Near real-time estimation of high spatiotemporal resolution rainfall from cloud top properties of the MSG satellite and commercial microwave link rainfall intensities

K.K. Kumah, B.H.P. Maathuis *, J.C.B. Hoedjes, Z. Su

Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, the Netherlands

ARTICLE INFO

Keywords:
Commercial Microwave Links
Meteosat Second Generation
High Spatiotemporal Resolution Rainfall Estimation

ABSTRACT

High spatiotemporal resolution rainfall is needed in predicting flash floods, local climate impact studies and agriculture management. Rainfall estimation techniques like satellites and the 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, Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) and European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Multisensor Precipitation Estimate (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.

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 is dependent 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).

Commercial microwave links (MWL) used by commercial telecom service providers for data transmission are capable of rainfall estimation (David et al. 2021; Leijnse et al. 2007; 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. 2016; 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 variation of raindrop sizes 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. 2021a; 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 and Berne 2008; Yan et al. 2021).

Satellites are spaceborne in low earth or geostationary (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 spatial range consisting of different kinds of 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 and Holleman 2009; Thies et al. 2008). Other retrieval techniques relate the MSG’s spectral features to cloud top properties and rainfall (Feidas and Giannakos 2010; Kumah et al. 2021b).

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 (Kumah et al. 2021b). However, the nonlinear and complex relationship between cloud top property and rainfall may be beyond the skill of parametric tests and conceptual models (Kühlein 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 nonlinear relations may be suitable for retrieving rainfall from the multivariate satellite data to overcome the limitations of the parametric techniques (Hu et al. 2019; Kühlein et al. 2014b). Several studies have successfully used machine learning algorithms such as the RF, artificial neural networks and deep-learning models for spatial rainfall estimation (Kühlein et al. 2014b; Laze et al. 2014; Meyer et al. 2016; Moraux et al. 2019). In particular, the RF machine learning algorithm (Breiman 2001) 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 nonlinear relations between predictor and target variables (Kühlein et al. 2014b). However, most of its applications to MSG data, such as (Kühlein 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 and MSG coverage.

Therefore, this study’s objective 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 such as those of (Kühlein et al. 2014b; Meyer et al. 2016), this study’s uniqueness is due to the following reasons:

1. We applied the RF algorithm for rainfall estimation in a topographically complex area in the Kenya Rift Valley, where gauge data is scarce.
2. For the first time, we trained the RF algorithm using MWL rainfall as the target variable.

2. Study area and dataset

Fig. 1 shows the study area in Kenya using ALOS World 3D 30 m (AW3D30) DEM (Caglar et al. 2018) to visualise 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 short 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 and 1525 mm (Kumah et al. 2021b; 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 The Trans-African Hydro-Meteorological Observatory (TAHMO) (van de Giesen et al. 2014) were available as 5 min rainfall accumulations. These computed the 15-min 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 Fig. 1 and illustrate a sparse distribution of ground data in the study area.

Safaricom provided the received signal levels (RLS) data for the set of MWL with arbitrary geometry (extending from areas close to the Aberdare mountains to Lake Naivasha in the centre of Fig. 1), frequency and length in the study area shown in Fig. 1. For the 2014 and 2019 periods, data from multiple MWL were available. In contrast, for the 2018 period, 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-min intervals and a 0.1 dBm resolution. This study focused on using the 15 and 23 GHz MWL for rainfall estimation because, at such frequencies, there is a nearly linear relationship between the attenuation and rainfall, which is less affected by non-uniform raindrop size distribution along the MWL’s signal transmission path (Atlas and Ulbrich 1977; Zinevich et al. 2010).

The infrared (IR) (IR10.8 μm and IR12.0 μm) and water vapour (WV) (WV6.2 μm and WV7.3 μm) channels used in this study were from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) radiometer onboard the Meteosat at 0° (2014 period) and 41.5° E (2018 2019 period). This corresponded to Meteosat 10 and 8 satellites, respectively (EUMETSAT 2016), when the data was acquired from (EUMETSAT 2020) at 3.5 km and 15 min spatial and temporal resolution. This spatial resolution is preserved over the study area. The IR and WV channels used in this study are sensitive to cloud top properties such as cloud top temperature and height. The data from the Meteosat at 0° were parallax corrected because of the satellite viewing angle, which causes displacement in the actual position of cloud tops depending on their location and height (Kumah et al. 2020; Roebeling and 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 from the MSG satellites’ repeat cycle, currently Meteosat 8, from the thermal IR channel. The MPE...
Algorithm relies on a weather-dependent monotonic function that relates the IR brightness temperatures to the passive microwave (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 h and in 5° × 5° 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 and Kerényi 2003). This study retrieved MPE data from EUMETSAT (2020) at 15 min and 3 × 3 km resolution for the evaluation period.

The IMERG version 6 (V06B) product is a level 3 globally gridded satellite precipitation estimate derived by intercalibrating, merging and interpolating precipitation estimates from several GPM constellation satellites and microwave calibrated IR estimates. The IMERG algorithm is run twice in near-real-time (Early (~4 h after observation time) and Late (~14 h after observation time)), and once (Final (~3.5 months after the observation time)), based on user requirements for latency and accuracy (Huffman et al. 2019; Jackson et al. 2019). The Final run product is calibrated using precipitation analysis from Global Precipitation Climatology Centre (GPCC) and the European Centre for Medium-Range Weather Forecasts (ECMWF) ancillary data, making it more reliable and suited for research (Moazami and Najafi 2021).

IMERG data has 4 precipitation fields: UnCalibrated precipitation (precipitationUnCal), Calibrated precipitation (precipitationCal), Infrared (IR) precipitation (IRprecipitation) and High-Quality precipitation (HQprecipitation). The precipitationUnCal and precipitationCal represent records before and after the gauge calibration post-processing step. The IRprecipitation is IR geostationary satellite-based precipitation, whereas the HQprecipitation is obtained by merging High-Quality Passive Microwave (PMW) precipitation estimates. This study utilised the Final run precipitationCal product because it is a research-grade product that is climatologically adjusted using GPCC ground data. Moreover, previous studies in the study area (Kumah et al. 2021b) 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 21 March 2022) at approximately 0.1° × 0.1° and 30 min of spatiotemporal resolution.

3. Method

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. Fig. 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 identified in step 1.

In this study, steps 1 and 2 were based on techniques described in previous studies (Kumah et al. 2020, 2021a; Kumah et al. 2021b), and step 3 was by using the RF algorithm. These steps agree with the optical rainfall retrieval approach that separates the rain area detection from rainfall retrievals (Bendix et al. 2010).

### 3.1.1. Detecting rain areas

The rain area identification technique was based on the approach described in (Kumah et al. 2021b). It relies on a parametric threshold model based on the conceptual idea 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 (2003). 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.

#### 3.1.2. Estimating rainfall intensities from the MWL data

This MWL rainfall estimation method is described in detail in (Kumah et al. 2020, 2021a); this section summarises the method. The approach retrieves rainfall intensity estimates from the mean MLS 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 MLS of the previous 24 h labelled as dry periods by the wet and dry classification step. Finally, the mean MLS data is corrected for the effect of the wet antenna (Schleiss et al. 2013) before retrieving attenuation and rainfall from eqs. 1 and 2, respectively.

\[
A = \frac{R - P}{L} \\
R = \left( \frac{A}{a} \right) \beta
\]

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 MLS, 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 2005) for the 15,23 GHz MWL, respectively.

#### 3.1.3. Estimating spatial rainfall intensities using RF

**3.1.3.1. Predictors 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 using 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 and Meijer 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. 2008), 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 1. The spectral features were derived from SEVIRI channels and differences. They are consistent with those used by previous studies (Kühnele et al. 2014b; Kumah et al. 2021b) to infer cloud top properties such as cloud top temperature and height for rain area and rain rate retrievals.

The gradient features indicate pixel gradients in the cloud top properties. This was computed based on the method described in (Kumah et al. 2021b). 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, characterised 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 average cloud patch pixel gradient to determine these distinct characteristics and improve the retrieved rainfall estimate.

**3.1.3.2. Compiling training and validation datasets.** This study utilised common machine learning techniques consisting of training and validation 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 set assessed the model’s ability to generalise 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 MWL and the corresponding MSG pixels covering the MWL that are shown in Fig. 3 for the raining (\( R > 1 \text{ mm h}^{-1} \)) and non-raining (\( R < 1 \text{ mm h}^{-1} \)) 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 Fig. 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.

### Table 1

| Spectral features | Pixel gradient features |
|-------------------|-------------------------|
| Cloud top properties | IR10.8 K | ΔIR10.8 K |
| Channels and differences | IR10.8–WV6.2 K | ΔIR10.8–WV6.2 K |
| Cloud top height | IR12.0–WV7.3 K | ΔIR12.0–WV7.3 K |
| Cloud height | WV6.2–WV7.3 K | ΔWV6.2–WV7.3 K |
3.1.3.3. 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 optimisation 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 tree node, 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 (Kühnlein et al. 2014a; Kühnlein et al. 2014b; 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). This package has over a dozen parameters to adjust to achieve a robust RF performance. However, this study focused on the number of decision trees (\( n_{\text{estimators}} \)) 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. Oversampling the minority class and downsampling the majority class are some approaches to balance the class distribution. 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 eq. 3. The model performance for the respective tuning values is the average of the MSEs from the hold one-out iterations.

\[
MSE = \frac{1}{N} \sum_{i=1}^{N-1} (R_{ti} - R_{\text{ref}})^2
\]

where:
- \( R_{\text{ref}} \) —represents all possible RF rainfall intensity estimates.
- \( R_{ti} \) —represents all possible target variable observations, and \( N \) is the number of samples.

The number of decision trees, \( n_{\text{estimators}} \), to grow in the forest is an important parameter to consider. According to Breiman (2001), the generalisation 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, \( n_{\text{estimators}} \) should be optimised to obtain a computationally feasible value. To determine the optimal value of the \( n_{\text{estimators}} \) parameter, many RF models were created using the training data for all possible values of \( n_{\text{estimators}} \) and max features. The maximum \( n_{\text{estimators}} \) were 2000 trees, whereas the max_features values ranged from 3 to 9 representing 30% to 90% of the total number of input features.

Fig. 3 exemplarily shows the effect of the number of trees with 3, 5, and 9 max_feature values on 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.
different max_features values. Based on these results, max_features
parameter to 100 in each scenario. Fig. 4 b presents the
- increases both. In practice, the max_features value is often treated as a
tuning parameter ( Kühnlein et al. 2014b ). To determine the optimal
- increases. The max_features parameter affects these two aspects such that reducing
the correlation and strength, whereas increasing it in
- increases the RF error rate. Breiman (2001) shows that the RF error rate largely depends on the
- increases 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
the correlation and strength, whereas increasing it in
increases both. In practice, the max_features value is often treated as a
tuning parameter ( 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, represen-
ting 30% to 90% of the number of input features while setting the
n_estimators parameter to 100 in each scenario. Fig. 4 b 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.

3.1.3.4. 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) and relative bias (RB) ( Kumah et al. 2020; Wilks
2006 ) described in eqs. 4 and 5 evaluated the RF model performance.

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

(4)

$$RB = \frac{\sum_{i=1}^{N} (R_{rfi} - R_{oi})}{\sum_{i=1}^{N} R_{oi}}$$

(5)

where:

- $R_{rfi}$ —represents all possible RF rainfall estimates.
- $R_{oi}$ —represents all possible gauge, IMERG and MPE rainfall esti-
mates, 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 estimate to assess the capability of the RF to
estimate path average rainfall intensities;
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 min and 3×3 km resolution. Also, the spatial and temporal
mismatch in the dataset was considered to ensure a comparison of
collocated rainfall intensity estimates. For this, the IMERG estimates
were spatially resampled using the nearest neighbourhood technique
that preserves the pixel values to the RF and MPE spatial resolution. On
the other hand, the gauge, RF, and MPE estimates were temporally
aggregated to IMERG’s 30 min temporal resolution by summing their
respective rainfall intensity estimates.

4. Results and discussion

4.1. Results

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 evalua-
tion is presented through visual and statistical analysis of the RF esti-
mates compared to gauge, MPE, and IMERG for rainfall events observed
from two locations in the study area. Secondly, a performance evalua-
tion 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.

Fig. 5 presents rainfall intensities of two rainfall events captured by
the RF, MPE, IMERG, and rain gauges. The gauge estimates are from
gauges TA00578 ( Fig. 5a ) and TA00986 ( Fig. 5b ), shown in brackets in
Fig. 5, situated at different locations within the study area. The RF, MPE,
and IMERG are estimates retrieved from the pixels containing the two
gauges. Fig. 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

Fig. 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; whis-
kers (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.
event. Moreover, the peak rainfall intensity captured by IMERG was above 30 mm per 30-min intervals, whereas the RF and MPE were comparable and below 30 mm per 30-min, compared to the gauge’s peak rainfall intensity below 10 mm per 30-min interval.

The rainfall events in Fig. 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. Also, the RF and MPE captured two comparable peaks below 25 mm per 30-min. However, IMERG’s event extends beyond 18:00 UTC and its peak rain intensity, like the gauge,
was below 15 mm per 30-min.

Fig. 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 (Fig. 6a) estimates were about 5 mm per 30-min, 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 (Fig. 6b) and MPE (Fig. 6c). Nonetheless, its outliers mostly below 30 mm per 30-min (Fig. 6a) compared to those of the IMERG and MPE, which mainly were below 50 mm per 30-min, may point to differences in their high rainfall intensity estimates.

Fig. 7 shows the density distribution of collocated rainfall intensity estimates from the gauge, RF, IMERG and MPE. The distribution of rainfall intensities in Fig. 7a suggests that compared to IMERG and MPE, the RF mostly overestimates the gauge rainfall intensities below 15 mm per 30-min. When this distribution is compared with that in Fig. 7b, it is evident that the RF underestimates the high rainfall intensities, judging by its estimates largely below 30 mm per 30-min. 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 prediction from all trees, which may not be a fair representation of the full range of the area’s rainfall estimates.

The better agreement between the RF’s mean and median and the MWL rainfall in intensities is because both represent average intensities over the MWL’s path. Furthermore, discrepancies in RF and MWL estimates that contribute to errors in Table 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.

4.1.2. Comparing the RF and MWL rainfall intensity estimates

We next compared the RF rainfall intensity estimates with estimates from independent MWL RSL data to assess the RF’s capability of path average rainfall estimates. Here, the RF’s mean, median, and maximum rainfall intensity over the MWL are included in the comparison to provide an idea of the range of rainfall intensities estimated by the RF over the MWL’s path and how it compares with the MWL’s rainfall estimates. Table 2 presents descriptive statistics of the absolute errors when comparing the RF’s mean, median, and maximum rainfall intensities to the MWL’s estimates, computed based on 920 15-min rainfall intensity data.

On average, absolute errors of the mean and median versus MWL rainfall intensities are around 4 mm h⁻¹ compared to about 7 mm h⁻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 75th 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 and 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 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.

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 is computed based on all rainfall intensity estimates by the RF, IMERG, and MPE during the validation, excluding the 0 mm h⁻¹ estimates.

Fig. 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.

| Absolute errors of RF versus MWL rainfall intensity estimates | 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     |

Fig. 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.
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°, which the RF and IMERG underestimated. Overall, it can be stated based on visual inspection of Fig. 8 that the rain areas in the RF are comparable to IMERG but with intensities that compare better with the MPE.

These discrepancies in rain areas and intensities in Fig. 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 (Kumah et al. 2021b), 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 intensity differences between the RF, IMERG, and MPE. On the other hand, IMERG uses spatiotemporal average rainfall from multiple microwave estimates, which may explain its low rainfall intensities in Fig. 8 compared to the RF and MPE.

Fig. 9 shows the spatial variability of MAE (Fig. 9a,b) and RB (Fig. 9c,d) computed from RF versus IMERG (Fig. 9a,c) and RF versus MPE (Fig. 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-min and 3, compared to the values of RF versus MPE, which were below 8 mm per 30-min 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 also Fig. 1).

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 (Kumah et al. 2021b) 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 min, 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-min), whereas the high intensities (above 30 mm per 30-min) are underestimated. On average, when compared to gauge estimates, the absolute errors were about 5 mm per 30-min, comparable to the IMERG and MPE versus gauge estimates, suggesting a good 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 higher 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 both representing 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 that 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 sensors and 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 and MPE represent average estimates of 3 × 3 km area, whereas the IMERG estimate represents a 10 × 10 km area. Thus, they are more likely to capture a rainfall event, though their intensity estimates may differ based on the measurement technique.

Additionally, the differences in the measurement techniques used by the RF, gauge, IMERG, and MPE may also explain the discrepancies in their rainfall estimates. For instance, the gauge rain gauge records rainfall accumulations continuous in time. The accuracy of the MWL rainfall data that trained (using tuned parameters) the RF model is affected by various 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 localised. Therefore, the MPE is likely to miss non-convective rainfall events. IMERG is a multisensor 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 cloud and rainfall formation (Adhikari and Behrangi 2022; Kumah et al. 2021b; Petković and 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 rainfall products, particularly in the many ungauged areas. The reason is that our rainfall intensity retrievals technique relies on MWL and MSG satellite data already existing in vast areas, including regions with sparse ground rainfall monitoring systems.

5. Conclusions

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-min, 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 is 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 (Kumah et al. 2021b) is transferable to other study areas and 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.

Overall, this study’s results demonstrate the potential of MWL and MSG data in a machine learning framework for high spatiotemporal rainfall retrievals. This is particularly beneficial for several applications since the MWL and geostationary satellites with SEVIRI capabilities like on MSG provide global data.

CRediT author statement

K. K. Kumah: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Software, Writing - original draft preparation. B. H. P. Maathuis: Data Curation, Writing - review and editing, Formal analysis, Investigation, Supervision, Z. Su: Writing -
review and editing, Formal analysis, Investigation, Supervision. J.C. B. Hoedjes: Data Curation, Supervision.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

This study was conducted within the framework of a PhD study conducted at the Department of Water Resources, Faculty of Geo-Information Science and earth Observation (ITC), University of Twente. The authors appreciate Naftaly Muhoro and Maurice Miriti of Safaricom for providing the MWL RSL data used in this study. We are also grateful to TAHMO for providing the study’s gauge station data.

References

Adhikari, A., Behrangi, A., 2022. Assessment of satellite precipitation products in relation with orographic enhancement over the Western United States. Earth Space Sci. 9, 1292–1307.

Arkin, P., Meinen, B., 1987. The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84. J. climatol. 7, 51–74.

Atmospheric Research 279 (2022) 106357

 Atlas, D., Ulbrich, C.W., 1977. Path- and area-integrated rainfall measurement by microwave attenuation in the 1–3 cm band. J. Appl. Meteorol. 16, 1232–1331.

Bendix, J., Nauß, T., Thies, B., Kühnlein, M., 2010. Rainfall-rate assignment using MSG SEVIRI data—a promising approach to spaceborne rainfall-rate retrieval for midlatitudes. J. Appl. Meteorol. Climatol. 49, 1477–1495.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Caglar, B., Becer, K., Meikı, C., Oterdi, M., 2018. On the vertical accuracy of the ALOS world 3D-30m digital elevation model. Remote Sens. Lett. 9, 607–615.

David, N., et al., 2021. On the power of microwave communication data to monitor rain for agricultural needs in Africa. WaterSA 15, 730.

EUMETSAT, 2016. Meteosat-8 satellite’s new position of 41.5E provides weather and climate view over the Indian Ocean, EUMETSAT, 2020. Archived Meteosat data. EUMETSAT.

Feidas, H., Giannakos, A., 2010. Identifying precipitating clouds in Greece using multispectral infrared Meteosat Second Generation satellite data. Theor. Appl. Climatol. 104, 25–42.

Gao, X., Sorooshian, S., Hsu, K.-L., Hong, Y., 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. J. Appl. Meteorol. 43, 1834–1853.

Gyasi-Agyei, Y., 2020. Identification of the optimum rain gauge network density for hydrological modelling based on radar rainfall analysis. WaterSA 12, 2852–2850.

Pedregosa, F., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Petrovíc, V., Kummerow, C.D., 2017. Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land. J. Appl. Meteorol. Climatol. 56, 597–614.

Roebeling, R.A., Holleman, I., 2009. SEVIRI rainfall retrieval and validation using weather radar observations. J. Geophys. Res. 114, 1–13.

Schleiss, M., Rieckermann, J., Berne, A., 2013. Quantification and modeling of wet-antenna attenuation for commercial microwave links. IEEE Geosci. Remote Sens. Lett. 10, 1195–1199.

Silver, M., Karnieli, A., Fredj, E., 2021. Improved gridded precipitation data derived from microwave link attenuation. Remote Sens. 13, 1–14.

Thies, B., Nauß, T., Bendix, J., 2008. Precipitation process and rainfall intensity differentiation using Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager data. J. Geophys. Res.-Atmos. 113, D23206–D23206.

Turini, N., Thies, B., Horna, N., Bendix, J., 2021. Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data. Eur. J. Remote Sens. 54, 117–139.

van de Giesen, N., Hut, R., Selker, J., 2014. the trans-african hydro-meteorological observatory (TAHMO). Water Resour. Res. 50, 137–155.

Wakachala, F.M., Shilenje, Z.W., Nguyo, J., Shaka, S., Apondo, W., 2015. Statistical differentiation using Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager images. Atmos. Res. 166, 139–155.

Wen, J., 2006. Statistical Methods in the Atmospheric Sciences, pp. 627–627.

Wolfensberger, D., Gabella, M., Bosacci, M., Germann, U., Berne, A., 2021. Random forest algorithm for quantitative precipitation estimation over Switzerland. Atmos. Meas. Tech. 14, 3193–3199.

Yang, X., 2013. Conditional simulation of spatial rainfall fields using random mixing: a study that implements full control over the stochastic process. Hydrolog. Earth Syst. Sci. 17, 3835–3835.

Zinoviev, A., Alpert, P., Messer, H., 2008. Estimation of rainfall fields using commercial microwave communication networks of variable density. Adv. Water Resour. 31, 1470–1480.

K.K. Kumah et al.

Lazri, M., Amour, S., Mohia, Y., 2014. Instantaneous rainfall estimation using neural network from multispectral observations of SEVIRI radiometer and its application in estimation of daily and monthly rainfall. Adv. Space Res. 53, 136–155.

Leijnse, H., Uijlenhoet, R., Stricker, J.N.M., 2007. Hydrometeorological application of a microwave link: 2. Precipitation. Water Resour. Res. 43, 1–9.

Lendy, J.M., Rosenfeld, D., 2003. A night-rain delineation algorithm for infrared satellite data based on microphysical considerations. J. Appl. Meteorol. 42, 1218–1226.

Li, X., et al., 2021. Leveraging machine learning for quantitative precipitation estimation from Fengyun-4 geostationary observations and ground meteorological measurements. Atmos. Meas. Tech. 14, 7007–7023.

Liu, Y., Chawla, N.V., Harper, M.P., Shirberg, E., Stolcke, A., 2006. A study in machine learning from imbalanced data for sentence boundary detection in speech. Comput. Speech Lang. 20, 468–494.

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. Biotechnol. Agron. Soc. Environ. 17, 392–406.

Messer, H., Zinoviev, A., Alpert, P., 2006. Environmental monitoring by wireless communication networks. Science 312, 713.

Messer, H., Goldbricht, O., Raytisfeld, A., Alpert, P., 2008. Recent results of rainfall mapping from cellular network measurements. Int. Conf. Acoust. Spec. 5157–5157.

Meyer, H., Kühnlein, M., Appelhans, T., Nauss, T., 2016. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmos. Res. 169, 424–433.

Michaelides, S., et al., 2009. Precipitation: Measurement, remote sensing, climatology and modeling. Atmos. Res. 94, 512–533.

Moazami, S., Najafi, M.R., 2021. A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. J. Hydrol. 594.

Morais, A., Dewitte, S., Cornelis, B., Manteau, A., 2019. Deep learning for precipitation estimation from satellite and rain gauges measurements. Remote Sens. 11, 2694.

Odongo, V.O., et al., 2015. Characterisation of hydroclimatological trends and variability in the Lake Naivasha basin, Kenya. Hydrol. Process. 29, 3276–3293.

Overeem, A., Leijnse, H., Uijlenhoet, R., 2016. Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network. Atmos. Meas. Tech. 9, 2425–2444.

Pedregosa, F., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2826–2850.

K.K. Kumah et al.

Reiher, M., 2014. Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. Hydrolog. Earth Syst. Sci. 12, 587–601.

van de Giesen, N., Hut, R., Selker, J., 2014. The trans-african hydro-meteorological observatory (TAHMO). Water Resour. Res. 50, 137–155.

Wakachala, F.M., Shilenje, Z.W., Nguyo, J., Shaka, S., Apondo, W., 2015. Statistical patterns of rainfall variability in the Great Rift Valley of Kenya. J. Environ. Agric. Sci. 5, 17–26.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences, pp. 627–627.

Wolfensberger, D., Gabella, M., Bosacci, M., Germann, U., Berne, A., 2021. RandomForest: a random forest algorithm for quantitative precipitation estimation over Switzerland. Atmos. Meas. Tech. 14, 3193–3199.

Yan, J., Liu, F., Bardossy, A., Tao, T., 2021. Conditional simulation of spatial rainfall fields using random mixing: a study that implements full control over the stochastic process. Hydrolog. Earth Syst. Sci. 25, 3819–3835.

Zinoviev, A., Alpert, P., Messer, H., 2008. Estimation of rainfall fields using commercial microwave communication networks of variable density. Adv. Water Resour. 31, 1470–1480.

Zinoviev, A., Messer, H., Alpert, P., 2010. Prediction of rainfall intensity measurement errors using commercial microwave communication links. Atmos. Meas. Tech. 3, 1385–1402.