Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data

Nazli Turini\textsuperscript{a}, Boris Thies\textsuperscript{a}, Natalia Horna\textsuperscript{b} and Jörg Bendix\textsuperscript{a}

\textsuperscript{a}Laboratory for Climatology and Remote Sensing, Faculty of Geography, Philipps-University Marburg, Marburg, Germany; \textsuperscript{b}Instituto Nacional de Meteorología e Hidrología Dirección de Estudios, Investigación y Desarrollo Hidrometeorológico, Quito, Ecuador

ABSTRACT
A new satellite-based algorithm for rainfall retrieval in high spatio-temporal resolution for Ecuador is presented. The algorithm relies on the precipitation information from the Integrated Multi-Satellite Retrieval for the Global Precipitation Measurement (GPM) (IMERG) and infrared (IR) data from the Geostationary Operational Environmental Satellite-16 (GOES-16). It was developed to (i) classify the rainfall area (ii) assign the rainfall rate. In each step, we selected the most important predictors and hyperparameter tuning parameters monthly. Between 19 April 2017 and 30 November 2017, brightness temperature derived from the GOES-16 IR channels and ancillary geo-information were trained with microwave-only IMERG-V06 using random forest (RF). Validation was done against independent microwave-only IMERG-V06 information not used for training. The validation results showed the new rainfall retrieval technique (multispectral) outperforms the IR-only IMERG rainfall product. This offers using the multispectral IR data can improve the retrieval performance compared to single-spectrum IR approaches. The standard verification score a median Heidke skill score of −0.6 for the rain area delineation and R between ~0.5 and ~0.62 for the rainfall rate assignment, indicating uncertainties for Andes’s high elevation. Comparison of RF rainfall rates in 2 km\(^2\) resolution with daily rain gauge measurements reveals the correlation of R = −0.33.

Introduction
Precipitation is necessary to study the hydrological cycle and to support water management. In areas characterized by a coarse network of operational rain gauge stations, highly variable spatiotemporal precipitation patterns cannot fully be captured. This particularly holds for regions with a complex climate and topography (Seidel et al., 2019). Ecuador is one of such areas; the Andes Mountains characterize the country with the highest peaks exceeding elevations of 6268 m above sea level (m.a.s.l.), bordered by the Amazon region in the east and the lowlands of the Pacific coast, where the southern coastal part has a semi-arid climate (Buytaert et al., 2006; Ochoa-Tocachi et al., 2016; Rollebeck & Bendix, 2011). Because the rain gauge network is coarse in Ecuador, satellite-based rainfall products represent an alternative source of reliable and area-wide information on rainfall.

Several products exist to derive rainfall from single or multispectral visible (VIS)/infrared (IR) and microwave (MW) sensors, and from merged MW and VIS/IR sensors (e.g., Climate Prediction Center MORPHing technique (CMORPH) (Joyce et al., 2004), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Behrangi et al., 2009; Hong et al., 2004), and Tropical Rainfall Measuring Mission (TRMM) MultiSatellite Precipitation Analysis (TMPA) (Huffman et al., 2007)). Concerning the combination of geostationary (GEO) IR with MW sensors on polar-orbiting satellites, empirical relationships between IR data and MW-based precipitation information can be derived for rainfall at a high spatiotemporal resolution (Adler et al., 1993; Kummerow & Giglio, 1995; Xu et al., 1999). Hong et al. (2004), Huffman et al. (2007), Kidd et al. (2003), Levizzani et al. (2007), and Todd et al. (2001) used the probability matching method between the cumulative distribution functions of the MW rain rate and the IR brightness temperature. Regression methods were applied by Kuligowski (2002), Martin et al. (1990), Miller et al. (2001), and Vicente et al. (1998), in which the MW-based rainfall estimates were related to coincident IR data. Huffman et al. (2007) merged TRMM rainfall information with MW and GEO IR data in the TMPA product with accurate precipitation estimates at a high spatiotemporal resolution. Typically, MW-based rainfall data are used to calibrate IR-based products and fill the spatial and temporal gaps where MW data is not available. The Integrated Multi-Satellite Retrieval for the Global Precipitation Measurement (GPM) (IMERG), the TRMM successor, the GPM (Huffman

et al., 2015) mission with its GPM core satellite and the so-called GPM constellation, provides a new global 30-min precipitation product with a 0.1° resolution (Huffman et al., 2015). Different studies have indicated the higher performance of the IMERG compared to other global satellite-based rainfall products (Chen & Li, 2016; Guo et al., 2016; Li et al., 2017; Manz et al., 2017; Prakash et al., 2016).

In Ecuador, several satellite-based rainfall products have been evaluated against gauge networks. PERSIANN (Hong et al., 2004) shows low consistency with daily rain gauge registration (Ward et al., 2011) in rain area detection. Manz et al. (2017) compared the performance of TMPA and the IMERG against gauge data at different temporal (hourly, three-hourly, daily) resolutions. The IMERG showed better rain area detection and rainfall estimation than TMPA, particularly in the high Andes. In another study by Erazo et al. (2018) in the high Andes, TRMM 3B43 Version 7 retrievals showed a high correlation of approximately 0.82 on a monthly scale compared to the interpolated gauge information at a spatial resolution of 0.25°.

Concerning the above-stated use of multiple sensors to increase the accuracy of satellite-based rainfall retrievals, new-generation GEO multispectral sensors, such as Meteosat Second-Generation Spinning Enhanced Visible and Infrared Imager (MSG SEVIRI), HIMAWARI, and the Geostationary Operational Environmental Satellite 16 (GOES-16), offer the potential to use more spectral bands for rainfall retrieval. This higher spectral information can improve the IR-only part of the IMERG product, relying solely on the 10.8 μm channel. Several studies have documented improved satellite-based rainfall estimation by integrating full spectral information compared to a single IR channel (Behrangi et al., 2010; Giannakos & Feidas, 2013; Kolbe et al., 2020; Kühnelein et al., 2014a, 2014b; Meyer et al., 2017; Turini et al., 2019).

On the regional scale, new GEO satellite systems offer the possibility to retrieve rainfall information at a higher spatiotemporal resolution compared to the IMERG product (GOES-16 (10 min, 2 km²), MSG SEVIRI (15 min, 3 km²), and HIMAWARI (10 min, 2 km²)). This would allow capturing smaller-scale rainfall events that occur due to topographic forcing and other local precipitation processes in a more accurate way (Abatzoglou et al., 2009; Derin et al., 2016; Manz et al., 2017; Satgê et al., 2016).

Although the remote sensing of rainfall at a high spatiotemporal resolution is becoming more prevalent, the main challenge of combining data from diverse sensors to make strides in inconsistency, precision, scope, and convenient rainfall estimates still remains (Behrangi et al., 2010). Recently, machine learning (ML) techniques have been used to retrieve rainfall from multispectral GEO data (Behrangi et al., 2009; Capacci & Conway, 2005; Grimes et al., 2003; Hong et al., 2004; Kolbe et al., 2020; Kühnelein et al., 2014a, 2014b; Meyer et al., 2017; Min et al., 2019; Turini et al., 2019). They offer high potential for dealing with nonlinear and complex relationships between a target variable, such as rainfall, and its predictors (Meyer et al., 2016).

The objective of this study was to provide accurate rainfall information at a high spatiotemporal resolution for Ecuador. We aimed to develop an algorithm to generate regionally adapted rainfall products for Ecuador based on the new GOES-16 system. The high spatiotemporal resolution of GOES-16 offers excellent potential to provide rainfall information in a quasi-continuous manner. The algorithm uses multispectral IR from GOES-16 to estimate surface rainfall rates based on RF (Breiman, 2001). Microwave-only precipitation information from IMERG-V06 was used as a reference to train the RF model (see the IMERG section for more detail). The developed algorithm includes two steps: (i) Classification of the rainfall area and (ii) assignment of the rainfall rate. In each step, feature selection and tuning of the RF model are implemented. We validated the RF model results against the independent microwave-based IMERG-V06 rainfall data not used for model training. To analyze a potential improvement in comparison to algorithms that rely only on one infrared channel we also validated the results of the IMERG IR-only dataset against the MW-based rainfall data. Additionally, we applied the RF models to an independent time period and validated the results with independent gauge data. In this step, we also validated the results of the IR-only IMERG products against the gauge data.

Methodology and data

The architecture and configuration of the rainfall retrieval model

We developed a new satellite-based rainfall retrieval method that uses a combination of GEO data and polar orbiting MW-based rainfall information to produce high spatio-temporal rainfall information. The novelty is the application of machine learning random forest methods for this purpose. Microwave-only precipitation information from the IMERG-V06 was used as a reference to train the RF models. The RF rainfall retrieval algorithm workflow is shown in Figure 1. The algorithm was developed in a processing framework in Python 2.7 using the Scikit-learn 0.20.2 package (Pedregosa et al., 2011).

In the first step, the predictors and the target variable (i.e., rainfall) are processed to generate training/validation and application datasets in the same spatiotemporal resolution. Rainfall area delineation is then
conducted by RF classification, where all cloudy pixels from the binary class of cloud mask (BCM) (Heidinger & Center, U. of W.–M. S. and E. & Research, N. N. C. for S. A, 2011) are classified into rainy or non-rainy. After the rainfall rate assignment is realized by RF regression, the respective RF regression model is applied to the second step’s RF classification model. The quality of both RF models (i.e., the RF classification and RF regression models) is assessed using the validation dataset. Finally, the RF models are applied to the independent time period and validated against the independent gauge data.

The RF classification and RF regression training is conducted based on the MW overpasses in the
IMERG, and the feature selection and tuning of the hyperparameters are performed monthly. The term “feature” used in the RF models has the same meaning as “predictor.” The number of available models for one day ranges from 1 to 4 on average, depending on the time of year (Figure 2). In regions with complex topography, the higher temporal resolution models lead to better precipitation estimates (Kolbe et al., 2019).

**Data processing – matching between GOES-16 and the microwave-based IMERG**

To train and validate the rainfall retrieval model, multiband GOES-16 Advanced Baseline Imager (ABI) IR data, IMERG data, and the digital elevation model (DEM) were downloaded and extracted for Ecuador (Figure 1 – data processing stage). The GOES-16 and IMERG data were processed in Python 2.7. Because of the different spatial resolutions of the GOES-IR bands and the BCM (~2 km²), IMERG (~11 km²), and DEM (~1 km²), we used the average resampling techniques in gdal (Contributors, 2018) to ensure pixel matching between the different datasets. All of the data were projected to WGS84 and resampled to the spatial resolution of the IMERG (~11 km²).

We used different sub-datasets in the IMERG product (Table A1 in Appendix A). We selected training and validation pixels from “precipitationCal” when the overpass of passive microwaves (PMWs) was available (“HQobservation”). We filtered the pixels where the “PrecipitationQualityIndex” was >0.6 (which indicates the current half-hour microwave swath data) (Huffman, 2019b).

In the pixels of the training and validation datasets, we also selected the “IRprecipitation” from the IMERG. This dataset provides level 3 IR retrieval from PERSIANN-CCS in the IMERG. These IR data are calibrated to the merged PMW-only estimates regionally (Huffman et al., 2019a). We used this dataset to compare it to the results of our rainfall retrieval.

To compile the most solid dataset for model training and validation, we defined the following criteria: (i) For temporal matching; we collected paired GOES-16 and the microwave-based IMERG where the difference in the observation time (scan time for GOES and “HQobservationTime” for the IMERG) between both datasets was within 7 min; (ii) the IMERG “PrecipitationQualityIndex” was considered to select

![Figure 2](image-url)  
**Figure 2.** The number of scenes with enough data for the RF model training and validation for Ecuador. Boxes show the 25th, 50th, and 75th percentiles. Whiskers extend to the most extreme data points between the 75th and 25th percentiles. Outliers are shown as diamonds.

**Table 1.** Overview of the predictor variables considered for the RF regression and the RF classification.

| GOES satellite | Derived data | Ancillary geoinformation |
|----------------|--------------|--------------------------|
| IR 3.9         | ΔΤ (all band combinations) | ELV |
| WV 6.2         | VAR (all bands) | TPI |
| WV 6.9         | MAD (all bands) | TRI |
| WV 7.3         | ROD (all bands) | Slope |
| IR 8.4         | CV (all band combinations) | Aspect |
| IR 9.6         | PCV (all band combinations) | |
| IR 10.3        |              | |
| IR 11.2        |              | |
| IR 12.3        |              | |
| IR 13.3        |              | |

Geostationary Operational Environmental Satellite (GOES) bands are shown with their central wavelength in micrometers (µm). ΔΤ indicates the temperature band differences. CV, cross-varioigram; EL, terrain elevation; IR, infrared; MAD, madogram; PCV, pseudo-CV; ROD, rodocram; TPI, Topographic Position Index; TRI, Topographic Ruggedness Index; VAR, varioigram; WV, water vapor.
pixels with the high-quality merged MW-only precipitation estimates (“HQprecipitation”); (iii) following Hou et al. (2014), we used a threshold of 0.2 mm/h to distinguish between rainy and non-rainy pixels in the IMERG product; (iv) we only considered cloudy pixels detected by the BCM. The predictor variables listed in Table 1 were extracted for the pixels fulfilling all criteria only, together with the corresponding MW-only precipitation estimate. Then the dataset is divided into training/validation dataset and application (temporally independent validation). The training/validation dataset was randomly divided into a training dataset (70%) and a validation dataset (30%).

To consider the complex diurnal precipitation processes in Ecuador, the above mentioned approach was applied on a scene-based, whenever possible. According to the criteria mentioned above, whenever enough rainy pixels of the MW-based IMERG rainfall product were available for a scene of the study area, we compiled a scene-based dataset. Of this scene-based dataset 70% of the pixels were used for model training and the remaining 30% of the pixels were used for validation.

Figure 2 provides an overview of the number of scenes used for the RF model training and validation in Ecuador for each month. The overall pixels used for the RF model training and validation were 1,202,213 and 516,145, respectively.

Training and validation of the RF models

Model training and hyperparameter selection for the RF models

For rain area delineation cloudy pixels were separated into rainy and non-rainy areas by the implementation of RF classification in Scikit-learn-0.20.2 (Pedregosa et al., 2011) in Python 2.7 (Figure 1 – Training). The classification training dataset was extremely unbalanced. The average ratio between the non-rainy and rainy pixels in a scene was approximately 10:1. To address this imbalance, we kept all of the observations belonging to the minority rainy class (microwave-based IMERG >0.2 mm/h) and randomly selected (with replacement) non-rainy pixels according to different ratios (Table A2 in Appendix A). Finally, we decided on five times more non-rainy pixels than rainy pixels, which provided the highest Heidke skill score (HSS) (Table A2 in Appendix A).

The first step for RF classification is determining the most important predictors and hyperparameter tuning (Figure 1 Training). Both steps are conducted monthly. Concerning feature selection, a common precipitation retrieval approach based on GEO IR data is defining a relationship between cloud-top brightness temperatures and reference rainfall amounts, assuming the rainfall rate increases as the cloud-top temperature decreases. As shown in several studies (Behrangi et al., 2009, 2010; Turini et al., 2019), other GEO-derived rainfall-relevant features could improve rainfall retrieval from satellites.

In this study, selecting the most important predictors was achieved by applying recursive feature elimination (Figure A1 in Appendix A). This process determines the best-performing predictor sets out of the initial predictor space (Table 1). The RF modeling was started with all 181 predictors, and the least significant predictors were removed based on the “feature importance” metrics using Scikit-learn for RF methods (Breiman, 2001). At the end of each iteration, before removing the least important predictor, the model’s quality was calculated using the out-of-bags (OOB) score (Breiman, 2001; Pedregosa et al., 2011). Ultimately, we filtered the predictors with the best average ranking based on feature importance and the highest OOB score from the original set of 181 monthly.

Then, to obtain robust RF performance results, we tuned and optimized the hyperparameters. We iteratively tuned the hyperparameters monthly, including the total number of decision trees and the number of input features (n) used at each node, to determine the optimal setting for the RF classification. Moreover, a grid search of the parameter options was conducted to identify the best combination of hyperparameters based on the OOB score and optimal processing time.

For addressing the imbalance between the non-rainy and rainy pixels monthly, we selected 2000 pixels randomly from the minority class (rainy pixels) and 10,000 pixels from the majority class (non-rainy pixels) for both the predictor selection and the hyperparameter tuning (Turini et al., 2019). Additionally, we used the ‘class_weight=“balanced”’ in RF from Scikit-learn-0.20.2 (Pedregosa et al., 2011) to include a balanced weight bootstrap subsample of the RF models. This balance mode automatically adjusts weights inversely proportional to the class frequencies (Pedregosa et al., 2011). To ensure spatially and temporally robust results, we repeated the predictor selection and hyperparameter tuning 50 times to produce 50 independent datasets of randomly selected pixels for each month (Turini et al., 2019).

For the rainfall rate assignment, the RF regression model was trained with rainfall information from the microwave-based IMERG (rainfall rate >0.2 mm/h). The feature selection and hyperparameter tuning were conducted in the same way as for the RF classification. As rainfall rates are not evenly distributed, unbalanced rainfall datasets are difficult to predict by RF. To obtain more balanced rainfall datasets, we applied two different undersampling methods. In this case, undersampling signifies the removal of low rainfall rates with high frequencies in favor of high rainfall rates with low frequencies.
The first undersampling was applied during the monthly feature selection and hyperparameter tuning. We classified the rainfall rate monthly with a range of 1 mm/h and estimated the probability of occurrence (PO) of every class. The average probability of occurrence (APO) was calculated by simply summing up all POs to the number of defined classes each month. If the PO in a class was higher than the APO, we randomly selected rainfall pixels from said class until the PO reached the APO. If the classes of precipitation rate ranges contained fewer pixels than the APO, we took all of these classes’ values.

We applied the second undersampling during the monthly feature selection and hyperparameter tuning and on the RF scene-based training. Here, we used the sample weight in the fit command of the RF package “compute sample weight,” which assigns a balanced weight to each rainfall class in the bootstrap sub-dataset (Pedregosa et al., 2011).

Table A3 in Appendix A summarizes the model’s performance with and without applying the class_weight or compute_sample_weight functions in RF.

Validation of rainfall area delineation and rainfall rate assignment

The result RF classification and RF regression was validated against the independent microwave-based IMERG rainfall from the validation dataset. Validation was done for the detected rainfall area and the assigned rainfall rate separately and the combined final rainfall product (RF-combined).

To analyze the performance of the RF classification, we calculated the HSS, probability of detection (POD), and false alarm ratio (FAR) in the study area over the study period (Table A4 in Appendix A). Meanwhile, to evaluate the performance of the RF regression, we used the correlation coefficient (R), the root-mean-square error (RMSE), and the mean absolute error (MAE) (Table A4 in Appendix A). The validation statistics were calculated for every scene for which enough validation pixels were available. To analyze a potential improvement of our approach compared to existing retrieval models that rely only on one IR channel, we validated the IR-only sub-dataset in the IMERG in the same way.

We have also used these validation metrics to analyze the spatial performance of the rain area delineation and rain rate assignment. We evaluated all metrics for each pixel in the study area. Moreover, to better understand the model performance with respect to the rainfall rate in the different regions, the relative RMSE and MAE were calculated by dividing their values in each pixel by the average MW-based IMERG rainfall rate over the whole period. Furthermore, the mean differences between the MW-based IMERG rainfall rate and the RF-combined model from the validation dataset were investigated. To calculate the mean difference, we subtracted the MW-based IMERG rainfall in each pixel from that of the RF-combined model and averaged it over the whole period.

Application and validation of the model in GOES native resolution

To generate the RF rainfall retrieval at a high spatio-temporal resolution (2 km², 15 min), we applied the models in the scenes where the microwave-based IMERG was available in the same time slot and the following scenes until the next microwave-based IMERG RF model was available.

To evaluate our approach’s potential improvement compared to the IR-only IMERG, we also generated the RF rainfall retrieval for 11 km² and 15 min. The rainfall area delineation and rainfall rate assignment’s overall performance was investigated against gauge stations (Figure 4) for the studied period using the HSS and R, respectively (Table A4 of Appendix A). For this purpose, the RF rainfall retrieval data (15 min and 2 km²/11 km²) were aggregated to the gauge data’s daily resolution. The aggregation was conducted when at least 92 scenes of the RF rainfall retrieval were available. The IR-only IMERG data were aggregated daily when at least 42 scenes from the IR-only IMERG were available. We used a threshold of 0.6 mm/day to delineate between rainy and non-rainy days. The R was calculated for days with rain in all datasets.

Dataset

GOES

GOES-16 is placed at 75.2°W longitude as its operational location (Goodman, 2020; Schmit et al., 2017). The primary instrument is the Advanced Baseline Imager (ABI), a multispectral passive imaging radiometer. ABI measures the Earth’s radiance in 16 spectral channels ranging from visible (0.47 μm) to longwave IR (13.3 μm) every 15 min. Included are ten IR bands between 3.9 and 13.4 μm with a nominal spatial resolution of 2 km² at the sub-satellite point. GOES-16 has been available since 19 April 2017, and this study used data from only one year, 19 April 2017 to 19 April 2018. The dataset from 19 April 2017 to 30 November 2017 is used for training and validation, and the dataset between 1 January to 19 April 2018 is used for application and temporally independent validation of the algorithm.

Several studies have shown that multispectral satellite data provide information about microphysical and optical cloud properties (Inoue, 1985; Lensky & Rosenfeld, 2003; Ou et al., 1993; Stone et al., 1990; Strabala et al., 1994); therefore, such data can improve
satellite-based rainfall retrievals (Kolbe et al., 2019, 2020; Kühnlein et al., 2014a, 2014b; Turini et al., 2019). Thies et al. (2008) demonstrated that the channel differences between IR bands provide information about the cloud phase, the cloud top, and the cloud water path, which can improve satellite-based rainfall retrievals (Kolbe et al., 2020; Kühnlein et al., 2014b; Turini et al., 2019). We, therefore, considered the IR bands and the IR channel differences as predictor variables in our RF-based approach.

Spatial information, such as the variance within spectral bands, which are important for capturing clouds’ heterogeneity, cannot be obtained by only looking at single pixel values (Egli et al., 2018). Information about the spatial variance within a given raster dataset has improves the output of satellite-based cloud detection and rainfall retrieval techniques (Egli et al., 2018; Kolbe et al., 2019, 2020; Schulz et al., 2017; Turini et al., 2019). To consider the spatial variability of raining clouds, texture features were calculated from these predictor variables using a $5 \times 5$ pixel moving window approach. For individual predictors, variograms (VARs), madograms (MADs), and rodograms (RODs) were calculated, while for each possible predictor variable combination, cross-variograms (CVs) and pseudo cross-variograms (PCVs) were calculated (Egli et al., 2018; Kolbe et al., 2019, 2020; Turini et al., 2019) (please refer to Schulz et al. (2017) for more information). In addition to the GOES-16 IR channels, the BCM was used to restrict the rainfall retrieval algorithm to cloudy pixels. A list of the predictors based on GOES spectral bands is shown in Table 1.

Ancillary geo-information

Due to their high weather variability, mountainous regions present a great challenge to satellite-based rainfall retrievals (Dinku et al., 2007, 2008). Therefore, in this study, the terrain elevation (ELV), Topographic Position Index (TPI), Topographic Roughness Index (TRI), slope, and aspect (orientation of the slope) were considered. The ELV was provided by the Global 30 Arc-Second Elevation DEM (ORNL DAAC, 2017). The TRI, slope, and aspect were retrieved from the ELV in GRASS GIS-7 (Neteler et al., 2012), and the TPI was calculated from the ELV in SAGA (Conrad et al., 2015). Table 1 provides an overview of the predictors. Initially, we considered 181 predictors for the development of the RF models.

IMERG

The IMERG is a product that merges and intercalibrates all available MW-based precipitation estimates, MW-calibrated IR estimates, and rain gauge measurements globally (Huffman et al., 2019a). Here, we used the IMERG-V06, which is the latest available version, for which Tan et al. (2019) demonstrated the general improvement in the precipitation field compared to version-05. The IMERG-V06 uses five PMW instruments to estimate rainfall. The PMW precipitation estimates were intercalibrated by applying the CMORPH algorithm (Joyce et al., 2004) to compute the motion vectors from the IR measurements and the diverse atmospheric variables in numerical models to produce the gridded rainfall at a fine resolution quasi-globally (11 km$^2$, 30 min). This gridded product was completed by microwave-calibrated rainfall estimates retrieved from an artificial neural network model from PERSIANN-CCS (Hong et al., 2004) and GEO IR. The final rainfall estimate was calibrated with monthly gauge data from the Global Precipitation Climatology Center (GPCC) (Huffman et al., 2019a) (for more information, refer to Huffman et al. (2019a)). The different sub-datasets in the IMERG final half-hourly product used in this study are shown in Table A1 in the Appendix A.

The microwave-based rainfall estimates from the IMERG-V06 product with the highest quality (see Data processing – matching between GOES-16 and the microwave-based IMERG section for more information) were used as a reference for developing the GOES-IR rainfall retrieval. The spatial distribution of the MW-based IMERG average rainfall rate for the study period is illustrated in Figure 3(a). Overall, the rainfall rate varied between 0.2 and >8 mm/h. The rainfall distributed unevenly over the country, depending on the altitude of the regions. The Amazon area was rainier compared to the Andes and the Pacific coastal plains. The rainfall in mountainous areas showed more variability in space, ranging between 0.2 and 6 mm/h. In transition areas from the Andes to the Amazon rainforest, the rainfall rate increased, mostly in the southeast of Ecuador. The rainfall in these regions was predominantly of the subscale convective systems (Bendix et al., 2009). For the semi-arid areas in the Southwest of Ecuador and the Northwest of Peru lower rainfall rates between 0.2 and 4 mm/h can be seen.

Gauges

We used a rain gauge network comprising 22 stations with daily resolution (Figure 4) to validate the final RF rainfall retrieval. The data were acquired from the National Institute of Meteorology and Hydrology (INAMHI) in Ecuador. These data are not part of the GPCC network and were thus not considered for the gauge-calibrated final IMERG product.
Results

Feature selection and the most important predictors

Figure 5(a) illustrates the optimum number of predictors for RF classification (orange) and RF regression (blue) with the respective optimum OOB score. The RF classification showed a number of features between 18 and 25 for an OOB score of approximately 0.85 and did not vary much between the months. The RF regression offered a higher variability between the months, with the best model performance for 25 features with an OOB score of ~0.6. Figure 5(b) shows the selected features and the feature importance for both the RF classification and the RF regression. Ancillary geoinformation played a vital role in modeling the rain area and the rainfall amount, which is not surprising, as elevation played a crucial role in the rain dynamics. The texture features were also identified as important predictors for the RF models. This especially holds for the PCV calculated for the different channel combinations.

Temporal validation of the RF rainfall retrieval

To assess the quality of the RF rainfall retrieval during the validation/training period (19 April 2017 to 31 November 2017), verification scores for the RF classification, the RF regression, and the RF-combined model were investigated. The verification scores are shown as a monthly boxplot to understand the model performance at different times of the study period (Figures 6–8). December 2017 was not included due to the low availability of GOES-16 data.

The boxes show the monthly 25th, 50th, and 75th percentiles of the verification indices. The whiskers extend to the maximum and minimum verification scores in a month between the 75th and 25th percentiles. Beyond the whiskers, the scores are considered outliers and are plotted as individual diamonds (Hunter, 2007).

Regarding the RF classification (Figure 6), the median of the POD per month varied between 0.55 and 0.63 in October 2017 and November 2018, respectively. The overall performance indicated by the HSS experienced slight changes between the different months, with higher amplitudes in April and September and an overall median of HSS of ~0.6. The average FAR per month ranged from ~0.21 to ~0.4. Compared to the IR-only IMERG, the RF classification models show better performance.

For the RF regression, the median of the R, RMSE, and MAE per month (Figure 7) was ~0.6, ~1.5 mm/h, and ~1.4 mm/h, respectively. The monthly variation in the values indicated no changes. As with the RF classification, the RF regression model outperformed the IR-only IMERG results.

Altogether, the RF-combined model validation resulted in R values between ~0.5 and ~0.62 (Figure 8). Compared to the validation results for the
independent RF regression (Figure 7), this is of lower quality. Again, the RF combined model outperformed the IR-only IMERG.

Figure 9 displays the monthly mean and maximum rainfall amounts for the RF combined model and the MW-based IMERG. Concerning the mean rainfall rate, the RF combined model showed an overestimation compared to the MW-based IMERG, although the annual cycle was in good accordance in both datasets. The extreme events were analyzed if the maximum rainfall amount per month in the MW-based IMERG could be reproduced by the RF combined model for the same location and the same time step. The graph indicates that the RF models had problems detecting high rainfall amounts compared to the MW-based IMERG and underestimated the rainfall amount.

Figure 4. The distribution of the gauges used in this study. INAMHI, National Institute of Meteorology and Hydrology.
Spatial validation of the RF rainfall retrieval

Figure 10 shows the spatial performance of the RF classification for the studied period. The HSS (Figure 10(a)) and POD (Figure 10(b)) share similarities in their spatial distribution. In the east of Ecuador, with the Amazon rainforest and a higher average rainfall rate (Figure 3(a)), the overall performance of the RF was better, proven by higher POD (Figure 10(b)), higher HSS (Figure 10(a)), and lower FAR (Figure 10(c)) values.

Along the Ecuadorian coast with lower average rainfall rates (Figure 3(a)), the HSS (Figure 10(a)) was between ~0.1 and ~0.8 with higher uncertainties near the cities of Jipijapa (in the south) and Muisne (in the north). The performance of the RF classification models seemed to be influenced by the topography (Figure 3(b)). The HSS values (Figure 10(a)) are lower in areas around the peak elevations near Azogues, Ambato (Chimborazo area), Cuenca (Cajas National Park), and Sangolqui-Quito (Pichincha volcano). The FAR (Figure 10(c)) showed comparatively higher values in these high-elevation areas.

The RF classification showed some uncertainties from the coastal transition areas to the Andes and the Andes to the Amazon rainforest. As an example, this was indicated by the high FAR value of approximately 0.75 at ~78.5°W and ~1.5°S (Imbabura Volcano), as well as the lower POD and HSS values for these regions. Altogether, the evaluation results indicate inaccuracies in the RF classification for the highest elevations around the volcanoes.
Figure 10(d) provides an overview of the RF classification performance, along with the altitude. A terrain elevation of approximately 0–500 m.a.s.l has the best performance; such heights are located in the Amazon and the coastal areas (Figure 3(b)). With increasing elevation, the rain area delineation’s performance decreased until 3500 m.a.s.l. For regions above 3500 m.s.a.l, the models performed poorly. Altogether, the graph confirms the low performance of the rain area delineation in the higher elevation and volcano regions of the Andes.

The evaluation results of the RF-combined model are displayed in Figure 11. The calculated mean differences between the average rainfall rates from the RF-combined model and the MW-based IMERG product indicates an overestimation by the RF-combined model over almost the entire Ecuador. In the mid-east of Ecuador and the southern parts of the Andes, the RF-combined model showed an underestimation. Moreover, across the eastern slopes of the Andes, with generally higher rainfall rates, the model underestimated the rainfall (Figure 3).

The relative MAE (Figure 11(b)) and relative RMSE (Figure 11(c)) show similar spatial performances, which seem to be related to the spatial distribution of the average rainfall rate in Figure 3(a). For areas with lower average rainfall rates in mountainous regions, higher relative MAE and RMSE values were calculated. This is also true for the western parts of the coastal plains and the semi-arid areas of northwestern Peru and southwestern Ecuador. The RF-combined model performed better in the Amazon basin. Figure 11(d) shows the better performance of the RF-combined model for higher rainfall rates concerning R and MAE. The mean differences between the
microwave-based IMERG and RF-combined model rainfall rates increased for higher rainfall rates.

The RMSE and MAE values for the study area are between 0 and 4 mm/h (Figure 12(a,b)). Higher RMSE and MAE values can be seen in the transition zones from lower to higher altitudes across the eastern and western slopes of the Andes. For example, high MAE and RMSE values occurred at –79.45° W and –1.15° S and at –77.3° W and 1.36° S. This corresponds to areas with higher mean rainfall rates (Figure 3(a)). The R values show a higher spatial variability in the RF-combined model’s performance, and lower R values can be found in some higher elevated areas and along the coast. Generally, the model performance slightly increases with elevation but decreases at very high elevations (Figure 12(d)).

**The overall performance of the RF rainfall retrieval compared to gauge data**

We compared the final product of the RF rainfall retrieval in two different spatial resolutions against independent daily gauge measurements from 1 January to 19 April 2018.

The results for the RF rainfall retrieval in 11 km² resolution showed a median HSS of 0.35 for all INAMHI rain gauge stations (Figure 13(a)), while the HSS for the IR-only IMERG product tends toward 0.2. The RF rainfall retrieval shows a better performance in estimating the rainfall rate (Figure 13(a)) with R values around ∼0.34 (Figure 13(a)).

Figure 13(c,d) displays the spatial distribution of the HSS and R values for the RF rainfall retrieval in 15 min and 2 km² resolution. The worst HSS was
Figure 8. Box plots of the verification scores for the rainfall rate assignment combined with the rain area delineation (RF-combined model), over the microwave swath in training/validation period. Boxes show 25th, 50th, and 75th percentiles. Whiskers extend to the most extreme data points between the 75th and 25th percentiles. Outliers are shown as diamonds.

Figure 9. The monthly mean and maximum rain rates for RF-combined model and the microwave-based IMERG in the training/validation period.
related to the transition areas from the Andes to the Amazon rainforest. The highest HSS of ~0.5 was obtained for the station near Portoviejo. The R value was also the most substantial for the station near Portoviejo, and was high (~0.4–~0.7) in the western/northern parts of Ecuador and the Amazon region. For the northern stations, lower R values were observed. The median R value for all the stations was ~0.33, and the best correlation was ~0.70. The HSS did not show much variability across the different regions.
Discussion

The feature selection results for the RF classification and the RF regression indicated that the models identified the close link between topography and rainfall. The feature importance of ancillary geoinformation was relatively high compared to the other predictors and was frequently selected. This shows that elevation plays a vital role in the rainfall model. The selected predictors for both models showed that the models preferred to use two bands in combination, where the

Figure 11. Spatial distribution of (a) the mean differences between the MW-based IMERG and the RF-combined model rain rates; (b) the relative MAE; (c) the relative RMSE; (d) the performance of the rainfall retrieval as box plots for low, medium, and high precipitation rates according to percentiles. The boxes display the 25th, 50th, and 75th percentiles. The relative MAE and RMSE were calculated by dividing the MAE and RMSE values in each pixel by the average MW-based IMERG rainfall rate over the training/validation period. Nan is the values in which no data from MW-based IMERG was available.
dominant texture metric selected in almost all months was the PCV. This was also shown by Egli et al. (2018) and Turini et al. (2019).

The models showed a high R (~0.64) with a low RMSE (~2.78 mm/h) and a low MAE (~1.66 mm/h). The variability of the R per month (Figure 8) indicated a connection to the detected rain area (Figure 6). Following the rain delineation models’ lower performance in July 2017 and October 2017, the RF-combined model’s quality was also worse for these months.

The verification scores for the rain area delineation showed an improvement in comparison to previous studies. On average, the HSS in our study was ~0.6, while in the studies by Huffman et al. (2019a) and Min et al. (2019) was between 0.2 and 0.5 and 0.53, respectively. This might be due to our study’s feature selection process, but not in the mentioned studies. For Ecuador, the study of Ward et al. (2011) revealed POD of ~0.36 and FAR of ~0.2 for the rain area in the TRMM 3B42 product at a daily timescale and a spatial resolution of 0.25° in a small basin of Paute in Ecuador near Cuenca. In this region, our RF models for the rain area delineation obtained POD between ~0.3 and ~0.48 and FAR between ~0.2 and ~0.3 in 15 min and 0.1°. This shows the application of new ML algorithms such as RF could improve MW–IR blending algorithms.

Concerning the rainfall rate assignment, the median R in our study was between ~0.5 and ~0.62 for half-hourly retrievals, while Kühnlein et al. (2014a) obtained R between 0.14 and 0.46 for Germany at an hourly resolution. In another study, Kühnlein et al. (2014b) achieved R between 0.69 and 0.72 on an hourly scale in Germany. Our product had a slightly lower performance than that of Kühnlein et al.

Figure 12. The (a) RMSE, (b) MAE, and (c) R in relation to latitude, longitude, and elevation. The (d) RMSE, MAE, and R spatially averaged in relation to elevation over the training/validation period.
which might be due to the more complex topography in Ecuador than Germany. Small-scale regional conventions are common in Ecuador (Bendix et al., 2009), while Germany has large-scale advective systems, particularly in winter, which are easier to detect. Subscale convective rainfall systems due to local topographic conditions (Bendix et al., 2009) are probably not captured by GOES data.
while the large-scale advective rainfall systems in Germany are easier to detect.

The validation of the TRMM in Ecuador with interpolated rainfall data from intense gauges revealed a mean R of 0.82 for a monthly spatial resolution of 0.25° (Erazo et al., 2018). Kühnlein et al. (2014b) showed that the performance of rainfall products increases by increasing the time interval.

Figure 9 indicates that the RF models have problems in predicting high rainfall rates, which was also shown by Kühnlein et al. (2014b). Since the applied RF regression model is a very low-order regression model (only the average of the observations over the leaves was used), extreme events were underestimated compared to the training dataset’s average values (Kühnlein et al., 2014b). Therefore, estimating the extreme events useful for flood management, e.g., under El Niño conditions (e.g., Bendix et al. (2017)) remains a challenge for the proposed RF-based rainfall retrieval.

The RF classification results indicated a relationship to the different climate zones in the study area. In the semi-arid regions in the southwest of Ecuador and the northwest of Peru, the RF models showed lower HSS and POD (Figure 10(a,b)) than the more rainy regions in the Amazon rainforest. For the RF-combined model, higher MAE and RMSE were found in the semi-arid areas with lower rainfall rates (Figures 3 and Figure 10). Training separate models for different climate zones might be an approach for further improvement.

The spatial details of the RF models’ performance enable us to identify better the error sources and their contributions to different climates. The spatial variability in the RF model performance could also be attributed to the fact that the training and validation pixels were selected randomly in each scene. As a result, the RF models were better able to capture the short-term variability of the rainfall distribution. The more extended training period provided more information from the MW swats in different regions and times (Behrangi et al., 2010; Turini et al., 2019). Another reason for the spatial differences in the verification score could be related to the different viewing geometries between the GEO and the polar-orbiting MW systems (Behrangi et al., 2010; Turini et al., 2019). These problems introduce some lags in rainfall events that might be problematic for rainfall structure analysis (Behrangi et al., 2010). Furthermore, differences between the MW systems considered in the GPM constellation might be an issue. According to Tan et al. (2019), MW sensors’ different properties could lead to different rainfall rates for the same rainfall event, even when the same retrieval scheme is applied.

The results confirmed the rainfall estimation limitations of satellites across regions with a complex topography (Figures 10(d) and Figure 11(d)). In Ecuador, high-elevation areas and volcanoes are covered by ice, which is erroneous in the MW-based IMERG (Huffman, 2019b); this is a surface-screening problem. The snow and ice on the ground weaken the upwelling microwave signal, erroneously considered the PMW retrieval's rainfall (Petkovic & Kummerow, 2017).

Compared to the IR-only IMERG, the RF models showed distinct improvements in rain area delineation (Figure 7) and rainfall rate assignment (Figure 8). This is in agreement with the results of Kolbe et al. (2019, 2020), and Turini et al. (2019). Also, we compared our product with INAMHI gauge measurements. The RF-based rainfall retrieval performed better on the 11 km² and 15 min scale than the IR-only IMERG (Figure 11(a)). This illustrates the higher potential of using multispectral GEO data rather than only one IR channel rainfall retrieval, as is done in the IMERG.

We found an R of ~0.33 and a HSS of ~0.27 for the RF-based rainfall retrieval (2 km², 15 min). Zubieta et al. (2017) examined TRMM-based products’ performance (TMPA V7 and TMPA RT) and the IMERG-V03 with a spatial resolution of 0.25° and 0.1°, respectively, in the Peruvian–Ecuadorian Amazon basin. In this study, the daily HSS varied from 0 to 0.2 in the Andean region. Our RF-based rainfall retrieval showed HSS values from 0.05 to 0.5 for the same areas at a 2 km² and 15 min resolution.

The evaluation results for the IMERG and TMPA obtained by Manz et al. (2017) at a daily scale and 0.25° revealed higher R in regions with higher precipitation, confirming our results. In the same study, the IMERG-V05 (R of ~0.2 to ~0.5) showed a better R than the TMPA product (R of ~0.2), which still confirms the better performance of our RF-based rainfall retrieval (R of ~0.33).

It must be noted that the evaluation of satellite-based precipitation products using a few gauges has uncertainties. One reason is the different measurements and viewing geometry. Another issue is the point observations from a scarce gauge network that cannot represent the spatial rainfall distribution. In regions with substantial precipitation variability at the local scale, such as in Ecuador, this is more relevant for validating satellite-based rainfall products (Tang et al., 2018). Another point is the daily temporal aggregation of our satellite-based rainfall product. Due to the scan cycle, some rain events might not be captured by our retrieval method. More rain gauges with a higher temporal resolution and a ground-based radar network would be ideal for a more realistic validation.

**Conclusion**

The validation showed that the RF models could retrieve good rainfall information over Ecuador with
higher accuracy than the IR-only IMERG. This illustrates the viability of our approach and the benefit of using multispectral IR data. Meanwhile, the spatial variability of the evaluation illustrates the influence of different climate zones and topography in Ecuador, which should be investigated in more detail in future studies.

High precipitation values were underestimated by the proposed algorithm, as mentioned earlier, due to RF regression problems with extreme values. This is an important issue for future research. We need more training datasets with extreme events; therefore, the ML algorithm can differentiate between heavy and light rain. A two-step classification approach could be used for defining extreme rainfall events. The first step delineates rainy and non-rainy areas. In the second step, the rainfall area is divided into “non-extreme rainfall events” and “extreme rainfall events.” A regionally defined threshold could separate non-extreme rainfall events and extreme rainfall events based on rainfall amounts. After classification, the RF regression models could be trained and applied separately for different classes. Another possibility could be using other ML techniques such as neural networks or extreme learning machine algorithms (Deo & Şahin, 2015).

This study was limited to the available microwave-only data of the IMERG. Therefore, some rainfall events might not have been recorded due to the overflight of the microwave satellite in Ecuador, which introduces uncertainty in our model’s training.

The RF-based rainfall retrieval showed medium performance against daily-scale ground-based rainfall measurements. To obtain ideas for further improvement of the algorithm, we are currently investigating the error structure in more detail using high-resolution gauge and weather radar data.

Acknowledgments

The authors gratefully acknowledge the financial support of the Seasonal Water Resource Management (SaWaM) by the Federal Ministry for Education and Research of Germany’s (BMBF) subproject Remote Sensing of Precipitation (02WGR1421C). The work was also supported by the DFG research unit FOR2730 RESPECT, subproject A1 BE1780/51-1. The authors also thank NOAA for providing the GOES-16 satellite data used in this study and the National Weather Service of Ecuador (INAMHI) for providing their rain gauge data. Additionally, the IMERG-V06 data were provided by the NASA/Goddard Space Flight Centers and PPS, which develop and compute IMERG as a contribution to GPM, archived at the NASA GES DISC. The authors kindly thank Johannes Drönner (University of Marburg) for providing the OpenCL code to calculate the geostatistical texture features.

Disclosure statement

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript, or in the decision to publish the results.

Funding

This work was supported by the Bundesministerium für Bildung und Forschung [02WGR1421C]; Deutsche Forschungsgemeinschaft [A1 BE1780/51-1].

References

Abatzoglou, J. T., Redmond, K. T., & Edwards, L. M. (2009). Classification of regional climate variability in the state of California. Journal of Applied Meteorology and Climatology, 48(8), 1527–1541. https://doi.org/10.1175/2009jamc2062.1

Adler, R. F., Negri, A. J., Keen, P. R., & Hakkarinen, I. M. (1993). Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. Journal of Applied Meteorology, 32(2), 335–356. https://doi.org/10.1175/1520-0450(1993)032<0335:comror>2.0.co2

Behrangi, A., Hsu, K.-I., Imam, B., Sorooshian, S., Huffman, G. J., & Kuligowski, R. J. (2009). PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. Journal of Hydrometeorology, 10(6), 1414–1429. https://doi.org/10.1175/2009jhm1139.1

Behrangi, A., Imam, B., Hsu, K., Sorooshian, S., Bellerby, T. J., & Huffman, G. J. (2010). REFAME: Rain estimation using forward-adjusted advection of microwave estimates. Journal of Hydrometeorology, 11(6), 1305–1321. https://doi.org/10.1175/2010jhm1248.1

Bendix, J., Fries, A., Zarate, J., Trachte, K., Rollenbeck, R., Pucha-Cofrep, F., Paladines, R., Palacios, I., Orellana, J., Naranjo, F. O.-V. C., Mendoza, L., Mejia, D., Guallpa, M., Gordillo, F., Gonzalez-Jaramillo, V., Dobbermann, M., Celleri, R., Carrillo, C., Araque, A., & Achilles, S. (2017). RadarNet-Sur-first weather radar network in tropical high mountains. Bulletin of the American Meteorological Society, 98(6), 1235–1254. https://doi.org/10.1175/BAMS-D-15-00178.1

Bendix, J., Trachte, K., Cermak, J., Rollenbeck, R., &Nauf, T. (2009). Formation of Convective Clouds at the Foothills of the Tropical Eastern Andes (South Ecuador). Journal of Applied Meteorology and Climatology, 48(8), 1682–1695. https://doi.org/10.1175/2009jamc2078.1

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buytaert, W., Celleri, R., Willems, P., De Bièvre, B., & Wyeure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. Journal of Hydrology, 329(3–4), 413–421. https://doi.org/10.1016/j.jhydrol.2006.02.031

Capacci, D., & Conway, B. J. (2005). Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling, 12(4), 291–305. https://doi.org/10.1017/S1350482705001787
Hayatbini, N., Kong, B., Hsu, K. L., Nguyen, P., Sorooshian, S., & Stephens, G. L. (2019). A Deep Learning Framework for Precipitation Estimation from GOES-16 multispectral satellite imagery - Application of the conditional Generative Adversarial Networks (cGANs). AGUFM, 2019, H31P-1978.

Heidinger, A., Center, U. of W.-M. S. S. and E., & Research, N. N. C. for S. A. (2011). Algorithm Theoretical Basis Document: ABI Cloud Mask. https://books.google.com/books/about/Algorithm_Theoretical_Basis_Document.html?hl=en&id=Kb_UoAECAAJ&lwp=2

Hong, Y., Hsu, K.-L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. Journal of Applied Meteorology, 43(12), 1834–1853. https://doi.org/10.1175/jam2173.1

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The global precipitation measurement mission. Bulletin of the American Meteorological Society, 95(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J. (2019b). IMERG V06 quality index. https://doi.org/10.5067/gpm/gpmv06_qi/pdf/IMERGv06_QI/pdflB-rzvQ

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2019a). NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document, Version 06. https://doi.org/10.5067/gpm/gpmv06_qi/pdf/IMERGv06_QI/pdflbIL3j

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., & Yoo, S.-H. (2015). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document, Version 4. 30. https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06_0.pdf?b1ExK

Huffman, G. I., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of Hydrometeorology, 8(1), 38–55. https://doi.org/10.1175/jhm560.1

Hunter, J. D. (2007). Matplotlib. A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Inoue, T. (1985). On the temperature measurements and effective cirrus emissivity clouds by window determination region of semi-transparent in the. Journal of the Meteorological Society of Japan, 63(1), 88–99. https://doi.org/10.2151/jmsj.1965.63_1_88

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (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. https://doi.org/10.1175/jhm560.1

Kidd, C., Kniveton, D. R., Todd, M. C., & Bellerby, T. J. (2003). Satellite rainfall estimation using combined passive microwave and infrared algorithms. Journal of Hydrometeorology, 4(6), 1088–1104. https://doi.org/10.1175/1525-7541(2004)004%044:ppm088:camptp2.0.co;2

Kolbe, C., Thies, B., Egli, S., Lehniert, L., Schulz, H., & Bendix, J. (2019). Precipitation retrieval over the tibetan plateau from the geostationary orbit—Part 1: Precipitation area delineation with Elektro-L2 and
Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (5), 2557–2570. https://doi.org/10.1109/TGRS.2018.2874950

Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling and Software*, 31, 124–130. https://doi.org/10.1016/j.envsoft.2011.11.014

Ochoa-Tocacchi, B. F., Buytaert, W., De Bièvre, B., Celleri, R., Crespo, P., Villács, M., Llerena, C. A., Acosta, L., Villazon, M., Gualpia, M., Gil-Ríos, J., Fuentes, P., Olaya, D., Viñas, P., Rojas, G., & Arias, S. (2016). Impacts of land use on the hydrological response of tropical Andean catchments. *Hydrological Processes*, 30 (22), 4074–4089. https://doi.org/10.1002/hyp.10980

ORNL DAAC. (2017). *Spatial Data Access Tool (SDAT)*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1388

 Ou, S. C., Liou, K. N., Gooch, W. M., & Takano, Y. (1993). Remote sensing of cirrus cloud parameters using advanced very-high-resolution radiometer 37- and 109-μm channels. *Applied Optics*, 32(12), 2171. https://doi.org/10.1364/AO.32.002171

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page

Petrovčić, V., & Kummerow, C. 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. https://doi.org/10.1175/JAMC-D-16-0174.1

Prakash, S., Mitra, A. K., Pai, D. S., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, 88, 1–7. https://doi.org/10.1016/j.adwate.2015.11.008

Rollenbeck, R., & Bendix, J. (2011). Rainfall distribution in the Andes of southern Ecuador derived from blending weather radar data and meteorological field observations. *Atmospheric Research*, 99(2), 277–289. https://doi.org/10.1016/j.atmosres.2010.10.018

Satгé, F., Bonnet, M.-P., Gosset, M., Molina, J., Lima, W. H. Y., Zolá, R. P., Timouк, F., & Garnier, J. (2016). Assessment of satellite rainfall products over the Andean plateau. *Atmospheric Research*, 167, 1–14. https://doi.org/10.1016/j.atmosres.2015.07.012

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebari, W. J. (2017). A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, 98(4), 681–698. https://doi.org/10.1175/bams-d-15-00230.1

Schulz, H. M., Li, C. F., Thies, B., Chang, S. C., & Bendix, J. (2017). Mapping the montane cloud forest of Taiwan using 12 year MODIS-derived ground fog frequency data. *PLoS ONE*, 12(2), 1–17. https://doi.org/10.1371/journal.pone.0172663

Seidel, J., Trachte, K., Orellana-Alvear, J., Figueroa, R., Celleri, R., Bendix, J., Fernandez, C., & Huggel, C. (2019). Precipitation characteristics at two locations in the tropical andes by means of vertically pointing micro-rain radar observations. *Remote Sensing*, 11(24), 2985. https://doi.org/10.3390/rs11242985
Appendix A

Figure A1. The architecture of recursive feature selection for selecting the most important predictors in the RF classification. IMERG, Integrated Multi-Satellite Retrieval for the Global Precipitation Measurement (GPM); OOB, out-of-bags.
Table A1. List of the different sub-datasets in the IMERG-V06 product (half-hourly data final run) and definitions used in this study.

| Sub-dataset               | Definitions                                      |
|---------------------------|--------------------------------------------------|
| precipitationCal          | Multi-satellite precipitation estimate with gauge calibration |
| HQObservationTime          | Microwave satellite observation time              |
| IRPrecipitation            | IR-only precipitation estimate                    |
| PrecipitationQualityIndex  | Quality Index for precipitationCal field          |

Table A2. Median HSS scores for the random forest (RF) classification with different class ratios and applying “class_weight = balanced” in RF over the training/validation period.

| Model name  | Ratio between non-rainy (majority class) and rainy (minority class) pixels | HSS  | POD  | FAR  |
|-------------|----------------------------------------------------------------------------|------|------|------|
| Scenario-0  | 1:1                                                                        | 0.41 | 0.87 | 0.63 |
| Scenario-1  | 2:1                                                                        | 0.53 | 0.77 | 0.54 |
| Scenario-2  | 3:1                                                                        | 0.57 | 0.70 | 0.40 |
| Scenario-3  | 4:1                                                                        | 0.58 | 0.65 | 0.35 |
| Scenario-4  | 5:1                                                                        | 0.6  | 0.62 | 0.31 |

Table A3. Median HSS and R for the RF classification (ratio of 1:5) and the RF regression with and without applying the class_weight function in the RF model.

| Median HSS with a 1:5 ratio distribution for the training/validation period | Median R with for training/validation period |
|----------------------------------------------------------------------------|---------------------------------------------|
| Including class weight balance 0.6                                        | Without class weight balance 0.58           |
| Without class weight balance 0.64                                        | Including class weight balance 0.64         |
| Without class weight balance 0.6                                          | Without class weight balance 0.6             |

Table A4. Validation metrics with equations, theoretical ranges, and optimum values.

| Name                        | Metrics equation | Range | Optimum | Validation     |
|-----------------------------|-------------------|-------|---------|----------------|
| Probability of detection    | POD = \( \frac{TP}{TP + FN} \) | [0,1] | 1       | Spatial and temporal |
| False alarm ratio           | FAR = \( \frac{FP}{FP + TN} \) | [0,1] | 0       | Spatial and temporal |
| Heidke skill score          | HSS = \( \sqrt{\frac{2(2TP - FP - FN)}{TP + FN + FP + TN}} \) | [0-1] | 1       | Spatial and temporal |
| Mean absolute error         | MAE = \( \frac{1}{n} \sum_{i=1}^{n} |P_i - O_i| \) | -     | -       | Spatial and temporal |
| Root-mean-square-error      | RMSE = \( \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2} \) | -     | -       | Spatial and temporal |
| Correlation coefficient     | \( R = \frac{\sum_{i=1}^{n} (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum_{i=1}^{n} (P_i - \bar{P})^2)(\sum_{i=1}^{n} (O_i - \bar{O})^2)}} \) | [-1,1] | 1       | Spatial and temporal |
| Mean differences            | Average (microwave-based IMERG minus RF-combined model) |       |         | spatial       |

TP, true positive; FN, false negative; FP, false positive; TN, true negative; P, rainfall from RF regression; O, microwave-based IMERG rainfall; n, number of pixels considered in the validation.