



## Article

# Real-Time Retrieval of Daily Soil Moisture Using IMERG and GK2A Satellite Images with NWP and Topographic Data: A Machine Learning Approach for South Korea

Soo-Jin Lee <sup>1</sup>, Eunha Sohn <sup>2</sup>, Mija Kim <sup>2</sup>, Ki-Hong Park <sup>2</sup>, Kyungwon Park <sup>3</sup> and Yangwon Lee <sup>4,\*</sup>

<sup>1</sup> Geomatics Research Institute, Pukyong National University, Busan 48513, Republic of Korea; sjlee610b@pukyong.ac.kr

<sup>2</sup> Satellite Planning Division, National Meteorological Satellite Center, Jincheon 27803, Republic of Korea; soneh0431@korea.kr (E.S.); kimmee74@korea.kr (M.K.); parkkihong@korea.kr (K.-H.P.)

<sup>3</sup> Climate Service and Research Division, APEC Climate Center, Busan 48058, Republic of Korea; kwpark@apcc21.org

<sup>4</sup> Department of Spatial Information Engineering, Pukyong National University, Busan 48513, Republic of Korea

\* Correspondence: modconfi@pknu.ac.kr

**Abstract:** Soil moisture (SM) is an indicator of the moisture status of the land surface, which is useful for monitoring extreme weather events. Representative global SM datasets include the National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP), the Global Land Data Assimilation System (GLDAS), and the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5 (ERA5), but due to their low spatial resolutions, none of these datasets well describe SM changes in local areas, and they tend to have a low accuracy. Machine learning (ML)-based SM predictions have demonstrated high accuracy, but obtaining semi-real-time SM information remains challenging, and the dependence of the validation accuracy on the data sampling method used, such as random or yearly sampling, has led to uncertainties. In this study, we aimed to develop an ML-based model for real-time SM estimation that can capture local-scale variabilities in SM and have reliable accuracy, regardless of the sampling method. This study was conducted in South Korea, and satellite image data, numerical weather prediction (NWP) data, and topographic data provided within one day were used as the input data. For SM modeling, 13 input variables affecting the surface SM status were selected: 10- and 20-day cumulative standardized precipitation indexes (SPI10 and SPI20), a normalized difference vegetation index (NDVI), downward shortwave radiation (DSR), air temperature (T<sub>air</sub>), land surface temperature (LST), soil temperature (T<sub>soil</sub>), relative humidity (RH), latent heat flux (LE), slope, elevation, topographic ruggedness index (TRI), and aspect. Then, SM models based on random forest (RF) and automated machine learning (AutoML) were constructed, trained, and validated using random sampling and leave-one-year-out (LOYO) cross-validation. The RF- and AutoML-based SM models had significantly high accuracy rates based on comparisons with in situ SM (mean absolute error (MAE) = 2.212–4.132%; mean bias error (MBE) = −0.110–0.136%; root mean square error (RMSE) = 3.186–5.384%; correlation coefficient (CC) = 0.732–0.913), while the AutoML-based SM model tended to have a higher accuracy than the RF-based SM model, regardless of the data sampling method used. In addition, when compared to in situ SM data, the SM models demonstrated the highest accuracy, outperforming both GLDAS and ERA5 SM data and well representing changes in the dryness/wetness of the land surface according to meteorological events (heatwave, drought, and rainfall). The SM models proposed in this study can, thus, offer semi-real-time SM data, aiding in the monitoring of moisture changes in the land surface, as well as short-term meteorological disasters, like flash droughts or floods.

**Keywords:** soil moisture; random forest; automated machine learning; leave-one-year-out cross-validation; real time



**Citation:** Lee, S.-J.; Sohn, E.; Kim, M.; Park, K.-H.; Park, K.; Lee, Y. Real-Time Retrieval of Daily Soil Moisture Using IMERG and GK2A Satellite Images with NWP and Topographic Data: A Machine Learning Approach for South Korea. *Remote Sens.* **2023**, *15*, 4168. <https://doi.org/10.3390/rs15174168>

Academic Editors: Hui Lu, Lun Gao, Xiaojun Li and Hongquan Wang

Received: 25 June 2023

Revised: 19 August 2023

Accepted: 22 August 2023

Published: 24 August 2023



**Copyright:** © 2023 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

Climate change is increasing the risk and severity of drought in California, the Pacific Northwest, the Western United States, and the Mediterranean region, among other areas [1]. Furthermore, the frequency of extreme heatwaves, which can accelerate droughts, has increased compared to the past, and this trend is expected to continue [2,3]. In particular, regions such as southern Pakistan, south–northern India, the Sahara, and southwestern Africa are predicted to experience even more severe heat stress in the future [2,3]. Conversely, the risk of flooding due to increased rainfall is predicted to increase in temperate regions of the Northern Hemisphere, western and eastern Eurasia, South Asia, Southeast Asia, and the western Amazon [4,5]. These hydrological disasters could result in intense damage and socioeconomic losses, including wildfires, decreased crop production, and energy issues related to hydropower [6–8]. Thus, monitoring the moisture status of the land surface is crucial. Soil moisture (SM), which is defined as the water present among soil particles, is a hydrological factor that impacts surface–atmosphere interactions [9]. SM interacts with the atmosphere through evaporation (from the land surface) and transpiration (from plants) and influences water circulation by determining runoff and the infiltration of precipitation [10,11]. Accurate estimation of SM is, thus, of major importance in terms of monitoring hydrometeorological disasters, such as droughts and floods, and assessing energy and water cycles.

In situ SM is measured using gravimetry, time domain reflectometry (TDR), and dielectric impedance. The gravimetric technique measures the gravimetric water content (mass of water per mass of dry soil) by weighing soil samples before and after removing moisture [12]. It is low cost but time consuming in nature because it requires skilled experience [13]. The TDR and dielectric impedance sensors measure volumetric moisture content (the water volume per soil volume) [12]. The TDR determines the dielectric constant by measuring the speed and time of electromagnetic waves [14]. Then, SM is found using a calibration equation derived from the empirical relationship between the dielectric constant and the volumetric moisture content of soils with different textures proposed by Topp et al. (1980) [15]. It is less sensitive to changes in soil salinity, temperature, and soil texture, and it can be non-destructively observed [16]. However, the installation cost is high, and in wet soil with high salinity, there is a possibility of error due to loss of reflection and increased conductivity [15,16]. The Hydraprobe sensor is a ratiometric coaxial impedance dielectric reflectometer operating at a frequency of 50 MHz [17]. It measures both the real and imaginary components of the dielectric permittivity using the ratio of the reflected signal over incident signal [17]. Then, it calculates SM using a calibration equation based on the real dielectric permittivity, unlike most other SM technologies based on the apparent permittivity [17]. It is less affected by salts and temperature than TDR sensors because of the delineation of the dielectric permittivity and operational frequency at 50 MHz [17]. However, given the natural fluctuations in SM levels over time, periodic updates to the calibration equations are required to ensure accurate SM estimation [18]. These soil moisture measurement techniques provide accurate SM information for each observation point, but it is difficult and expensive to provide spatially continuous information.

Representative global SM datasets include those of the National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP), the Global Land Data Assimilation System (GLDAS), and the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5 (ERA5) (Table 1). SMAP provides daily global SM data derived using the brightness temperature measured via a passive microwave (L-band) radiometer and the tau–omega radiative transfer model (RTM) [19]. The RTM is a model used to determine brightness temperature based on ground and vegetation emissivity [19,20]. Surface emissivity changes at microwave frequencies are a result of variations in the dielectric properties between dry and wet soils [20]. Thus, SM is inferred via the reverse application of the RTM using brightness temperature [19]. GLDAS provides land products such as SM, soil temperature, snowfall, and rainfall through the integration of satellite- and ground-based observational data via advanced land surface modeling and data as-

simulation [21]. ERA5, which is the fifth-generation ECMWF reanalysis for global climate and weather, provides hourly estimates of key atmospheric, ocean wave, and land surface parameters through the assimilation of model data and worldwide observations [22]. The SMAP, GLDAS, and ERA5 SM datasets are useful for observing soil moisture in global regions because they comprise spatially and temporally continuous global SM data with a short temporal resolution (hourly to daily) (Table 1). However, these datasets have a relatively low spatial resolutions (9–36 km), limiting their ability to capture land surface SM changes in local areas, and the GLDAS is only updated monthly, which makes it difficult to quickly capture the moisture status of the land surface (Table 1). And, besides the low spatial resolution, they also have limitations in terms of their accuracy. The GLDAS data showed low accuracy in studies assessing SM accuracy in regard to the Korean [23] and global regions [24]. The SMAP data had lower accuracy than the GLDAS data in the Korean region [25]. The ERA5 data also had low accuracy in studies conducted in the Tibetan Plateau region [26] and Jiangsu Province, China [27]. This result indicates that these datasets are unsuitable for use in local SM monitoring due to their high inherent uncertainties. As a result, such global-level data cannot be sufficiently applied to the local-scale variabilities due to complex and heterogeneous land surfaces in countries like South Korea [28].

**Table 1.** Representative global soil moisture data.

Soil Moisture Data	SMAP <sup>1</sup>	GLDAS <sup>2</sup>	ERA5 <sup>3</sup>
Source	NASA	NASA	ECMWF
Coverage	Global	Global	Global
Update frequency	≤50 h	Monthly	Daily
Resolution (temporal/spatial)	Daily/9–36 km	Daily/0.25° (≈27.75 km)	Hourly/0.25° (≈27.75 km)

<sup>1</sup> Soil Moisture Active Passive; <sup>2</sup> Global Land Data Assimilation System; <sup>3</sup> ECMWF Reanalysis 5.

Meanwhile, with the development of machine learning (ML), it is possible to train large amounts of data for prediction or classification [29]. As a result, it is possible to generate estimates of SM not only in bare soil, but also in vegetation areas that are difficult to calculate via physical techniques. In Liu et al. (2020) [30], SM in cropland was estimated with high accuracy using ML methods, including a generalized regression neural network (GRNN), random forest (RF), support vector regression (SVR), and deep neural networks (DNN), with Sentinel-1 and Sentinel-2 data used as inputs (RMSE = 0.005–0.040 cm<sup>3</sup>/cm<sup>3</sup>, CC ≥ about 0.900) [30]. A higher accuracy was obtained using RF and DNN than using GRNN and SVR. In Lee et al. (2019) [23], SM modeling was performed using DNN and various data (e.g., solar insolation, outgoing longwave radiation, broadband albedo, the normalized difference vegetation index (NDVI), and precipitation). The model using random sampling (RMSE = 3.644%, CC = 0.895) showed higher accuracy than the model using yearly sampling (RMSE = 8.745%, CC = 0.473) [23]. However, there were fears that the accuracy may greatly vary depending on the data sampling methods used, even though it was a deep learning model. Automated machine learning (AutoML) is a recently developed ML method that offers high productivity by automating time-consuming and repetitive tasks involved in ML model development, such as selecting the best ML model among various models and optimizing hyperparameters [31]. Babaeian et al. (2021) [32] generated an AutoML-based SM model using NDVI and near-infrared transformed reflectance (NTR) derived from unmanned aerial system (UAS) data and soil properties. The model had high accuracy (RMSE < 0.020 m<sup>3</sup>/m<sup>3</sup>, CC > 0.900) [32]. However, because the results were obtained through training and validation using the random sampling method, with only nine images being acquired during winter (December–February) and early spring (March), additional validation with sufficient data was necessary. Apart from that study, there has been little discussion of AutoML-based SM prediction methods. In addition, as previous

ML-based studies used input data that were not provided in semi-real time, the discussion of real-time soil moisture information is still limited.

Therefore, in this study, we aimed to retrieve SM grid data that could accurately capture local-scale variabilities in SM for complex and heterogeneous land surfaces in countries such as South Korea. We also sought to develop a model based on ML techniques (RF and AutoML) that could estimate daily SM in real time and calculate SM with stable accuracy not only via a random sampling method, given the high accuracy demonstrated in the previous studies [23,30,32], but also via a yearly sampling method. This study was conducted in Korean regions for which in situ SM data were available. The RF- and AutoML-based SM models used satellite image data, numerical weather prediction (NWP) data, and topographic data provided within one day as the input data, and these data were trained and validated using random sampling and yearly sampling methods. The remainder of this paper is organized as follows: Section 2 describes the data used, pre-processing, the input variables used in modeling, the model structure based on RF and AutoML, and the performance evaluation method used. Section 3 presents an evaluation of quantitative accuracy of SM predicted via the constructed models using in situ SM and global SM data, and an evaluation of the qualitative accuracy of the results for extreme weather events.

## 2. Data and Methods

### 2.1. Study Area

The study areas encompassed the country of South Korea, ranging from latitude 33.0°N–38.7°N to longitude 125.0°E–129.6°E (Figure 1). South Korea is situated in the mid-latitude temperate climate zone and experiences four distinct seasons [33]. Spring and autumn typically have clear and dry weather due to the influence of the migratory high pressure, whereas summer is humid and hot due to the effects of the North Pacific high pressure [33]. Winter is cold and dry due to the influence of continental high pressure [33]. Over a period of 106 years (from 1912 to 2017), the average annual temperature was 13.2 °C, and the annual precipitation rate was about 1237.4 mm [34]. Geographically, the region is characterized by the significant presence of the Taebaek Mountains in the east [35], resulting in higher elevation in the east and relatively lower elevation in the west (Figure 1a). Most of the topsoil consists of sandy loam, loam, silt loam, clay loam, silty clay loam, and loamy sand (Figure 1b).

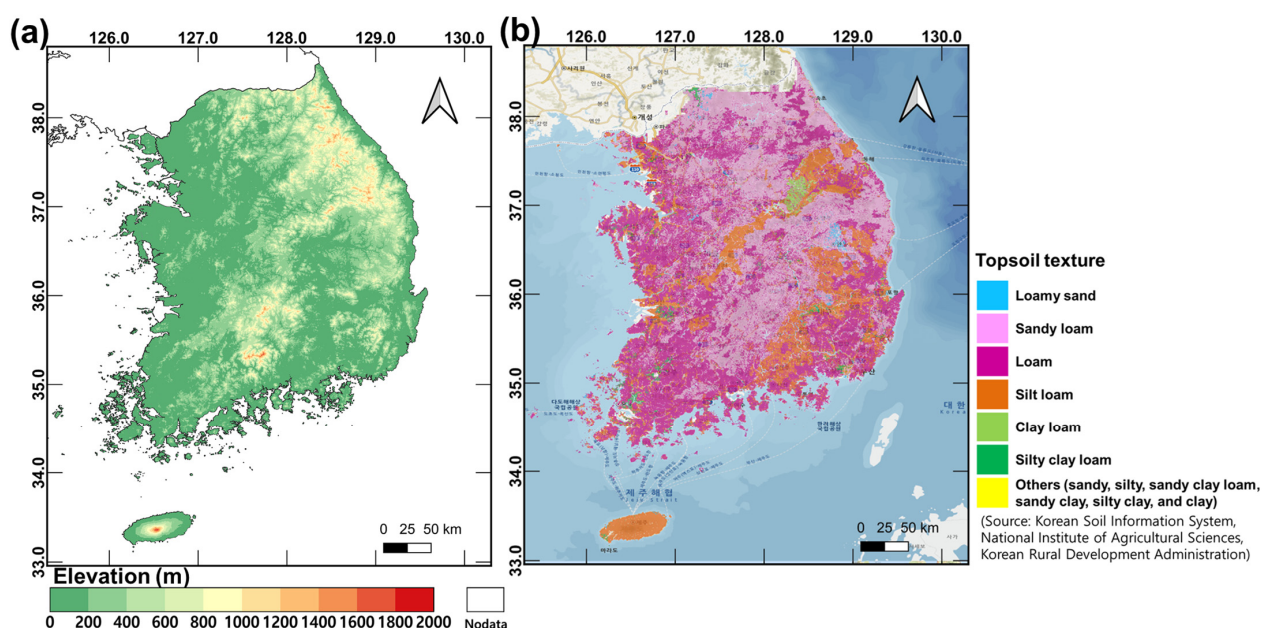


Figure 1. Map of (a) elevation and (b) topsoil texture [36] in the study area.

## 2.2. Materials

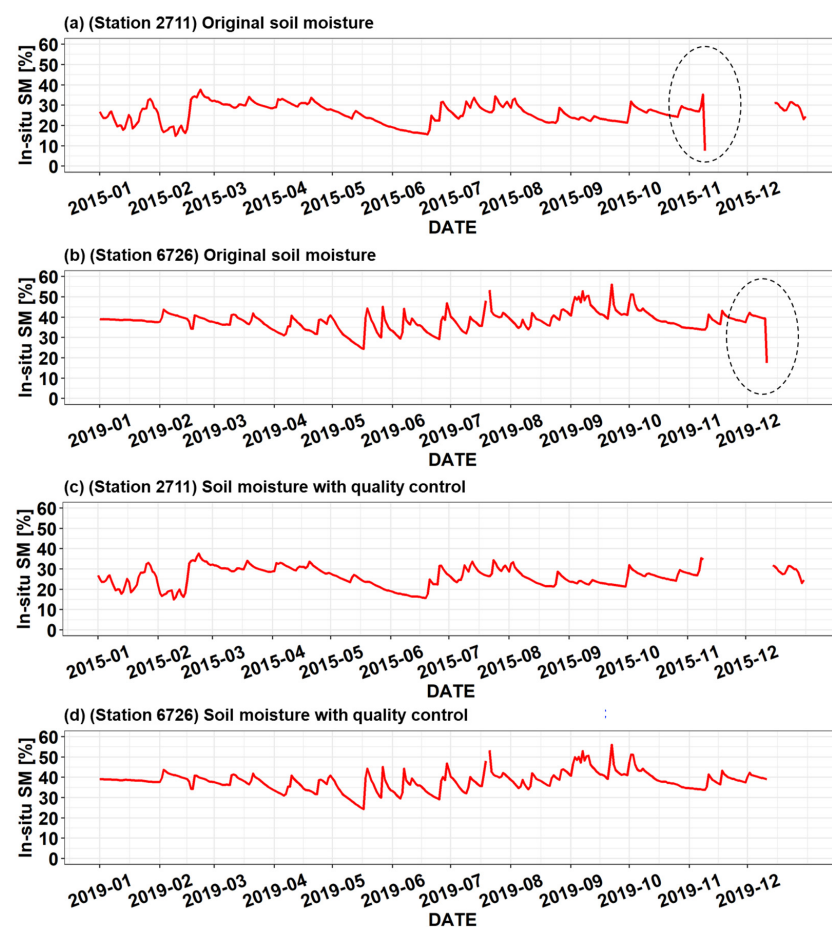
To calculate daily SM in real time, we used satellite images, NWP data, and topographic data provided within 1 day. The data period lasted from March to November and from 2014 to 2021 (Table 2). The winter season (December, January, and February) was excluded because the accuracy of in situ SM data could be reduced due to snow cover and soil freezing [37]. The Korean Rural Development Administration (RDA) in South Korea provided SM data (unit: %) collected at 10 min intervals [38]. The SM data were measured by a time domain reflectometry (TDR) sensor at a depth of 10 cm for more than 73 stations (as of November 2021). However, some of the measurements showed an abrupt decrease in value (Figure 2a,b), and some SM stations were located in unsuitable places, such as buildings and parking lots (Figure 3). For these reasons, only 23 SM stations were selected by screening unsuitable places. Then, through a quality control (QC) process, extremely low SM values less than negative 3 standard deviations from the daily mean SM were eliminated (Figure 2c,d). The refined daily averages for the 23 stations were used to develop and evaluate the SM models.

**Table 2.** Data used for soil moisture modeling.

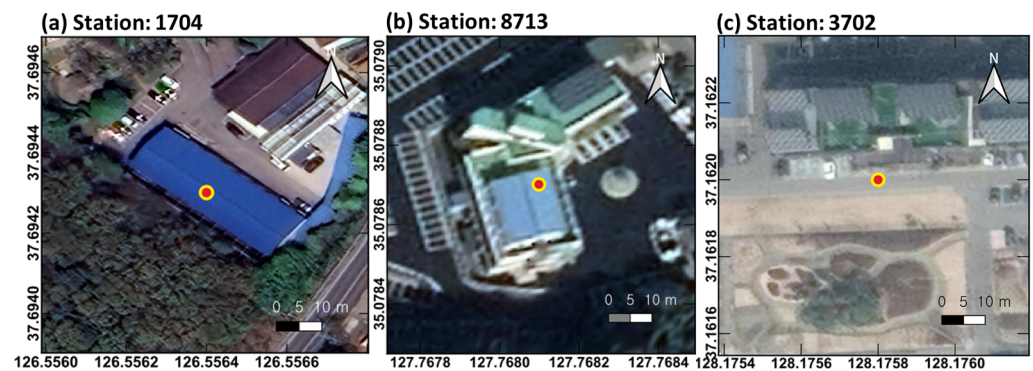
Data Type	Input Variables	Spatial Resolution	Temporal Resolution	Update Frequency	Source
In situ data	Soil moisture (SM) (Depth: 0~10 cm)	Point	10 min	10 min	RDA
Satellite data	Rainfall	0.1°	Daily	12 h	IMERG
	Normalized difference vegetation index (NDVI)	2 km	Daily	Daily	GK2A
		1 km	8 days	-	VIIRS
	Downward shortwave radiation (DSR)	2 km	10 min	10 min	GK2A
Numerical weather prediction Data	Total downward surface shortwave flux (DSSF)	1.5 km	3 h	8 times per day	LDAPS
	Air temperature (Tair)				
	Land surface temperature (LST)				
	Soil temperature (Tsoil) (Depth: 0~10 cm)				
	Relative humidity (RH)				
	Latent heat flux (LE)				
Topographic data	Slope, elevation, topographic ruggedness index (TRI), and aspect	30 m	-	-	SRTM DEM

The Integrated Multi-Satellite Retrievals for GPM (IMERG) late precipitation data and the Geo-Kompsat-2A (GK2A) NDVI and downward shortwave radiation (DSR) data were used as satellite data. IMERG provides global cumulative daily precipitation data (unit: mm) with a 0.1-degree resolution, which was estimated using microwave data obtained from several passive microwave satellites through the Global Precipitation Measurement (GPM) mission of NASA and the Japan Aerospace Exploration Agency (JAXA) (Table 2) [39]. GK2A is Korea's geostationary meteorological satellite, which was launched on 5 December 2018, and it is equipped with Advanced Meteorological Imager (AMI), which has 16 channels for meteorological and space weather-monitoring missions [40]. GK2A provides daily NDVI and 10-minute DSR data with a resolution of 2 km [41,42] (Table 2). As GK2A NDVI data underestimate noise, which can be less than 0.3 in summer due to meteorological factors, such as clouds, heavy rain, and monsoon [43], the data were used after applying the real-time noise improvement approach to the GK2A daily NDVI product, as proposed by Lee et al. (2022) [43]. This method corrects GK2A NDVI data through processes such as time series correction reflecting the growth cycle of vegetation, the removal of outliers using long-term Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI data, and missing pixel restoration using a penalized least squares

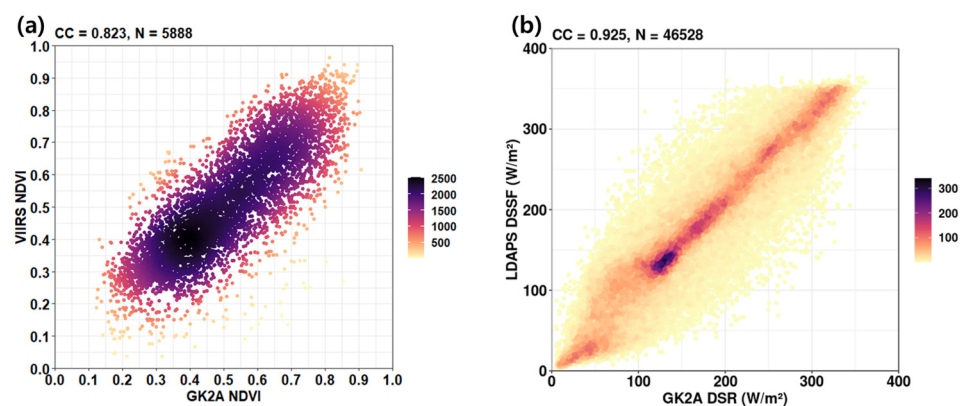
regression based on discrete cosine transform (DCT-PLS) (Figure A1) [43]. This approach effectively improved the underestimation via meteorological factors (Figures A2 and A3), and it showed that the correlation with MODIS NDVI was higher than the original value in all seasons and the difference decreased [43]. GK2A DSR data were used as daily average values, being expressed as  $W/M^2$  [42]. However, because the GK2A products were only supplied from July 2019 onwards, insufficient training data were available for stable SM modeling. Thus, to perform SM modeling using as much data as possible, for the period before 2020 (2014–2019), we included the NASA Visible Infrared Imaging Radiometer Suite (VIIRS) 16-day composite NDVI product (VNP13A2), which provided data every 8 days [44], and the Korea Meteorological Administration (KMA) Local Data Assimilation and Prediction System (LDAPS) total downward surface shortwave flux (DSSF) data [45]. We then confirmed whether these data could be used as substitute data through correlation analysis. Both the relationships between GK2A NDVI and VIIRS NDVI and between GK2A DSR and LDAPS DSSF showed high correlation ( $\geq 0.823$ ) (Figure 4). Therefore, VIIRS NDVI and LDAPS DSSF data were used before 2020, and GK2A NDVI and DSR data were used after 2020. LDAPS is a local forecasting model that predicts weather in the Korean Peninsula and provides meteorological and surface data every 3 h (8 times per day) at a resolution of 1.5 km (Table 2) [45–47]. DSSF, air temperature ( $T_{air}$ ), land surface temperature (LST), soil temperature ( $T_{soil}$ ), relative humidity (RH), and latent heat flux (LE) data from LDAPS were used to perform SM modeling. As noted above, DSSF data served as substitute data for GK2A DSR, and daily average values were used. LDAPS data for the other variables were obtained at UTC 03:00.



**Figure 2.** Time series distribution of in situ soil moisture (SM) before and after quality control (QC): (a) SM in 2015 at station 2711, (b) SM in 2019 at station 6726, (c) SM with QC in 2015 at station 2711, and (d) SM with QC in 2019 at station 6726. The dotted circles in (a,b) showed an abrupt decrease of SM values.



**Figure 3.** Stations in improper locations such as (a,b) buildings and (c) road (the figure is based on Google maps).



**Figure 4.** (a) The correlation between 16-day composite values of GK2A NDVI and VIIRS NDVI, and (b) the correlation between the daily average values of GK2A DSR and LDPAS DSSF; both correlations were determined from 1 January 2020 to 31 December 2021.

The topographic data included slope, elevation, topographic ruggedness index (TRI), and aspect, all of which were considered to be constant variables. The data were extracted from Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) data at an approximate resolution of 30 m, which was jointly produced and provided by NASA and the U.S. National Geospatial-Intelligence Agency (NGA) [48]. The satellite, NWP, and topographic data were resampled to a 500-m resolution and adjusted to the coordinate reference system of the World Geodetic System 1984 (WGS84). In addition, these data were cropped to cover the area of South Korea (latitude 33.0°N to 38.7°N, longitude 125.0°E to 129.6°E), where SM maps were created.

### 2.3. Input Variables Used to Perform Soil Moisture Modeling

The 13 explanatory variables used to perform SM modeling consisted of factors that affect SM status (Figure 5): 10-day cumulative standardized precipitation index (SPI), 20-day cumulative SPI, NDVI, DSR, Tair, LST, Tsoil, RH, LE, slope, elevation, TRI, and aspect. The 10- and 20-day cumulative SPI (SPI10 and SPI20) were related to precipitation, which supplies water to the land surface. SPI is a widely used drought index that is recommended by the World Meteorological Organization (WMO) for the diagnosis and prediction of meteorological drought [49]. As SM status is affected by cumulative precipitation [50], the SPI based on n-days cumulative precipitation was used. To calculate SPI10 and SPI20, the cumulative 10- and 20-day precipitation extracted from the IMERG daily precipitation data, respectively, were calculated, after which step the cumulative probability of the precipitation was determined using the empirical cumulative density function (ECDF), with the values being transformed into z-scores. NDVI is calculated by dividing the difference between near-infrared and red reflectance by their sum and

represents the degree of vegetation health. As the latter factor tends to decrease under water stress, NDVI is used to indirectly monitor SM and drought conditions [51,52]. DSR, Tair, LST, Tsoil, RH, and LE, corresponding to the red arrows in the Figure 5, are related to SM loss. DSR, which refers to the amount of short-wave radiation derived from the sun, affects evaporation and transpiration, i.e., water movement from the land surface to the air, and is the energy source for the land surface [53]. Tair, LST, and Tsoil (depth: 0–10 cm) are temperature factors that usually cause evaporation on the land surface and transpiration in vegetation, thus affecting SM [54–56]. RH is the ratio of the current atmospheric water vapor amount to the saturation water vapor amount at the current temperature, expressed as a percentage (%). The RH influences evapotranspiration and, therefore, SM, with higher rates of evapotranspiration occurring on days with a low RH [57]. LE is the amount of heat released or absorbed when a substance changes state without a change in temperature [58], with absorption occurring when water changes to vapor via evapotranspiration [59]. LE is, thus, used to calculate evapotranspiration and related to SM loss [60]. Topographic factors, such as slope, elevation, TRI, and aspect, were selected because topography, as a determinant of the direction of water movement, plays an important role in the amount and distribution of SM [61,62]. Slope was the rate of change in elevation for each DEM cell and has a value between 0 and 90°. Elevation was the height of an object or place above the sea, and its unit is m. TRI was the amount of elevation difference between the adjacent cells of a DEM [63]. Aspect was the downslope direction of the maximum rate of change in value between each cell and adjacent cells [64].

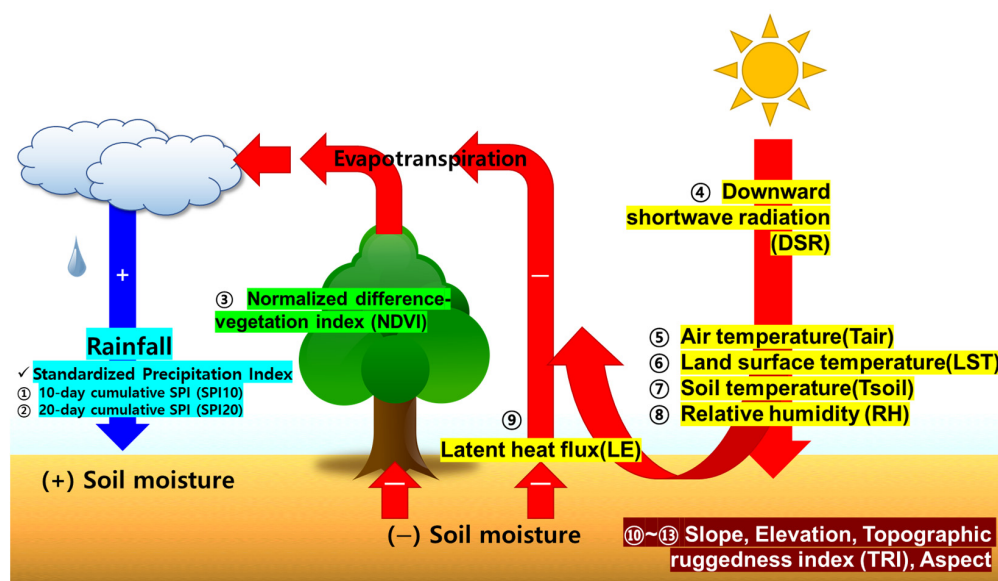


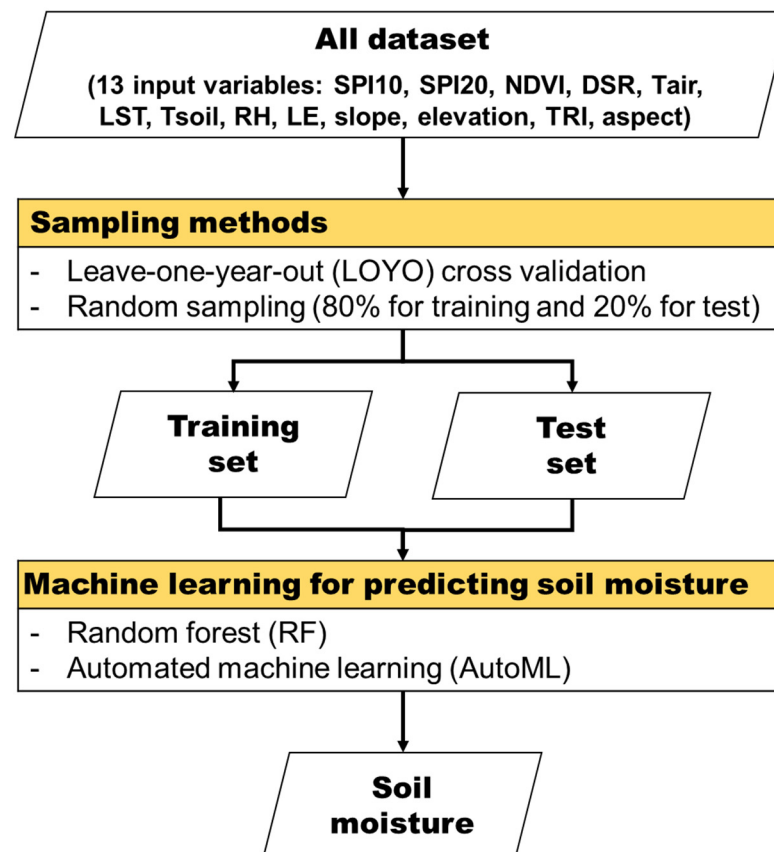
Figure 5. The 13 input variables used to model soil moisture.

#### 2.4. Model Development and Performance Evaluation

A daily matchup dataset consisting of the 13 aforementioned input variables affecting SM and in situ SM data was constructed, and SM models were then built using the ML algorithms RF and AutoML (Figure 6). The RF model classifies or predicts (regresses) via the ensemble of many decision trees built during the training process and is an improved technique used to create a single decision tree with a high probability of overfitting [65,66]. As RF randomly constructs many individual decision trees during training, it mitigates the bias of the model and enhances generalization [67]. The RF-based SM model consisted of 50 decision trees and had a depth of 20. Default hyperparameters set using H<sub>2</sub>O, which is a Java-based ML/AI platform, were used. AutoML enhances productivity by automating time- and resource-consuming processes (such as algorithm selection, modeling, hyperparameter optimization, and comparisons between dozens of models) performed to obtain models with high accuracy [31]. The AutoML library provided by H<sub>2</sub>O internally



provides ML models, such as distributed random forest (DRF), a generalized linear model (GLM), extreme gradient boosting (XGBoost), gradient boosting machine (GBM), and DNN, and allows the construction of an ensemble model from those models [31]. The constructed SM models were evaluated using leave-one-year-out (LOYO) cross-validation and random sampling. In LOYO, the data from a particular year were designated for validation, while the data from all other years were used to train the model. This process was repeated for each year to evaluate the model's performance. Therefore, with a dataset of 8 years, eight training and validation processes were performed using the LOYO method (Figure 7). This method could evaluate the average accuracy of models that could be expected in terms of predicting SM at a future point in time. In the random sampling method, the matchup dataset was randomly divided into a training set (80%) and a test set (20%), and the models used to perform SM estimation were then trained and validated. This process was repeated five times to determine the average performance. The accuracy of SM estimation was quantitatively evaluated using in situ SM data and four performance indices (mean bias error (MBE), mean absolute error (MAE), RMSE, and CC) (Table 3), and the results were compared to GLDAS SM (depth: 0–10 cm) and ERA5 SM (depth: 0–7 cm) data derived from 2020 to 2021. In addition, a qualitative evaluation was conducted by comparing SM maps derived from the constructed model based on the extreme climate report jointly published by the KMA and various governmental organizations of Korea.



**Figure 6.** Flow chart used to perform modeling of soil moisture. The 13 input variables used to perform modeling were 10-day cumulative SPI (SPI10), 20-day cumulative SPI (SPI20), normalized difference vegetation index (NDVI), downward shortwave radiation (DSR), air temperature (Tair), land surface temperature (LST), soil temperature (Tsoil), relative humidity (RH), latent heat flux (LE), slope, elevation, topographic ruggedness index (TRI), and aspect. Two sampling strategies, i.e., leave-one-year-out (LOYO) cross-validation and random sampling (80% for training and 20% for testing), were employed during the modeling process. The machine learning techniques used to estimate SM were random forest (RF) and automated machine learning (AutoML).

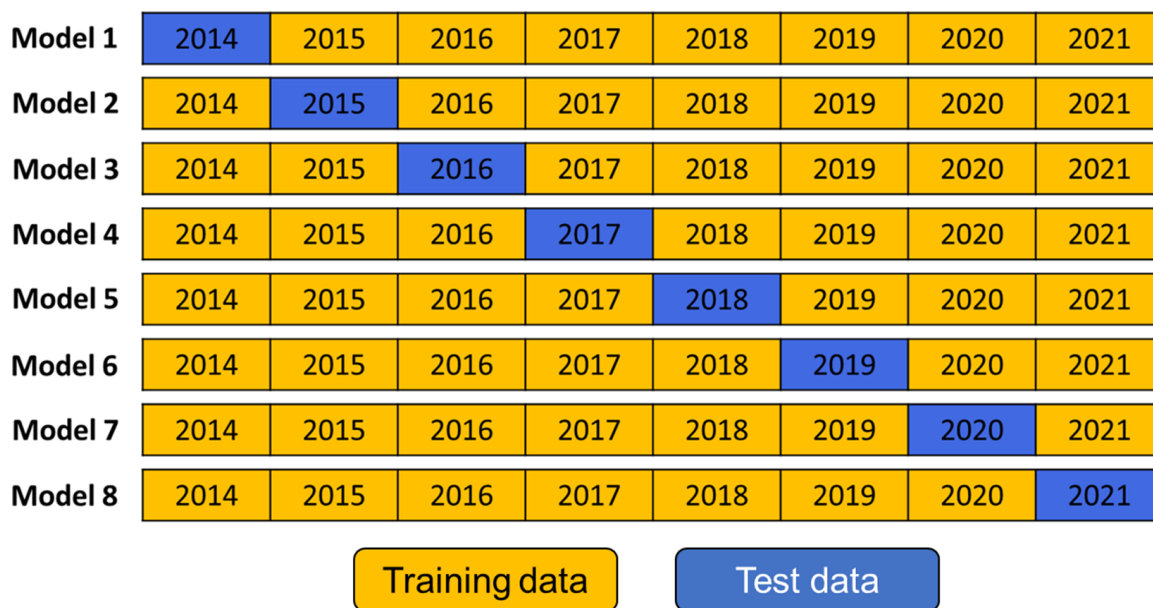


Figure 7. Leave-one-year-out (LOYO) cross-validation.

Table 3. Performance indices used to evaluate soil moisture models.

Performance Indices	Equation
Mean bias error (MBE)	$\frac{1}{N} \sum_{i=1}^N (\text{predicted SM} - \text{insitu SM})$
Mean absolute error (MAE)	$\frac{1}{N} \sum_{i=1}^N  \text{predicted SM} - \text{insitu SM} $
Root mean square error (RMSE)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (\text{predicted SM} - \text{insitu SM})^2}$
Correlation coefficient (CC)	$X = \text{predicted SM}, Y = \text{insitu SM},$ $\frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}$

### 3. Results and Discussion

#### 3.1. Model Performance Evaluation

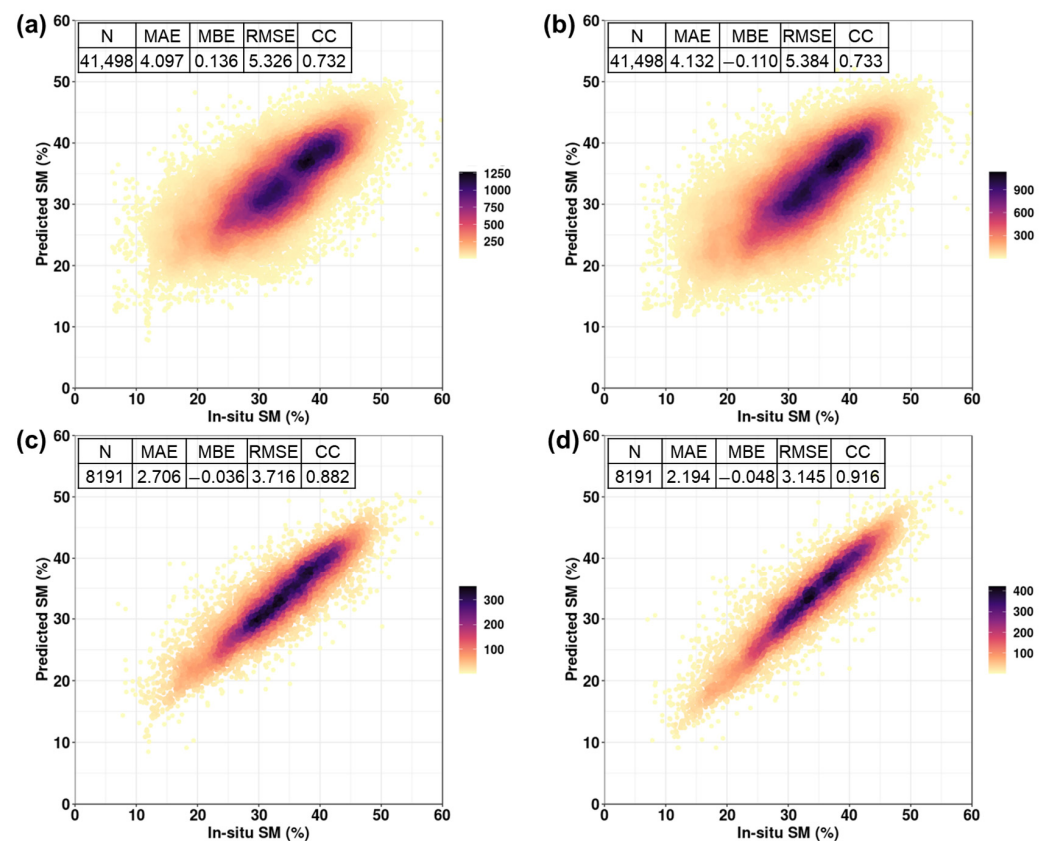
In this study, SM models based on RF and AutoML were trained and validated using LOYO and random sampling methods. When evaluated using the LOYO method, the two models had similar accuracy rates. On average, the MAE between the predicted and in situ SMs was about 4.100%, and the MBE was close to zero (Tables 4 and 5). The RMSE was approximately 5.300%, and the CC of about 0.730 indicated a strongly positive linear relationship (Tables 4 and 5; Figure 8a,b). There were no significant differences between the accuracy rates of the SM models trained using RF and AutoML between years, and there was no temporal dependency (Tables 4 and 5). SM estimation in future years (e.g., 2022, 2023, etc.) will, therefore, have a similar accuracy. Moreover, a stable accuracy will be obtained even if these models are trained by mixing different data of the same type when there is not enough data, such as in the case of NDVI (from VIIRS and GK2A) and DSR (from LDAPS and GK2A).

**Table 4.** LOYO accuracy of the soil moisture model based on random forest (RF).

Year	N	MAE	MBE	RMSE	CC
2014	6209	3.782	−0.386	4.814	0.740
2015	6219	3.776	−0.378	4.908	0.776
2016	5764	5.130	0.982	6.739	0.667
2017	5004	4.159	0.817	5.351	0.685
2018	4849	3.592	−0.191	4.864	0.792
2019	4666	3.692	0.576	4.755	0.747
2020	4091	3.936	0.543	5.017	0.795
2021	4696	4.712	−0.879	6.158	0.653
Avg.	41,498	4.097	0.136	5.326	0.732

**Table 5.** LOYO accuracy of the soil moisture model based on AutoML.

Year	N	MAE	MBE	RMSE	CC
2014	6209	3.670	−0.816	4.737	0.765
2015	6219	3.847	−0.957	5.030	0.775
2016	5764	5.158	0.872	6.780	0.664
2017	5004	4.235	0.688	5.446	0.679
2018	4849	3.686	−0.362	5.005	0.791
2019	4666	3.737	0.323	4.825	0.744
2020	4091	3.997	0.340	5.086	0.783
2021	4696	4.723	−0.969	6.162	0.660
Avg.	41,498	4.132	−0.110	5.384	0.733

**Figure 8.** Scatter plot between the in situ soil moisture (SM) and predicted SM using the SM models based on (a) LOYO and RF, (b) LOYO and AutoML, (c) random sampling and RF at round 4, or (d) random sampling and AutoML at round 4.

In the evaluation using the random sampling method, the matchup dataset was divided accordingly, and training and validation of the RF- and AutoML-based SM models were performed five times to determine their respective average performances. Both models used the same training and verification datasets generated on a new random sampling basis every round. In all rounds, the performance was similar (Tables 6 and 7), and a strong linear relationship between the predicted and in situ SMs was confirmed in the scatterplots (Figure 8c,d). SM models based on AutoML had a higher accuracy (MAE = 2.212%, MBE = −0.003%, and RMSE = 3.186%) and correlation (CC = 0.913) than those based on RF (Tables 6 and 7). This finding means that AutoML can build models with better predictive capabilities by automating algorithm selection, modeling, and hyperparameter optimization. A tendency toward a higher accuracy based on random sampling than the LOYO method was also determined. The primary reason for such difference lies in the nature of random sampling, which aims to obtain a representative sample from a population through an unbiased selection process [68]. Given the 41,498 matchup datasets, the randomly chosen test set (20%) will likely share a bias distribution similar to that of the training set (80%). Based on the LOYO approach, however, the statistical characteristics of the validation dataset of a specific year may or may not be identical to those of the other years used in training because of significant differences in weather conditions due to climate change and extreme weather every year. For this reason, random sampling is more likely to have higher performance scores than LOYO, but LOYO is closer to realistic conditions recorded nowadays, considering climate change.

As a result of model training, RF provides variable importance. Figure 9 shows the average variable importance provided by the RF-based SM models constructed using the LOYO method. The most important variables used to estimate SM were SPI10 and SPI20, i.e., precipitation affecting the water supply to the land surface, followed by topographical variables (elevation, slope, TRI, and aspect) that affect the distribution of SM, the vegetation variable (NDVI), and SM loss variables (Tsoil, Tair, LST, RH, LE, and DSR).

**Table 6.** Random sampling accuracy of the soil moisture models based on random forest (RF).

Round	N	MAE	MBE	RMSE	CC
1	8191	2.708	−0.032	3.781	0.877
2	8191	2.719	0.030	3.749	0.882
3	8191	2.730	−0.045	3.793	0.877
4	8191	2.706	−0.036	3.716	0.882
5	8191	2.701	−0.027	3.749	0.876
Avg.	8191	2.713	−0.022	3.758	0.879

**Table 7.** Random sampling accuracy of the soil moisture models based on AutoML.

Round	N	MAE	MBE	RMSE	CC
1	8191	2.185	0.010	3.174	0.914
2	8191	2.268	0.052	3.236	0.912
3	8191	2.218	−0.022	3.212	0.912
4	8191	2.194	−0.048	3.145	0.916
5	8191	2.195	−0.006	3.165	0.912
Avg.	8191	2.212	−0.003	3.186	0.913

### 3.2. Comparison with Other Soil Moisture Data

A comparison between the daily SM data derived from the global representative datasets (GLDAS and ERA5) and the RF- and AutoML-based SM models developed in this study, using the in situ SM data for a 2-year period (2020–2021), showed that the latter models have higher accuracy rates and stronger correlations than the former datasets (MAE = 4.350–4.385%, RMSE = 5.656–5.686%, CC ≈ 0.720) (Figure 10a,b). The accuracy rates of and correlations between GLDAS and ERA5 SM and the in situ data were low

(MAE = 6.878–7.472%, RMSE = 8.575–9.524%, CC = 0.219–0.407) (Figure 10c,d). Thus, the models developed in this study were better at predicting SM in Korea.

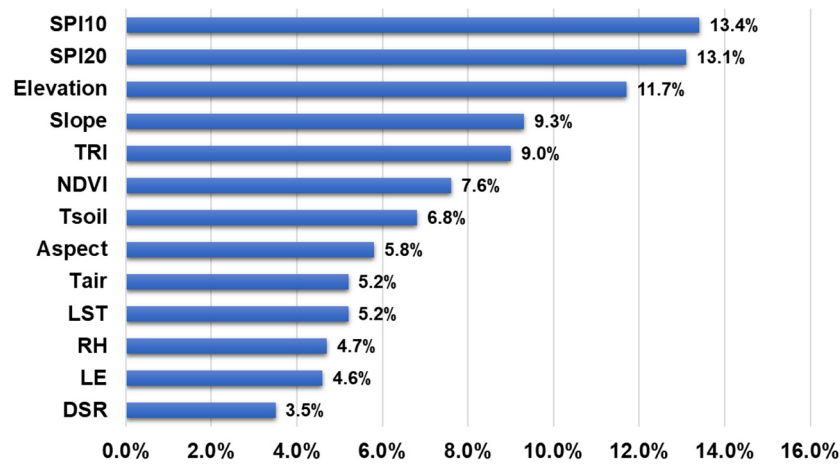


Figure 9. Variable importance calculated using the soil moisture model based on LOYO and RF.

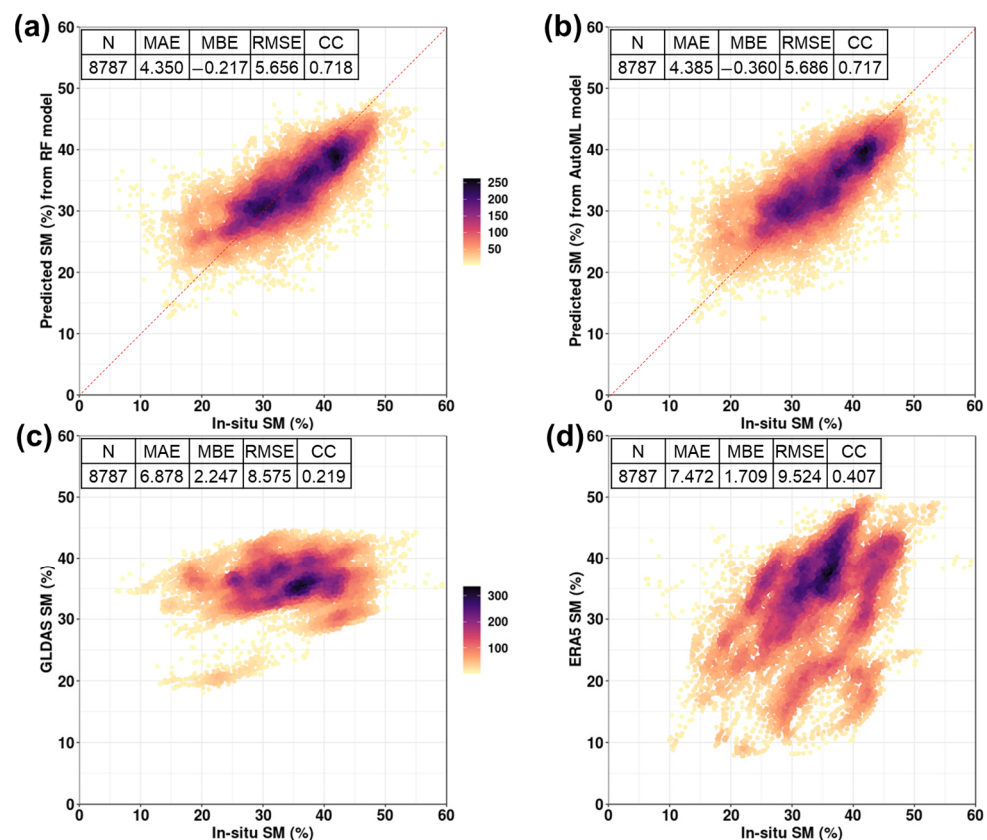


Figure 10. Scatter plot between in situ soil moisture (SM) and the predicted SM: (a) LOYO and RF model, (b) LOYO and AutoML model, (c) GLDAS SM, and (d) ERA5 SM for 2 years (2020–2021).

### 3.3. Qualitative Evaluation with Extreme Weather Events

A qualitative evaluation was performed using the SM maps produced via the SM model based on AutoML, which showed the best performance during the accuracy evaluation using the LOYO and random sampling methods, as well as the extreme weather events that occurred in 2020 and 2021, when heat waves and heavy rain were frequent in Korea. In June 2020, a heat wave lasted for almost a month (Figure 11a); in July, there was a long rainy season and a temperature inversion that resulted in lower temperatures

than those measured in June (Figure 11a,b) [69,70]. These weather events affected the amount and distribution of SM. Specifically, in June 2020, the low SM status continued due to the heat wave (Figure 12), while beginning on June 29, which marked the start of heavy rainfall, there was an increasing trend of SM with the rainy season, as well as a temperature inversion (Figure 13). In addition, the distribution of SM in July was related to a deviation in precipitation. Areas with higher-than-average precipitation tended to show SM distributions  $\geq 30\%$ , whereas areas with less-than-average precipitation tended to have SM distributions  $< 30\%$  (Figures 13 and 14).

In 2020, between 18 and 21 November, significant rainfall of up to 36.53 mm fell in the Seoul metropolitan area and southern South Korea (Figure 15a). In areas where heavy rain had fallen since 18 November, SM increased (Figure 15b).

From the middle to the end of July 2021, ground conditions became significantly dry due to heat waves and low precipitation [71]. Beginning on 11 July 2021, the persistent above-average temperatures and below-average precipitation (Figure 16) marked a gradual decrease in SM, particularly toward the end of July (Figure 17). These results show that SM estimated via the model built in this study well represented the dry/wet changes in the land surface that occurred in response to weather events.

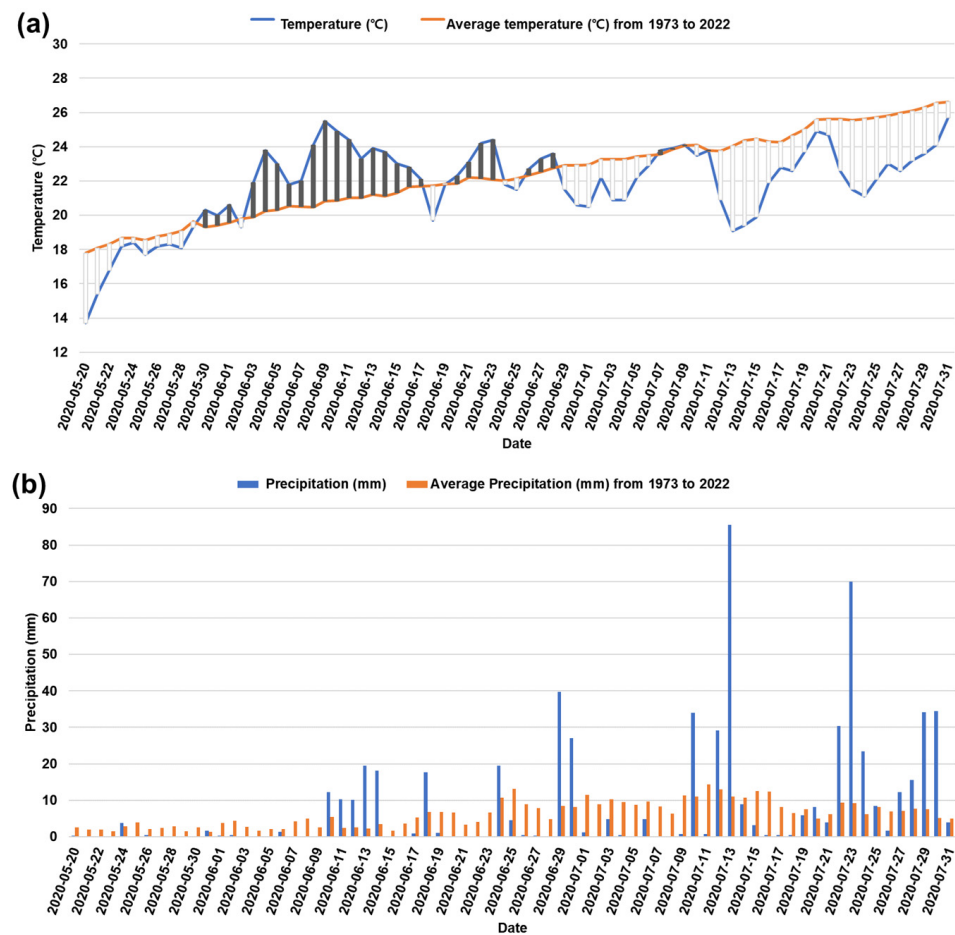


Figure 11. Distribution of (a) temperature and (b) precipitation from 20 May 2020 to 31 July 2020.

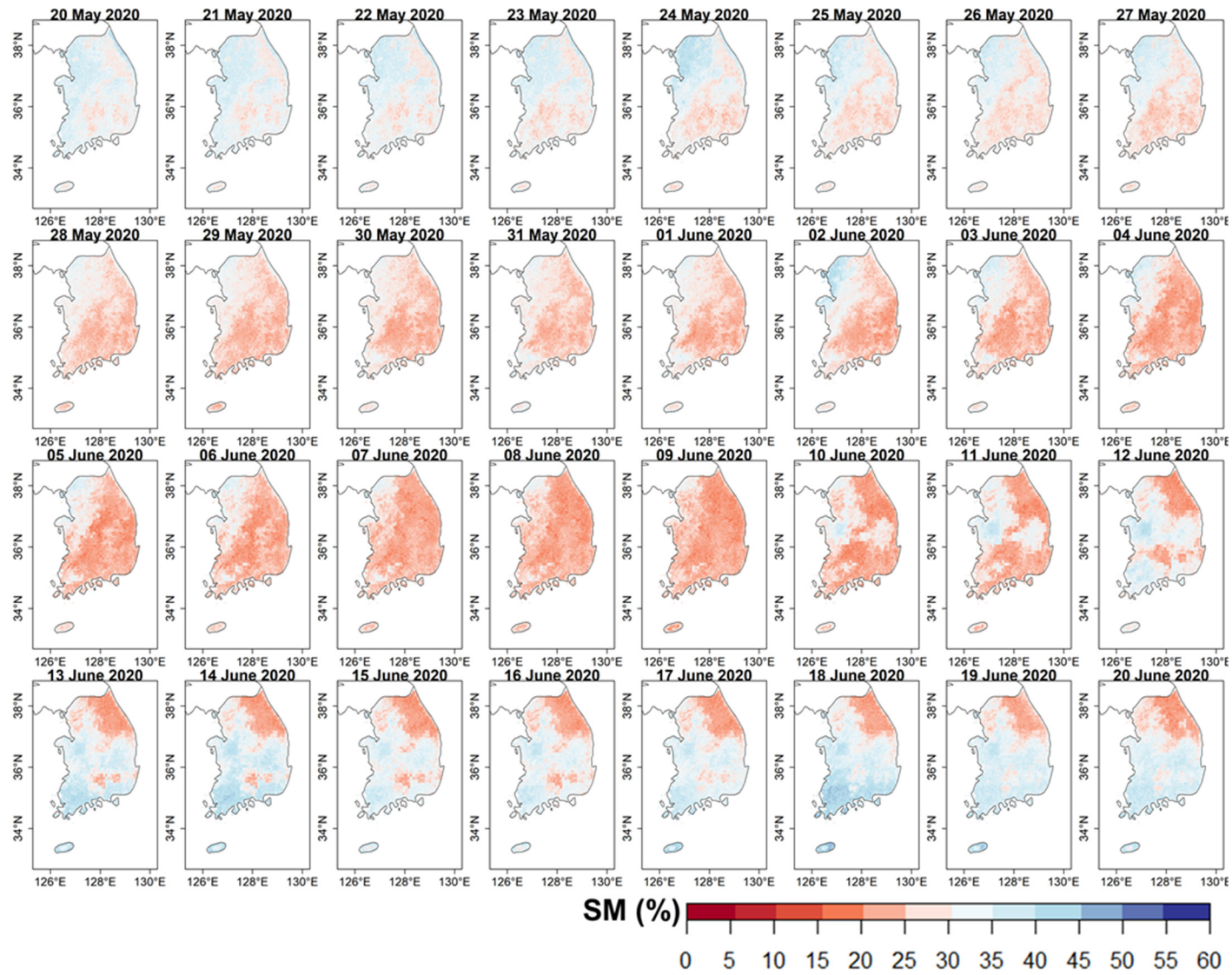
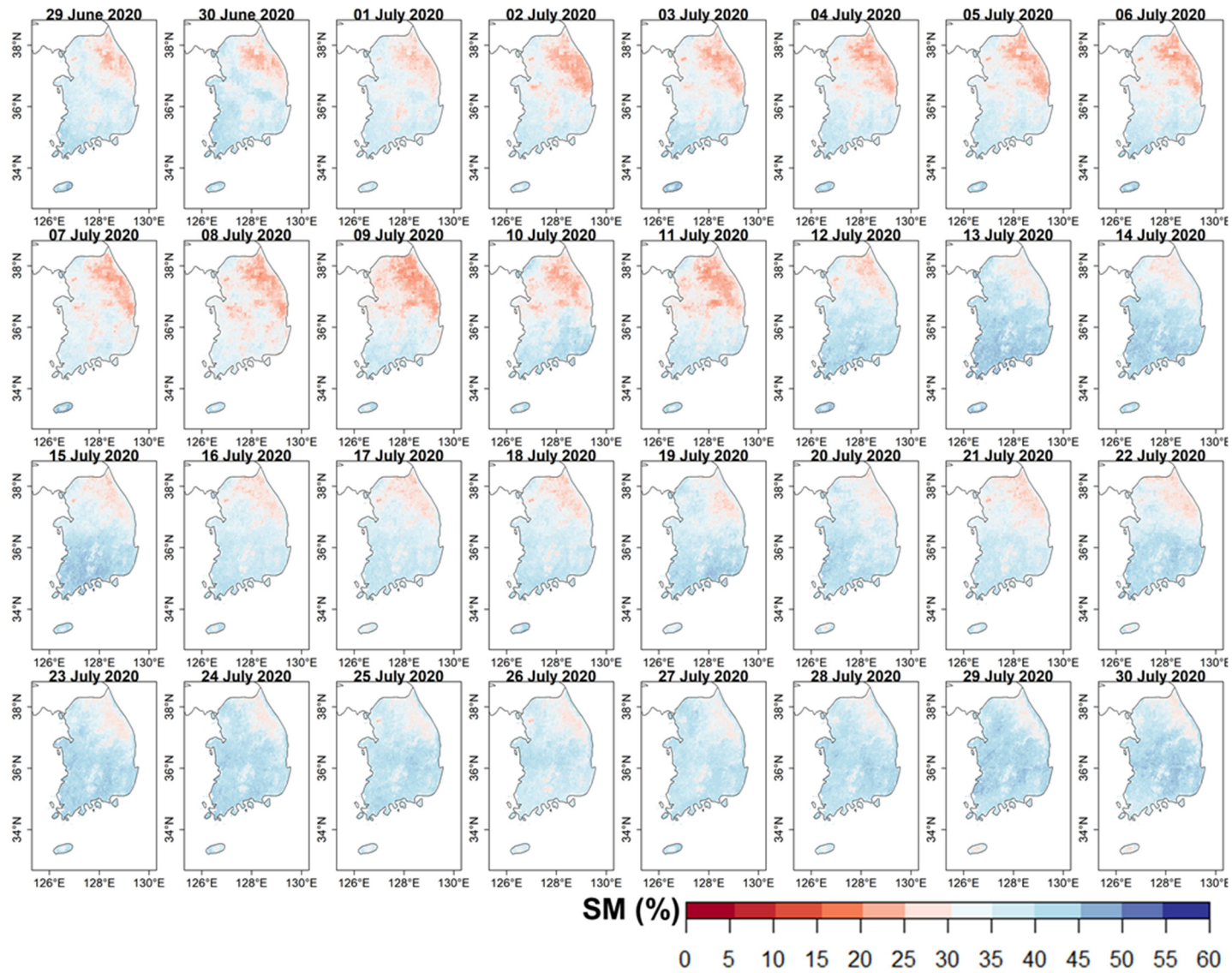


Figure 12. Soil moisture maps calculated using the AutoML-based soil moisture model from 20 May 2020 to 20 June 2020.



**Figure 13.** Soil moisture maps calculated using the AutoML-based soil moisture model from 29 June 2020 to 30 July 2020.



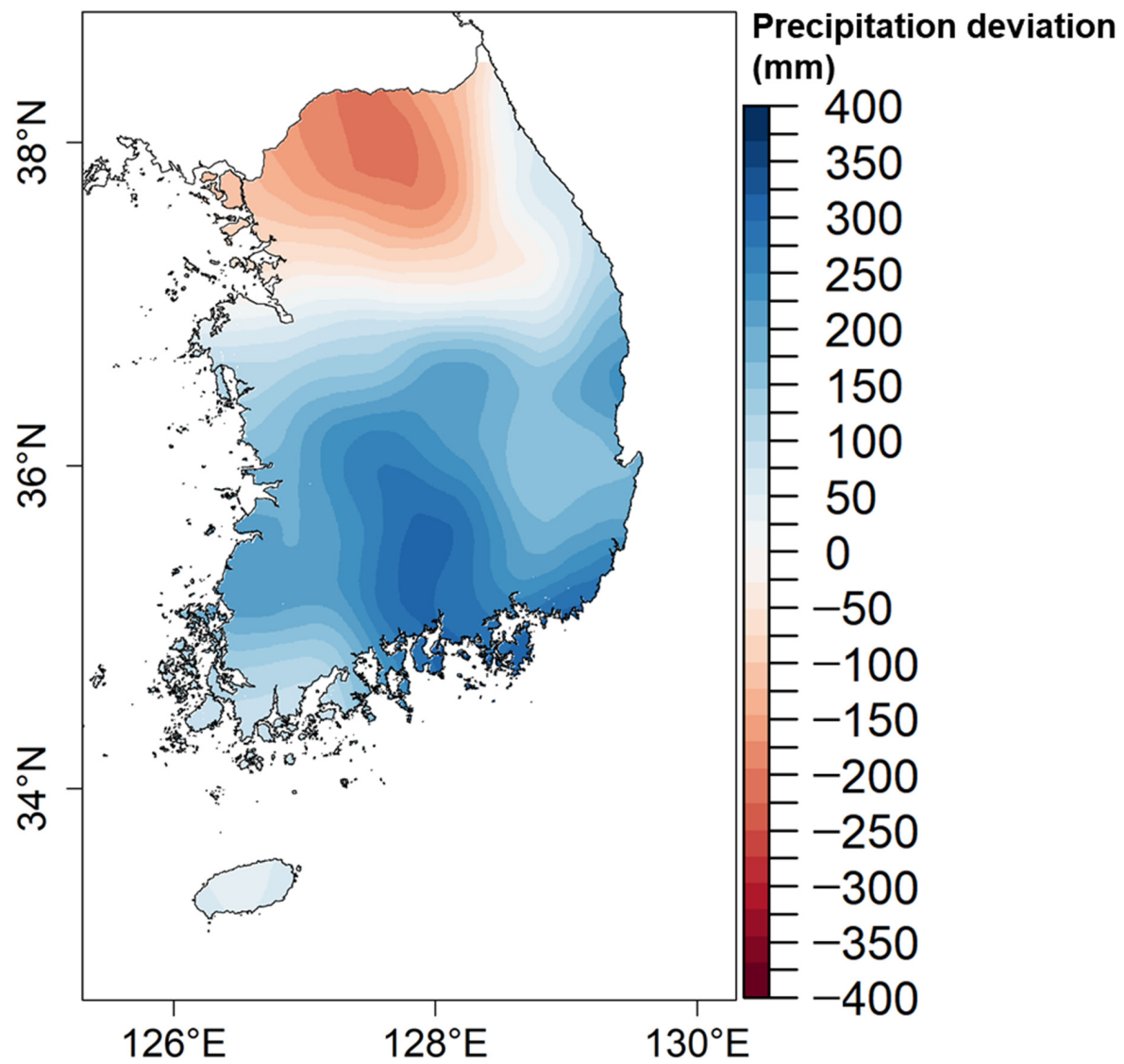
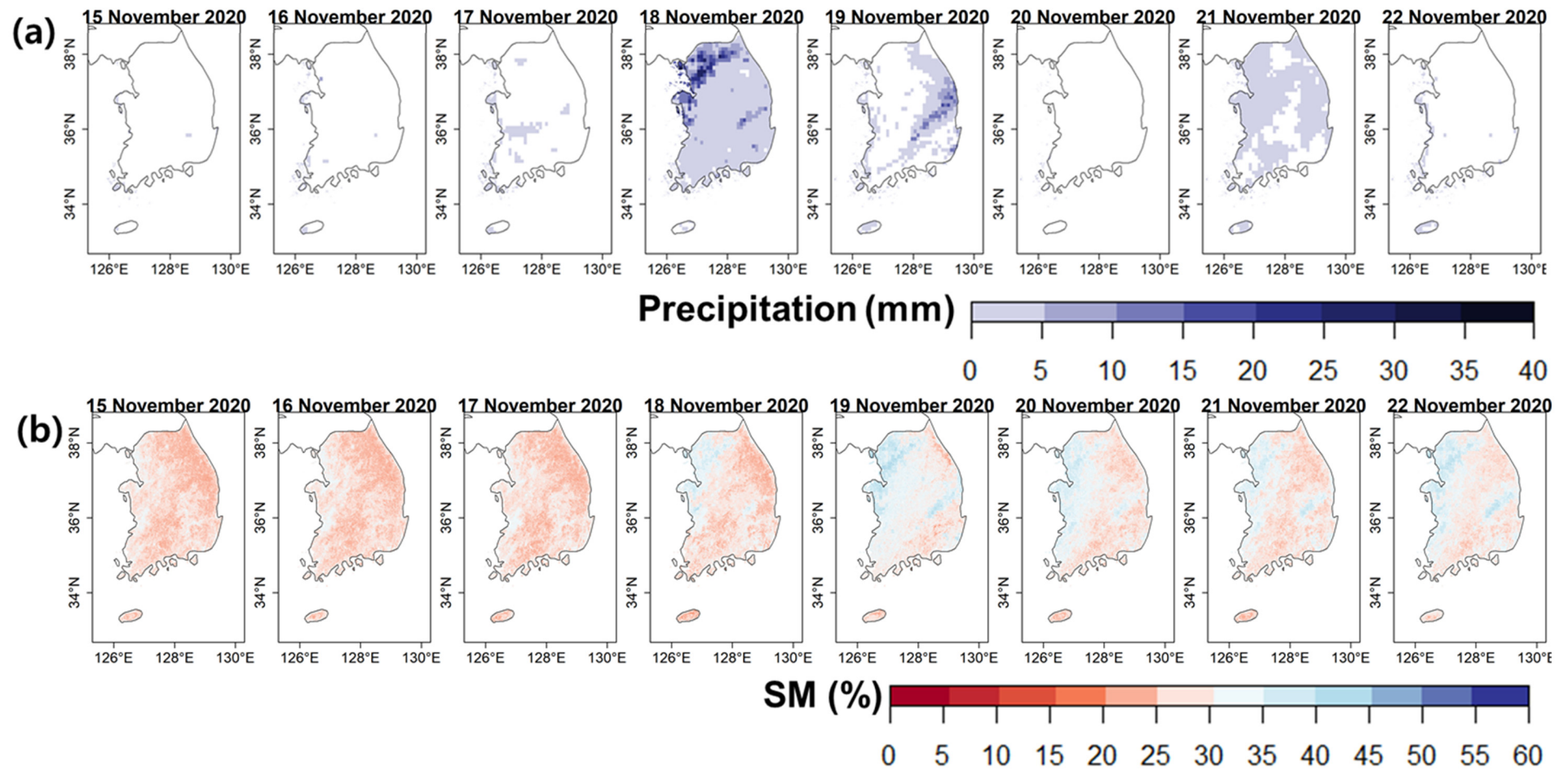


Figure 14. Distribution of precipitation deviation (mm) in July 2020 [72].



**Figure 15.** Distribution of (a) precipitation and (b) soil moisture calculated using the AutoML-based soil moisture model from 15 November 2020 to 22 November 2020.

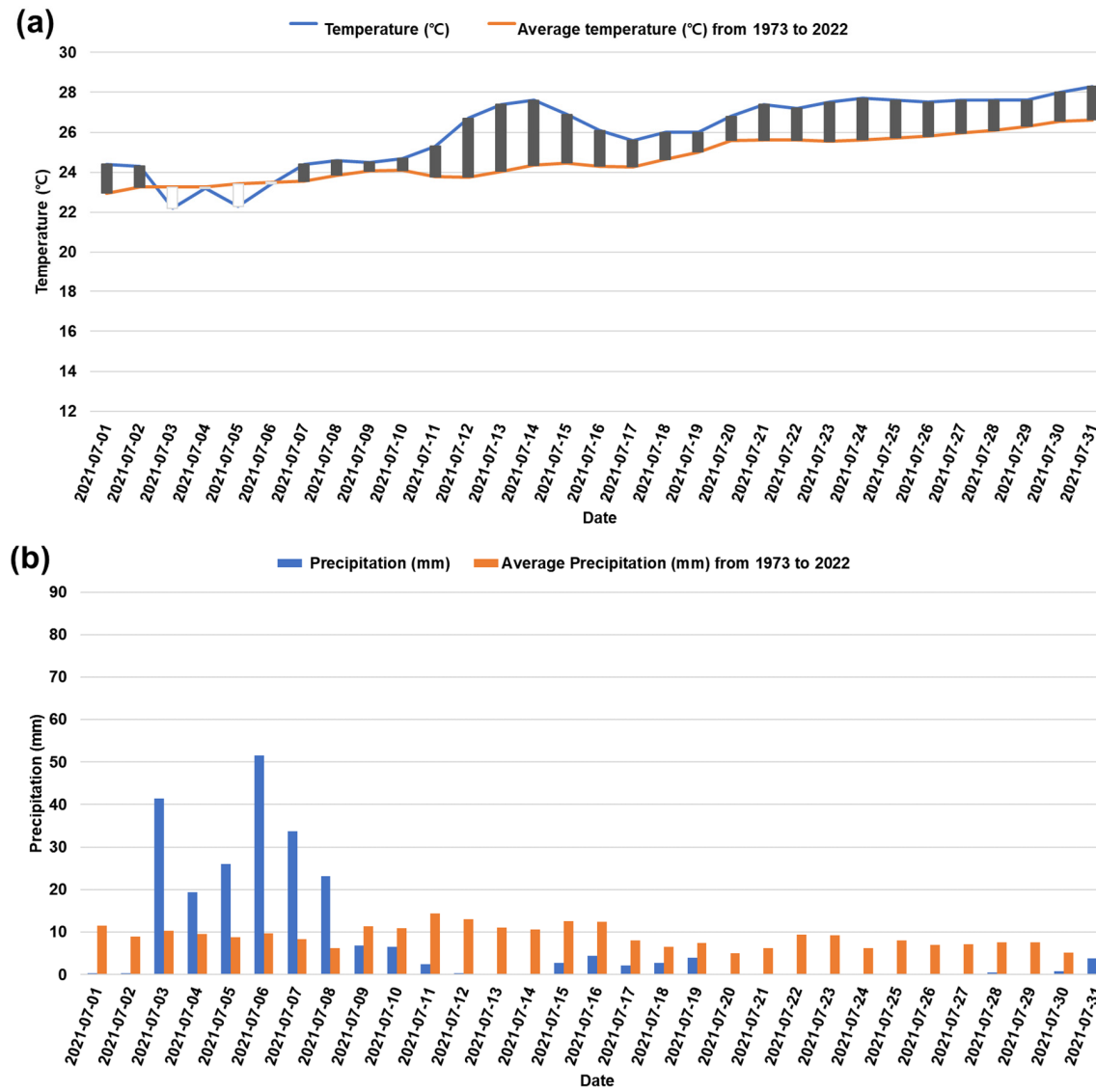


Figure 16. Distribution of (a) temperature and (b) precipitation from 1 July 2021 to 31 July 2021.

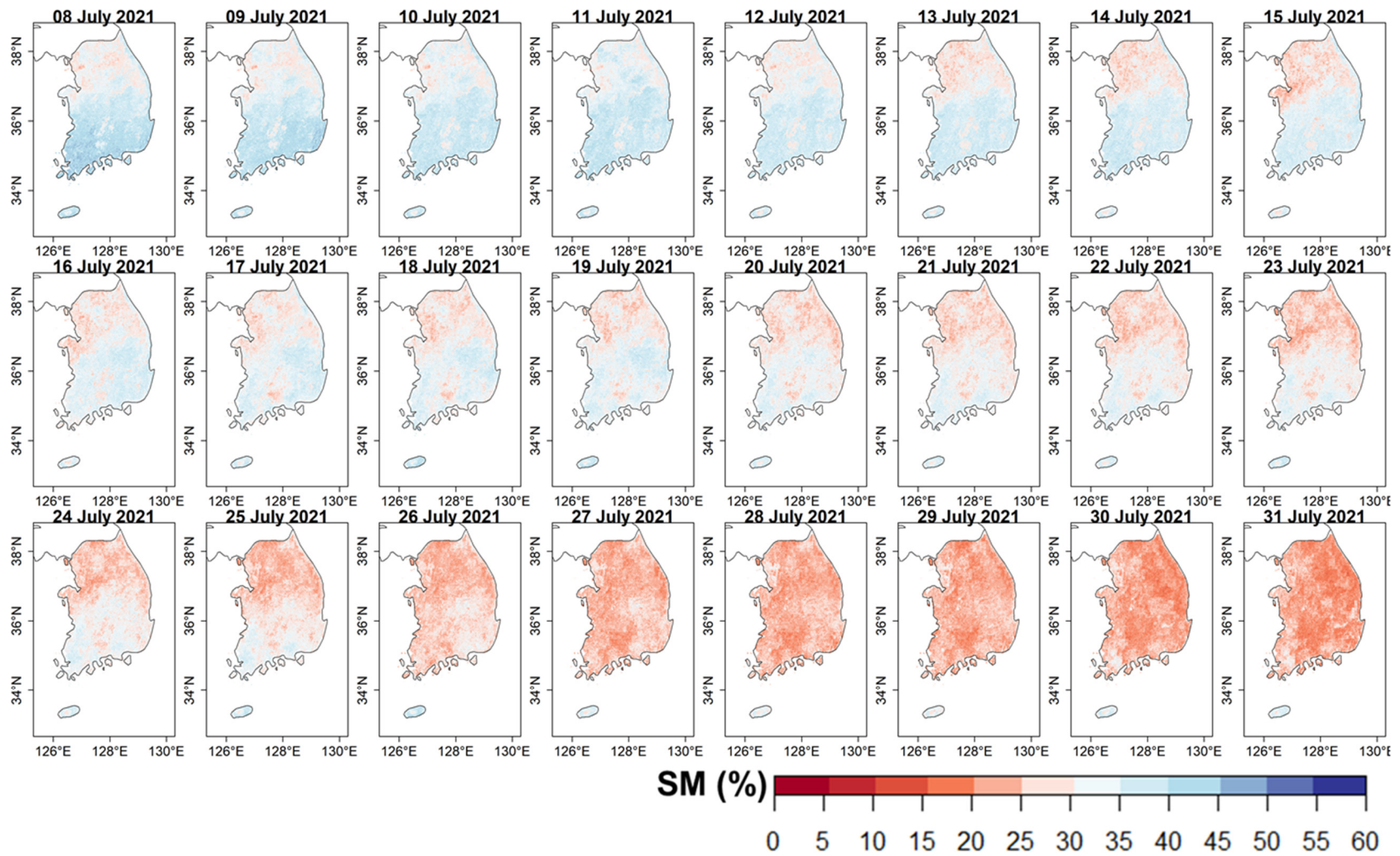


Figure 17. Soil moisture maps calculated using the AutoML-based soil moisture model from 8 July 2021 to 31 July 2021.

#### 4. Conclusions

RF- and AutoML-based SM models were constructed using various explanatory variables derived from satellite image data, NWP data, and topographical data provided within one day to estimate daily SM in real time. SM estimations obtained via both models had high accuracy rates, as demonstrated by comparisons with the in situ SM (MAE = 2.212–4.132%, MBE = −0.110–0.136%, RMSE = 3.186–5.384%, CC = 0.732–0.913), and a higher accuracy was obtained via the AutoML-based SM model, regardless of the data sampling method (random sampling and LOYO) used. As in previous studies, models using random sampling had a higher accuracy than those using LOYO, but the latter approach had a stable accuracy (MAE = 4.097–4.132%, MBE = −0.110–0.136%, RMSE = 5.326–5.384%, CC = 0.732–0.733). These results indicate that the input variables and model structures were of low temporal dependence and a high accuracy can be obtained when predicting SM at any point in time. Furthermore, SM data obtained via the proposed models had a higher accuracy than GLDAS and ERA5 SM data and, accordingly, enabled much more accurate monitoring of local-scale SM in Korea. In examinations of spatiotemporal variations in SM, using daily SM maps to predict extreme weather events in South Korea from 2020 to 2021 (including the heatwave in June 2020, the long rainy season and temperature inversion in July 2020, the heavy rain in November 2020, and the low precipitation accompanied by a heatwave in July 2021), it was found that the model effectively captured the dry/wet changes in the land surface related to the changes in precipitation and temperature.

In this study, we confirmed the possibility of predicting SM in real time with high accuracy, regardless of the sampling method, using ML models and various variable data related to SM supplied within one day, as well as the possibility of producing SM data that could monitor the change in moisture in the local areas considered. Therefore, it is expected to be useful in terms of analyzing short-term extreme weather events, such as flash droughts or floods, and it can assist in swift decision-making for disaster/water management and crop cultivation.

**Author Contributions:** Conceptualization, S.-J.L. and Y.L.; methodology, S.-J.L. and Y.L.; formal analysis, S.-J.L.; data curation, S.-J.L.; writing—original draft preparation, S.-J.L. and Y.L.; writing—review and editing, S.-J.L., E.S., M.K., K.-H.P., K.P. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R111A1A01073185). This research was supported by the R&D Program of Korea Meteorological Administration (KMA2020-00120). This work was carried out with the support of the “Cooperative Research Program for Agriculture Science & Technology Development (PJ014787042023)”, Rural Development Administration, the Republic of Korea.

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

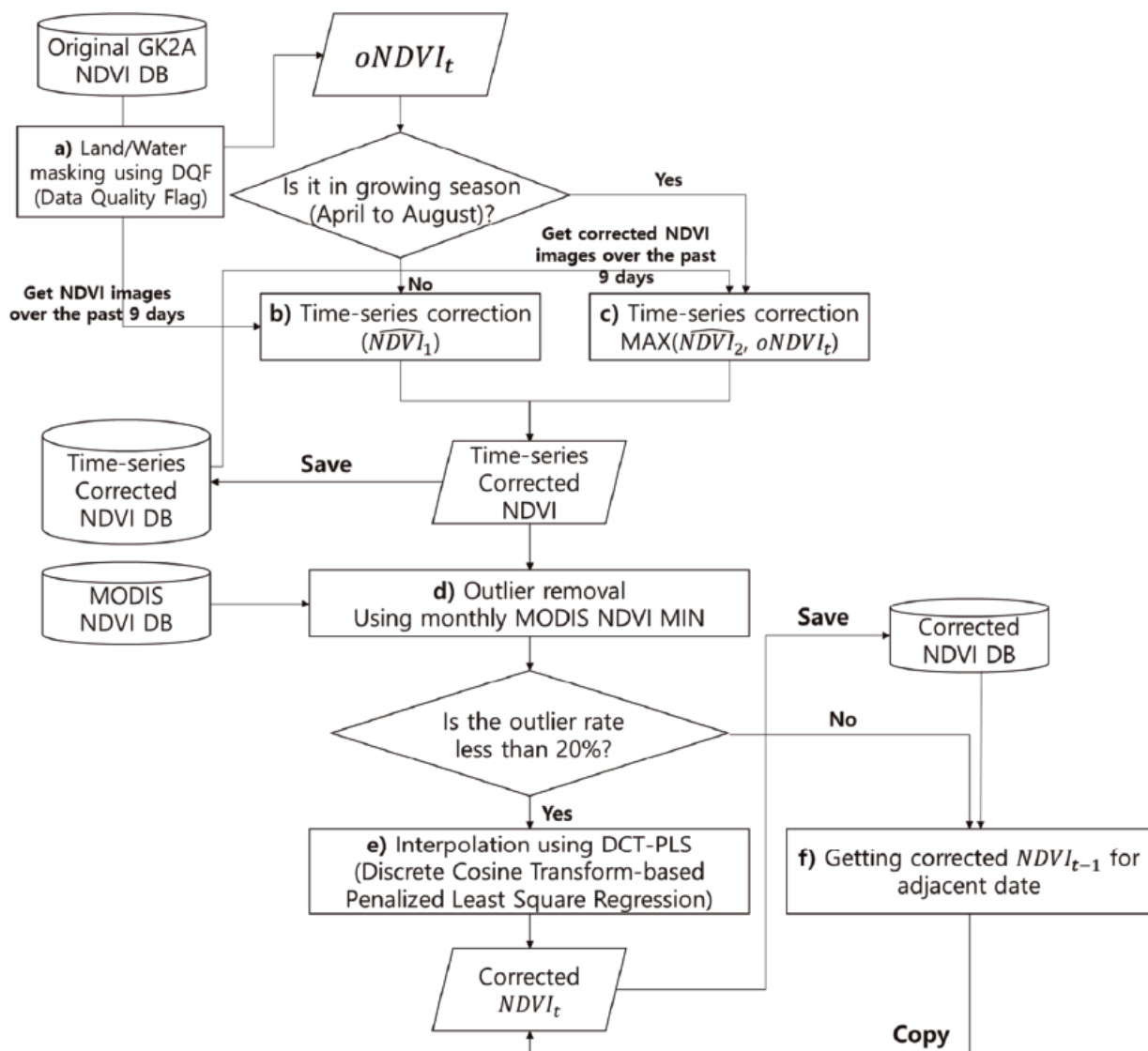
#### Abbreviations

AMI	Advanced Meteorological Imager
AutoML	Automated machine learning
CC	Correlation coefficient
DCT-PLS	Penalized least square regression based on discrete cosine transform
DEM	Digital Elevation Model
DNN	Deep neural networks
DRF	Distributed random forest
DSR	Downward shortwave radiation
DSSF	Total downward surface shortwave flux
ECDF	Empirical cumulative density function
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF Reanalysis 5
GBM	Gradient boosting machine

---

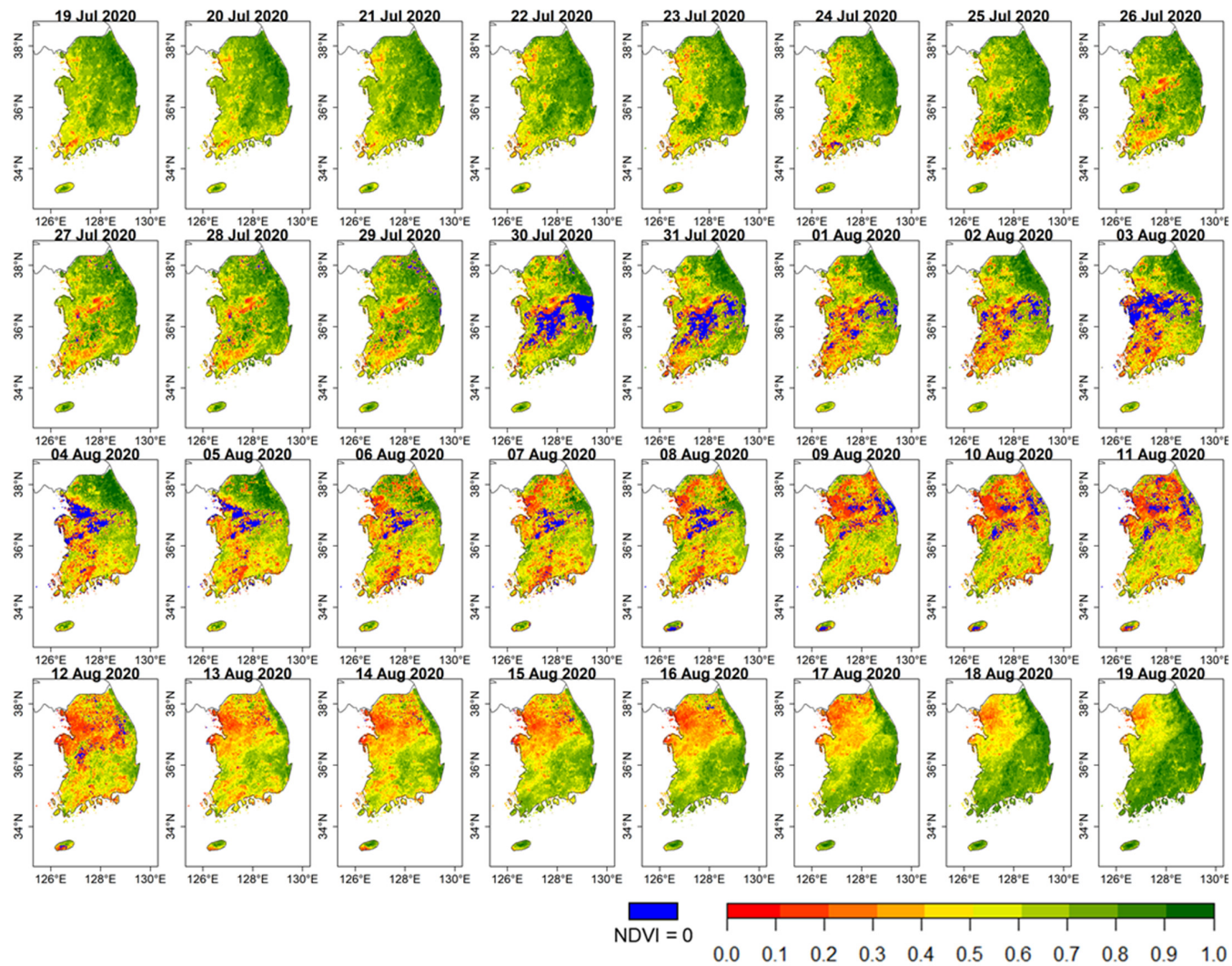
GK2A	GEO-KOMPSAT-2A
GLDAS	Global Land Data Assimilation System
GLM	Generalized linear model
GPM	Global Precipitation Measurement
GRNN	Generalized regression neural network
IMERG	The Integrated Multi-Satellite Retrievals for GPM
JAXA	Japan Aerospace Exploration Agency
KMA	Korea Meteorological Administration
LDAPS	Local Data Assimilation and Prediction System
LE	Latent heat flux
LST	Land surface temperature
LOYO	Leave-one-year-out cross-validation
MAE	Mean absolute error
MBE	Mean Bias Error
ML	Machine learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NGA	U.S. National Geospatial-Intelligence Agency
NTR	Near-infrared transformed reflectance
NWP	Numerical weather prediction
QC	Quality control
RMSE	Root mean square error
SMAP	Soil Moisture Active Passive

## Appendix A



- $oNDVI_t$ : Original GK2A NDVI image of the date you want to process
- $\overline{NDVI}_1$ : Average of original GK2A NDVI images from  $t - 9$  to  $t$
- $\overline{NDVI}_2$ : Average of  $oNDVI_t$  and time-series corrected NDVI images from  $t - 9$  to  $t - 1$

**Figure A1.** Real-time process used to correct the underestimation noise in the GK2A daily NDVI [43]. The six major steps in this process are as follows: (a) Land/water masking of GK2A NDVI is performed using the data quality flag (DQF) information provided in the GK2A NDVI dataset. (b) Based on the growth cycle of the vegetation, a moving average-based time series correction is applied to the declining phase of vegetation growth, which, in this case, extends from September to March. (c) During the vegetation growth phase (April to August), a time series correction that combines moving averages and maximum value composite (MVC) is applied. (d) Based on the monthly minimum value extracted from the long-term MODIS NDVI data between 2012 and 2021, if the (MODIS NDVI minimum value-corrected GK2A NDVI) value for the same month is  $\geq 0.1$ , it is considered to be an outlier and removed. (e) The ratio of outlier pixels to land pixels is calculated; for an image with  $< 20\%$  outliers, the missing value is restored using the discrete cosine transform-based penalized least square regression (DCT-PLS) method. (f) For an image containing  $\geq 20\%$  outliers, applying DCT-PLS is difficult, and the value is instead replaced using a corrected image derived from the previous data.



**Figure A2.** Original GK2A NDVI maps created from 19 July to 19 August 2020, which contain noise, in which the NDVI values drop below  $\sim 0.3$  due to meteorological factors, such as clouds and heavy rain [43].



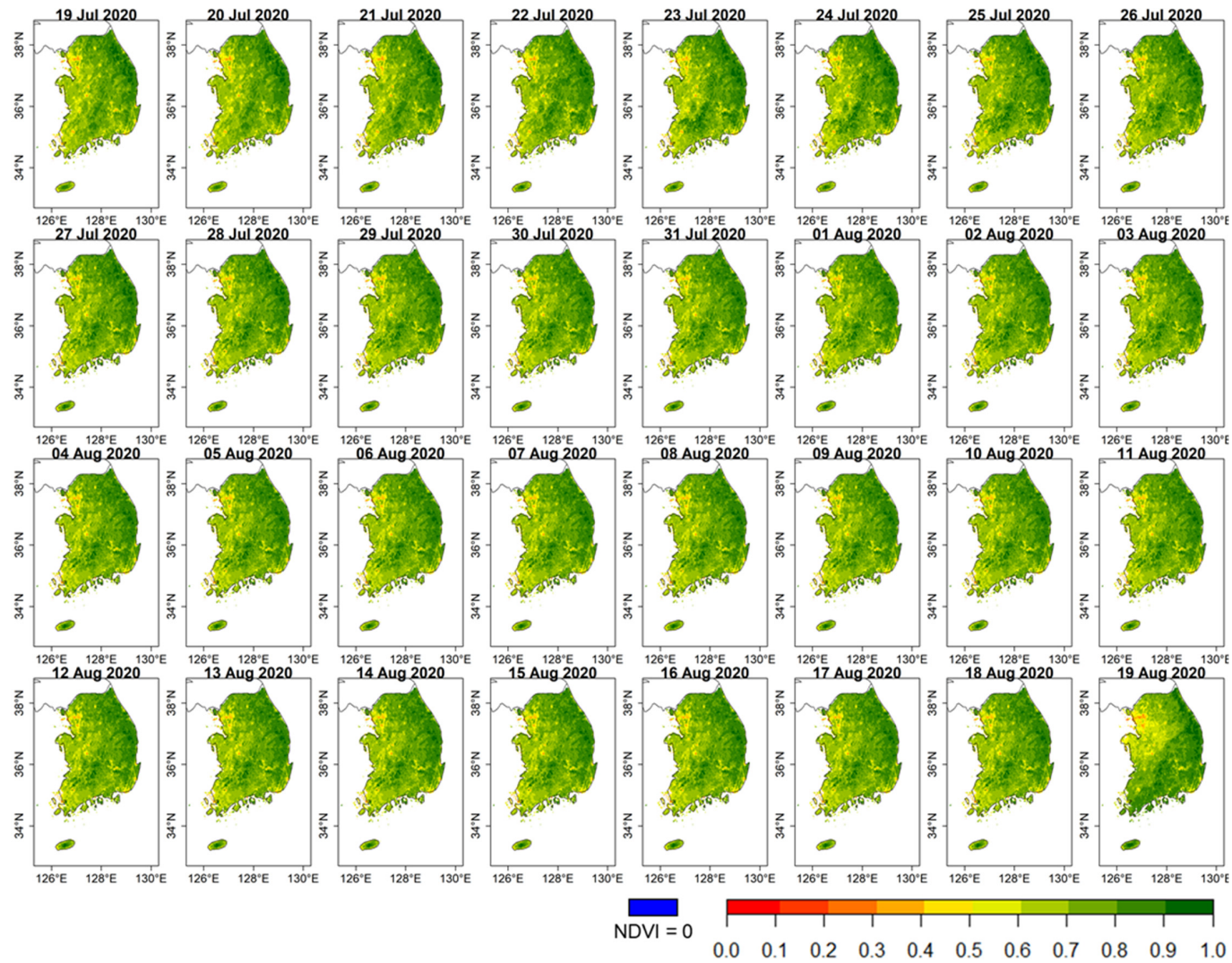


Figure A3. Corrected GK2A NDVI maps created from 19 July to 19 August 2020 [43].

## References

1. Cook, B.I.; Mankin, J.S.; Anchukaitis, K.J. Climate Change and Drought: From Past to Future. *Curr. Clim. Chang. Rep.* **2018**, *4*, 164–179. [CrossRef]
2. Iyakaremye, V.; Zeng, G.; Yang, X.; Zhang, G.; Ullah, I.; Gahigi, A.; Vuguziga, F.; Asfaw, T.G.; Ayugi, B. Increased high-temperature extremes and associated population exposure in Africa by the mid-21st century. *Sci. Total Environ.* **2021**, *790*, 148162. [CrossRef] [PubMed]
3. Ullah, I.; Saleem, F.; Iyakaremye, V.; Yin, J.; Ma, X.; Syed, S.; Hina, S.; Asfaw, T.G.; Omer, A. Projected Changes in Socioeconomic Exposure to Heatwaves in South Asia Under Changing Climate. *Earth's Future* **2022**, *10*, e2021EF002240. [CrossRef]
4. Paik, S.; Min, S.K.; Zhang, X.; Donat, M.G.; King, A.D.; Sun, Q. Determining the Anthropogenic Greenhouse Gas Contribution to the Observed Intensification of Extreme Precipitation. *Geophys. Res. Lett.* **2020**, *47*, e2019GL086875. [CrossRef]
5. Eccles, R.; Zhang, H.; Hamilton, D. A review of the effects of climate change on riverine flooding in subtropical and tropical regions. *J. Water Clim. Chang.* **2019**, *10*, 687–707. [CrossRef]
6. Littell, J.S.; Peterson, D.L.; Riley, K.L.; Liu, Y.; Luce, C.H. A Review of the Relationships between Drought and Forest Fire in the United States. *Glob. Chang. Biol.* **2016**, *22*, 2353–2369. [CrossRef] [PubMed]
7. Lee, S.-J.; Kim, N.; Lee, Y. Development of Integrated Crop Drought Index by Combining Rainfall, Land Surface Temperature, Evapotranspiration, Soil Moisture, and Vegetation Index for Agricultural Drought Monitoring. *Remote Sens.* **2021**, *13*, 1778. [CrossRef]
8. da Silva, R.C.; de Marchi Neto, I.; Seifert, S.S. Electricity Supply Security and the Future Role of Renewable Energy Sources in Brazil. *Renew. Sustain. Energy Rev.* **2016**, *59*, 328–341. [CrossRef]
9. Petropoulos, G.P.; Griffiths, H.; Dorigo, W.; Xaver, A.; Gruber, A. Surface soil moisture estimation: Significance, controls and conventional measurement techniques. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*; Petropoulos, G.P., Ed.; Taylor and Francis: Oxford, UK, 2013; Chapter 2; pp. 29–48.
10. Li, X.; Liu, L.; Duan, Z.; Wang, N. Spatio-Temporal Variability in Remotely Sensed Surface Soil Moisture and Its Relationship with Precipitation and Evapotranspiration during the Growing Season in the Loess Plateau, China. *Environ. Earth Sci.* **2014**, *71*, 1809–1820. [CrossRef]
11. McColl, K.A.; Alemohammad, S.H.; Akbar, R.; Konings, A.G.; Yueh, S.; Entekhabi, D. The Global Distribution and Dynamics of Surface Soil Moisture. *Nat. Geosci.* **2017**, *10*, 100–104. [CrossRef]
12. Jim, B. *Soil Water Status: Content and Potential*; App. Note: 2S-I; Campbell Scientific, Inc.: Logan, UT, USA, 2001.
13. Jaria, F. Soil Moisture Measurement. Available online: [https://www.mcgill.ca/globalfoodsecurity/files/globalfoodsecurity/2012\\_soilmoisture.pdf](https://www.mcgill.ca/globalfoodsecurity/files/globalfoodsecurity/2012_soilmoisture.pdf) (accessed on 12 August 2023).
14. Noborio, K. Measurement of Soil Water Content and Electrical Conductivity by Time Domain Reflectometry: A Review. *Comput. Electron. Agric.* **2001**, *31*, 213–237. [CrossRef]
15. Topp, G.C.; Davis, J.L.; Annan, A.P. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.* **1980**, *16*, 574–582. [CrossRef]
16. Rasheed, M.W.; Tang, J.; Sarwar, A.; Shah, S.; Saddique, N.; Khan, M.U.; Imran Khan, M.; Nawaz, S.; Shamshiri, R.R.; Aziz, M.; et al. Soil Moisture Measuring Techniques and Factors Affecting the Moisture Dynamics: A Comprehensive Review. *Sustainability* **2022**, *14*, 11538. [CrossRef]
17. Stevens Water Monitoring Systems Inc. *HydraProbe Soil Sensor, User's Manual*; Stevens Water Monitoring Systems Inc.: Portland, OR, USA, 2018. Available online: [https://www.stevenswater.com/resources/documentation/hydraprobe/HydraProbe\\_Manual\\_Jan\\_2018.pdf](https://www.stevenswater.com/resources/documentation/hydraprobe/HydraProbe_Manual_Jan_2018.pdf) (accessed on 12 August 2023).
18. Rowlandson, T.L.; Berg, A.A.; Bullock, P.R.; Hanis-Gervais, K.; Ojo, E.R.; Cosh, M.H.; Powers, J.; McNairn, H. Temporal Transferability of Soil Moisture Calibration Equations. *J. Hydrol.* **2018**, *556*, 349–358. [CrossRef]
19. Entekhabi, D.; Yueh, S.; O'Neill, P.E.; Kellogg, K.H.; Allen, A.; Bindlish, R.; Brown, M.; Chan, S.; Colliander, A.; Crow, W.T.; et al. *SMAP Handbook Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space*; JPL Publication: Pasadena, CA, USA, 2014.
20. Neelam, M.; Mohanty, B.P. Global Sensitivity Analysis of the Radiative Transfer Model. *Water Resour. Res.* **2015**, *51*, 2428–2443. [CrossRef]
21. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [CrossRef]
22. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; et al. ERA5 hourly data on single levels from 1940 to present. In *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*; ECMWF Reading: London, UK, 2018. Available online: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview> (accessed on 1 May 2022).
23. Lee, C.S.; Sohn, E.; Park, J.D.; Jang, J.D. Estimation of Soil Moisture Using Deep Learning Based on Satellite Data: A Case Study of South Korea. *GISci. Remote Sens.* **2019**, *56*, 43–67. [CrossRef]
24. Deng, Y.; Wang, S.; Bai, X.; Wu, L.; Cao, Y.; Li, H.; Wang, M.; Li, C.; Yang, Y.; Hu, Z.; et al. Comparison of soil moisture products from microwave remote sensing, land model, and reanalysis using global ground observations. *Hydrol. Process.* **2020**, *34*, 836–851. [CrossRef]

25. Kim, Y.; Lee, S.-J.; Kim, J.; Lee, Y. Deep Learning-based Retrieval of Daily 500-m Soil Moisture for South Korea. *J. Korean Cartogr. Assoc.* **2017**, *17*, 109–121. [[CrossRef](#)]
26. Cheng, M.; Zhong, L.; Ma, Y.; Zou, M.; Ge, N.; Wang, X.; Hu, Y. A Study on the Assessment of Multi-Source Satellite Soil Moisture Products and Reanalysis Data for the Tibetan Plateau. *Remote Sens.* **2019**, *11*, 1196. [[CrossRef](#)]
27. Fan, L.; Xing, Z.; De Lannoy, G.; Frappart, F.; Peng, J.; Zeng, J.; Li, X.; Yang, K.; Zhao, T.; Shi, J.; et al. Evaluation of Satellite and Reanalysis Estimates of Surface and Root-Zone Soil Moisture in Croplands of Jiangsu Province, China. *Remote Sens. Environ.* **2022**, *282*, 113283. [[CrossRef](#)]
28. Kim, N.; Kim, K.; Lee, S.; Cho, J.; Lee, Y. Retrieval of Daily Reference Evapotranspiration for Croplands in South Korea Using Machine Learning with Satellite Images and Numerical Weather Prediction Data. *Remote Sens.* **2020**, *12*, 3642. [[CrossRef](#)]
29. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine Learning on Big Data: Opportunities and Challenges. *Neurocomputing* **2017**, *237*, 350–361. [[CrossRef](#)]
30. Liu, Y.; Qian, J.; Yue, H. Combined Sentinel-1A with Sentinel-2A to Estimate Soil Moisture in Farmland. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1292–1310. [[CrossRef](#)]
31. LeDell, E.; Poirier, S. H<sub>2</sub>O automl: Scalable automatic machine learning. In Proceedings of the AutoML Workshop at ICML, Online, 17–18 July 2020; Volume 2020.
32. Babaeian, E.; Paheding, S.; Siddique, N.; Devabhaktuni, V.K.; Tuller, M. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Remote Sens. Environ.* **2021**, *260*, 112434. [[CrossRef](#)]
33. Korea Meteorological Administration (KMA) Climate Characteristics of Korea. Available online: <https://www.weather.go.kr/w/obs-climate/climate/statistics/korea-char.do> (accessed on 10 August 2023).
34. National Institute of Meteorological Sciences Climate Change Report of the Korean Peninsula over the Last 100 Years (11-1360620-000132-01). Available online: [http://www.nims.go.kr/flexer/view.jsp?FileDir=/DF978&SystemFileName=20180816153001\\_0.pdf&ftype=pdf&FileName=%ED%95%9C%EB%B0%98%EB%8F%84%20100%EB%85%84%EC%9D%98%20%EA%B8%B0%ED%9B%84%EB%B3%80%ED%99%94\\_%EB%B3%B4%EA%B3%A0%EC%84%9C.pdf&org=KOR\\_OP\\_DF\\_MV\\_2&idx=4168&c\\_idx=755&seq=0](http://www.nims.go.kr/flexer/view.jsp?FileDir=/DF978&SystemFileName=20180816153001_0.pdf&ftype=pdf&FileName=%ED%95%9C%EB%B0%98%EB%8F%84%20100%EB%85%84%EC%9D%98%20%EA%B8%B0%ED%9B%84%EB%B3%80%ED%99%94_%EB%B3%B4%EA%B3%A0%EC%84%9C.pdf&org=KOR_OP_DF_MV_2&idx=4168&c_idx=755&seq=0) (accessed on 10 August 2023).
35. Jung, Y.Y.; Shin, W.J.; Seo, K.H.; Koh, D.C.; Ko, K.S.; Lee, K.S. Spatial Distributions of Oxygen and Hydrogen Isotopes in Multi-level Groundwater across South Korea: A Case Study of Mountainous Regions. *Sci. Total Environ.* **2022**, *812*, 151428. [[CrossRef](#)] [[PubMed](#)]
36. National Institute of Agricultural Sciences of Korean Rural Development Administration Korean Soil Information System—Topsoil Texture. Available online: <http://soil.rda.go.kr/geoweb/soilmain.do#> (accessed on 10 August 2023).
37. Zhao, L.; Yang, K.; Qin, J.; Chen, Y.; Tang, W.; Lu, H.; Yang, Z.-L. The scale-dependence of SMOS soil moisture accuracy and its improvement through land data assimilation in the central Tibetan Plateau. *Remote Sens. Environ.* **2014**, *152*, 345–355. [[CrossRef](#)]
38. National Institute of Agricultural Sciences of Korean Rural Development Administration Agricultural Weather Information Service. Available online: <http://weather.rda.go.kr/w/analysis/inquiry.do> (accessed on 2 March 2023).
39. NASA GLOBAL PRECIPITATION MEASUREMENT—IMERG: Integrated Multi-satellite Retrievals for GPM. Available online: <https://gpm.nasa.gov/data/imerg> (accessed on 12 April 2023).
40. National Meteorological Satellite Center of Korea Meteorological Administration (NMSC of KMA) Geo-KOMPSAT-2A Overview. Available online: <https://nmsc.kma.go.kr/enhome/html/base/cmm/selectPage.do?page=satellite.gk2a.fact> (accessed on 1 January 2023).
41. Han, K.-S.; Seong, N.-H. *GK-2A AMI Algorithm Theoretical Basis Document (Vegetation Index/Fractiona Vegetation Cover) (Version 1.0)*; National Meteorological Satellite Center: Jincheon-gun, Republic of Korea, 2019; pp. 1–52. Available online: <https://nmsc.kma.go.kr/homepage/html/base/cmm/selectPage.do?page=static.edu.atbdGk2a> (accessed on 13 August 2023).
42. Jang, J.; Lee, K.-T. *GK-2A AMI Algorithm Theoretical Basis Document: RSR, DSR, and ASR (Version 1.1)*; National Meteorological Satellite Center: Jincheon-gun, Republic of Korea, 2019; pp. 1–42. Available online: <https://nmsc.kma.go.kr/homepage/html/base/cmm/selectPage.do?page=static.edu.atbdGk2a> (accessed on 12 August 2023).
43. Lee, S.-J.; Youn, Y.; Sohn, E.; Kim, M.; Lee, Y. A Real-time Correction of the Underestimation Noise for GK2A Daily NDVI. *Korean J. Remote Sens.* **2022**, *38*, 1301–1314.
44. NASA’s The Land Processes Distributed Active Archive Center (LP DAAC) VNP13A2 v001. Available online: <https://lpdaac.usgs.gov/products/vnp13a2v001/> (accessed on 12 August 2023).
45. Numerical Modeling Center of the Korea Meteorological Administration. *Numerical Forecasting Takes Responsibility for the Weather and Climate Industries!—Utilization Guide of Numerical Weather Prediction Model Data for Activation of the Weather Industry*; Publication Report Number: 11–1360395-000252-01; Numerical Modeling Center of the Korea Meteorological Administration: Seoul, Republic of Korea, 2013.
46. Cho, D.; Yoo, C.; Im, J.; Cha, D.-H. Comparative Assessment of Various Machine Learning-Based Bias Correction Methods for Numerical Weather Prediction Model Forecasts of Extreme Air Temperatures in Urban Areas. *Earth Space Sci.* **2020**, *7*, e2019EA000740. [[CrossRef](#)]
47. Kim, D.-J.; Kang, G.; Kim, D.-Y.; Kim, J.-J. Characteristics of LDAPS-Predicted Surface Wind Speed and Temperature at Automated Weather Stations with Different Surrounding Land Cover and Topography in Korea. *Atmosphere* **2020**, *11*, 1224. [[CrossRef](#)]

48. USGS EROS Archive—Digital Elevation—Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. Available online: [https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science_center_objects=0#qt-science_center_objects) (accessed on 12 August 2023).
49. Svoboda, M.; Hayes, M.; Wood, D. *Standardized Precipitation Index User Guide*; World Meteorological Organization: Geneva, Switzerland, 2012.
50. Gwak, Y.S.; Kim, Y.T.; Won, C.H.; Kim, S.H. The Relationships between Drought Indices (SPI, API) and In-situ Soil Moisture in Forested Hillslopes. *WIT Trans. Ecol. Environ.* **2017**, *220*, 217–224.
51. Chen, T.; de Jeu, R.A.M.; Liu, Y.Y.; van der Werf, G.R.; Dolman, A.J. Using Satellite Based Soil Moisture to Quantify the Water Driven Variability in NDVI: A Case Study over Mainland Australia. *Remote Sens. Environ.* **2014**, *140*, 330–338. [[CrossRef](#)]
52. Gu, Y.; Hunt, E.; Wardlow, B.; Basara, J.B.; Brown, J.F.; Verdin, J.P. Evaluation of MODIS NDVI and NDWI for Vegetation Drought Monitoring Using Oklahoma Mesonet Soil Moisture Data. *Geophys. Res. Lett.* **2008**, *35*, L22401. [[CrossRef](#)]
53. Zuluaga, C.F.; Avila-Diaz, A.; Justino, F.B.; Wilson, A.B. Climatology and Trends of Downward Shortwave Radiation over Brazil. *Atmos. Res.* **2021**, *250*, 105347. [[CrossRef](#)]
54. Jiang, Y.; Weng, Q. Estimation of Hourly and Daily Evapotranspiration and Soil Moisture Using Downscaled LST over Various Urban Surfaces. *GISci. Remote Sens.* **2017**, *54*, 95–117. [[CrossRef](#)]
55. Rahimikhoob, A. Estimation of Evapotranspiration Based on Only Air Temperature Data Using Artificial Neural Networks for a Subtropical Climate in Iran. *Theor. Appl. Climatol.* **2009**, *101*, 83–91. [[CrossRef](#)]
56. Feldhake, C.M.; Boyer, D.G. Effect of Soil Temperature on Evapotranspiration by C3 and C4 Grasses. *Agric. For. Meteorol.* **1986**, *37*, 309–318. [[CrossRef](#)]
57. Eagleman, J.R. Pan Evaporation, Potential and Actual Evapotranspiration. *J. Appl. Meteorol. Climatol.* **1967**, *6*, 482–488. [[CrossRef](#)]
58. Cabeza, L.F.; Roca, J.; Noguès, M.; Mehling, H.; Hiebler, S. Immersion Corrosion Tests on Metal-Salt Hydrate Pairs Used for Latent Heat Storage in the 48 to 58 °C Temperature Range. *Mater. Corros.* **2002**, *53*, 902–907. [[CrossRef](#)]
59. Schneider, T.; O’Gorman, P.A.; Levine, X.J. Water Vapor and the Dynamics of Climate Changes. *Rev. Geophys.* **2010**, *48*, RG3001. [[CrossRef](#)]
60. Shuttleworth, W.J. Evapotranspiration Measurement Methods. *Southwest Hydrol.* **2008**, *7*, 22–23.
61. Western, A.W.; Grayson, R.B.; Blöschl, G.; Willgoose, G.R.; McMahon, T.A. Observed Spatial Organization of Soil Moisture and Its Relation to Terrain Indices. *Water Resour. Res.* **1999**, *35*, 797–810. [[CrossRef](#)]
62. Qiu, Y.; Fu, B.J.; Wang, J.; Chen, L.D. Soil Moisture Variation in Relation to Topography and Land Use in a Hillslope Catchment of the Loess Plateau, China. *J. Hydrol.* **2001**, *240*, 243–263. [[CrossRef](#)]
63. Riley, S.J.; DeGloria, S.D.; Elliot, R. Index That Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
64. Chen, Y.C.; Chang, K.T.; Wang, S.F.; Huang, J.C.; Yu, C.K.; Tu, J.Y.; Wang, C.W.; Liu, C.C. Controls of Preferential Orientation of Earthquake-and Rainfall-Triggered Landslides in Taiwan’s Orogenic Mountain Belt. *Earth Surf. Process. Landf.* **2019**, *44*, 1661–1674. [[CrossRef](#)]
65. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
66. Prasad, A.; Iverson, L.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
67. Xue, L.; Liu, Y.; Xiong, Y.; Liu, Y.; Cui, X.; Lei, G. A Data-Driven Shale Gas Production Forecasting Method Based on the Multi-Objective Random Forest Regression. *J. Pet. Sci. Eng.* **2021**, *196*, 107801. [[CrossRef](#)]
68. Sharma, G. Pros and cons of different sampling techniques. *Int. J. Appl. Res.* **2017**, *3*, 749–752.
69. Korean Meteorological Administration (KMA) 2020 Extreme Climate Report. Available online: [http://www.climate.go.kr/home/n\\_search/search\\_view.php?dist=&bbs\\_name=&no\\_bbs=6494&act=&go=&skind=all&sword=&bname=&s\\_keyword=%EC%9D%B4%EC%83%81%EA%B8%B0%ED%9B%84%EB%B3%B4%EA%B3%A0%EC%84%9C](http://www.climate.go.kr/home/n_search/search_view.php?dist=&bbs_name=&no_bbs=6494&act=&go=&skind=all&sword=&bname=&s_keyword=%EC%9D%B4%EC%83%81%EA%B8%B0%ED%9B%84%EB%B3%B4%EA%B3%A0%EC%84%9C) (accessed on 2 May 2022). (In Korean).
70. Min, S.K.; Jo, S.Y.; Seong, M.G.; Kim, Y.H.; Son, S.W.; Byun, Y.H.; Lott, F.C.; Stott, P.A. Human Contribution to the 2020 Summer Successive Hot-Wet Extremes in South Korea. *Bull. Am. Meteorol. Soc.* **2022**, *103*, S90–S97. [[CrossRef](#)]
71. Korean Meteorological Administration (KMA) 2021 Extreme Climate Report. Available online: [http://www.climate.go.kr/home/n\\_search/search\\_view.php?dist=&bbs\\_name=&no\\_bbs=6609&act=&go=&skind=all&sword=&bname=&s\\_keyword=%EC%9D%B4%EC%83%81%EA%B8%B0%ED%9B%84%EB%B3%B4%EA%B3%A0%EC%84%9C](http://www.climate.go.kr/home/n_search/search_view.php?dist=&bbs_name=&no_bbs=6609&act=&go=&skind=all&sword=&bname=&s_keyword=%EC%9D%B4%EC%83%81%EA%B8%B0%ED%9B%84%EB%B3%B4%EA%B3%A0%EC%84%9C) (accessed on 2 May 2022). (In Korean).
72. Korean Meteorological Administration (KMA) Distribution of Precipitation Deviation in July 2020. Available online: <https://data.kma.go.kr/stcs/grnd/grndRnDmap.do?pgmNo=207> (accessed on 10 May 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.