, 114*(D1), 1-13. doi:10.1029/2008jd010579

Kawamoto, Kazuaki, Nakajima, Teruyuki, & Nakajima, Takashi Y. (2001). A Global Determination of Cloud Microphysics with Avhrr Remote Sensing. *Journal of Climate, 14*(9), 2054-2068. doi:10.1175/1520-0442(2001)014<2054:Agdocm>2.0.Co;2

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., . . . Spencer, R. W. (2003). The Advanced Microwave Scanning Radiometer for the Earth Observing System (Amsr-E), Nasda's Contribution to the Eos for Global Energy and Water Cycle Studies. *IEEE Transactions on Geoscience and Remote Sensing, 41*(2), 184-194. doi:10.1109/tgrs.2002.808331

KERICHO, COUNTY GOVERNMENT OF. (2015). *County Government of Kericho Annual Development Plan 2016-2017*. Retrieved from Nairobi:

Kerkmann, Jochen, Rosenfeld, D., Lutz, H.J., Prieto, J., & König, M. (2014). Applications of Meteosat Second Generation (Msg): Meteorological Use of the Seviri Ir3.9 Channel. *MSG Interpretation Guide*. Retrieved on 17 March 2021 from https://www.eumetsat.int/msg-interpretation-guide

Kharin, Viatcheslav V., & Zwiers, Francis W. (2003). On the Roc Score of Probability Forecasts. *Journal of Climate, 16*(24), 4145-4150. doi:10.1175/1520-0442(2003)016<4145:Otrsop>2.0.Co;2

Kidd, C. (2001). Satellite Rainfall Climatology: A Review. *International Journal of Climatology, 21*(9), 1041-1066. doi:10.1002/joc.635

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bull Am Meteorol Soc, 98*(1), 69-78. doi:10.1175/BAMS-D-14-00283.1

Kidd, C., & Levizzani, V. (2011). Status of Satellite Precipitation Retrievals. *Hydrology and Earth System Sciences, 15*(4), 1109-1116. doi:10.5194/hess-15-1109-2011

Kidd, Chris, Levizzani, Vincenzo, & Laviola, Sante. (2010a). Quantitative Precipitation Estimation from Earth Observation Satellites. *Rainfall: State of the Science*, 127-158. doi:https://doi.org/10.1029/2009GM000920

Kidd, Chris, Levizzani, Vincenzo, & Laviola, Sante. (2010b). Section Ii: Rainfall Measurement and Estimation. In *Rainfall: State of the Science* (pp. 127-158).

Kidder, Stanley Q., Kankiewicz, J. Adam, & Eis, Kenneth E. (2005). *Meteosat Second Generation Cloud Algorithms for Use at Afwa*. Retrieved from BACIMO 2005, Monterey, CA:

Kilavi, Mary, MacLeod, Dave, Ambani, Maurine, Robbins, Joanne, Dankers, Rutger, Graham, Richard, . . . Todd, Martin. (2018). Extreme Rainfall and Flooding over Central Kenya Including Nairobi City During the Long-Rains Season 2018: Causes, Predictability, and Potential for Early Warning and Actions. *Atmosphere, 9*(12), 472-472. doi:10.3390/atmos9120472

Kimani, Margaret, Hoedjes, Joost, & Su, Zhongbo. (2017). An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa. *Remote Sensing, 9*(5). doi:10.3390/rs9050430

Kingsley, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., Rwasoka, D. T., Retsios, B. V., & Su, B. Z. (2021). Rain Area Detection in South-Western Kenya by Using Multispectral Satellite Data from Meteosat Second Generation. *Sensors (Basel), 21*(10). doi:10.3390/s21103547

Kokhanovsky, A. A. (2003). A Semianalytical Cloud Retrieval Algorithm Using Backscattered Radiation in 0.4–2.4 Mm Spectral Region. *Journal of Geophysical Research, 108*(D1), 4008-4008. doi:10.1029/2001jd001543

Kotchoni, D. O. Valerie, Vouillamoz, Jean-Michel, Lawson, Fabrice M. A., Adjomayi, Philippe, Boukari, Moussa, & Taylor, Richard G. (2018). Relationships between Rainfall and Groundwater Recharge in Seasonally Humid Benin: A Comparative Analysis of Long-Term Hydrographs in Sedimentary and Crystalline Aquifers. *Hydrogeology Journal, 27*(2), 447-457. doi:10.1007/s10040-018-1806-2

Kühnlein, Meike, Appelhans, Tim, Thies, Boris, & Nauß, Thomas. (2014a). Precipitation Estimates from Msg Seviri Daytime, Nighttime, and Twilight Data with Random Forests. *Journal of Applied Meteorology and Climatology, 53*(11), 2457-2480. doi:10.1175/jamc-d-14-0082.1

Kühnlein, Meike, Appelhans, Tim, Thies, Boris, & Nauss, Thomas. (2014b). Improving the Accuracy of Rainfall Rates from Optical Satellite Sensors with Machine Learning — a Random Forests-Based Approach Applied to Msg Seviri. *Remote Sensing of Environment, 141*, 129-143. doi:10.1016/j.rse.2013.10.026

Kumah, Kingsley K., Hoedjes, Joost C. B., David, Noam, Maathuis, Ben H. P., Gao, H. Oliver, & Su, Bob Z. (2020). Combining Mwl and Msg Seviri Satellite Signals for Rainfall Detection and Estimation. *Atmosphere, 11*(9). doi:10.3390/atmos11090884

Kumah, Kingsley K., Hoedjes, Joost C. B., David, Noam, Maathuis, Ben H. P., Gao, H. Oliver, & Su, Bob Z. (2021). The Msg Technique: Improving Commercial Microwave Link Rainfall Intensity by Using Rain Area Detection from Meteosat Second Generation. *Remote Sensing, 13*(16). doi:10.3390/rs13163274

Kumjian, Matthew R. (2018). Weather Radars. In C. Andronache (Ed.), *Remote Sensing of Clouds and Precipitation* (pp. 15-63). Cham: Springer International Publishing.

Kummerow, Christian, Barnes, William, Kozu, Toshiaki, Shiue, James, & Simpson, Joanne. (1998). The Tropical Rainfall Measuring Mission (Trmm) Sensor Package. *Journal of Atmospheric and Oceanic Technology, 15*(3), 809-817. doi:10.1175/1520-0426(1998)015<0809:Ttrmmt>2.0.Co;2

Kummerow, Christian, Hong, Y., Olson, W. S., Yang, S., Adler, R. F., McCollum, J., . . . Wilheit, T. T. (2001). The Evolution of the Goddard Profiling Algorithm (Gprof) for Rainfall Estimation from Passive Microwave Sensors. *Journal of Applied Meteorology, 40*(11), 1801-1820. doi:10.1175/1520-0450(2001)040<1801:Teotgp>2.0.Co;2

Kummerow, Christian, Masunaga, Hirohiko, & Bauer, Peter. (2007). A Next-Generation Microwave Rainfall Retrieval Algorithm for Use by Trmm and Gpm. In V. Levizzani, P. Bauer, & T. F. Joseph (Eds.), *Measuring Precipitation from Space* (pp. 235-252).

Kunkee, David B., Poe, Gene A., Boucher, Donald J., Swadley, Steven D., Hong, Ye, Wessel, John E., & Uliana, Enzo A. (2008). Design and Evaluation of the First Special Sensor Microwave Imager/Sounder. *IEEE Transactions on Geoscience and Remote Sensing, 46*(4), 863-883. doi:10.1109/tgrs.2008.917980

Lahat, Dana, Adali, Tulay, & Jutten, Christian. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE, 103*(9), 1449-1477. doi:10.1109/jproc.2015.2460697

Lazri, Mourad, Ameur, Soltane, Brucker, Jean Michel, Testud, Jacques, Hamadache, Bachir, Hameg, Slimane, . . . Mohia, Yacine. (2013). Identification of Raining Clouds Using a Method Based on Optical and Microphysical Cloud Properties from Meteosat Second Generation

Daytime and Nighttime Data. *Applied Water Science, 3*(1), 1-11. doi:10.1007/s13201-013-0079-0

Lazri, Mourad, Ameur, Soltane, & Mohia, Yacine. (2014). Instantaneous Rainfall Estimation Using Neural Network from Multispectral Observations of Seviri Radiometer and Its Application in Estimation of Daily and Monthly Rainfall. *Advances in Space Research, 53*(1), 138-155. doi:10.1016/j.asr.2013.10.005

Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007a). Hydrometeorological Application of a Microwave Link: 2. Precipitation. *Water Resources Research, 43*(4), 1-9. doi:10.1029/2006wr004989

Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007b). Rainfall Measurement Using Radio Links from Cellular Communication Networks. *Water Resources Research, 43*(3). doi:10.1029/2006wr005631

Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2008). Microwave Link Rainfall Estimation: Effects of Link Length and Frequency, Temporal Sampling, Power Resolution, and Wet Antenna Attenuation. *Advances in Water Resources, 31*(11), 1481-1493. doi:10.1016/j.advwatres.2008.03.004

Lengfeld, K., Clemens, M., Münster, H., & Ament, F. (2014). Performance of High-Resolution X-Band Weather Radar Networks – the Pattern Example. *Atmospheric Measurement Techniques, 7*(12), 4151-4166. doi:10.5194/amt-7-4151-2014

Lensky, Itamar M., & Rosenfeld, Daniel. (1997). Estimation of Precipitation Area and Rain Intensity Based on the Microphysical Properties Retrieved from Noaa Avhrr Data. *Journal of Applied Meteorology, 36*(3), 234-242. doi:10.1175/1520-0450(1997)036<0234:Eopaar>2.0.Co;2

Lensky, Itamar M., & Rosenfeld, Daniel. (2003a). A Night-Rain Delineation Algorithm for Infrared Satellite Data Based on Microphysical Considerations. *Journal of Applied Meteorology, 42*(9), 1218-1226. doi:10.1175/1520-0450(2003)042<1218:Andafi>2.0.Co;2

Lensky, Itamar M., & Rosenfeld, Daniel. (2003b). Satellite-Based Insights into Precipitation Formation Processes in Continental and Maritime Convective Clouds at Nighttime. *Journal of Applied Meteorology, 42*(9), 1227-1233. doi:10.1175/1520-0450(2003)042<1227:Siipfp>2.0.Co;2

Levizzanni, Vincenzo, P, Bauer, & J, Turk F. (2007). *Measuring Precipitation from Space* (Vol. 28): Springer Netherlands.

Li, Xinyan, Yang, Yuanjian, Mi, Jiaqin, Bi, Xueyan, Zhao, You, Huang, Zehao, . . . Li, Wanju. (2021). Leveraging Machine Learning for Quantitative Precipitation Estimation from Fengyun-4 Geostationary Observations and Ground Meteorological Measurements. *Atmospheric Measurement Techniques, 14*(11), 7007-7023. doi:10.5194/amt-14-7007-2021

Liberman, Y., Samuels, R., Alpert, P., & Messer, H. (2014). New Algorithm for Integration between Wireless Microwave Sensor Network and Radar for Improved Rainfall Measurement and Mapping. *Atmospheric Measurement Techniques, 7*(10), 3549-3563. doi:10.5194/amt-7-3549-2014

Liu, Yang, Chawla, Nitesh V., Harper, Mary P., Shriberg, Elizabeth, & Stolcke, Andreas. (2006). A Study in Machine Learning from Imbalanced Data for Sentence Boundary Detection in Speech. *Computer Speech & Language, 20*(4), 468-494. doi:10.1016/j.csl.2005.06.002

Lutz, Hans-Joachim, Inoue, Toshiro, & Schmetz, Johannes. (2003). Comparison of a Split-Window and a Multi-Spectral Cloud Classification for Modis Observations. *Journal of the Meteorological Society of Japan. Ser. II, 81*(3), 623-631. doi:10.2151/jmsj.81.623

Ly, S., Charles, C., & Degre, A. (2013). Different Methods for Spatial Interpolation of Rainfall Data for Operational Hydrology and Hydrological Modeling at Watershed Scale. A Review. *Biotechnologie Agronomie Societe Et Environnement, 17*(2), 392-406. Retrieved from <Go to ISI>://WOS:000320130900010

Macharia, Joseph M., Ngetich, Felix K., & Shisanya, Chris A. (2020). Comparison of Satellite Remote Sensing Derived Precipitation Estimates and Observed Data in Kenya. *Agricultural and Forest Meteorology, 284*. doi:10.1016/j.agrformet.2019.107875

Maidment, Ross I., Allan, Richard P., & Black, Emily. (2015). Recent Observed and Simulated Changes in Precipitation over Africa. *Geophysical Research Letters, 42*(19), 8155-8164. doi:10.1002/2015gl065765

Maranan, Marlon, Fink, Andreas H., Knippertz, Peter, Amekudzi, Leonard K., Atiah, Winifred A., & Stengel, Martin. (2020). A Process-Based Validation of Gpm Imerg and Its Sources Using a Mesoscale Rain Gauge Network in the West African Forest Zone. *Journal of Hydrometeorology, 21*(4), 729-749. doi:10.1175/jhm-d-19-0257.1

Marshall, J. S., Langille, R. C., & Palmer, W. Mc K. (1947). Measurement of Rainfall by Radar. *Journal of Meteorology, 4*(6), 186-192. doi:10.1175/1520-0469(1947)004<0186:Morbr>2.0.Co;2

Martin, Fencl, & Vojtech, Bares. (2018). *Quantifying Hardware Related Attenuation from the Analysis of Nearby Microwave Links*. Paper presented at the 2018 IEEE Workshop on Statistical Signal Processing (SSP), Germany, Freiburg.

Mehrabi, Zia, McDowell, Mollie J., Ricciardi, Vincent, Levers, Christian, Martinez, Juan Diego, Mehrabi, Natascha, . . . Jarvis, Andy. (2020). The Global Divide in Data-Driven Farming. *Nature Sustainability, 4*(2), 154-160. doi:10.1038/s41893-020-00631-0

Mendes, Thiago Augusto, Pereira, Sávio Aparecido dos Santos, Rebolledo, Juan Félix Rodriguez, Gitirana, Gilson de Farias Neves, Melo, Maria Tereza da Silva, & Luz, Marta Pereira da. (2021). Development of a Rainfall and Runoff Simulator for Performing Hydrological and Geotechnical Tests. *Sustainability, 13*(6). doi:10.3390/su13063060

Messer, H., Goldshtein, O., Rayitsfeld, A., & Alpert, P. (2008). Recent Results of Rainfall Mapping from Cellular Network Measurements. *2008 Ieee International Conference on Acoustics, Speech and Signal Processing, Vols 1-12*(1), 5157-+. doi:Doi 10.1109/Icassp.2008.4518820

Messer, H., Zinevich, A., & Alpert, P. (2006). Environmental Monitoring by Wireless Communication Networks. *Science, 312*(5774), 713. doi:10.1126/science.1120034

Meyer, Hanna, Kühnlein, Meike, Appelhans, Tim, & Nauss, Thomas. (2016). Comparison of Four Machine Learning Algorithms for Their Applicability in Satellite-Based Optical Rainfall Retrievals. *Atmospheric Research, 169*, 424-433. doi:10.1016/j.atmosres.2015.09.021

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, Remote Sensing, Climatology and Modeling. *Atmospheric Research, 94*(4), 512-533. doi:10.1016/j.atmosres.2009.08.017

Michaelides, Silas. (2008). *Precipitation: Advances in Measurement, Estimation and Prediction*: Springer Berlin Heidelberg.

Milford, J. R., McDougall, V. D., & Dugdale, G. (1994). *Rainfall Estimation from Cold Cloud Duration: Experience of the Tamsat Group in West Africa*. Niamey. http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_6/colloques2/010008087.pdf

Minda, & Nakamura. (2013). Method for Wet Antenna Correct at 50 Ghz. *Journal of Chemical Information and Modeling, 53*(9), 1689-1699. doi:10.1017/CBO9781107415324.004

MoAlf. (2017). *Climate Risk Profile for Kericho County.Kenya County Climate Risk Profile Series*. Retrieved from Nairobi, Kenya.:

Moazami, S., & Najafi, M. R. (2021). A Comprehensive Evaluation of Gpm-Imerg V06 and Mrms with Hourly Ground-Based Precipitation Observations across Canada. *Journal of Hydrology, 594*. doi:10.1016/j.jhydrol.2020.125929

Moraux, Arthur, Dewitte, Steven, Cornelis, Bruno, & Munteanu, Adrian. (2019). Deep Learning for Precipitation Estimation from Satellite and Rain Gauges Measurements. *Remote Sensing, 11*(21). doi:10.3390/rs11212463

Nicholson, Sharon E. (2017). Climate and Climatic Variability of Rainfall over Eastern Africa. *Reviews of Geophysics, 55*(3), 590-635. doi:10.1002/2016rg000544

NOAA. (2009). *User's Guide for Building and Operating Environmental Satellite Receiving Stations*. Retrieved from Suitland:

Nystuen, Jeffrey A. (1998). Temporal Sampling Requirements for Automatic Rain Gauges. *Journal of Atmospheric and Oceanic Technology, 15*(6), 1253-1260. doi:10.1175/1520-0426(1998)015<1253:Tsrfar>2.0.Co;2

Nystuen, Jeffrey A., Proni, John R., Black, Peter G., & Wilkerson, John C. (1996). A Comparison of Automatic Rain Gauges. *Journal of Atmospheric and Oceanic Technology, 13*(1), 62-73. doi:10.1175/1520-0426(1996)013<0062:Acoarg>2.0.Co;2

O'Neil, T. R., Hu, K., Truong, N. R., Arshad, S., Shacklett, B. L., Cunningham, A. L., & Nasr, N. (2021). The Role of Tissue Resident Memory Cd4 T Cells in Herpes Simplex Viral and Hiv Infection. *Viruses, 13*(3), 75-84. doi:10.3390/v13030359

Odongo, Vincent Omondi, van der Tol, Christiaan, van Oel, Pieter R., Meins, Frank M., Becht, Robert, Onyando, Japheth, & Su, Zhongbo. (2015). Characterisation of Hydroclimatological Trends and Variability in the Lake Naivasha Basin, Kenya. *Hydrological Processes, 29*(15), 3276-3293. doi:10.1002/hyp.10443

Ogwang, Bob Alex, Chen, Haishan, Li, Xing, & Gao, Chujie. (2014). The Influence of Topography on East African October to December Climate: Sensitivity Experiments with Regcm4. *Advances in Meteorology, 2014*, 1-14. doi:10.1155/2014/143917

Ojo, Olumuyiwa Idowu, & Ilunga, Masengo Francois. (2017). The Rainfall Factor of Climate Change Effects on the Agricultural Environment: A Review.

American Journal of Applied Sciences, 14(10), 930-937. doi:10.3844/ajassp.2017.930.937

Olsen, R., Rogers, D., & Hodge, D. (1978). The Ar≪Sup≫B≪/Sup≫Relation in the Calculation of Rain Attenuation. *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, 26*(2), 318-329. doi:10.1109/tap.1978.1141845

Ostrometzky, Jonatan, & Messer, Hagit. (2014). *Accumulated Rainfall Estimation Using Maximum Attenuation of Microwave Radio Signal*. Paper presented at the 2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM). http://www.scopus.com/inward/record.url?eid=2-s2.0-84907391739&partnerID=tZOtx3y1

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2011). Measuring Urban Rainfall Using Microwave Links from Commercial Cellular Communication Networks. *Water Resources Research, 47*(12), 1-16. doi:10.1029/2010wr010350

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2013). Country-Wide Rainfall Maps from Cellular Communication Networks. *Proc Natl Acad Sci U S A, 110*(8), 2741-2745. doi:10.1073/pnas.1217961110

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2016a). Two and a Half Years of Country-Wide Rainfall Maps Using Radio Links from Commercial Cellular Telecommunication Networks. *Water Resources Research, 52*(10), 8039-8065. doi:10.1002/2016wr019412

Overeem, Aart, Leijnse, Hidde, & Uijlenhoet, Remko. (2016b). Retrieval Algorithm for Rainfall Mapping from Microwave Links in a Cellular Communication Network. *Atmospheric Measurement Techniques, 9*(5), 2425-2444. doi:10.5194/amt-9-2425-2016

Overeem, Aart, Leijnse, Hidde, van Leth, Thomas C., Bogerd, Linda, Priebe, Jan, Tricarico, Daniele, . . . Uijlenhoet, Remko. (2021). Tropical Rainfall Monitoring with Commercial Microwave Links in Sri Lanka. *Environmental Research Letters, 16*(7). doi:10.1088/1748-9326/ac0fa6

Paredes Trejo, Franklin Javier, Álvarez Barbosa, Humberto, Peñaloza-Murillo, Marcos A., Moreno, Maria Alejandra, & Farias, Asdrubal. (2016). Intercomparison of Improved Satellite Rainfall Estimation with Chirps Gridded Product and Rain Gauge Data over Venezuela. *Atmósfera, 29*(4), 323-342. doi:10.20937/atm.2016.29.04.04

Pastorek, J., Fencl, M., Rieckermann, J., & Bares, V. (2019). Commercial Microwave Links for Urban Drainage Modelling: The Effect of Link Characteristics and Their Position on Runoff Simulations. *J Environ Manage, 251*, 109522. doi:10.1016/j.jenvman.2019.109522

Patrick, Stephen. (2020). Microwave Link: Gigabit Microwave Connectivity. Retrieved on 13 July 2020 from https://www.microwave-link.com/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825-2830. Retrieved from <Go to ISI>://WOS:000298103200003

Petković, Veljko, & Kummerow, Christian D. (2017). Understanding the Sources of Satellite Passive Microwave Rainfall Retrieval Systematic Errors over Land. *Journal of Applied Meteorology and Climatology, 56*(3), 597-614. doi:10.1175/jamc-d-16-0174.1

Rahimi, A. R., Upton, G. J. G., & Holt, A. R. (2004). Dual-Frequency Links—a Complement to Gauges and Radar for the Measurement of Rain. *Journal of Hydrology, 288*(1-2), 3-12. doi:10.1016/j.jhydrol.2003.11.008

Rios, Gaona, Aart, Overeem, Hidde, Leijnse, Marc, Bierkens, & Remko, Uijlenhoet. (2012a). Rainfall Field Estimation Using Simulated Microwave Link Information. In: Wageningen University.

Rios Gaona, M. F., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2015). Measurement and Interpolation Uncertainties in Rainfall Maps from Cellular Communication Networks. *Hydrology and Earth System Sciences, 19*(8), 3571-3584. doi:10.5194/hess-19-3571-2015

Rios Gaona, Manuel F., Overeem, Aart, Brasjen, A. M., Meirink, Jan Fokke, Leijnse, Hidde, & Uijlenhoet, Remko. (2017). Evaluation of Rainfall Products Derived from Satellites and Microwave Links for the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing, 55*(12), 6849-6859. doi:10.1109/tgrs.2017.2735439

Rios, Gaona Manuel Felipe, Aart, Overeem, Hidde, Leijnse, Marc, Bierkens, Remko, & Uijlenhoet. (2012b). *Rainfall Field Estimation Using Simulated Microwave Link Information*. Retrieved from http://igitur-archive.library.uu.nl/student-theses/2012-0116-200717

library@uu.nl

Roe, Gerard H. (2005). Orographic Precipitation. *Annual Review of Earth and Planetary Sciences, 33*(1), 645-671. doi:10.1146/annurev.earth.33.092203.122541

Roebeling, R. A., & Holleman, I. (2009). Seviri Rainfall Retrieval and Validation Using Weather Radar Observations. *Journal of Geophysical Research, 114*(D21), 1-13. doi:10.1029/2009jd012102

Rong, Guangzhi, Li, Kaiwei, Han, Lina, Alu, Si, Zhang, Jiquan, & Zhang, Yichen. (2020). Hazard Mapping of the Rainfall–Landslides Disaster Chain Based on Geodetector and Bayesian Network Models in Shuicheng County, China. *Water, 12*(9). doi:10.3390/w12092572

Rosenfeld, Daniel. (2007). Cloud Top Microphysics as a Tool for Precipitation Measurements. In V. Levizzani, P. Bauer, & F. J. Turk (Eds.), *Measuring Precipitation from Space* (Vol. 1, pp. 61-77). Dordrecht: Springer.

Rosenfeld, Daniel, & Gutman, Garik. (1994). Retrieving Microphysical Properties near the Tops of Potential Rain Clouds by Multispectral Analysis of Avhrr Data. *Atmospheric Research, 34*(1-4), 259-283. doi:10.1016/0169-8095(94)90096-5

Rosenfeld, Daniel, & Lensky, Itamar M. (1998). Satellite–Based Insights into Precipitation Formation Processes in Continental and Maritime Convective Clouds. *Bulletin of the American Meteorological Society, 79*(11), 2457-2476. doi:10.1175/1520-0477(1998)079<2457:Sbiipf>2.0.Co;2

Ross, Maidment, David, Grimes, Allan, Richard P., Tarnavsky, Elena, Stringer, Marc, Hewison, Tim, . . . Emily, Black. (2014). The 30 Year Tamsat African Rainfall Climatology and Time Series (Tarcat) Data Set. *Journal of Geophysical Research: Atmospheres, 119*, 10619-10644. doi:10.1002/2014JD021927.Received

Rowe, Robert D. (1976). The Effects of Aggregation over Time on T-Ratios and R 2 'S. *International Economic Review, 17*(3). doi:10.2307/2525801

Safont, G., Salazar, A., & Vergara, L. (2019). Multiclass Alpha Integration of Scores from Multiple Classifiers. *Neural computation, 31*(4), 806-825. doi:10.1162/neco_a_01169

Scheidegger, Andreas, & Rieckermann, Jörg. (2014). *Bayesian Assimilation of Rainfall Sensors with Fundamentally Different Integration Characteristics*. Paper presented at the International Symposium Weather Radar and Hydrology (WRaH), Washington, DC.

Schleiss, Marc, & Berne, Alexis. (2010). Identification of Dry and Rainy Periods Using Telecommunication Microwave Links. *IEEE Geoscience and Remote Sensing Letters, 7*(3), 611-615. doi:10.1109/lgrs.2010.2043052

Schleiss, Marc, Rieckermann, J., & Berne, A. (2013). Quantification and Modeling of Wet-Antenna Attenuation for Commercial Microwave Links. *IEEE Geoscience and Remote Sensing Letters, 10*(5), 1195-1199. doi:10.1109/lgrs.2012.2236074

Schmetz, J., Tjemkes, S. A., Gube, M., & van de Berg, L. (1997). Monitoring Deep Convection and Convective Overshooting with Meteosat. *Advances in Space Research, 19*(3), 433-441. doi:10.1016/s0273-1177(97)00051-3

Schmetz, Johannes, Ackermann, Jörg, Bonekamp, Hans, Clérigh, Eoin Ó, von Engeln, Axel, Luntama, Juha-Pekka, . . . Klaes, K. Dieter. (2007). An Introduction to the Eumetsat Polar System. *Bulletin of the American Meteorological Society, 88*(7), 1085-1096. doi:10.1175/bams-88-7-1085

Schmetz, Johannes, Pili, Paolo, Tjemkes, Stephen, Just, Dieter, Kerkmann, Jochen, Rota, Sergio, & Ratier, Alain. (2002). Supplement to an Introduction to Meteosat Second Generation (Msg). *Bulletin of the American Meteorological Society, 83*(7), 992-992. doi:10.1175/BAMS-83-7-Schmetz-2

Schneider, T., Bischoff, T., & Haug, G. H. (2014). Migrations and Dynamics of the Intertropical Convergence Zone. *Nature, 513*(7516), 45-53. doi:10.1038/nature13636

Sebastianelli, S., Russo, F., Napolitano, F., & Baldini, L. (2013). On Precipitation Measurements Collected by a Weather Radar and a Rain Gauge Network. *Natural Hazards and Earth System Sciences, 13*(3), 605-623. doi:10.5194/nhess-13-605-2013

Sene, Kevin. (2013). *Flash Floods* (1 ed.): Springer Dordrecht.

Seyama, Eric S., Masocha, Mhosisi, & Dube, Timothy. (2019). Evaluation of Tamsat Satellite Rainfall Estimates for Southern Africa: A Comparative Approach. *Physics and Chemistry of the Earth, Parts A/B/C, 112*, 141-153. doi:10.1016/j.pce.2019.02.008

Sieck, Lisa, Steiner, Matthias, Burges, Stephen, Smith, James, & Alonso, Carlos. (2003). *The Challenges of Measuring Rainfall: Observations Made at the Goodwin Creek Research Watershed.* Paper presented at the First Interagency Conference on Research in the Watersheds, Benson, Arizona.

Silver, Micha, Karnieli, Arnon, & Fredj, Erick. (2021). Improved Gridded Precipitation Data Derived from Microwave Link Attenuation. *Remote Sensing*, *13*(15). doi:10.3390/rs13152953

Stengel, M., Kniffka, A., Meirink, J. F., Lockhoff, M., Tan, J., & Hollmann, R. (2014). Claas: The Cm Saf Cloud Property Data Set Using Seviri. *Atmospheric Chemistry and Physics*, *14*(8), 4297-4311. doi:10.5194/acp-14-4297-2014

Strabala, Kathleen I., Ackerman, Steven A., & Menzel, W. Paul. (1994). Cloud Properties Inferred from 8-12-µm Data. *Journal of Applied Meteorology*, *33*(2), 212-229. doi:10.1175/1520-0450(1994)033<0212:Cpifd>2.0.Co;2

Strangeways, Ian. (2010). A History of Rain Gauges. *Weather*, *65*(5), 133-138. doi:10.1002/wea.548

Stuhlmann, Rolf, Holmlund, Kenneth, Schmetz, Johannes, Roquet, Hervé, Tjemkes, Stephen, Grandell, Jochen, . . . Aminou, Donny. (2017). Observations Depuis L'orbite Géostationnaire Avec Meteosat Troisième Génération (Mtg). *La Météorologie*(97). doi:10.4267/2042/62167

Tan, Jackson, Huffman, George J., Bolvin, David T., & Nelkin, Eric J. (2019). Imerg V06: Changes to the Morphing Algorithm. *Journal of Atmospheric and Oceanic Technology*, *36*(12), 2471-2482. doi:10.1175/jtech-d-19-0114.1

Tang, Guoqiang, Ma, Yingzhao, Long, Di, Zhong, Lingzhi, & Hong, Yang. (2016). Evaluation of Gpm Day-1 Imerg and Tmpa Version-7 Legacy Products over Mainland China at Multiple Spatiotemporal Scales. *Journal of Hydrology*, *533*, 152-167. doi:10.1016/j.jhydrol.2015.12.008

Tazalika, Lukiya, Stordal, Frode, Büchner, Matthias, Panitz, Hans-Jürgen, Nikulin, Grigory, Ketiem, Patrick, . . . Endris, Hussen Seid. (2013). Assessment of the Performance of Cordex Regional Climate Models in Simulating East African Rainfall. *Journal of Climate*, *26*(21), 8453-8475. doi:10.1175/jcli-d-12-00708.1

Thies, B., & Nauss, T. (2008). *A New Technique for Detecting Precipitation at Mid-Latitudes During Daytime Using Meteosat Second Generation Seviri*. Retrieved from Darmstadt, Germany: http://www.eumetsat.int/groups/cps/documents/document/pdf_conf_p_s10_60_thies_v.pdf

Thies, B., Nauss, T., & Bendix, J. (2008a). Discriminating Raining from Non-Raining Cloud Areas at Mid-Latitudes Using Meteosat Second

Generation Seviri Night-Time Data. *Meteorological Applications, 15*(2), 219-230. doi:10.1002/met.56

Thies, B., Nauss, T., & Bendix, J. (2008b). Discriminating Raining from Non-Raining Clouds at Mid-Latitudes Using Meteosat Second Generation Daytime Data. *Atmospheric Chemistry and Physics, 8*(9), 2341-2349. doi:10.5194/acp-8-2341-2008

Thies, B., Nauss, T., & Bendix, J. (2008c). First Results on a Process-Oriented Rain Area Classification Technique Using Meteosat Second Generation Seviri Nighttime Data. *Advances In Geosciences, 16*(April), 63-72. doi:10.5194/adgeo-16-63-2008

Thies, Boris, Nauß, Thomas, & Bendix, Jörg. (2008d). Precipitation Process and Rainfall Intensity Differentiation Using Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager Data. *Journal of Geophysical Research, 113*(D23), D23206-D23206. doi:10.1029/2008jd010464

Thurai, M., Petersen, W. A., Tokay, A., Schultz, C., & Gatlin, P. (2011). Drop Size Distribution Comparisons between Parsivel and 2-D Video Disdrometers. *Advances In Geosciences, 30*, 3-9. doi:10.5194/adgeo-30-3-2011

Todini, E. (2001). A Bayesian Technique for Conditioning Radar Precipitation Estimates to Rain-Gauge Measurements. *Hydrology and Earth System Sciences, 5*(2), 187-199. doi:10.5194/hess-5-187-2001

Turdukulov, Ulanbek D., Kraak, Menno-Jan, & Blok, Connie A. (2007). Designing a Visual Environment for Exploration of Time Series of Remote Sensing Data: In Search for Convective Clouds. *Computers & Graphics, 31*(3), 370-379. doi:10.1016/j.cag.2007.01.028

Turini, Nazli, Thies, Boris, Horna, Natalia, & Bendix, Jörg. (2021). Random Forest-Based Rainfall Retrieval for Ecuador Using Goes-16 and Imerg-V06 Data. *European Journal of Remote Sensing, 54*(1), 117-139. doi:10.1080/22797254.2021.1884002

Uijlenhoet, R., & Berne, A. (2008). Stochastic Simulation Experiment to Assess Radar Rainfall Retrieval Uncertainties Associated with Attenuation and Its Correction. *Hydrology and Earth System Sciences, 12*(2), 587-601. doi:10.5194/hess-12-587-2008

Uijlenhoet, Remko, Leijnse, Hidde, & Berne, Alexis. (2010). Errors and Uncertainties in Microwave Link Rainfall Estimation Explored Using

Drop Size Measurements and High-Resolution Radar Data. *Journal of Hydrometeorology, 11*(6), 1330-1344. doi:10.1175/2010jhm1243.1

Uijlenhoet, Remko, Overeem, Aart, & Leijnse, Hidde. (2018). Opportunistic Remote Sensing of Rainfall Using Microwave Links from Cellular Communication Networks. *Wiley Interdisciplinary Reviews-Water, 5*(4). doi:10.1002/wat2.1289

Upton, G. J. G., Holt, A. R., Cummings, R. J., Rahimi, A. R., & Goddard, J. W. F. (2005). Microwave Links: The Future for Urban Rainfall Measurement? *Atmospheric Research, 77*(1-4), 300-312. doi:10.1016/j.atmosres.2004.10.009

Uttal, Taneil, Heidinger, Andrew K., & Pavolonis, Michael J. (2005). Daytime Global Cloud Typing from Avhrr and Viirs: Algorithm Description, Validation, and Comparisons. *Journal of Applied Meteorology, 44*(6), 804-826. doi:10.1175/jam2236.1

van de Giesen, Nick, Hut, Rolf, & Selker, John. (2014). The Trans-African Hydro-Meteorological Observatory (Tahmo). *Wiley Interdisciplinary Reviews-Water, 1*(4), 341-348. doi:10.1002/wat2.1034

van het Schip, T. I., Overeem, A., Leijnse, H., Uijlenhoet, R., Meirink, J. F., & van Delden, A. J. (2017). Rainfall Measurement Using Cell Phone Links: Classification of Wet and Dry Periods Using Geostationary Satellites. *Hydrological Sciences Journal, 62*(9), 1343-1353. doi:10.1080/02626667.2017.1329588

Villarini, Gabriele, & Krajewski, Witold F. (2009). Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics, 31*(1), 107-129. doi:10.1007/s10712-009-9079-x

Villarini, Gabriele, Mandapaka, Pradeep V., Krajewski, Witold F., & Moore, Robert J. (2008). Rainfall and Sampling Uncertainties: A Rain Gauge Perspective. *Journal of Geophysical Research, 113*(D11). doi:10.1029/2007jd009214

Wakachala, Francis M., Shilenje, Zablon W., Nguyo, John, Shaka, Saumu, & Apondo, William. (2015). Statistical Patterns of Rainfall Variability in the Great Rift Valley of Kenya. *Journal of Environmental and Agricultural Sciences, 5*, 17-26.

Walther, Bruno A., & Moore, Joslin L. (2005). The Concepts of Bias, Precision and Accuracy, and Their Use in Testing the Performance of Species Richness Estimators, with a Literature Review of Estimator

Performance. *Ecography, 28*(6), 815-829. doi:10.1111/j.2005.0906-7590.04112.x

Wang, Jingyu, Houze, Jr Robert A., Fan, Jiwen, Brodzik, Stacy R., Feng, Zhe, & Hardin, Joseph C. (2019). The Detection of Mesoscale Convective Systems by the Gpm Ku-Band Spaceborne Radar. *Journal of the Meteorological Society of Japan. Ser. II, 97*(6), 1059-1073. doi:10.2151/jmsj.2019-058

Wang, Z., Schleiss, M., Jaffrain, J., Berne, A., & Rieckermann, J. (2012). Using Markov Switching Models to Infer Dry and Rainy Periods from Telecommunication Microwave Link Signals. *Atmospheric Measurement Techniques, 5*(7), 1847-1859. doi:10.5194/amt-5-1847-2012

Watts, P. D., Bennartz, R., & Fell, F. (2011). Retrieval of Two-Layer Cloud Properties from Multispectral Observations Using Optimal Estimation. *Journal of Geophysical Research, 116*(D16), 1-22. doi:10.1029/2011jd015883

Wilheit, Thomas T., Chang, Alfred T. C., & Chiu, Long S. (1991). Retrieval of Monthly Rainfall Indices from Microwave Radiometric Measurements Using Probability Distribution Functions. *Journal of Atmospheric and Oceanic Technology, 8*(1), 118-136. doi:10.1175/1520-0426(1991)008<0118:Romrif>2.0.Co;2

Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences* (Vol. 14).

Wilson, James W., & Brandes, Edward A. (1979). Radar Measurement of Rainfall— a Summary. *Bulletin of the American Meteorological Society, 60*(9), 1048-1058. Retrieved from http://www.jstor.org/stable/26219220

Wolfensberger, Daniel, Gabella, Marco, Boscacci, Marco, Germann, Urs, & Berne, Alexis. (2021). Rainforest: A Random Forest Algorithm for Quantitative Precipitation Estimation over Switzerland. *Atmospheric Measurement Techniques, 14*(4), 3169-3193. doi:10.5194/amt-14-3169-2021

Wolff, David B., Nelkin, Eric J., Bolvin, David T., Huffman, George J., Adler, Robert F., Gu, Guojun, . . . Stocker, Erich 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. doi:10.1175/jhm560.1

Yan, H., Liang, C., Li, Z., Liu, Z., Miao, B., He, C., & Sheng, L. (2015). Impact of Precipitation Patterns on Biomass and Species Richness of Annuals in a

Dry Steppe. *PLoS One, 10*(4), e0125300. doi:10.1371/journal.pone.0125300

Yan, Hao, & Yang, Song. (2007). A Modis Dual Spectral Rain Algorithm. *Journal of Applied Meteorology and Climatology, 46*(9), 1305-1323. doi:10.1175/jam2541.1

Yan, Jieru, Li, Fei, Bárdossy, András, & Tao, Tao. (2021). Conditional Simulation of Spatial Rainfall Fields Using Random Mixing: A Study That Implements Full Control over the Stochastic Process. *Hydrology and Earth System Sciences, 25*(7), 3819-3835. doi:10.5194/hess-25-3819-2021

Young, Alisa H., Bates, John J., & Curry, Judith A. (2013). Application of Cloud Vertical Structure from Cloudsat to Investigate Modis-Derived Cloud Properties of Cirriform, Anvil, and Deep Convective Clouds. *Journal of Geophysical Research: Atmospheres, 118*(10), 4689-4699. doi:10.1002/jgrd.50306

Zhang, W., Furtado, K., Wu, P., Zhou, T., Chadwick, R., Marzin, C., . . . Sexton, D. (2021). Increasing Precipitation Variability on Daily-to-Multiyear Time Scales in a Warmer World. *Sci Adv, 7*(31). doi:10.1126/sciadv.abf8021

Zinevich, A., Messer, H., & Alpert, P. (2010). Prediction of Rainfall Intensity Measurement Errors Using Commercial Microwave Communication Links. *Atmospheric Measurement Techniques, 3*(5), 1385-1402. doi:10.5194/amt-3-1385-2010

Zinevich, Artem, Alpert, Pinhas, & Messer, Hagit. (2008). Estimation of Rainfall Fields Using Commercial Microwave Communication Networks of Variable Density. *Advances in Water Resources, 31*(11), 1470-1480. doi:10.1016/j.advwatres.2008.03.003

Zipser, E. J., Cecil, Daniel J., Liu, Chuntao, Nesbitt, Stephen W., & Yorty, David P. (2006). Where Are the Most Intense Thunderstorms on Earth? *Bulletin of the American Meteorological Society, 87*(8), 1057-1072. doi:10.1175/bams-87-8-1057

Author's biography and list of publications

Kwabena Kingsley Kumah was born in Kumasi, Ghana, on December 15, 1987. He acquired his BSc in Environmental Science from the University of Cape Coast with a final year thesis that researched the suitability of two methods in determining the point of zero charge of adsorbent–a research that is of fundamental relevance in surface science. Kwabena began his MSc (2014) and PhD (2016) at the University of Twente, Faculty of Geo-information Science and Earth Observation, Department of Water Resources. His final year theses investigated, developed and validated new methods for retrieving rainfall's spatiotemporal state from commercial telecommunication networks and geostationary satellite data. His research interests lie in spatial rainfall detection, estimation, and mapping from ground and satellite data. Specifically, he focuses on the relationship between the satellite's spectral characteristics, cloud microphysical properties, and rainfall. He employs parametric (with underlying conceptual models) and machine learning techniques to achieve accurate rainfall retrieval. He also aims for accurate near-real-time (extreme) rainfall estimation from convective systems. Kwabena's research has been thoroughly reviewed by experts in environmental and atmospheric sciences and published in leading scientific journals such as Atmospheric Research, Remote Sensing, Sensors and more. Also, he has contributed to online articles and participated in conferences, symposiums and workshops (see below). His research works are also featured (together with other experts) in the large circulation magazine – 'Physics' of the American Physical Society (https://physics.aps.org/articles/v14/161).

Publications

ISI Journal articles

Kumah, K. K., Hoedjes, J. C. B., David, N., Maathuis, B. H. P., Gao, H. O., & Su, B. Z. (2020). Combining MWL and MSG SEVIRI Satellite Signals for Rainfall Detection and Estimation. *Atmosphere, 11*(9). doi:10.3390/atmos11090884

Kumah, K. K., Hoedjes, J. C. B., David, N., Maathuis, B. H. P., Gao, H. O., & Su, B. Z. (2021a). The MSG Technique: Improving Commercial Microwave Link Rainfall Intensity by Using Rain Area Detection from Meteosat Second Generation. *Remote Sensing, 13*(16). doi:10.3390/rs13163274

Kumah, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., Rwasoka, D. T., Retsios, B. V., & Su, B. Z. (2021b). Rain Area Detection in South-Western Kenya by Using Multispectral Satellite Data from Meteosat Second Generation. *Sensors (Basel), 21*(10). doi:10.3390/s21103547

David, N., Liu, Y., **Kumah, K. K.**, Hoedjes, J. C. B., Su, B. Z., & Gao, H. O. (2021). On the Power of Microwave Communication Data to Monitor Rain for Agricultural Needs in Africa. *Water, 13*(5), 730. doi:10.3390/w13050730

Kumah, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., & Su, Z. (2022). Near real-time estimation of high spatiotemporal resolution rainfall from cloud top properties of the MSG satellite and commercial microwave link rainfall intensities. Atmospheric Research, 279. doi:10.1016/j.atmosres.2022.106357

Manuscript(s) deposited at a not-for-profit preprint server

K. K. Kumah, B. H. P. Maathuis, J.C.B. Hoedjes and Z. Su. Near Real-Time Estimation of High Spatiotemporal Resolution Rainfall from Cloud Top Properties of the MSG satellite and Commercial Microwave Link Rainfall Intensities. ESSOAr [Preprint]. April 6, 2022. Retrieved from https://www.essoar.org/doi/10.1002/essoar.10511022.1, on May 2, 2022

Conference paper (peer-reviewed)

David, N., Gao H Oliver, **Kumah, K.K**, Hoedjes C.B. Joost, Su Z., & Liu, Y. (2019). Microwave communication networks as a sustainable tool of rainfall monitoring for agriculture needs in Africa. Paper presented at the 16th International Conference on Environmental Science and Technology, Rhodes, Greece.

Conferences, Symposium and workshop presentations

Kumah, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., & Su, Z. (2022). High Spatial and Temporal Resolution Rainfall Estimation from Cloud Top Properties of MSG Using Random Forest Algorithm Trained with Microwave Link Derived Rainfall. Presented at EUMETSAT Meteorological Satellite Conference 2022, Brussels, Belgium. https://program-eumetsat2022.kuoni-congress.info/2022-09-23

Kumah, K.K., Maathuis, B. H. P., Hoedjes, J. C. B., & Su, B. Z. (2017). Rainfall Estimation using Signal from Telecommunication Networks and Geostationary Satellite Data- Potential of satellite data for wet/dry classification of link signals Presented at the BOUSSINESQ LECTURE 2017: KUMMEROW, Delft, Netherlands. https://boussinesqcenter.nl/2019/03/01/boussinesq-lecture-2017-kummerow/

Kumah, K. K., Maathuis, B. H. P., Hoedjes, J. C. B., & Su, B. Z. (2019). Opportunistic use of commercial microwave links and Meteosat Second Generation data for rainfall studies in Kenya. Presented at Symposium on the hydrometeorological usage of data from commercial microwave link networks, Garmisch-Partenkirchen. https://indico.scc.kit.edu/event/570/sessions/1547/

Online article

Kingsley, K. K. (2016). Mobile Telecommunication Network Signals Could Improve Rainfall Observation in East Africa. [Online article] Retrieved from https://www.prescouter.com/2016/07/telecommunication-signals-help-improve-rainfall-observation-east-africa/