



Article

Application of Random Forest Algorithm for Merging Multiple Satellite Precipitation Products across South Korea

Giang V. Nguyen ¹, Xuan-Hien Le ^{2,3} , Linh Nguyen Van ¹, Sungho Jung ¹, Minh Yeon ¹ and Giha Lee ^{1,*}

- ¹ Department of Advanced Science and Technology Convergence, Kyungpook National University, Sangju 37224, Korea; giangnv@knu.ac.kr (G.V.N.); linhnguyen@knu.ac.kr (L.N.V.); sh1202@knu.ac.kr (S.J.); alsgh2620@knu.ac.kr (M.Y.)
- ² Disaster Prevention Emergency Management Institute, Kyungpook National University, Sangju 37224, Korea; hienlx@knu.ac.kr
- ³ Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Hanoi 100000, Vietnam
- * Correspondence: leegiha@knu.ac.kr; Tel.: +82-10-4057-5302

Abstract: Precipitation is a crucial component of the water cycle and plays a key role in hydrological processes. Recently, satellite-based precipitation products (SPPs) have provided grid-based precipitation with spatiotemporal variability. However, SPPs contain a lot of uncertainty in estimated precipitation, and the spatial resolution of these products is still relatively coarse. To overcome these limitations, this study aims to generate new grid-based daily precipitation based on a combination of rainfall observation data with multiple SPPs for the period of 2003–2017 across South Korea. A Random Forest (RF) machine-learning algorithm model was applied for producing a new merged precipitation product. In addition, several statistical linear merging methods have been adopted to compare with the results achieved from the RF model. To investigate the efficiency of RF, rainfall data from 64 observed Automated Synoptic Observation System (ASOS) installations were collected to analyze the accuracy of products through several continuous as well as categorical indicators. The new precipitation values produced by the merging procedure generally not only report higher accuracy than a single satellite rainfall product but also indicate that RF is more effective than the statistical merging method. Thus, the achievements from this study point out that the RF model might be applied for merging multiple satellite precipitation products, especially in sparse region areas.

Keywords: precipitation; machine learning; random forest; merging; South Korea



Citation: Nguyen, G.V.; Le, X.-H.; Van, L.N.; Jung, S.; Yeon, M.; Lee, G. Application of Random Forest Algorithm for Merging Multiple Satellite Precipitation Products across South Korea. *Remote Sens.* **2021**, *13*, 4033. <https://doi.org/10.3390/rs13204033>

Academic Editor: Silas Michaelides

Received: 26 August 2021

Accepted: 5 October 2021

Published: 9 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Precipitation has a significant role in supporting human life on earth. It directly affects our daily life and production activities. Therefore, information about the variability of precipitation, such as intensity, duration, and frequency, is extremely important [1–3]. Currently, precipitation information is collected by three main methods: ground-based observation systems, weather radar systems, and satellite monitoring systems [4]. The rain gauge station is the primary method to obtain rainfall information with high reliability. To monitor the spatial distribution of rain in a given area, the number of stations needs to satisfy certain requirements. However, mountainous areas where there is a significant change in topography often face a significant challenge; that is, the density of measuring stations is sparse, discrete, and unevenly distributed. In addition, developing and maintaining a dense network of measurements is a major financial obstacle for developing countries [5–7]. Precipitation data from radar measurement systems have high uncertainty inaccuracies, such as systematic bias and random errors, in electronic signals in the difficult operating environment of complicated terrain [1,8]. Fortunately, over the decades, the breakthrough development of satellite-based observation technology has emerged. Satellite precipitation products (SPPs) have provided unprecedented opportunities in earth monitoring and address the aforementioned limitations [9–11]. Several SPPs are currently widely used for

water resource management, including Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN) [12,13], Climate Prediction Center Morphing (CMORPH) [14], Tropical Rainfall Measurement Mission Multi-satellite Precipitation Analysis (TMPA) [15], Climate Hazards Group (CHG) Infrared Precipitation with Station data (CHIRPS) [16], Global Satellite Mapping of Precipitation (GSMaP) [17], Soil Moisture to Rain (SM2RAIN) [18], and Multi-Source Weighted-Ensemble Precipitation (MSWEP) [12]. SPPs indicate that they might be a helpful tool and open up new horizons for us to observe the earth from a distance, especially for controlling and mitigating extreme weather events in sparse data areas. Nevertheless, many evaluations from previous studies have pointed out that SPPs contain a lot of uncertainties and biases (e.g., false precipitation events, systematic or random errors, underestimates of the intensity of precipitation events), which might originate from retrieval algorithms or due to indirect measurements [19–23]. Therefore, in recent years numerous studies have been implemented to tackle these outstanding issues and find out how to improve the performance and reduce biases of SPPs in terms of intensity precipitation estimation and the detection of rain events.

Merging is a method to blend useful information from several precipitation products into one product that has high accuracy [24]. A lot of approaches ranging from simple to complex have been put forward in a variety of regions of the world; each has certain advantages and disadvantages. Shen et al. [25] concluded that the accuracy of the merging product could be enhanced by exploiting the strengths of a single satellite-based precipitation product through the one-outlier-removed method. Other studies demonstrated that the inverse error variance weighting method is a robust approach for improving the precision of precipitation estimated from SPPs and effectively reducing errors [26,27]. However, the main limitation of traditional merging methods lies in the fact that the weighted values are estimated by the arithmetic mean method. This might have detrimental effects on the merging product's rainfall intensity estimation and ability to detect precipitation events [28,29]. To overcome these problems, Bayesian model averaging (BMA), a more sophisticated ensemble method that uses observation data to derive optimal weighted values for each SPP, has been successfully implemented in various regions around the world [28,30–32]. The performance of SPPs not only depends on the variability of topography but also on the diversity of climate, and these have certain impacts on the accuracy of the merging product. Therefore, a variation of BMA called Dynamic Bayesian Model Averaging (DBMA) [33], in which the weight of each product will vary by region and over time, has been used to account for the spatiotemporal changes in precipitation. Despite these challenges, studies have shown encouraging results in improving performance and reducing the errors of SPPs. Nevertheless, the mentioned methods still exhibit several drawbacks because most methods are based on assumptions (e.g., ignore the influence of topography on the rainfall intensity, local climate conditions, or precipitation data satisfy a Gaussian distribution in the case of the BMA-based approach), but the assumptions might be not true in practice. Various research has shown that the accuracy of SPP data is heavily affected in areas with sudden topography changes [34–38]. Meanwhile, in flat regions or lower elevation areas, the performance of SPPs is more likely better than in other regions. Topographic data must also be considered in the merging process [29]. In recent years, it has been a fact that Machine Learning (ML) and Deep Learning (DL) algorithms have achieved extraordinary performance in data science fields. These algorithms have received much attention and gained momentum for applications in hydrology, remote sensing, and water resources management [23,39,40]. ML and DL proved highly suitable for merging multiple SPPs with ground observation data to map the influence of individual members in merging products and define the nonlinear relationship between topography and rainfall intensity. For instance, Baez-Villanueva et al. [29] compared the performance of Random Forest (RF) against one-outlier-removed averaged (OORA) and Kriging with external drift for blending multiple SPPs in Chile. In addition, several complicated approaches have been tried for merging purposes. A double ML was proposed by Zhang et al. [41], and a spatiotemporal deep fusion method was implemented by Wu et al. [42] across the mainland

of China. The final results from these studies all indicated that merging products obtained by a data-driven approach have significant improvements both in terms of precipitation intensity and in the ability to detect rainfall events. The resounding success of ML and DL in these studies has encouraged us to apply a data-driven approach to improve the performance of merging processing across the region of interest. RF is a method belonging to the group of ML algorithms [43,44]. RF can cope well with data that have small data sets and with high-dimensional data [44]. Furthermore, previous results have claimed the following: (a) RF has good capability in mapping nonlinear relationships between input and output variables, (b) RF can be applied for merging multiple SPPs with ground observation data, and (c) RF proved more flexible in the combination of multiple types of explanatory variables [29,45]. RF has been applied successfully in several areas of the world; however, the application of RF for merging in other regions is still relatively limited, especially in areas with sparse observational data. Furthermore, RF is a data-driven method that means the performance of the model will more or less rely on the quality of the input data, so what type of data will be forced to the model requires more attention because it will have effects on the results obtained, and the quality of satellite-based precipitation depends on each certain area, due to the influence of topographic and climatic factors of that area. Therefore, RF's capability for merging multiple SPPs should be more explored at different specific areas, before it can be widely applied to future studies.

The aim is to investigate the applicability and reliability of the RF algorithm in merging high accuracy of ground observation with continuous spatial distribution from multiple SPPs. In this study, South Korea was selected as a case study for our research. Because every year in this area, there are various types of natural disasters, so information about rainfall is essential. In addition, in this region, there is a relatively good observational data system to serve to evaluate the results obtained from merging products accurately. The rest of the paper is organized as follows. Section 2 introduces the study area and precipitation data as well as the methodology for merging multiple SPPs. Section 3 presents the evaluation results and discussion. Finally, the conclusions are summarized in Section 4.

2. Data and Methodology

2.1. Study Area and Data

2.1.1. Study Area

South Korea is located in the northeastern part of Asia, between 33–39°N and 124–130°E with a total area of approximately is 99,373 km². Due to its location and geographical characteristics, every year, South Korea suffers from many storms moving from the sea to the mainland, causing a lot of severe damage to people and property. In addition, the topography characteristics in South Korea are very complex, which have a significant impact on the rainfall distribution and climate (Figure 1). As a result, collecting rainfall information in remote areas or at high altitudes is frequently tricky [1]. There are four distinct seasons in South Korea (spring, summer, autumn, and winter). However, because of its division into two mountain ranges, Taebaeksanmaek extending from the north to near the middle of the country in the eastern region and Sobaesanmaek cutting across from east to west in the southern area [1], there are different climate subregions in this area. The northern regions have lower temperatures, especially in the winter. Meanwhile, the southern areas have consistently higher temperatures in all seasons. The average annual temperature in South Korea ranges from 10 °C to 15 °C [46]; August is the hottest month, with an average monthly temperature of about 25 °C and January is the coldest month, with an average temperature of −7 °C [1]. The rainy season lasts from June to September, with the monsoon and typhoons accounting for 70% of annual precipitation. The dry season lasts from October to February and the snow-dominated season lasts from November to February [1,47]. The annual precipitation is approximately 1000–1850 mm/year [1].

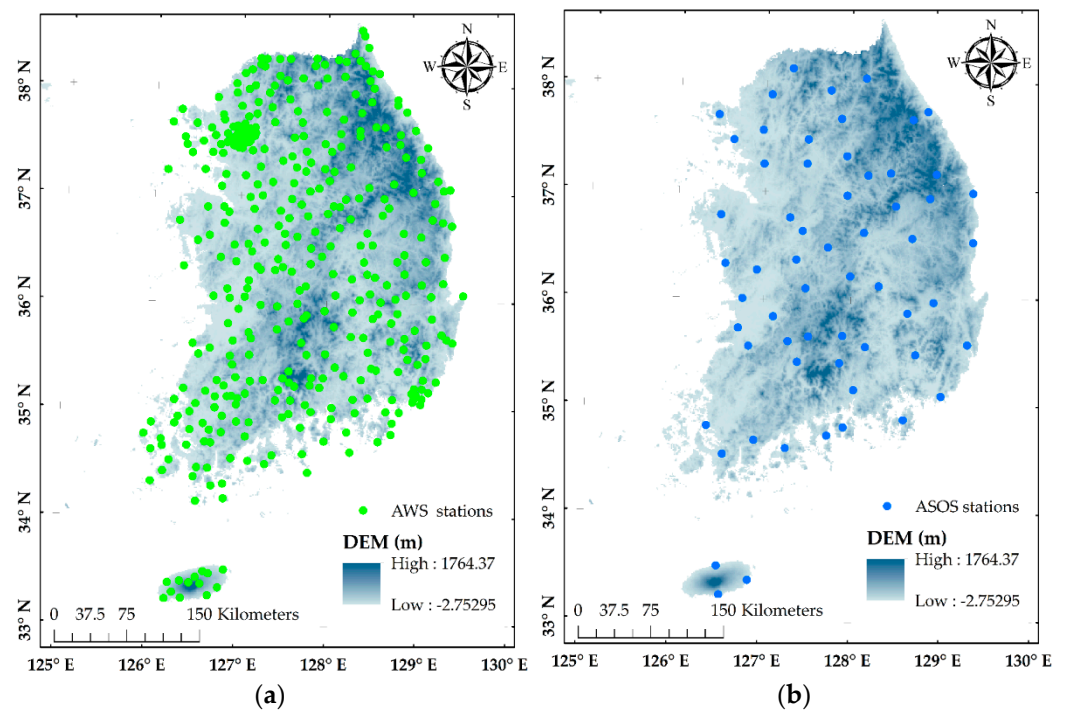


Figure 1. The elements of this study; (a) Automatic Weather Stations (AWS), (b) Automated Synoptic Observation System (ASOS).

2.1.2. Observation Data

The observation precipitation data used in this study are obtained from Automatic Weather Station (AWS; Figure 1a) and Automated Synoptic Observing System (ASOS; Figure 1b) locations and are provided by the Korea Meteorological Administration. AWS stations are installed at about 510 locations on a national scale at high density and are automatically observed to understand local meteorological phenomena. ASOS data are ground observation data like AWS, which refers to data observed at the same time by all stations to know the atmospheric conditions at a specified time. Except for a few observational elements, ASOS provides all elements related to the atmosphere, and the instruments are installed at 102 points across South Korea. The aforementioned data (AWS, ASOS) can be downloaded at the Korea Meteorological Administration's Weather Data Opening Portal <https://data.kma.go.kr/cmmn/main.do> (accessed on 9 August 2021) [48].

2.1.3. Satellite-Based Precipitation Products

The CHIRPS product refers to the Climate Hazards Group Infrared Precipitation with Station is a satellite-gauge precipitation product with quasi-global coverage (50°N–50°S). It was developed by the U.S. Geological Survey (USGS) and the CHG at the University of California, Santa Barbara, by incorporating multiple precipitation sources. The CHIRPS precipitation data is available from 1981 to the present at 5 km spatial resolution. There are three main steps to construct CHIRPS precipitation products. First, rainfall observations collected from FAO and CHCN were combined with thermal infrared (TIR) data to generate monthly precipitation climatology data (CHPclim). Second, to obtain the CHIRPS product and reduce systematic error, the long-term mean precipitation based on Cold Cloud Duration information was multiplied with CHPclim. Third, rainfall from observation data is fused with this information to generate the CHIRPS by using the modified inverse distance weighting algorithm, a smart interpolation method [16,38]. In this study, the CHIRPS version 2 (CHIRPSv2) product with $0.05^\circ \times 0.05^\circ$ spatial resolution at a daily temporal scale in the period from 2003 to 2017 was chosen. The data can be obtained at <https://data.chc.ucsb.edu/> (accessed on 22 April 2021) [49].

The GSMaP product, which stands for the Global Satellite Mapping of Precipitation, was developed by the Japan Aerospace Exploration Agency (JAXA) [17,50]. The GSMaP precipitation product currently maintains three versions of rainfall data: standard precipitation products (GSMaP_M), near-real-time precipitation products (GSMaP_N), and gauge-calibrated standard precipitation products (GSMaP_G), at high temporal (1 h) and spatial (0.10°) resolution for the quasi-global coverage of 60°N – 60°S . Initially, the backward and forward morphing technique and a Kalman filter within the Microwave–IR Combined Algorithm was used to generate the infrared (IR) information. After that, IR data were coordinated with passive microwaves (PMW) to produce the GSMaP_M from 2002. In 2007, by using only forward motion vectors in the PWM–IR process, the GSMaP_N precipitation product was generated, which can provide rainfall information in near-real-time; it is a simplified version of GSMaP_M. The final product is the GSMaP_G that blends observation data from global gauge analysis (CPC) with GSMaP_M. For this study, a daily temporal scale is generated from the latest version of 24-h GSMaP_G data. GSMaP data were obtained from the Earth Observation Research Center of JAXA at <https://sharaku.eorc.jaxa.jp/> (accessed on 22 April 2021) [50].

Global Precipitation Measurement (GPM) was created by a collaboration between the National Aeronautics and Space Administration (NASA) and the JAXA. The original purpose of this product was to provide high-resolution spatial and temporal information to the world. GPM not only inherited the Tropical Rainfall Measuring Mission (TRMM) satellite's benefits in detecting precipitation in tropical climate regions but also has two primary sensors, the Dual-Frequency Precipitation Radar (DPR) and GPM Microwave Imager (GMI) that are used to significantly improve the identification of solid precipitation and microprecipitation [51]. Precipitation was created from the GPM Level 3 product using the IMERG algorithm. Currently, three main types of precipitation products were provided by IMERG, including the near-real-time “Early Run” (IMERG_E) using forward morphing, “Late Run” (IMERG_L) using both forward morphing and backward morphing, and the “Final Run” (IMERG_F), which also includes monthly gauges analyses in addition to the combination of forwarding and backward morphing [51,52]. IMERG offers quasi-global coverage from 50°N – 50°S to 60°N – 60°S at $0.1^\circ \times 0.1^\circ$ and temporal resolution of 30 min. The IMERG_F product at daily temporal scale was used in this study and downloaded from NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC, <https://disc.gsfc.nasa.gov/> (accessed on 22 April 2021)).

The TRMM was created for quantitative observation of precipitation in tropical and subtropical areas by cooperation between NASA and JAXA. Various high-quality information of precipitation was estimated by blending microwave data from the TRMM Microwave Imager (TMI), Special Sensor Microwave Imager, Special Sensor Microwave Imager/Sounder, Advanced Microwave Scanning Radiometer-EOS (AM-SR-E), Advanced Microwave Sounding Unit-B (AMSU-B), and Microwave Humidity Sounder with rain-gauge analysis from the Global Precipitation Climatology Project (GPCP) [15,26,53] to generate the TRMM Multi-satellite Precipitation Analysis (TMPA) product with near-global coverage (50°S – 50°N), which was used in this study. Data with $0.25^\circ \times 0.25^\circ$ spatial resolution at a daily temporal scale were downloaded from NASA's GES DISC at <https://disc.gsfc.nasa.gov/> (accessed on 22 April 2021)).

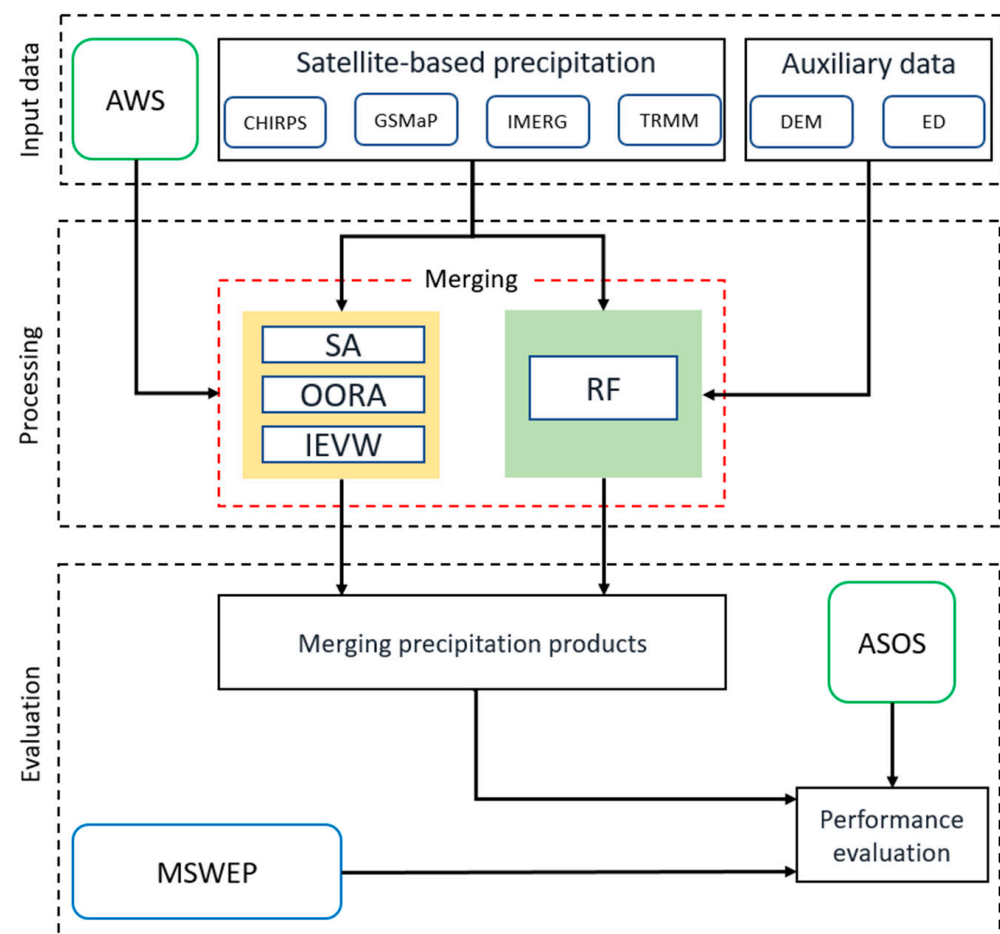
The MSWEP is a whole global coverage precipitation product. With the aim of providing accurate rainfall information on a worldwide scale, an optimal procedure was conducted to take advantage of the strengths of different data sources, including measured-, satellite-, and reanalysis-based data. MSWEP has been validated at a global scale using observation data from ~70,000 gauges and hydrological modeling for ~9000 catchments [54]. The MSWEP product is freely available from 1979 to the present at <https://www.gloh2o.org/> (accessed on 22 April 2021). Detailed information on MSWEP can be found in [7,54]. In this study, MSWEP data with daily temporal and 0.10° spatial resolution were downloaded from the above website. A brief summary of SPPs used in this study is represented in Table 1.

Table 1. Data used in this study from 2003 to 2017.

Data	Resolution		Coverage		Sources
	Spatial	Temporal	Spatial	Temporal	
CHIRPSv2	0.05°	daily	Global 50°N-S	1981-present	[16]
GSMaP	0.1°	daily	Global 60°N-S	2000-present	[55]
IMERG	0.1°	daily	Global 60°N-S	2000-present	[51]
TRMM	0.25°	daily	Global 50°N-S	1998-present	[15]
MSWEP	0.10°	daily	Global 60°N-S	1979-present	[7]

2.2. Methods

The overall workflow describing the process for merging multiple SPPs in this study is represented in Figure 2. Initially, multiple SPPs, including CHIRPS, GSMaP, IMERG, and TRMM, and several auxiliary data, were processed before these data were used as input of merging methods. After that, an RF model was constructed to map the relationship between SPPs, auxiliary data against observation data from AWS. To evaluate the robustness of the RF algorithm, several statistical-based methods, namely simple average (SA), one-outlier-remove average (OORA), and inverse error variance weighted (IEVW), were also employed. It should be noted that the auxiliary data are only used for the RF model, while it will be excluded from the merging process by statistical-based methods.

**Figure 2.** The overall workflow in this study.

In the final step, several continuous indices (e.g., MAE, RMSE, and KGE) and categorical indicators, including the probability of detection (POD), false alarm rate (FAR), and critical success index (CSI), chose to evaluate the accuracy and reliability of merging

products against an independent observation data collected from ASOS system. Moreover, the merging products developed in this study were also compared with an existing merging product called MSWEP. Achieving high accuracy and reliability for the precipitation product is a subject for further research in the future.

2.2.1. Processing Data

For merging purposes, SPPs at a daily scale were downloaded and preprocessed in ArcGIS and Python environments before these datasets were used as inputs for the different merging methods. Due to discrepancies in terms of spatial resolution, the first step is to align the data sources to a similar spatial resolution. In addition, the spatial resolution of SPPs is still too coarse (mainly ranging from 0.05° to 0.25°), which may not be suitable for further analysis in the future, for instance, for hydrological analysis or climate assessment at a regional scale [56,57]. For that reason, in this study, the simple but efficient nearest-neighbor interpolation method was used to downscale the original resolution of SPPs to the same spatial resolution (0.01°), which guarantees the retention of the original value and does not introduce more errors [29,42], and the final products were generated from merging method will have the spatial resolution at 0.01° . Additionally, as mentioned in Section 1, the influence of the topography needs to be considered during the merging process because it produces a significant effect on the spatial distribution of precipitation. To address this issue, the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) dataset version 4 at 3 arc-second (about 90 m) was up-scaled to 0.01° to match downscaled satellite-based precipitation data. Moreover, several previous studies claimed that geographical proximity information should be incorporated with the merging product [24]. However, Zhang et al. [41] indicated that data such as slope, aspect, or terrain shadows do not have much influence on the merging product. In addition, traditional merging approaches sometimes exclude information about observation locations. This has resulted in the final result from the merging process being biased, suboptimal, or even unnatural distribution of precipitation of the merged product. For that reason, in this study, besides multiple SPPs and DEM, the Euclidean distance (ED) was exploited as a covariate in the RF model to account for spatial autocorrelation, which was proved an effective variable in past studies [29,45]. The locations of 384 observation stations from the AWS system were utilized for the construction of the ED grid before this information was forced into the RF model.

2.2.2. Random Forest

RF is an ensemble model, which belongs to a supervised learning algorithm, and it has gained more attention in recent years due to its ability and versatility. Basically, the RF model was constructed based on multiple decision trees (DTs) to overcome with shortcomings that remain in the DTs model, such as the results from DTs are typically relatively poor performance or especially DTs model easy prone to the unstable situation when having a slight change in the data, and reduce overfitting issues [43]. A general RF model is shown in Figure 3. The randomness of the RF model is primarily reflected in two aspects: the first is in the process of generating subset feature samples by applying the bootstrap sampling method to draw the random subset samples from the training dataset, and the second is the randomly selected features or predictors at splitting nodes in each DT [44]. So RF can assess the relevance of each characteristic during classification or regression problems, generate an impartial estimate, and deal with situations when there are a lot of outliers and missing data. Additionally, RF can handle with high dimension data, and it might be easy to implement for merging multiple SPPs purposes. However, the primary drawback of RF is that it might become sluggish and inappropriate for real-time forecasts if there are a lot of trees.

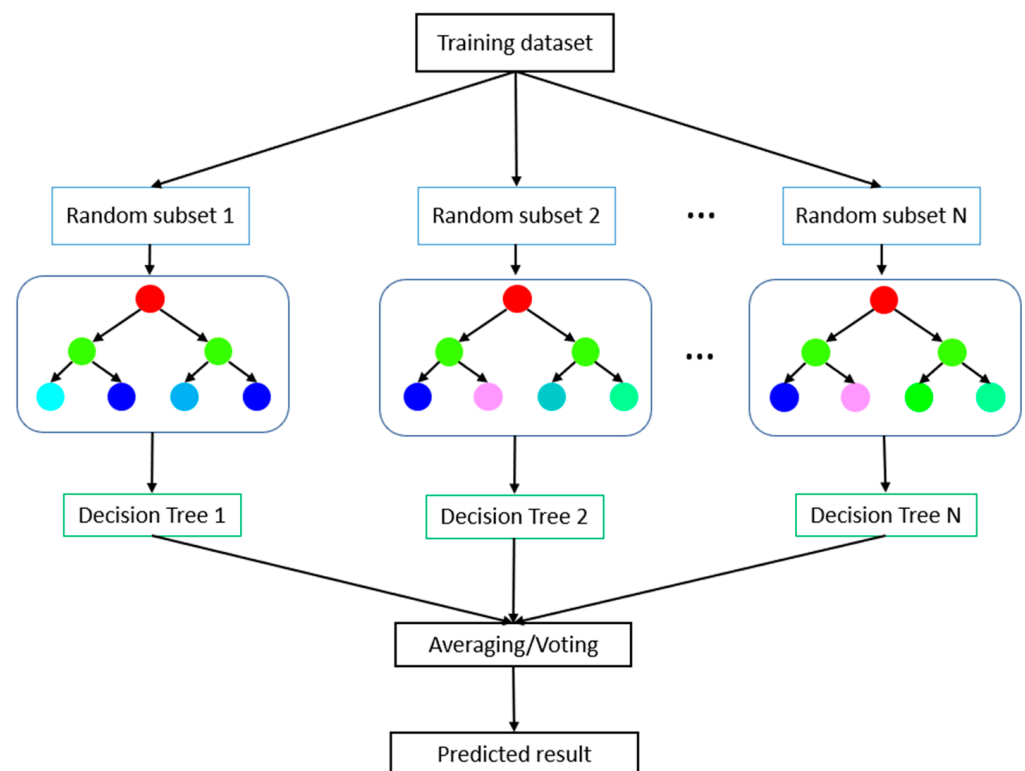


Figure 3. A general random forest model.

For this study, by taking the average of results from various DTs, the outcome will be estimated as represented in Equation (1):

$$\theta^{\Lambda B}(x) = \frac{1}{B} \sum_{b=1}^B t_b^*(x) \quad (1)$$

where $\theta^{\Lambda B}$ is the final result from the RF model, b is the individual bootstrap sample, B is the total number of decision trees, and t_b^* is the individual decision tree.

Previous studies pointed out that there are numerous hyperparameters in the RF model; however, the number of trees (`n_estimators`), the number of randomly selected variables at each decision split (`max_features`), and the minimum number of samples at a leaf node (`min_samples_leaf`) are more sensitive to the final results [29,41,58]. The trade-off between computation efficiency and the reliability of the results was carefully considered so that through trial-and-error processing, the set of hyperparameters (`n_estimators` = 100, `max_features` = 20, and `min_samples_leaf` = 5) was determined to balance the two factors mentioned above in our study.

2.2.3. Statistical-Based Methods

In addition to the RF model, three statistical-based methods, including simple average (SA), one-outlier-removed average (OORA), and inverse error variance weighting (IEVW), were also carried out for blending multiple SPPs. The formulas for each method are described as follows:

$$P_{SA} = \frac{1}{N} \sum_i^N S_i \quad (2)$$

$$P_{OORA} = \frac{1}{N-1} \sum_{i=1}^{N-1} S_i \quad (3)$$

$$P_{IEVW} = \frac{1}{\sum_{i=1}^N 1/e_i^2} \sum_{i=1}^N \frac{1}{e_i^2} S_i \text{ with } e = \frac{1}{\sigma_i^2} \quad (4)$$

where P_{SA} , P_{OORA} , and P_{IEVW} are the merging precipitation products from simple average, one-outlier-removed average, and inverse error variance weighting approach, respectively. N is the number of SPPs, S_i is the i^{th} SPP, and e is the error variance, σ is the error square between SPPs and observation data. The merged product obtained from the SA method (Equation (2)) stems from the assumption that each SPP product has the same weight. The SA only depends on the number of SPPs, while in the OORA, the root mean square error is used as an objective function to estimate the difference between the observed precipitation data with the precipitation data were extracted from the satellite. Then, the product that has the biggest error will be removed, and the final product will be estimated as in Equation (3) [22]. The IEVW method is actually applied to the GPCP [22,59]. Before the result in Equation (4) is determined, the IEVW will find out. The performance of each product is different, so if one product outperforms the others, that product will have a more significant impact on the final result.

2.2.4. Performance Evaluation

To evaluate the reliability of precipitation products, in this study, a wide approach for satellite-based precipitation analysis, namely point-to-pixel, was adopted to extract the estimated precipitation products. The assumption of the point-to-pixel method is that the rainfall of a grid cell can be represented by the rainfall of an observation station corresponding to that grid cell. After that, several continuous indicators were used to assess the discrepancies between estimation and observation data, including the Kling–Gupta efficiency (KGE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \text{ with } \alpha = \sigma_s/\sigma_o \text{ and } \beta = \mu_s/\mu_o \quad (5)$$

$$MAE = \frac{\sum_{i=1}^N |S_i - O_i|}{N} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (S_i - O_i)^2}{N}} \quad (7)$$

where O_i and S_i are the precipitation observed at gauges and satellite-based precipitation, respectively; i and N indicate time step and total length of the data, and σ and μ are the standard deviations and mean values of observed and estimated precipitation data. The optimal value of the KGE is one, and it is 0 for MAE and RMSE. Furthermore, to assess the capability of merging products for detecting rainfall events, several categorical indices were also adopted at 64 observation stations, which include the probability of detection (POD), the false alarm rate (FAR), and the critical success index (CSI). Due to the frequency of precipitation of varying intensities is an essential characteristic that has a significant impact on surface runoff and flood modeling [60] so that categorical skill indicators were applied for five classes of precipitation intensity, including no rain ([0, 1) mm/d), light rain ([1, 5) mm/d), moderate rain ([20, 40) mm/d), heavy rain ([20, 40) mm/d) and violent rain (≥ 40 mm/d) [21,29,56]. The formula of categorical indices is as follows:

$$POD = \frac{Hit}{Hit + Miss} \quad (8)$$

$$FAR = \frac{False}{Hit + False} \quad (9)$$

$$CSI = \frac{Hit}{Hit + Miss + False} \quad (10)$$

where *Hit*, *False*, and *Miss* were identified from a contingency table (Table 2). The best performance will be reached when *POD* and *CSI* are 1, whereas this value is 0 for *FAR*.

Table 2. The contingency table was used to identify the categorical performance.

Satellite Product	Observation Data		
	Yes	No	Total
Yes	Hit (H)	False alarm (F)	H + F
No	Miss (M)	Correct negative (C)	M + C
Total	H + M	F + C	N = H + F + M + C

3. Results and Discussion

3.1. Temporal Evaluation of the Precipitation Products

The performance of primary precipitation products, as well as the results obtained from the RF merging approach, were compared with observation data from the ASOS system by using several continuous indices, presented in Table 3.

Table 3. Error statistics for four precipitation products and the RF merging product at daily scale from 2003 to 2017.

SPPs	MAE (mm/d)	RMSE (mm/d)	CC	β	γ	KGE
CHIRPSv2	4.65	13.83	0.46	0.96	0.97	0.46
GSMaP	3.96	12.25	0.50	1.21	1.09	0.42
IMERG	4.27	12.52	0.53	1.02	0.88	0.51
TRMM	4.51	13.73	0.47	0.95	0.95	0.45
RF-MERGE	1.09	4.44	0.95	1.09	1.04	0.86

The primary SPP exhibits relative biases over the region of interest in this study (Table 3). For instance, the values of daily statistical metrics range from 3.96 mm/d to 4.65 mm/d, 12.25 mm/d to 13.83 mm/d, and 0.46 to 0.53 for MAE, RMSE, and correlation coefficient (CC), respectively. More specifically, CHIRPSv2 generally has shown the worst performance with the highest median MAE (4.65 mm/d), followed by TRMM (median MAE 4.51 mm/d). Both of these products also have the lowest CC values, with a median CC of 0.46 for CHIRPSv2 and 0.47 for TRMM. The GSMaP and IMERG seem to be more accurate than the two aforementioned products; as can be seen, the median MAE values of GSMaP and IMERG are 3.96 mm/d and 4.27 mm/d, respectively. Additionally, from the values of the data given in Table 3, it is difficult to determine the best product for the study area in terms of all evaluation metrics. For instance, the highest median value of KGE before conducting the merging processing belongs to IMERG, although the errors of this product, estimated by MAE and RMSE, are relatively large. Good agreement is seen between precipitation estimated from RF-MERGE and rainfall observation data. The merging product demonstrates a significant improvement.

As we can see from Figure 4, after combining multiple satellite products using the RF approach, the MAE and RMSE of the merged product decreased to 0.69–4.87 mm/d and 2.94–15.18 mm/d, respectively. Furthermore, Figure 4c–e reveal the individual components of the KGE metric at a daily scale. The CC values of multiple SPP and the merging product with observed data obtained at corresponding grid cells are indicated in Figure 4c. In general, the RF-MERGE had the highest correlation when it was used for comparing with observation data, ranging from 0.63 to 0.98 at the daily scale. In contrast, the correlation coefficients between the estimated precipitation from satellite images and the measured data are relatively low; most of the correlations of range from 0.2 to 0.65. Figure 4d plots the bias ratio (β) of the Kling–Gupta efficiency for the comparisons between a variety of SPP against rain-gauge data at a daily scale. Contrary to the tendency of most satellite products, the β value of CHIRPSv2 and TRMM tends to underestimate, ranging from 0.81 to 1.22 for

CHIRPSv2 and 0.82 to 1.29 for TRMM. Meanwhile, this value is 0.97–1.59, 0.85–1.37, and 0.78–1.40 for GSMaP, IMERG, and RF-MERGE, respectively.

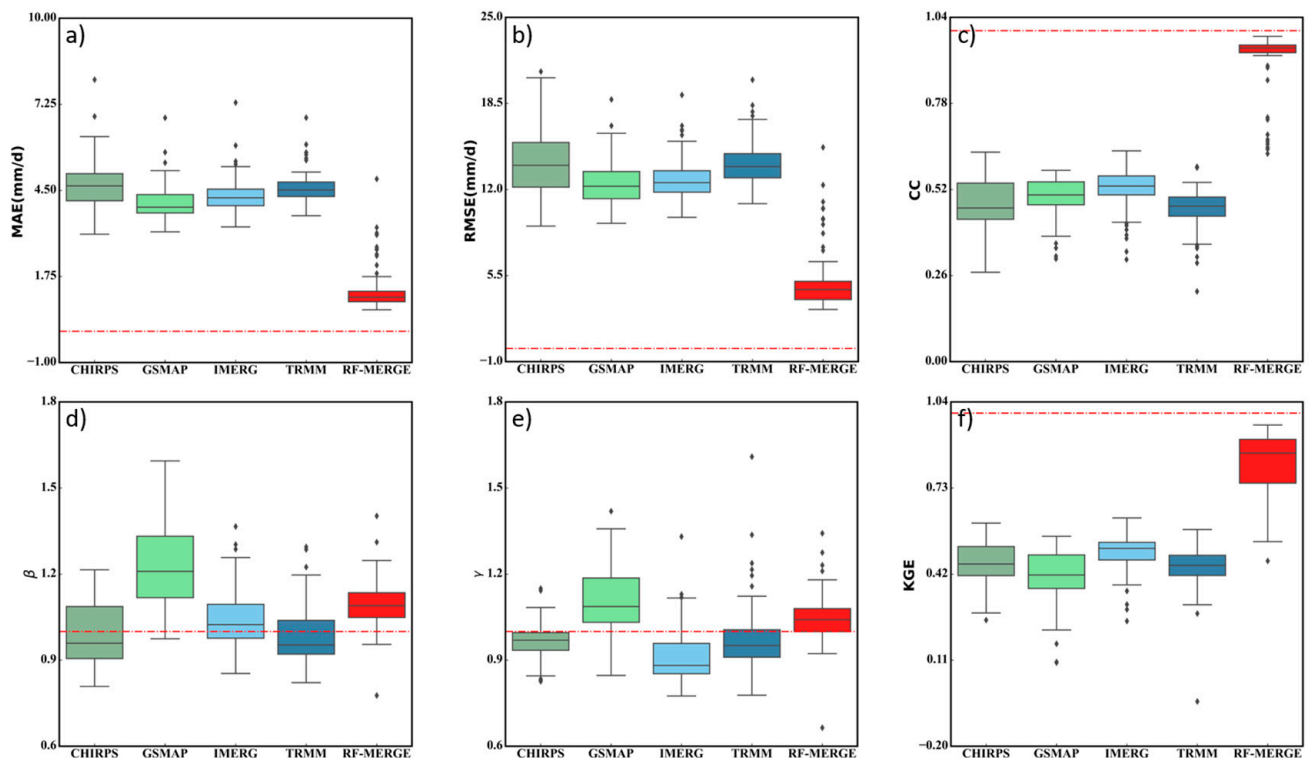


Figure 4. Boxplots of (a) MAE, (b) RMSE, (c) CC, (d) β , (e) γ , and (f) KGE at daily scale from 2003 to 2017. The optimal value is indicated by the red dashed line.

From the information of the proportion of variability (γ) between SPPs and rain-gauge data shown in Figure 4e, most of the products exhibit a general tendency to underestimate the variability of precipitation at a daily temporal scale. Figure 4f represents the KGE values for daily precipitation of original SPP and merging products at the daily temporal scale. Most of the products, except RF-MERGE, have fairly similar KGE values, ranging from -0.04 to 0.62 . Simultaneously, we can also observe that the KGE value of the RF-MERGE product is fairly high with a median value of 0.96 , indicating that the product obtained from the merging process has improved significantly compared to the original data, which is consistent with the findings of several earlier studies [29,34,42].

Regarding the capability of multiple SPPs and merging products for detecting the occurrence of precipitation at a daily scale, from Figure 5, it is clear that not only do most of the primary precipitation products show similar results at all metrics with slight fluctuation, but it is also revealed that in the original precipitation datasets, an underestimation in the determination of rain/no rain event was observed for different precipitation intensities in the original precipitation datasets. We also note that the IMERG product displays the best performance for all five rainfall intensities, followed by GSMaP, TRMM, and, lastly, CHIRPSv2. For example, from the information provided in Figure 5, IMERG showed a slight advantage in detecting light precipitation. Several reasons can explain the discrepancy in these results. The GMI sensor is more capable of capturing light rain than the TMI [52,61,62]. Compared to the nine channels (10–85.5 GHz) on the TMI, on the GMI four channels (10–183 GHz) were added for sensors. Moreover, the DPR onboard the GPM uses the Ku band (35.5 GHz) and the Ka band (35.3 GHz), while the PR onboard the GPM satellite only uses the Ku band. One more reason why IMERG has better performance than other products is the improvement of temporal and spatial resolution with IMERG. A remarkable performance was obtained by the RF-MERGE product, with higher values of POD and CSI and a lower value of FAR when compared with other products. The average

values of POD and CSI for RF-MERGE were 0.77 and 0.63, respectively. Meanwhile, the mean value of FAR was only 0.23. From this analysis, we can see that combining information from various precipitation data sources through the RF approach can get more accurate rainfall information.

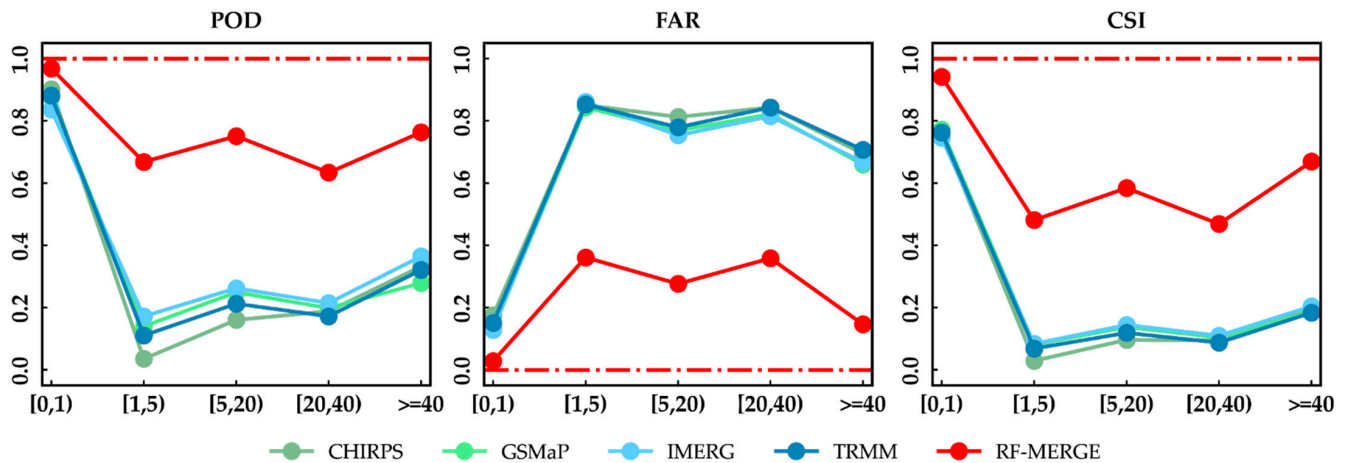


Figure 5. Categorical indices: POD, FAR, and CSI, at five precipitation intensity classes (in mm/d). The optimal value of each index is indicated by the red dashed line.

3.2. Spatial Evaluation of the Precipitation Products

Figure 6 reveals information related to the investigation of spatial distribution error metrics over South Korea during the period 2003–2017. As can be seen from Figure 6a, the accuracy of each rainfall product was estimated with ground rain gauges in terms of the KGE metric. IMERG displays slightly better performance than the other products, with the median value of KGE ranging from 0.25 to 0.62. Meanwhile, the accuracy of the rest of the products is only in the range of -0.04 – 0.58 . Figure 6b,c display the distribution of the MAE and RMSE error criteria at 64 observed stations. The CHIRPSv2 product reports the highest error for almost all stations (with median values ranging from 3.10 mm/d to 8.04 mm/d for MAE and 9.24 mm/d to 20.91 mm/d for RMSE), while GSMaP, IMERG, and TRMM have roughly similar error distributions (with median values in the range of 3.18 mm/d to 7.31 mm/d and 9.45 mm/d to 18.80 mm/d for MAE and RMSE, respectively). Overall, most primary rainfall products present inconsistent spatial distribution performances with changes in topography and climate. The results presented in Figure 6 indicate that SPPs have poor performance in the West, East, and southern coastal regions. In the winter in South Korea, there will generally be cold air masses moving from high latitudes to low latitudes. However, due to the influence of the topography, these air masses will be blocked by two mountain ranges (Taebaesanmaek and Sobaesanmaek) and unable to move down to the southern area. Meanwhile, the southern region is often affected by low-pressure systems coming from the southern sea. Consequently, throughout the winter season, the West and East regions are dominated by snow, and the South region is a rain-dominated area, with little snow in the winter [1,63]. According to Stampoulis et al. [64] claimed that over areas that have the cold surface cover or where have a predominance of snowfall, satellite precipitation products tend to have lower accuracy compared with other regions. In addition, several studies pointed out that some areas, such as the coastal area, have less correlation with rainfall observation data [1,65]. Therefore, the complexity of the topography and the diversity of the climate in South Korea are the reasons why SPP, generally, fails to achieve accuracy over the area of interest. Figure 6 also demonstrates that the final results achieved from the RF merging method can well capture the rainfall estimation, with the value of KGE increasing to 0.47–0.96, while the error of merging precipitation was significantly reduced to 0.69–4.87 mm/d and 2.94–15.18 mm/d for MAE and RMSE, respectively. However, in some regions, the performance of merging is still

relatively modest, which might be ascribed to the poor performance of the four primary precipitation products.

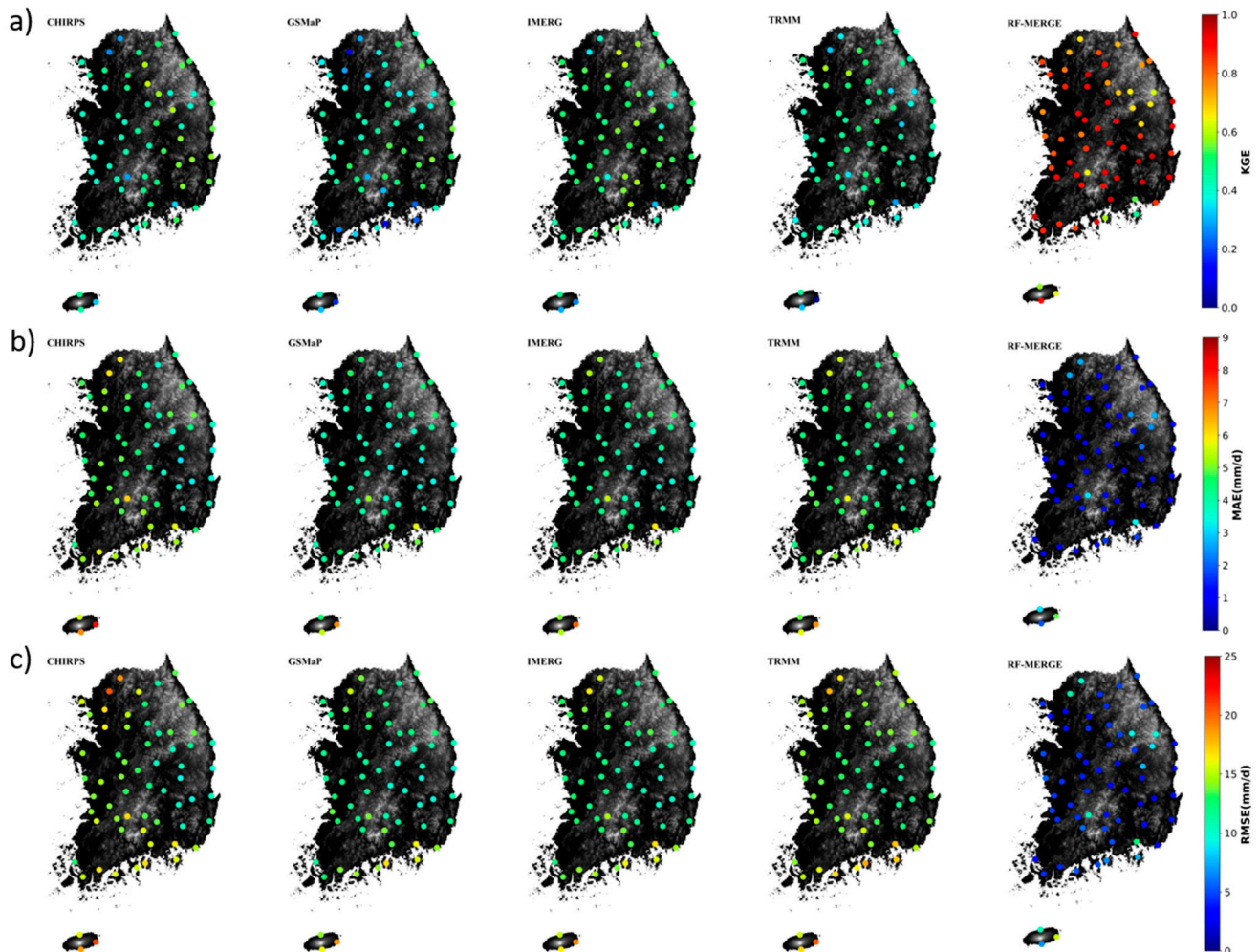


Figure 6. Evaluation of daily precipitation with four original precipitation products (CHIRPSv2, GSMaP, IMERG, and TRMM) and RF-MERGE at 64 stations by (a) KGE, (b) MAE and (c) RMSE during the period of 2003–2017 over South Korea.

3.3. Comparison between RF and Different Merging Methods

To further clarify the superiority of the RF for merging multiple SPPs, besides several simple merging approaches, this study also utilized a merging precipitation product (MSWEP) for comparison purposes. Figure 7 plots the performance of RF-MERGE compared to SA, OORA, IEVW, and MSWEP. Overall, the statistically based methods indicate similar trends with high errors and low accuracy. The values of MAE, RMSE, and KGE range from 3.097–6.953 mm/d, 8.914–18.697 mm/d, and 0.131–0.578, respectively, for SA, OORA, and IEVW. Meanwhile, MSWEP shows a trend of heterogeneity in the error and accuracy indicators. Specifically, from the information represented in Figure 7, when compared with SA, OORA, and IEVW, the MSWEP has lower errors and also lower accuracy. The values of MAE, RMSE, and KGE for MSWEP range from 2.580 to 6.646 mm/d, 8.097 to 18.089 mm/d, and 0.112 to 0.612, respectively. As can be seen, RF-MERGE exhibits better performance than other merging methods. In general, RF-MERGE is outstanding with respect to the aspects of reducing errors and improving accuracy. The values of MAE and RMSE decreased to 0.692–4.858 mm/d and 2.950–15.193 mm/d, and the KGE value increased to 0.46–0.960. These analyses demonstrated the robustness of RF in comparison with a statistical method or MSWEP precipitation product for integrating multiple SPP over the region of interest.

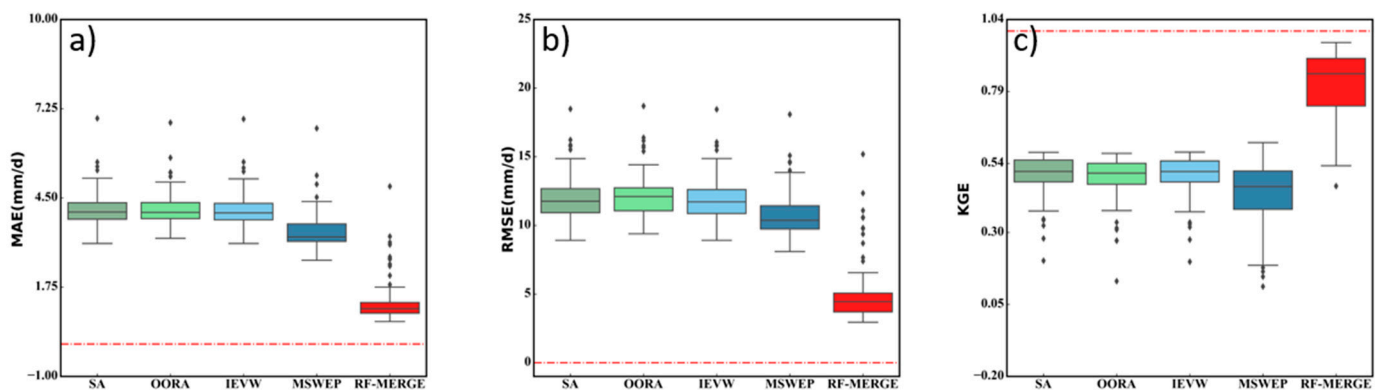


Figure 7. Evaluation of the performance of difference merging approach using (a) MAE, (b) RMSE, and (c) KGE at daily temporal scale over South Korea.

4. Conclusions

In this study, multiple sources of satellite precipitation data were merged with observed data using the RF machine-learning algorithm with the aim of improving the accuracy of rainfall estimation, especially in sparse data regions. South Korea was selected as a case study for this purpose. The performance of the RF approach was not only evaluated with a separate observation dataset but also was compared with existing merging precipitation products that use other methods. The analysis of the results showed the following:

- (i) The reliability and accuracy of the data obtained from the RF method have improved significantly compared to the original data in terms of both precipitation intensity and the ability to distinguish rain events.
- (ii) RF is proven to have outstanding performance against other merging approaches.
- (iii) The RF merging product can be used for other purposes such as hydrological modeling, drought, or even data reconstruction.

Although the analysis of results revealed the robustness of RF, the evaluation also indicates that there are several limitations in this study. For instance, the northern regions and the southern parts of RF-MERGE showed poor performance in rainfall intensity estimation or failed to detect rainfall events. The lack of primary precipitation data is one of the most important reasons to explain the lower accuracy of the final product after the merging process. In addition, the capability of the RF model should be considered to improve the accuracy of merging precipitation. Furthermore, the assumption that rainfall at the station is equivalent to the rainfall of a cell at a grid scale might not be valid in reality and could add uncertainty to the final results. This assumption needs to be carefully considered in future lines of research.

The finding in our research not only highlighted the capability and reliability of RF in merging satellite-based precipitation data with ground-based observation data but also pointed out this method is completely applicable to other regions of the world, especially those with sparse data sources. In addition, the analysis in this study also represents the potential advantage of merging multiple SPPs together, and we can obtain a new product with high reliability than original products.

Author Contributions: Conceptualization, G.V.N. and G.L.; data curation, G.V.N., S.J.; formal analysis, G.V.N.; methodology, G.V.N., G.L., X.-H.L.; supervision, G.L.; visualization, G.V.N.; writing—original draft, G.V.N.; and writing—review and editing, G.V.N., G.L., X.-H.L., S.J., M.Y., and L.N.V. All authors have read and agreed to the published version of the manuscript.

Funding: This subject is supported by Korea Ministry of Environment as “The SS project;201900283001”.

Data Availability Statement: The satellite precipitation products used in this study are all publicly available in the data sources described in Section 2.1.3. Regarding the datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kim, J.; Han, H. Evaluation of the CMORPH High-Resolution Precipitation Product for Hydrological Applications over South Korea. *Atmos. Res.* **2021**, *258*, 105650. [[CrossRef](#)]
- Chang, F.-J.; Guo, S. Advances in Hydrologic Forecasts and Water Resources Management. *Water* **2020**, *12*, 1819. [[CrossRef](#)]
- Wehbe, Y.; Temimi, M.; Adler, R.F. Enhancing Precipitation Estimates Through the Fusion of Weather Radar, Satellite Retrievals, and Surface Parameters. *Remote Sens.* **2020**, *12*, 1342. [[CrossRef](#)]
- Sun, Q.; Miao, C.; Duan, Q.; Ashouri, H.; Sorooshian, S.; Hsu, K.-L. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Rev. Geophys.* **2018**, *56*, 79–107. [[CrossRef](#)]
- Cai, Y.; Jin, C.; Wang, A.; Guan, D.; Wu, J.; Yuan, F.; Xu, L. Comprehensive Precipitation Evaluation of TRMM 3B42 with Dense Rain Gauge Networks in a Mid-Latitude Basin, Northeast, China. *Theor. Appl. Climatol.* **2016**, *126*, 659–671. [[CrossRef](#)]
- Sharifi, E.; Saghafian, B.; Steinacker, R. Downscaling Satellite Precipitation Estimates With Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *J. Geophys. Res. Atmos.* **2019**, *124*, 789–805. [[CrossRef](#)]
- Beck, H.E.; Wood, E.F.; Pan, M.; Fisher, C.K.; Miralles, D.G.; van Dijk, A.I.J.M.; McVicar, T.R.; Adler, R.F. MSWEP V2 Global 3-Hourly 0.1 Precipitation: Methodology and Quantitative Assessment. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 473–500. [[CrossRef](#)]
- Ren, M.; Xu, Z.; Pang, B.; Liu, W.; Liu, J.; Du, L.; Wang, R. Assessment of Satellite-Derived Precipitation Products for the Beijing Region. *Remote Sens.* **2018**, *10*, 1914. [[CrossRef](#)]
- Golian, S.; Moazami, S.; Kirstetter, P.-E.; Hong, Y. Evaluating the Performance of Merged Multi-Satellite Precipitation Products Over a Complex Terrain. *Water Resour. Manag.* **2015**, *29*, 4885–4901. [[CrossRef](#)]
- Yan, G.; Liu, Y.; Chen, X. Evaluating Satellite-Based Precipitation Products in Monitoring Drought Events in Southwest China. *Int. J. Remote. Sens.* **2018**, *39*, 3186–3214. [[CrossRef](#)]
- Wang, N.; Liu, W.; Sun, F.; Yao, Z.; Wang, H.; Liu, W. Evaluating Satellite-Based and Reanalysis Precipitation Datasets with Gauge-Observed Data and Hydrological Modeling in the Xihe River Basin, China. *Atmos. Res.* **2020**, *234*, 104746. [[CrossRef](#)]
- Hsu, K.-L.; Gao, X.; Sorooshian, S.; Gupta, H.V. Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteorol.* **1997**, *36*, 15. [[CrossRef](#)]
- Ashouri, H.; Hsu, K.-L.; Sorooshian, S.; Braithwaite, D.K.; Knapp, K.R.; Cecil, L.D.; Nelson, B.R.; Prat, O.P. PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 69–83. [[CrossRef](#)]
- Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P. CMORPH: A Method That Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J. Hydrometeorol.* **2004**, *5*, 17. [[CrossRef](#)]
- Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J.; Wolff, D.B.; Adler, R.F.; Gu, G.; Hong, Y.; Bowman, K.P.; Stocker, E.F. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.* **2007**, *8*, 38–55. [[CrossRef](#)]
- Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes. *Sci. Data* **2015**, *2*, 150066. [[CrossRef](#)] [[PubMed](#)]
- Ushio, T.; Sasashige, K.; Kubota, T.; Shige, S.; Okamoto, K.; Aonashi, K.; Inoue, T.; Takahashi, N.; Iguchi, T.; Kachi, M.; et al. A Kalman Filter Approach to the Global Satellite Mapping of Precipitation (GSMaP) from Combined Passive Microwave and Infrared Radiometric Data. *JMSJ* **2009**, *87A*, 137–151. [[CrossRef](#)]
- Ciabatta, L.; Massari, C.; Brocca, L.; Gruber, A.; Reimer, C.; Hahn, S.; Paulik, C.; Dorigo, W.; Kidd, R.; Wagner, W. SM2RAIN-CCI: A New Global Long-Term Rainfall Data Set Derived from ESA CCI Soil Moisture. *Earth Syst. Sci. Data* **2018**, *10*, 267–280. [[CrossRef](#)]
- Schreiner-McGraw, A.P.; Ajami, H. Impact of Uncertainty in Precipitation Forcing Data Sets on the Hydrologic Budget of an Integrated Hydrologic Model in Mountainous Terrain. *Water Resour. Res.* **2020**, *56*, e2020WR027639. [[CrossRef](#)]
- Reynolds, J.E.; Halldin, S.; Seibert, J.; Xu, C.Y.; Grabs, T. Flood Prediction Using Parameters Calibrated on Limited Discharge Data and Uncertain Rainfall Scenarios. *Hydrol. Sci. J.* **2020**, *65*, 1512–1524. [[CrossRef](#)]
- Baez-Villanueva, O.M.; Zambrano-Bigiarini, M.; Ribbe, L.; Nauditt, A.; Giraldo-Osorio, J.D.; Thinh, N.X. Temporal and Spatial Evaluation of Satellite Rainfall Estimates over Different Regions in Latin-America. *Atmos. Res.* **2018**, *213*, 34–50. [[CrossRef](#)]
- Valdés-Pineda, R.; Demaría, E.M.C.; Valdés, J.B.; Wi, S.; Serrat-Capdevilla, A. Bias Correction of Daily Satellite-Based Rainfall Estimates for Hydrologic Forecasting in the Upper Zambezi, Africa. *Hydrol. Earth Syst. Sci. Discuss.* **2016**, 1–28, [preprint]. [[CrossRef](#)]
- Le, X.-H.; Lee, G.; Jung, K.; An, H.; Lee, S.; Jung, Y. Application of Convolutional Neural Network for Spatiotemporal Bias Correction of Daily Satellite-Based Precipitation. *Remote Sens.* **2020**, *12*, 2731. [[CrossRef](#)]

24. Chao, L.; Zhang, K.; Li, Z.; Zhu, Y.; Wang, J.; Yu, Z. Geographically Weighted Regression Based Methods for Merging Satellite and Gauge Precipitation. *J. Hydrol.* **2018**, *558*, 275–289. [[CrossRef](#)]
25. Shen, Y.; Xiong, A.; Hong, Y.; Yu, J.; Pan, Y.; Chen, Z.; Saharia, M. Uncertainty Analysis of Five Satellite-Based Precipitation Products and Evaluation of Three Optimally Merged Multi-Algorithm Products over the Tibetan Plateau. *Int. J. Remote Sens.* **2014**, *35*, 6843–6858. [[CrossRef](#)]
26. Khairul, I.; Mastrantonas, N.; Rasmy, M.; Koike, T.; Takeuchi, K. Inter-Comparison of Gauge-Corrected Global Satellite Rainfall Estimates and Their Applicability for Effective Water Resource Management in a Transboundary River Basin: The Case of the Meghna River Basin. *Remote Sens.* **2018**, *10*, 828. [[CrossRef](#)]
27. Mastrantonas, N.; Bhattacharya, B.; Shibuo, Y.; Rasmy, M.; Espinoza-Dávalos, G.; Solomatine, D. Evaluating the Benefits of Merging Near-Real-Time Satellite Precipitation Products: A Case Study in the Kinu Basin Region, Japan. *J. Hydrometeorol.* **2019**, *20*, 1213–1233. [[CrossRef](#)]
28. Ma, Y.; Hong, Y.; Chen, Y.; Yang, Y.; Tang, G.; Yao, Y.; Long, D.; Li, C.; Han, Z.; Liu, R. Performance of Optimally Merged Multisatellite Precipitation Products Using the Dynamic Bayesian Model Averaging Scheme Over the Tibetan Plateau. *J. Geophys. Res. Atmos.* **2018**, *123*, 814–834. [[CrossRef](#)]
29. Baez-Villanueva, O.M.; Zambrano-Bigiarini, M.; Beck, H.E.; McNamara, I.; Ribbe, L.; Nauditt, A.; Birkel, C.; Verbist, K.; Giraldo-Osorio, J.D.; Xuan Thinh, N. RF-MEP: A Novel Random Forest Method for Merging Gridded Precipitation Products and Ground-Based Measurements. *Remote Sens. Environ.* **2020**, *239*, 111606. [[CrossRef](#)]
30. Wang, Q.J.; Schepen, A.; Robertson, D.E. Merging Seasonal Rainfall Forecasts from Multiple Statistical Models through Bayesian Model Averaging. *J. Clim.* **2012**, *25*, 5524–5537. [[CrossRef](#)]
31. Fu, Y.; Xia, J.; Yuan, W.; Xu, B.; Wu, X.; Chen, Y.; Zhang, H. Assessment of Multiple Precipitation Products over Major River Basins of China. *Theor. Appl. Climatol.* **2016**, *123*, 11–22. [[CrossRef](#)]
32. Zhu, J.; Kong, F.; Ran, L.; Lei, H. Bayesian Model Averaging with Stratified Sampling for Probabilistic Quantitative Precipitation Forecasting in Northern China during Summer 2010. *Mon. Weather Rev.* **2015**, *143*, 3628–3641. [[CrossRef](#)]
33. Rahman, K.U.; Shang, S.; Shahid, M.; Wen, Y.; Khan, Z. Application of a Dynamic Clustered Bayesian Model Averaging (DCBA) Algorithm for Merging Multisatellite Precipitation Products over Pakistan. *J. Hydrometeorol.* **2020**, *21*, 17–37. [[CrossRef](#)]
34. Woldemeskel, F.M.; Sivakumar, B.; Sharma, A. Merging Gauge and Satellite Rainfall with Specification of Associated Uncertainty across Australia. *J. Hydrol.* **2013**, *499*, 167–176. [[CrossRef](#)]
35. Zambrano, F.; Wardlow, B.; Tadesse, T.; Lillo-Saavedra, M.; Lagos, O. Evaluating Satellite-Derived Long-Term Historical Precipitation Datasets for Drought Monitoring in Chile. *Atmos. Res.* **2017**, *186*, 26–42. [[CrossRef](#)]
36. Katsanos, D.; Retalis, A.; Michaelides, S. Validation of a High-Resolution Precipitation Database (CHIRPS) over Cyprus for a 30-Year Period. *Atmos. Res.* **2016**, *169*, 459–464. [[CrossRef](#)]
37. Lai, C.; Zhong, R.; Wang, Z.; Wu, X.; Chen, X.; Wang, P.; Lian, Y. Monitoring Hydrological Drought Using Long-Term Satellite-Based Precipitation Data. *Sci. Total Environ.* **2019**, *649*, 1198–1208. [[CrossRef](#)] [[PubMed](#)]
38. Duan, Z.; Liu, J.; Tuo, Y.; Chiogna, G.; Disse, M. Evaluation of Eight High Spatial Resolution Gridded Precipitation Products in Adige Basin (Italy) at Multiple Temporal and Spatial Scales. *Sci. Total Environ.* **2016**, *573*, 1536–1553. [[CrossRef](#)]
39. Le, X.-H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. [[CrossRef](#)]
40. Sit, M.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A Comprehensive Review of Deep Learning Applications in Hydrology and Water Resources. *arXiv* **2020**, arXiv:2007.12269. [[CrossRef](#)]
41. Zhang, L.; Li, X.; Zheng, D.; Zhang, K.; Ma, Q.; Zhao, Y.; Ge, Y. Merging Multiple Satellite-Based Precipitation Products and Gauge Observations Using a Novel Double Machine Learning Approach. *J. Hydrol.* **2021**, *594*, 125969. [[CrossRef](#)]
42. Wu, H.; Yang, Q.; Liu, J.; Wang, G. A Spatiotemporal Deep Fusion Model for Merging Satellite and Gauge Precipitation in China. *J. Hydrol.* **2020**, *584*, 124664. [[CrossRef](#)]
43. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
44. Biau, G.; Scornet, E. A Random Forest Guided Tour. *TEST* **2016**, *25*, 197–227. [[CrossRef](#)]
45. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ* **2018**, *6*, e5518. [[CrossRef](#)] [[PubMed](#)]
46. Park, S.; Shin, M.; Im, J.; Song, C.-K.; Choi, M.; Kim, J.; Lee, S.; Park, R.; Kim, J.; Lee, D.-W.; et al. Estimation of Ground-Level Particulate Matter Concentrations through the Synergistic Use of Satellite Observations and Process-Based Models over South Korea. *Atmos. Chem. Phys.* **2019**, *19*, 1097–1113. [[CrossRef](#)]
47. Joo, J.; Kim, S.; Park, M.; Kim, J. Evaluation and Calibration Method Proposal of RCP Daily Precipitation Data. *J. Korean Soc. Hazard. Mitig.* **2015**, *15*, 79–91. [[CrossRef](#)]
48. Open MET Data Portal. Available online: <https://data.kma.go.kr/cmmn/main.do> (accessed on 9 August 2021).
49. CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center—UC Santa Barbara. Available online: <https://chc.ucsb.edu/data/chirps> (accessed on 22 April 2021).
50. JAXA Global Rainfall Watch (GSMaP). Available online: <https://sharaku.eorc.jaxa.jp/GSMaP/> (accessed on 22 April 2021).
51. Zhou, Z.; Guo, B.; Xing, W.; Zhou, J.; Xu, F.; Xu, Y. Comprehensive Evaluation of Latest GPM Era IMERG and GSMaP Precipitation Products over Mainland China. *Atmos. Res.* **2020**, *246*, 105132. [[CrossRef](#)]
52. Tan, M.L.; Duan, Z. Assessment of GPM and TRMM Precipitation Products over Singapore. *Remote Sens.* **2017**, *9*, 720. [[CrossRef](#)]

53. Wang, R.; Chen, J.; Wang, X. Comparison of IMERG Level-3 and TMPA 3B42V7 in Estimating Typhoon-Related Heavy Rain. *Water* **2017**, *9*, 276. [[CrossRef](#)]
54. Beck, H.E.; Vergopolan, N.; Pan, M.; Levizzani, V.; van Dijk, A.I.J.M.; Weedon, G.P.; Brocca, L.; Pappenberger, F.; Huffman, G.J.; Wood, E.F. Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6201–6217. [[CrossRef](#)]
55. Aonashi, K.; Awaka, J.; Hirose, M.; Kozu, T.; Kubota, T.; Liu, G.; Shige, S.; Kida, S.; Seto, S.; Takahashi, N.; et al. GSMaP Passive Microwave Precipitation Retrieval Algorithm: Algorithm Description and Validation. *JMSJ* **2009**, *87A*, 119–136. [[CrossRef](#)]
56. Chen, S.; Xiong, L.; Ma, Q.; Kim, J.-S.; Chen, J.; Xu, C.-Y. Improving Daily Spatial Precipitation Estimates by Merging Gauge Observation with Multiple Satellite-Based Precipitation Products Based on the Geographically Weighted Ridge Regression Method. *J. Hydrol.* **2020**, *589*, 125156. [[CrossRef](#)]
57. Chen, Y.; Huang, J.; Sheng, S.; Mansaray, L.R.; Liu, Z.; Wu, H.; Wang, X. A New Downscaling-Integration Framework for High-Resolution Monthly Precipitation Estimates: Combining Rain Gauge Observations, Satellite-Derived Precipitation Data and Geographical Ancillary Data. *Remote Sens. Environ.* **2018**, *214*, 154–172. [[CrossRef](#)]
58. Fan, Z.; Li, W.; Jiang, Q.; Sun, W.; Wen, J.; Gao, J. A Comparative Study of Four Merging Approaches for Regional Precipitation Estimation. *IEEE Access* **2021**, *9*, 33625–33637. [[CrossRef](#)]
59. Azam, M.; Park, H.K.; Maeng, S.J.; Kim, H.S. Regionalization of Drought across South Korea Using Multivariate Methods. *Water* **2018**, *10*, 24. [[CrossRef](#)]
60. Tian, Y.; Peters-Lidard, C.D.; Choudhury, B.J.; Garcia, M. Multitemporal Analysis of TRMM-Based Satellite Precipitation Products for Land Data Assimilation Applications. *J. Hydrometeorol.* **2007**, *8*, 1165–1183. [[CrossRef](#)]
61. Ma, Y.; Tang, G.; Long, D.; Yong, B.; Zhong, L.; Wan, W.; Hong, Y. Similarity and Error Intercomparison of the GPM and Its Predecessor-TRMM Multisatellite Precipitation Analysis Using the Best Available Hourly Gauge Network over the Tibetan Plateau. *Remote Sens.* **2016**, *8*, 569. [[CrossRef](#)]
62. Peng, F.; Zhao, S.; Chen, C.; Cong, D.; Wang, Y.; Ouyang, H. Evaluation and Comparison of the Precipitation Detection Ability of Multiple Satellite Products in a Typical Agriculture Area of China. *Atmos. Res.* **2020**, *236*, 104814. [[CrossRef](#)]
63. Kim, K.; Park, J.; Baik, J.; Choi, M. Evaluation of Topographical and Seasonal Feature Using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmos. Res.* **2017**, *187*, 95–105. [[CrossRef](#)]
64. Stampoulis, D.; Anagnostou, E.N. Evaluation of Global Satellite Rainfall Products over Continental Europe. *J. Hydrometeorol.* **2012**, *13*, 588–603. [[CrossRef](#)]
65. Kubota, T.; Ushio, T.; Shige, S.; Kida, S.; Kachi, M.; Okamoto, K. Verification of High-Resolution Satellite-Based Rainfall Estimates around Japan Using a Gauge-Calibrated Ground-Radar Dataset. *J. Meteorol. Soc. Jpn. Ser. II* **2009**, *87A*, 203–222. [[CrossRef](#)]