Delft University of Technology

Enhancing Rainfall Estimates in East Africa by Integrating TAHMO Ground Data and Satellite Rainfall Products

Hoogelander, Vincent; Hut, Rolf; Le Coz, Camille; Dong, Jianzhi; VAN DE GIESEN, Nick

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Enhancing Rainfall Estimates in East Africa by Integrating TAHMO Ground Data and Satellite Rainfall Products

Vincent Hoogelander,[a] Rolf Hut,[a] Camille Le Coz,[b] Jianzhi Dong,[c] and Nick van de Giesen[a]

[a] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands
[b] Laboratoire de Météorologie Dynamique, Ecole Polytechnique, CNRS, Palaiseau, France
[c] Institute of Surface-Earth System Science, Tianjin University, Tianjin, China

ABSTRACT: East Africa relies heavily on satellite-based rainfall estimates due to the lack of in situ data. However, satellite rainfall products often perform poorly in this region. In this study, data from the Trans-African Hydrometeorological Observatory (TAHMO) were used to build a regional rainfall product in East Africa based on the Soil Moisture to Rain (SM2Rain) algorithm. Subsequently, this regional product was merged with a reanalysis product (ERA5) and two microwave (MW)/infrared (IR)-based rainfall products (IMERG and CHIRPS) based on the Statistical Uncertainty Analysis-Based Precipitation Merging (SUPER) framework. Within this framework, merging weights are derived from the error variances of the rainfall products determined from quadruple collocation on a pixel-to-pixel basis. The merged and individual products are evaluated using data from individual TAHMO stations. We tested SUPER with various interproduct dependency assumptions and found that, in the best-performing configuration, IMERG contributed the most to the merged product, followed by CHIRPS, ERA5, and SM2Rain. SM2Rain showed performance comparable to other rainfall products but is more useful for detecting the offset of the rainy season in drier climates and less reliable under wet conditions. The findings indicated that the merged product outperforms the individual products in most performance metrics. Additionally, we demonstrated the importance of comparing satellite and ground-measured precipitation time series, alongside evaluating performance metrics. The ultimate goal of this study is to develop a workflow to enhance the accuracy of rainfall measurements in East Africa by leveraging information from TAHMO data and different existing products, contributing to the improvement of satellite-based rainfall estimates in East Africa.

KEYWORDS: Africa; Rainfall; Gauges; Remote sensing; Satellite observations; Statistical techniques

1. Introduction

Rainfall plays a crucial role in East Africa, particularly as a determinant of both flood early warning systems and agricultural productivity (Kilavi et al. 2018; MacLeod et al. 2021; Palmer et al. 2023). East Africa's predominantly rain-fed agriculture relies heavily on timely rainfall for crop development. Additionally, due to the region's vulnerability to flash floods, there is a need for robust early warning systems to mitigate risks to human lives and livelihoods. Therefore, there is a critical need for timely and accurate rainfall data in this region. However, East Africa currently lacks a dense network of rainfall radars for making spatial and temporal high-resolution rainfall maps. Consequently, rain gauges remain the most reliable method for measuring rainfall. The limitation of using solely rain gauges for rainfall mapping is that in most regions, the number of such rain gauges is small and unevenly distributed, a situation particularly prevalent in East Africa (Kimani et al. 2017). This makes it difficult to create accurate spatial maps over large areas. Besides, the sparse distribution of gauges negatively affects their reliability of identifying localized rainfall events, such as convective rains, which are common in East Africa.

The lack of precipitation data in these data-scarce regions can be overcome by using satellite and reanalysis products for precipitation estimation. The advantage of using these products is that they provide a continuous and spatially extensive coverage of rainfall data. However, present satellite-derived and reanalysis rainfall products perform poorly over Africa. One possible reason is that these products are typically developed and tested in regions with a high density of ground data. Validation studies of commonly used rainfall products generally observe performance differences between areas with dense and sparse ground data networks (Liu et al. 2024; Beck et al. 2017; Ageet et al. 2022). Higher gauge density typically correlates with better satellite product performance. This difference may be caused by regional variations in the relationship between areal and point measurements of rainfall (Skaugen 1997). Current satellite-derived and reanalysis rainfall products primarily rely on data from regions with a higher density of weather stations, as they combine satellite and ground data to estimate rainfall. Some satellite-based rainfall products are specifically developed for data-scarce areas, such as CHIRPS and Tropical Applications of Meteorology Using Satellite (TAMSAT) (Funk et al. 2015b; Maidment et al. 2017). However, the number of rain gauges from Africa used in these products is limited and even declining (Dinku 2019), reducing their reliability.

The Trans-African Hydrometeorological Observatory (TAHMO) aims to tackle the ground data gap by installing and operating a dense network of weather stations in sub-Saharan Africa (SSA). Currently, TAHMO has installed over 600 weather stations in SSA, with major concentrations in Ghana and Kenya.

In this paper, data from TAHMO are used, together with publicly available satellite data, to develop a workflow to produce accurate rainfall estimates in East Africa. The workflow is mainly based on the Statistical Uncertainty Analysis-Based Precipitation Merging (SUPER) framework, a framework designed by Dong et al. (2022). Within this framework, four different satellite rainfall products are merged using error statistics for each individual pixel, considering cross-product dependencies. The reason why the SUPER framework is used is because it does not require gauge data, which is an advantage in data-sparse regions such as East Africa. The framework requires the availability of three independent products. To meet this requirement, a reanalysis product (ERA5), soil moisture–estimated product [Soil Moisture to Rain (SM2Rain)], and two satellite thermal infrared/passive microwave rainfall products (IMERG + CHIRPS) are used. However, the SM2Rain product is not operationally implemented in the EUMETSAT Operational Hydrology and Water Management Satellite Application Facility (H SAF) framework yet, and its data record is only available from 2007 until 2022. Therefore, it was also necessary to develop an operational version of SM2Rain for this region. This paper presents the methods and results related to the development of this regional version of SM2Rain and the merged rainfall product. This will be done by validating the merged and individual products against TAHMO station data. Additionally, the analysis will include an examination of the effect of the number of stations used in the SM2Rain development and the intermediate steps in the merging algorithm, as well as a comparison between ground and satellite data by analyzing individual pixel and point data. The developed workflow presented in this paper was refined through an iterative design process.

2. Methods and data

This section describes the study area, the available ground data, the methodologies used for the regional SM2Rain implementation, the merging algorithm, and the validation of the different satellite rainfall products.

a. TAHMO stations and study area

The primary study area in this paper covers Kenya, Uganda, and Rwanda. This region was selected due to the availability of a relatively dense network of TAHMO stations. Additionally, another reason for focusing on this region is that the study includes the results from a workshop held in Nairobi, Kenya, where scientists from eight countries developed methods for improving rainfall estimates in this region by merging satellite and ground data. The results of this workshop formed the starting point of the work presented in this paper. We use data from 2018 to 2022 due to broad data availability across stations in these years. To expand the coverage of stations, we also included a couple of stations located in Northern Tanzania. As of writing, TAHMO provides hydrometeorological ground data from a total of about 227 stations in these countries. Rainfall from TAHMO stations is measured using a drip counter (van de Giesen et al. 2014), and data are available at a 5-min resolution. TAHMO has a quality assessment (QA) and quality control (QCo) team that ensures data quality through cross calibration

with nearby stations and satellite data. Additionally, we filtered out certain stations and time periods with very dubious data that the QA/QCo team did not remove. All data have been aggregated to a daily resolution to match the temporal resolution of the Advanced Scatterometer (ASCAT) soil moisture data. We use TAHMO data in both the SM2Rain implementation and the validation of all satellite rainfall products. For the validation of SM2Rain, we separate the stations used for calibration and validation by only calibrating for 2 years, allowing us to validate using the remaining 3 years. Figure 1 highlights the study area and the network of TAHMO stations across the selected region. In addition to the uniqueness of the dataset, Schunke et al. (2021) emphasized the applicability of TAHMO observation data for hydrometeorological applications in the African context, finding good agreement between TAHMO data and synoptic stations in Burkina Faso. Similarly, Muita et al. (2021) found strong correlations between meteorological variables measured by TAHMO and a synoptic weather station in Nairobi.

b. Regional SM2Rain implementation

SM2Rain converts soil moisture data to produce quantitative rainfall estimations. In this paper, we will refer to the regional SM2Rain implementation as SM2Rain as well. It should be noted that while the algorithm remains the same, the data used in this version differ from those of the original product developed by Brocca et al. (2019). While the existing SM2Rain mainly relies on ERA5 data as reference products to optimize the essential parameters for the global data record, the version of this study uses ground measurements from TAHMO for the calibration process.

1) ASCAT DATASET

SM2Rain requires soil moisture data that are assimilated from satellite observations. In this study, we use the EUMETSAT H SAF Near-Real-Time Scatterometer Root Zone Soil Moisture Product (RZSM-ASCAT-NRT-10 km/H26), available at a daily resolution with a spatial scale of 10 km. This dataset consists of soil moisture data at four different soil layers between 0 and 289 cm. In this study, we only use data from the top soil layer (0–7 cm). The temporal extent of the H26 product spans from 8 January 2021 to the present. To include more years of data, the H141 and H142 products are also used. The H141 product is a reprocessed data record of the soil moisture in the rootzone from 1992 to 2018, and the H142 product is an extension of this product from 2019 to 2021. These datasets have the same characteristics as the H26 product. A more extensive description of the used datasets can be found in the product user manuals provided by EUMETSAT (EUMETSAT H SAF 2021b).

2) ALGORITHM

In this section, we provide a concise overview of the SM2Rain algorithm. SM2Rain converts changes in the soil moisture content to rainfall, using basic soil water balance equations:

$$nZ\frac{\Delta S(t)}{\Delta t} = P(t) - q(t) - R(t) - E(t), \qquad (1)$$

Study area and rainfall classification derived from TAHMO stations (2018-2022)
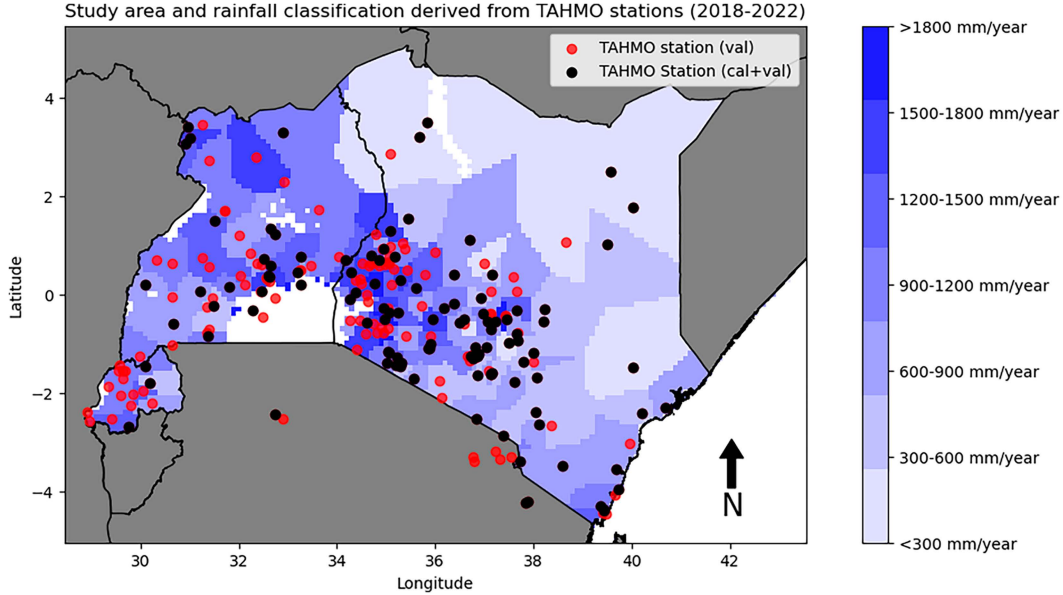

FIG. 1. Locations of selected TAHMO stations in this study, used for SM2Rain calibration and validation of the satellite rainfall products. The blue shading in the map represents the different regions derived from the rainfall classification. SM2Rain is calibrated with "cal + val" stations, and the "val" stations are used for the individual SM2Rain validation analysis. For the validation of SUPER, we used both cal + val and val stations.

where n is the effective porosity (—), Z is the soil layer depth (mm), S is the relative soil moisture content (—), t is the time (days), $P(t)$ is the rainfall rate (mm day^{-1}), $q(t)$ is the percolation and subsurface runoff (mm day^{-1}), $R(t)$ is the overland runoff rate (mm day^{-1}), and $E(t)$ is the actual evapotranspiration (mm day^{-1}). In the SM2Rain algorithm, the overland runoff is neglected (see Brocca et al. 2019), so that

$$P(t) \approx nZ\frac{\Delta S(t)}{\Delta t} + aS(t)^b. \tag{2}$$

Here, a (mm day^{-1}) and b (—) represent the parameters that describe the nonlinear relationship between the rate of drainage and evaporation combined and the level of soil saturation. Last, the parameters n and Z are expressed in a single parameter Z^*(mm), which represents the soil water capacity, so that

$$P(t) = Z^*\frac{\Delta S(t)}{\Delta t} + aS(t)^b. \tag{3}$$

Prior to generating the precipitation time series with Eq. (3), a modified exponential filter as proposed in Brocca et al. (2013) is applied to the soil moisture data to reduce soil moisture noise. This is done using the following equations:

$$G_i = \frac{G_{i-1}}{G_{i-1} + e^{-\Delta t/W_i}}, \tag{4}$$

where G_i is the gain at time step i (—), Δt is the time difference (days), and W_i is the time constant at day i (days). The filtered soil moisture data are then calculated with

$$S_{i,\text{filtered}} = S_{i-1} + G_i(S_i - S_{i-1}). \tag{5}$$

And the time constant is updated with

$$W_i = T \times S_i^{(-c)}, \tag{6}$$

with $G_0 = 1$. By implementing this filter, two additional parameters, T (days) and c (—), are required in the calibration process, next to the parameters a, b, and Z^*. Parameter T represents the minimum value of W_i associated with saturated soil conditions, whereas c is a parameter that causes W_i to increase as the soil moisture content decreases. In other words, higher values of T and c result in stronger filtering effects.

The five SM2Rain parameters are calibrated by minimizing the root-mean-square error (RMSE) between the SM2Rain estimation and the ground data from TAHMO. The parameters are optimized with a truncated Newton (TNC) algorithm (Dembo and Steihaug 1983). Rather than calibrating at each individual station, the parameters are calibrated within seven regions categorized by rainfall classification. These classifications are derived from TAHMO stations having more than 90% of data in a year, for which the classes are derived from the mean annual rainfall with TAHMO data from 2018 to 2022. The classes are extrapolated using a nearest-neighbor approach. The result of this classification can be found in Fig. 1. Note that this figure does not represent a long-term climatology of the region due to the limited time frame of 5 years and the variability in station availability across different years. As some stations may be used in 1 year, but not in another, this may actually lead to the misrepresentation of the true climatology.

For the calibration process, only stations having more than 90% of valid data during the calibration period, spanning from 2020 to 2021, are considered. These years were selected as these are relatively a dry and wet year, respectively. In

total, 110 stations are available for calibration, and their locations are indicated in Fig. 1. To analyze the effect of the number of stations used in SM2Rain calibration, an additional validation analysis is performed by calibrating SM2Rain with varying numbers of stations. Specifically, the calibrations are conducted using 10, 25, 50, and 75 stations. This analysis is performed to assess the extent to which using a larger number of stations improves the performance of SM2Rain and to determine whether it is beneficial to include as many stations as possible for optimal accuracy. Among these stations, we ensure that at least one station is located in each of the seven rainfall zones, as derived from the rainfall classification. These simulations are validated with a standard set of stations for 2018 (80 stations), 2019 (105 stations), and 2022 (126 stations), each with at least 80% data coverage per year. We emphasize that data from the selected stations in these years are not included in those used for calibration. To minimize the selection uncertainty, this process is repeated 10 times for each number of stations used, resulting in a total of 40 calibrations in total for this test. The validation metrics can be found in section 2d. After calibrating the parameters at the individual locations of the selected TAHMO stations, the parameters are assigned to a 0.1° grid that matches the corresponding rain classes. The rainfall data are smoothed using a Gaussian filter, using a smoothing parameter of $\sigma = 5$, to preserve spatial coherence and continuity in the rainfall maps. This parameter value is determined based on the visual inspection of the rainfall maps. After assigning all SM2Rain parameters to the grid, the rainfall time series are generated for each individual grid cell using Eq. (3).

c. Merging algorithm

Using the SUPER framework algorithm, the regional SM2Rain product is merged with a reanalysis product (ERA5) and two satellite thermal infrared (IR)/microwave (MW) rainfall products [IMERG (IR/MW) and CHIRPS (IR)]. This section gives a brief overview of this algorithm. A more detailed description of SUPER can be found in the paper by Dong et al. (2022).

1) SUPER

The SUPER framework is a merging tool to enhance rainfall estimations in regions with limited ground data (Dong et al. 2022). Within the algorithm, rainfall estimates are made using the weighted averaging of individual rainfall products. The weighting factors are based on the error variances of the products that are derived using quadruple collocation (QC). Subsequently, a rain/no-rain correction is implemented using the categorical triple collocation merging (CTC-M) algorithm.

As a first preprocessing step, all rainfall products are regridded to a 0.1° × 0.1° resolution, using the nearest-neighbor method for products having a higher resolution. Then, a monthly correction is applied to all individual products with CHPclim, a monthly rainfall climatology dataset (Funk et al. 2015a). After preprocessing, the datasets are merged using merging weight factors derived from QC. QC is a statistical method used to refine the accuracy of remotely sensed rainfall products by comparing at least three independent products (x_1, x_2, and x_3). Other than the commonly used TC (Stoffelen 1998), QC allows for a fourth rainfall product x_4, which shares dependency with another product x_3. SUPER uses the QC error estimator from Gruber et al. (2016). The error variances of the rainfall products are estimated by using the following linear system:

$$\overline{\mathbf{y}} = \begin{bmatrix} \sigma_a^2 \\ \sigma_b^2 \\ \sigma_c^2 \\ \sigma_d^2 \\ \sigma_{ab} \\ \sigma_{ac}\sigma_{ad}/\sigma_{cd} \\ \sigma_{bc}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ac}\sigma_{cd}/\sigma_{ad} \\ \sigma_{bc}\sigma_{cd}/\sigma_{bd} \\ \sigma_{ad}\sigma_{cd}/\sigma_{ac} \\ \sigma_{bd}\sigma_{cd}/\sigma_{bc} \\ \sigma_{ac}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ad}\sigma_{bc}/\sigma_{cd} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1&0&0&0&0&1&0&0&0&0 \\ 0&1&0&0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0&1&0&0 \\ 0&0&0&1&0&0&0&0&1&0 \\ 0&0&0&0&1&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \end{bmatrix}, \quad \overline{\mathbf{x}} = \begin{bmatrix} \beta_a^2\sigma_\Theta^2 \\ \beta_b^2\sigma_\Theta^2 \\ \beta_c^2\sigma_\Theta^2 \\ \beta_d^2\sigma_\Theta^2 \\ \beta_a\beta_b\sigma_\Theta^2 \\ \sigma_{\varepsilon a}^2 \\ \sigma_{\varepsilon b}^2 \\ \sigma_{\varepsilon c}^2 \\ \sigma_{\varepsilon d}^2 \\ \sigma_{\varepsilon a}\sigma_{\varepsilon b} \end{bmatrix}, \quad \overline{\mathbf{y}} = \mathbf{A}\overline{\mathbf{x}}.$$

In this system, σ_i^2 is the dataset variance of product i, σ_{ij}^2 is the dataset covariance between product i and j, $\beta_i^2\sigma_\Theta$ is the signal variance of product i, $\sigma_{\varepsilon i}^2$ is the error variance of product i, and $\sigma_{\varepsilon i}\sigma_{\varepsilon j}^2$ is the interproduct error covariance between i and j.

To get the error variances and interproduct covariance, the system can be solved using the common least squares solution:

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\overline{\mathbf{y}}. \tag{7}$$

A more detailed derivation of the least squares solution, showing how the elements of $\overline{\mathbf{y}}$ are mathematically related to $\overline{\mathbf{x}}$, is provided in appendix A. The error covariances of each product (indexed by i) are then rescaled to product a and are subsequently used in the final error matrix:

$$\sigma_{\epsilon i,s}^2 = \sigma_{\epsilon i}^2 \left(\frac{\beta_a^2 \sigma_{\Theta}^2}{\beta_i^2 \sigma_{\Theta}^2} \right), \tag{8}$$

$$\mathbf{E} = \begin{bmatrix} \sigma_{\epsilon a,s}^2 & \sigma_{\epsilon a}\sigma_{\epsilon b,s} & 0 & 0 \\ \sigma_{\epsilon a}\sigma_{\epsilon b,s} & \sigma_{\epsilon b,s}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\epsilon c,s}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\epsilon d,s}^2 \end{bmatrix}.$$

Here, the error matrix ($\mathbf{E}$) includes the error variances of the different products and interproduct error cross correlations (ECCs). Using $\mathbf{E}$, the weight factors of the rainfall products can be determined:

$$\lambda_i = \frac{\sum_i e_{ij}}{\sum_i \sum_j e_{ij}}, \tag{9}$$

where $i, j \in [1, \ldots, 4]$ indicates the rainfall product and e_{ij} are the elements of the error matrix. The merged precipitation estimates are calculated with

$$x_m = \sum_{i=1}^{N} \lambda_i x_i^s, \tag{10}$$

where x_m is the merged rainfall estimate and x_i^s is the scaled rainfall estimate of product i. Last, the QC-based product-truth correlation and ECC can be estimated with

$$R_i^{qc} = \frac{\beta_a^2 \sigma_{\Theta}^2}{\beta_a^2 \sigma_{\Theta}^2 + \sigma_{\epsilon i}^2}, \tag{11}$$

$$R_{\epsilon a\epsilon b}^{qc} = \frac{\sigma_{\epsilon a}\sigma_{\epsilon b}}{\sqrt{\sigma_{\epsilon a}^2 \sigma_{\epsilon b}^2}}. \tag{12}$$

For a "perfect" rainfall product, R_i^{qc} would be 1, and for perfectly error-independent products, $R_{\epsilon a\epsilon b}^{qc}$ would be 0.

To enhance the accuracy, the SUPER algorithm applies a rain/no-rain filter via CTC-M (Dong et al. 2020). This filter requires three independent rainfall products and assesses the classification skill of capturing rain/no-rain of each product through CTC. Based on this assessment, the products are merged within a probabilistic framework to maximize the likelihood of accurate rain/no-rain estimates at every time step, which is then applied as an additional filter to the merged product from Eq. (10). This study follows the steps outlined in Dong et al. (2020, 2022). Likewise, a rain/no-rain threshold of 0.5 mm day^{-1} and a merging parameter of $n = 1.5$ are used, as suggested in Dong et al. (2022).

2) RAINFALL PRODUCTS

As explained in section 2c(1), SUPER requires at least three independent rainfall products. The study of Chen et al. (2021) has shown that reanalysis, soil moisture–estimated, and thermal infrared/passive microwave rainfall products are relatively independent, which is also confirmed by TC analysis in the study of Dong et al. (2022). Consequently, these products are well suited for use within the SUPER algorithm scheme. However, Chen et al. (2021) have tested this assumption solely for Europe, and their analysis mostly focused on testing the interdependency between SM2Rain and the H SAF H23 daily precipitation product. Therefore, this assumption is tested again by using various combinations in the QC analysis. For each combination, both the QC-derived and traditionally derived product-truth correlations are compared. "Traditionally derived" refers to deriving correlations using ground data from TAHMO. Using a combination with wrongly assumed error independence will result in biased ECC and product-truth correlations. This test will provide insight in the validity of the assumptions made. In addition, since the H26 product relies on the land surface model HTESSEL, which is also used in ERA5, this may introduce an extra error dependency between SM2Rain and ERA5. To check this, an additional QC analysis is conducted in appendix B, using SM2Rain derived from ASCAT H119/H120 (EUMETSAT H SAF 2021a), to assess the influence of different soil moisture products on SUPER.

Next to SM2Rain, ERA5, IMERG (late), and CHIRPS are chosen as rainfall products for this study. ERA5 is a reanalysis product providing global atmospheric data. IMERG is a product from the Global Precipitation Measurement (GPM) that uses MW and IR data to make rainfall estimates on a half-hourly resolution. Additionally, CHIRPS is selected as an additional satellite rainfall product. CHIRPS is based on IR data and is bias-corrected using monthly data from ground stations. In the study of Dong et al. (2022), IMERG and CHIRPS are assumed to have interdependency as these have both IR-based retrieval. All individual products are bias corrected on a monthly scale using CHPclim, a monthly climatology dataset, with corrections applied to each pixel after regridding CHPclim. Table 1 summarizes the selected rainfall products used in SUPER.

Figure 2 presents an overview of the total workflow developed in this study. To summarize, the process is first started with the regional implementation of SM2Rain. The most important input datasets for this are the ASCAT soil moisture data and rainfall data from TAHMO. After calibration per rainfall class using TNC optimization, parameter grids are made using a nearest-neighbor method. Subsequently, SM2Rain time series are produced over the grid. Following this step, SM2Rain, along with CHIRPS, IMERG, and ERA5, is corrected using CHPclim and is regridded to a 0.1 grid. Finally, the four rainfall datasets are merged using the SUPER framework.

d. Validation of satellite rainfall estimates

The rainfall estimates from both the merged product and the individual rainfall products are validated on a pixel-to-

TABLE 1. Selected datasets for the merged product.

	Spatial resolution	Temporal resolution	Latency	Type	Reference
SM2Rain	0.1°	1 day	12 h	MW	Brocca et al. (2019)
ERA5-Land	0.1°	1 h	5 days[a]	Reanalysis	Muñoz Sabater et al. (2021)
IMERG late (V06B)	0.1°	30 min	10 h	IR + MW	Huffman et al. (2019)
CHIRPS (V2.0)	0.05°	1 day	3 weeks	IR	Funk et al. (2015b)
CHPclim (V1.0)	0.05°	1 month	—	Climatology	Funk et al. (2015a)

[a] ERA5-Land is available 5 days after the assimilation, but the final data may be different 2 or 3 months later in case serious errors are detected.

station basis using the ground data from TAHMO as the reference dataset. The evaluation is based on the RMSE, Kling–Gupta efficiency (KGE), mean absolute bias, Spearman's correlation ρ, probability of detection (POD), false alarm ratio (FAR), and the Heidke skill score (HSS). The RMSE is quantified based only on rain events: Data are considered only when both the station and satellite measure rainfall. The POD and FAR metrics measure the product's ability to capture true events and the false alarms, while the HSS is an overall true/false detection metric. An overview of all evaluation metrics can be found in Table 2.

The validation is performed by first assessing the mean performance across all stations over the entire period. For this, only stations having >50% data coverage over the entire period (2018–22) are included. Subsequently, annual performance is analyzed for both daily and monthly resolutions. For this, only stations having >80% data coverage in the specific year are included. The detection abilities of different rainfall intensities are assessed using the POD, FAR, and HSS.

Additionally, we validate the detection ability of the rainy season offset. For this, we follow the flexible rainy season definition as proposed in the study of Seregina et al. (2019). Here, the rainy season is defined as the period of at least five consecutive pentads exceeding a location-specific determined rainfall threshold. To determine this, a long climatology is needed to reliably establish a robust threshold for the onset. A long-term climatology is not available for TAHMO stations, and since the period 2018–22 may be too short, we included the years 2016 and 2017 and considered stations with more than 80% data availability between 2016 and 2022 to derive the

most reliable approximation of the threshold possible with TAHMO data. This resulted in a selection of eight stations in total. These stations detected 38 rainy seasons altogether for the period 2018–22. For the validation, we assess whether the onset of the rainy season detected in different rainfall products aligns with the onset identified by TAHMO, using buffers of 2, 5, and 15 days. Additionally, we analyze the results separately for the four drier stations and four wetter stations by categorizing them based on their rainfall class. A more detailed overview of the rainy season analysis can be found in appendix C.

3. Results

This section presents the results from SM2Rain and SUPER and is organized as follows. First, the calibration results of the regional SM2Rain implementation are presented in section 3a. Next, section 3b provides the QC analysis of the SUPER framework. Finally, the validation results for all individual products are presented in section 3c.

a. SM2Rain

Figure 3 illustrates the spatial distribution of SM2Rain parameters across the region. As a result of the Gaussian filter, the parameter fields show smoother transitions than the rainfall classification of Fig. 1. Parameter a has higher values within wet regions, while b has higher values in dry areas. These parameters describe the drainage within the soil water balance. A higher a indicates more drainage for a given soil moisture content, and a higher b value implies more sensitive responses to
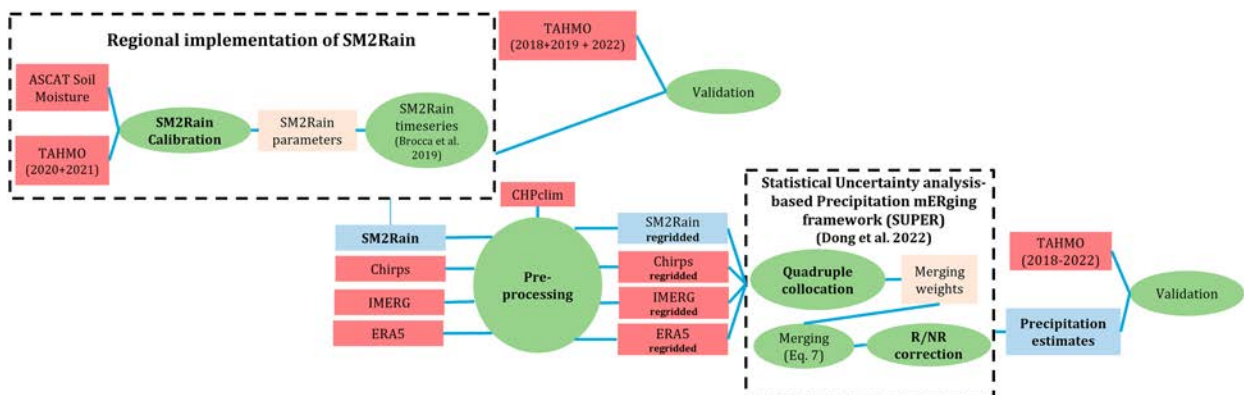

FIG. 2. The complete workflow for making rainfall estimates.

TABLE 2. Performance metrics used for product evaluations. The X_i and Y_i are the observed and predicted values, respectively, for the ith day, N is the total number of data points, r is the linear correlation coefficient between X_i and Y_i, α is the ratio of the standard deviation between X_i and Y_i, β is the ratio of the mean of X_i and Y_i, A is the number of true positive predictions (correctly predicted events), B is the number of false positive predictions (incorrectly predicted events), C is the number of false negative predictions (missed events), D is the number of true negative predictions (correctly predicted nonevents), and d_i is the difference between the ranks of the observed and predicted values for the ith data point.

	Full name	Definition	Range	Optimal value
RMSE	Root-mean-square error	$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2}$	0 to ∞	0
KGE	Kling–Gupta efficiency	$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$	$-\infty$ to 1	1
Bias	Bias	$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i)$	$-\infty$ to 0 to ∞	0
ρ	Spearman's correlation	$\rho = 1 - \frac{6\sum_{i=1}^{N}d_i^2}{N(N^2-1)}$	-1 to 1	1
POD	Probability of detection	$\text{POD} = \frac{A}{A+C}$	0 to 1	1
FAR	False alarm ratio	$\text{FAR} = \frac{B}{A+B}$	0 to 1	0
HSS	Heidke skill score	$\text{HSS} = \frac{2(AD - BC)}{(A+C)(C+D)+(A+B)(B+D)}$	$-\infty$ to 1	1

changes in soil moisture. The parameters Z^* and T have less spatial coherence with respect to wet and dry areas; however, their values are generally higher in wetter areas. Parameter c is generally lower in wet areas. Table 3 presents the parameter statistics from the calibration result for the regional SM2Rain product.

Figure 4 illustrates the impact of varying the number of stations used for SM2Rain calibration, showing the spread of performance metrics among stations when comparing different SM2Rain calibration settings against TAHMO station data from different years. It can be observed that there is no large change in the performance of SM2Rain with increasing numbers of stations. This indicates the robustness of SM2Rain's

calibration using rainfall classes and suggests that a very dense network of stations is not necessarily needed to improve its performance. To support this, Fig. 5 shows that the parameter ranges of all variables over the pixels do not significantly change when using a different number of stations in the SM2Rain calibration.

When conducting a quick comparison between the SM2Rain calibrated with TAHMO versus the "original" SM2Rain product (V1.5) (Brocca et al. 2019) in the year 2019 with stations independent from the calibration years, we find the following metrics: RMSE (10.85 vs 10.70 mm day^{-1}) and KGE (-0.20 vs -0.41) for daily resolution, and RMSE (2.33 vs 2.94 mm day^{-1}) and KGE (-0.44 vs -1.39) for monthly resolution. These

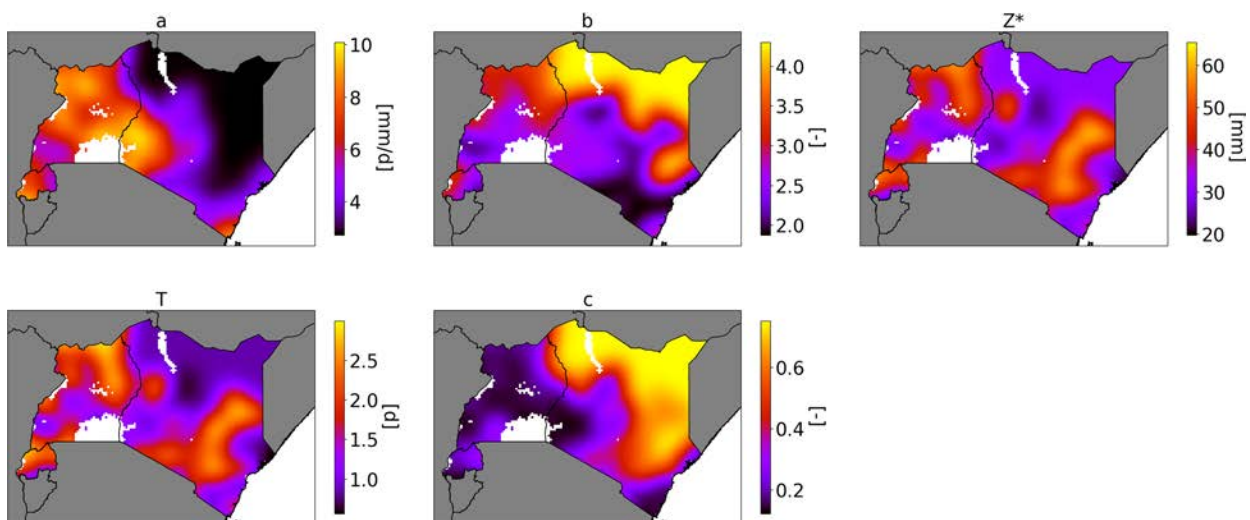

FIG. 3. SM2Rain parameter fields, based on the calibration with TAHMO data from 2019 to 2020. Parameters a, b, and Z^* are used to make rainfall estimates based on the conversion of the soil moisture water balance [Eq. (2)]. Parameters c and T are used in the modified exponential filter that is applied to the soil moisture data [Eq. (6)].

TABLE 3. Summary statistics of SM2Rain parameters in this study, based on the average calibration results over the 5-yr period.

Parameter (unit)	Median	Min	Max
a (mm day^{-1})	4.70	2.74	10.09
b (—)	2.91	1.88	4.30
Z^*(mm)	35.09	19.76	65.39
T (day)	1.40	0.56	3.00
c (—)	0.34	0.12	0.75

results indicate a small improvement in mean performance metrics, suggesting a benefit from incorporating local rain stations when calibrating rainfall products. Additional validation results of SM2Rain simulations with ASCAT H119/H120, calibrated with TAHMO, can be found in appendix B.

b. SUPER

Using Eqs. (11) and (12), the QC-based product-truth correlations and ECCs are quantified to test the assumptions underlying the SUPER merging. Figure 6 shows the QC-based ECC of the different products on a pixel-to-pixel basis from different QC combinations. Since the absolute value of ECC cannot theoretically exceed 1, any data from different combinations that exceed this limit are filtered out. From this figure, it can be seen that combinations (Figs. 6c–f) exhibit absolute ECCs larger than 1, which is physically impossible, indicating that using these combinations likely result in biased estimates. Therefore, these combinations are excluded from further analysis. From the remaining combinations, both CHIRPS–

IMERG (Fig. 6a) and ERA5–SM2Rain (Fig. 6b) show positive ECCs, with higher ECCs observed in drier areas. Generally, ERA5–SM2Rain has higher ECCs than CHIRPS–IMERG. Since combinations (Figs. 6a and b) are the only ones with realistic ECC values, they are considered for further analysis.

Figure 7 presents maps displaying the spatial distribution of weight factors λ_1 of each product for combinations (Figs. 7a and b), together with boxplots of the λ_1 distribution over all grid cells. For combination (Fig. 7a), ERA5 has clearly the highest total contribution in most regions. The contributions of both IMERG and CHIRPS are relatively low, as highlighted in the boxplots of the λ_1 distribution per grid cell. For combination (Fig. 7b), it is the other way around; IMERG has the highest contribution, followed by CHIRPS. However, the difference between these two is smaller than the difference between SM2Rain and ERA5 in combination (Fig. 7a). Both ERA5 and SM2Rain have significantly lower contributions in combination (Fig. 7b).

For both combinations, the QC-based product-truth correlations are quantified using Eq. (11). The results are summarized in Table 4. Based on this table, it is observed that the QC-based Pearson correlations from combination (SM-ER) have a closer match with the correlations found with TAHMO, compared to combination (IM-CH). In the study by Chen et al. (2021), some ECC was also found between SM2Rain and ERA5 in Europe, but this did not lead to large biases in the QC. The ECC between SM2Rain and ERA5 may potentially be caused by the use of HTESSEL land surface model in both ERA5 and H26
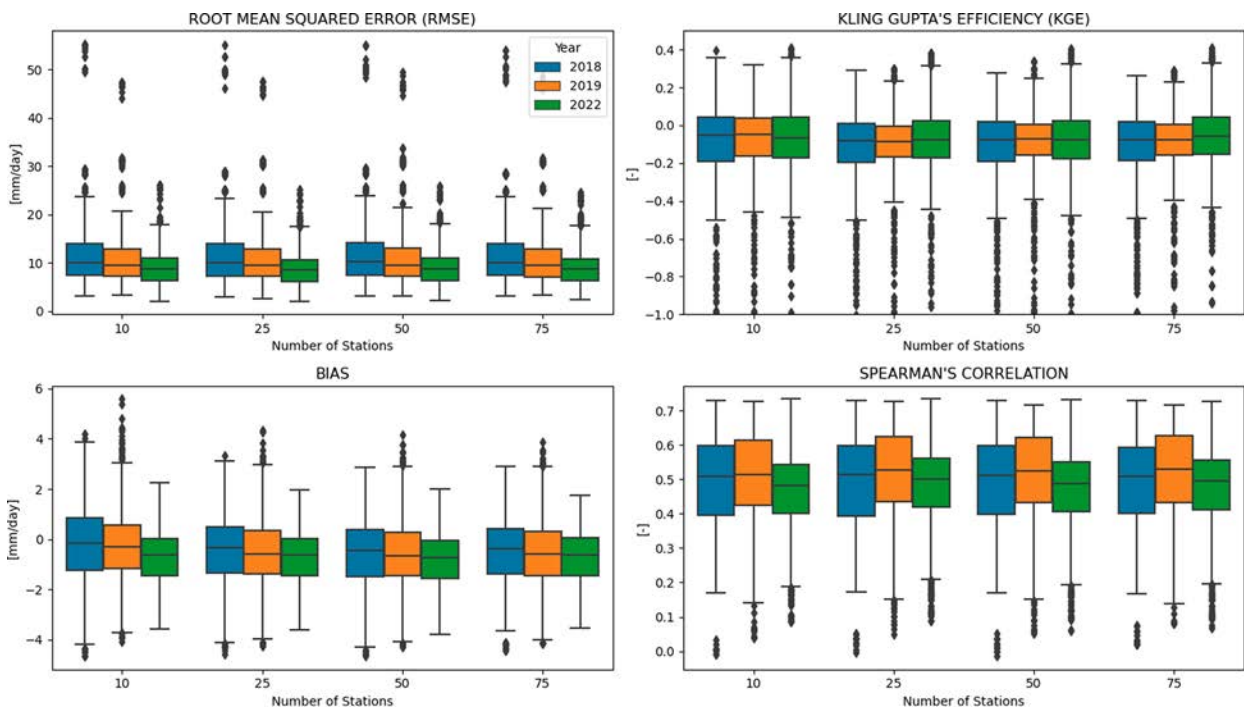

FIG. 4. Validation results of SM2Rain produced with varying numbers of random stations used in calibration. Each calibration setting was repeated 10 times for a unique set of stations each year. The validation data include 80 stations in 2018, 105 stations in 2019, and 126 stations in 2022.
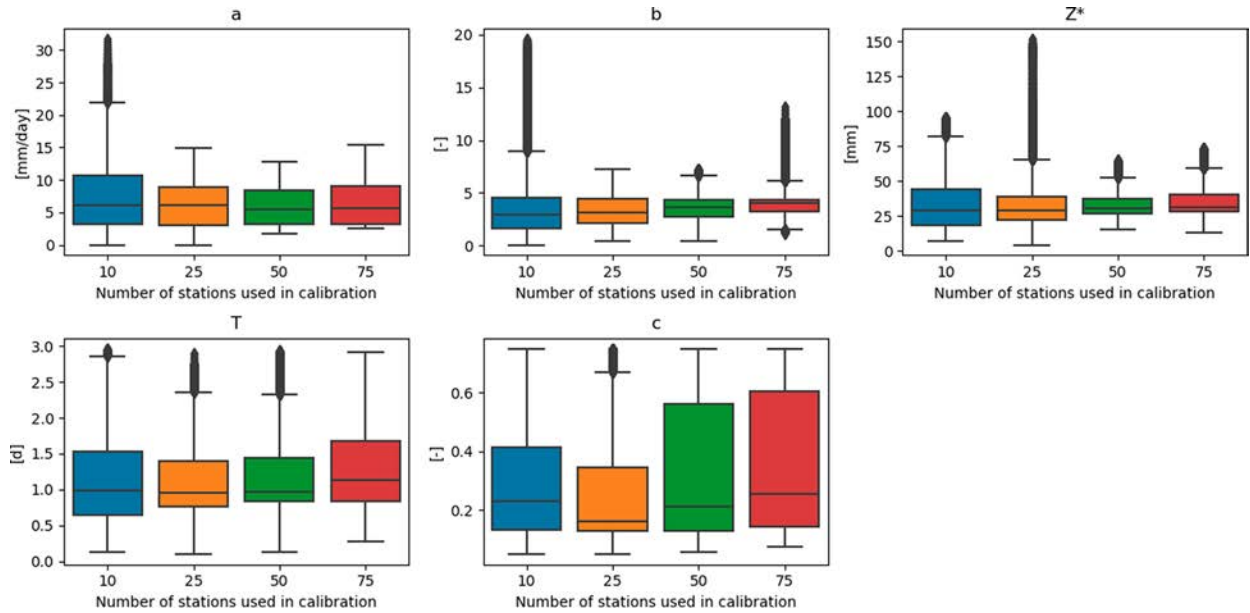
FIG. 5. Spread of SM2Rain parameters from 10 calibrations, each using a varying number of randomly selected stations. Each boxplot represents data points from 10 repeated calibrations times the number of grid points ($n = 304\,084$).

soil moisture data. It is important to note that there may be uncertainties associated with these comparisons due to the relatively limited number of data points. These uncertainties are related to potential biases among the stations, making it challenging to directly compare the quality QC-derived and traditionally derived product-truth correlations. We validated SUPER using both combinations IM-CH(CH-IM) and (SM-ER).

Additionally, we tested both combinations for the rain/no-rain classification.

c. Validation of rainfall products

This section presents the outcomes of the validation of SUPER, SM2Rain, and the other individual rainfall products, all of which are assessed by comparing pixel data against station
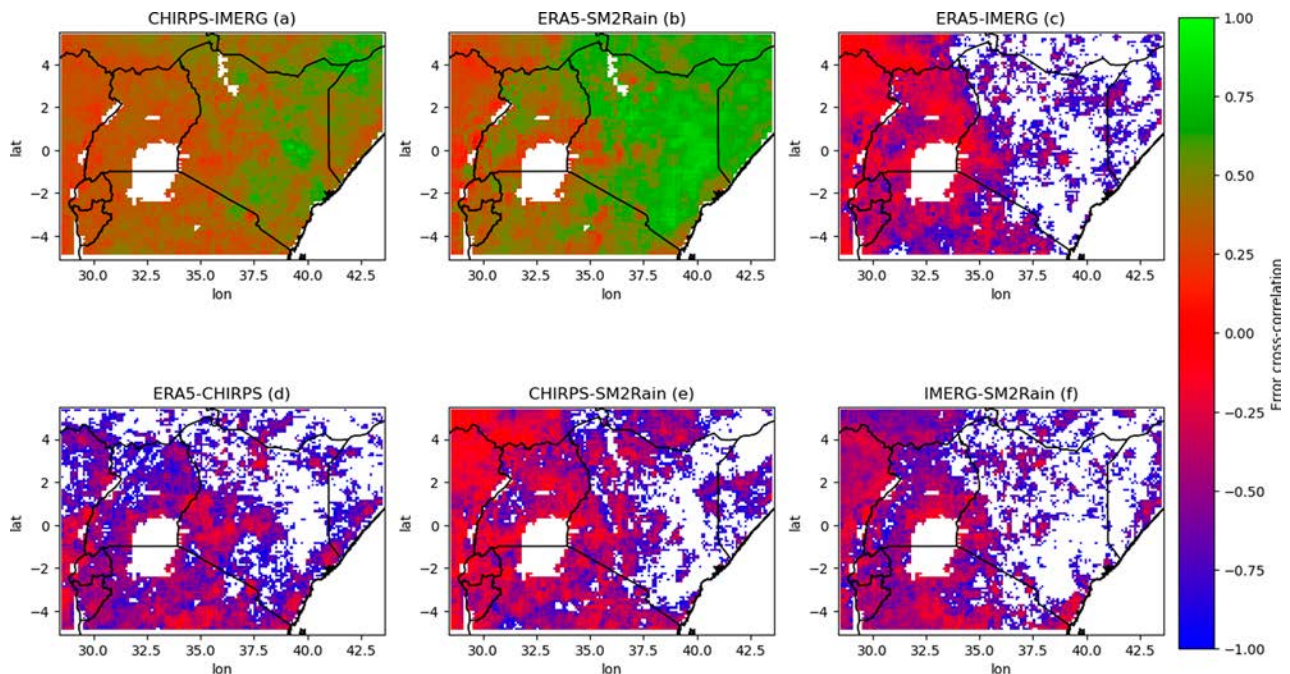

FIG. 6. QC-based ECCs between products using different SUPER configurations. Pixels having |ECC| > 1 are excluded in the maps.
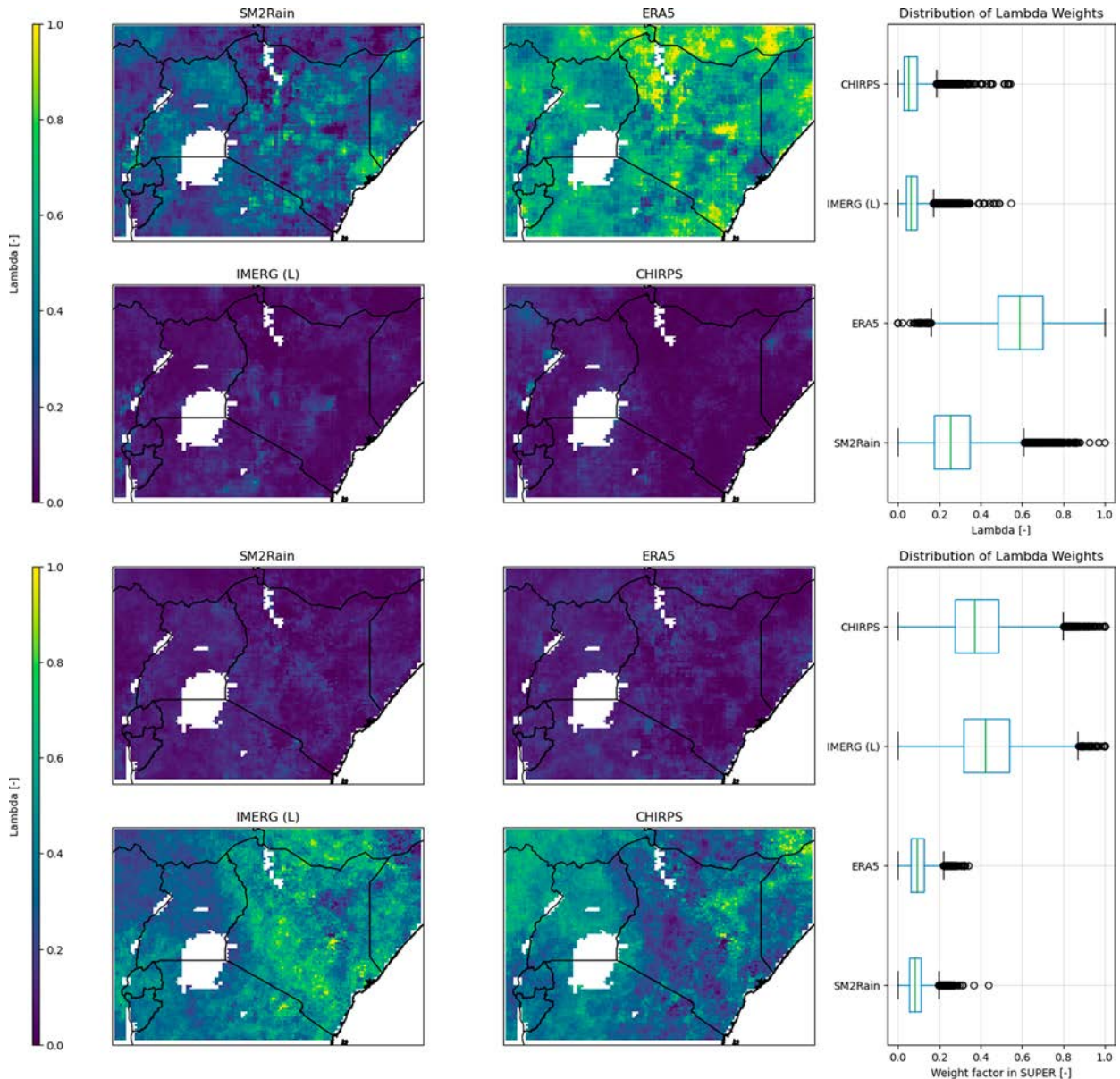
FIG. 7. Maps of the lambda (merging) weights of the rainfall products in SUPER, together with boxplots showing the distribution of merging weights across all pixels for each product. (top) IMERG-CHIRPS (IM-CH) error dependency assumption. (bottom) ERA5-SM2Rain (SM-ER) error dependency assumption.

data. Table 5 shows the performance metrics of all products at daily scale, considering all stations averaged over the considered period (2018–22). Among the individual products, SM2Rain performs best in terms of RMSE and bias, while IMERG has the highest KGE and Spearman correlation. CHIRPS shows the highest RMSE and also has the lowest correlation among the products. ERA5 exhibits both the highest absolute bias and the lowest KGE.

We tested SUPER considering both combinations (IM-CH and SM-ER) as error-dependent rainfall products with lambda weights from Fig. 7. For both combinations, we also separately tested the rain/no-rain classification (CTC-M). We found that

CTC-M using ERA5–IMERG–SM2Rain (HSS = 0.36) is a better classifier than ERA5–IMERG–CHIRPS (HSS = 0.27), or any other possible combination. Therefore, we used this CTC-M configuration for both SUPER combinations.

SUPER combination SM-ER has comparable RMSE and correlation to IM-CH, but outperforms IM-CH in terms of KGE and bias. This combination generally outperforms the other individual products as well. It has the highest KGE and the lowest bias. Along with IMERG, it has the highest correlation, and after SM2Rain, it has the lowest RMSE.

Looking at a monthly scale, we found that combination SM-ER generally outperforms IM-CH, although the differences

TABLE 4. Mean QC-based and traditionally based product-truth Pearson's correlations.

	CHIRPS–IMERG	ERA5–SM2Rain	TAHMO
SM2Rain	0.78	0.54	0.65
ERA5	0.78	0.55	0.65
IMERG	0.53	0.76	0.73
CHIRPS	0.49	0.71	0.67

between the configurations are small. SUPER outperforms all individual products in terms of RMSE, KGE, and correlation, but SM2Rain retains the lowest absolute bias. Notably, CHIRPS demonstrates a significantly improved performance compared to the daily resolution, and it has the lowest RMSE and highest KGE and correlation after SUPER. This good performance is likely attributed to the monthly correction applied in CHIRPS.

Figure 8 provides a deeper insight into the spread of performance among stations and years by showing the boxplots of different metrics at a daily resolution. Generally, our findings align with those in Table 5 over the years. However, the figure highlights a large variability in performance among stations. Especially for the RMSE and KGE, we observe a skewed distribution with the median positioned toward the edges of the box. This suggests that a majority of the stations consistently match the satellite products, although a minor part exhibits more variable or poorer performance. Specifically for KGE, there are a large number of outliers among the stations. This could also indicate that some stations might be biased compared to the observed pixel average due to various reasons. ERA5 demonstrates a considerable drop in RMSE performance in 2022 compared to previous years and other products. This decline is attributed to a substantial increase in bias during this specific year, which is significantly higher than in other years. All state-of-the-art rainfall products (ERA5, CHIRPS, and IMERG) consistently overestimate true rainfall, as measured by TAHMO, as indicated by the positive biases across all years for these products. The correlation remains relatively constant over the years for all products.

For the monthly scale (Fig. 9), we again found relatively consistent results compared to Table 6. SUPER shows a relatively smaller spread in performance compared to other products. The correlations at the monthly scale are higher compared to daily resolution, despite the presence of more outliers among the stations for this metric. The high mean RMSE and bias of ERA5 in Tables 5 and 6 are caused by an outlier in these metrics for the year 2022.

The detection abilities of the rainfall products to identify specific rainfall intensities are illustrated in Fig. 10. IMERG and both SUPER combinations perform best in detecting no rain (<0.5 mm day^{-1}). For light rain (0.5–5 mm day^{-1}), SM2Rain has the highest POD. CHIRPS performs poorly in detecting light rain and generally has the highest number of rain events when there is no rain measured by TAHMO stations. IMERG is the best at detecting heavy rainfall (>20 mm day^{-1}). For other intensities, SUPER (SM-ER) has the best detection abilities, followed by IMERG.

Last, Fig. 11 visualizes the ability of different rainfall products to detect the onset of rainy seasons. A good performance at a low number of buffer days indicates that the product accurately captures the timing of the rainy season offset. It can be clearly seen that detection percentages increase significantly as the number of buffer days increases. When all eight stations are considered, CHIRPS and SUPER (ER–SM) show the highest detection percentages across all buffer days. However, for the drier stations, SM2Rain has the highest detection percentage for the smallest buffer, followed by CHIRPS and SUPER (IR-CH). The SUPER (SM-ER) combination and CHIRPS exceed 100% detection for a 15-day buffer, suggesting that these products recorded a drop below the threshold during an ongoing rainy season, followed by a new consecutive five pentads of threshold exceedance. While IMERG and ERA5 show relatively low detection percentages for dry stations, they perform better for wet stations, whereas SM2Rain struggles to detect the rainy seasons of the wet stations.

d. Case studies

By comparing time series from satellite and station measurements, we gain a better understanding of what is actually happening on the ground rather than relying on validation metrics only. In this section, we present different case studies comparing time series data from satellite and station measurements at daily and monthly resolutions across six different stations (three for daily and three for monthly). For this, only SUPER, SM2Rain, and IMERG are considered. For SUPER, we considered combination (SM-ER) assuming error cross correlation between SM2Rain and ERA5. We have selected stations based on data availability and their rainfall classification, representing dry, medium, and wet areas, as shown in Fig. 1. To provide a more detailed view, the daily time series focuses on data from 2020 only, and the monthly time series includes data from both 2020 and 2021. We arbitrarily selected the stations for this analysis, although we ensured that they were

TABLE 5. Performance metrics of SUPER, SM2Rain, and other rainfall products on a daily resolution, averaged over all stations and the entire 5-yr period. The ρ is the Spearman correlation.

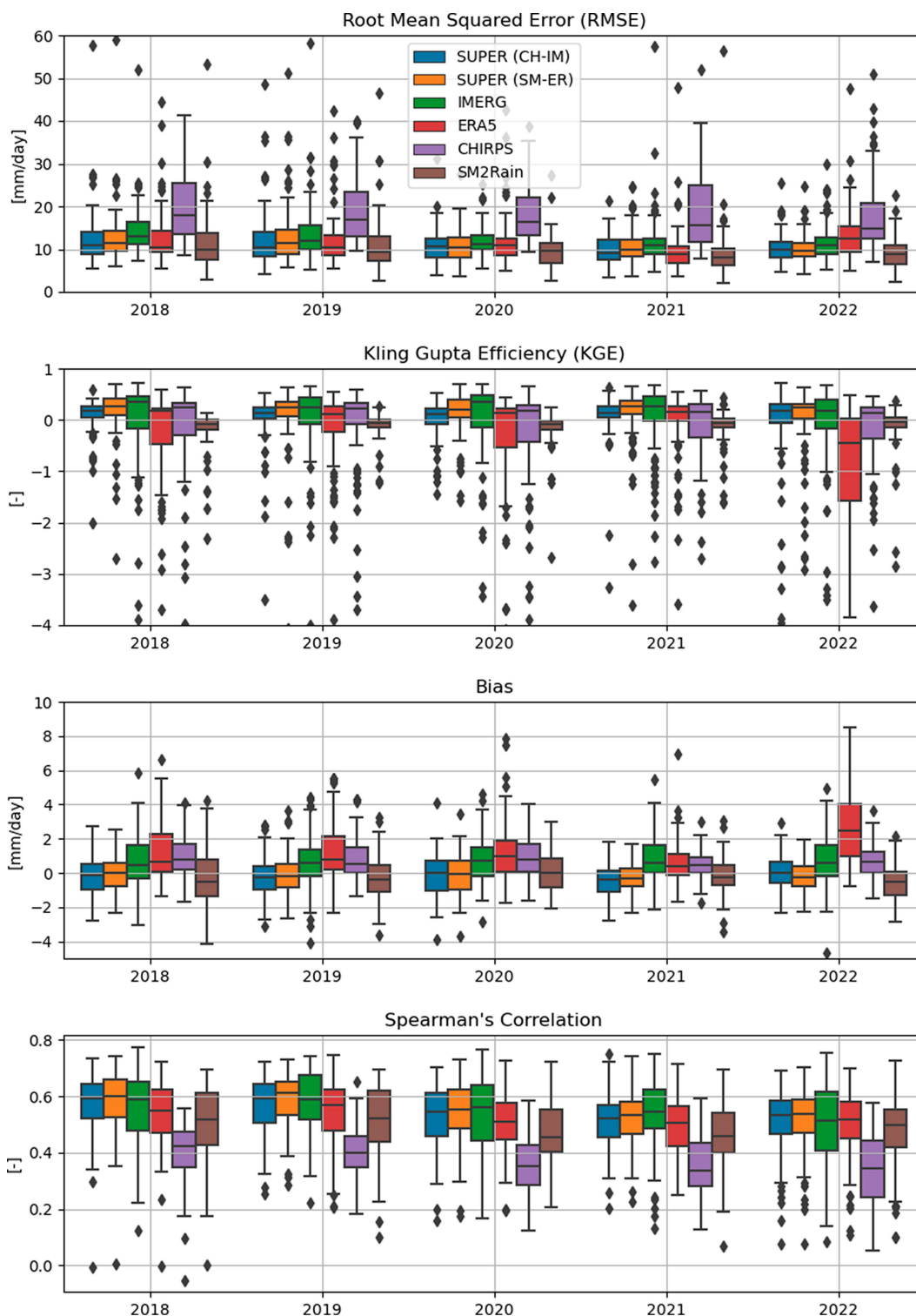|	RMSE (mm day^{-1})	KGE (—)	Bias (mm day^{-1})	ρ
SUPER (IM-CH)	11.44	0.15	−0.30	0.54
SUPER (SM-ER)	11.69	0.26	−0.21	0.55
SM2Rain	10.25	−0.05	−0.26	0.47
ERA5	11.91	−0.14	1.60	0.51
IMERG (late)	12.47	0.22	0.82	0.55
CHIRPS	18.33	0.08	0.73	0.38

FIG. 8. Yearly performance metrics of SUPER, SM2Rain, and other rainfall products on a point-to-pixel basis on a daily resolution, spread across all stations considered per year.
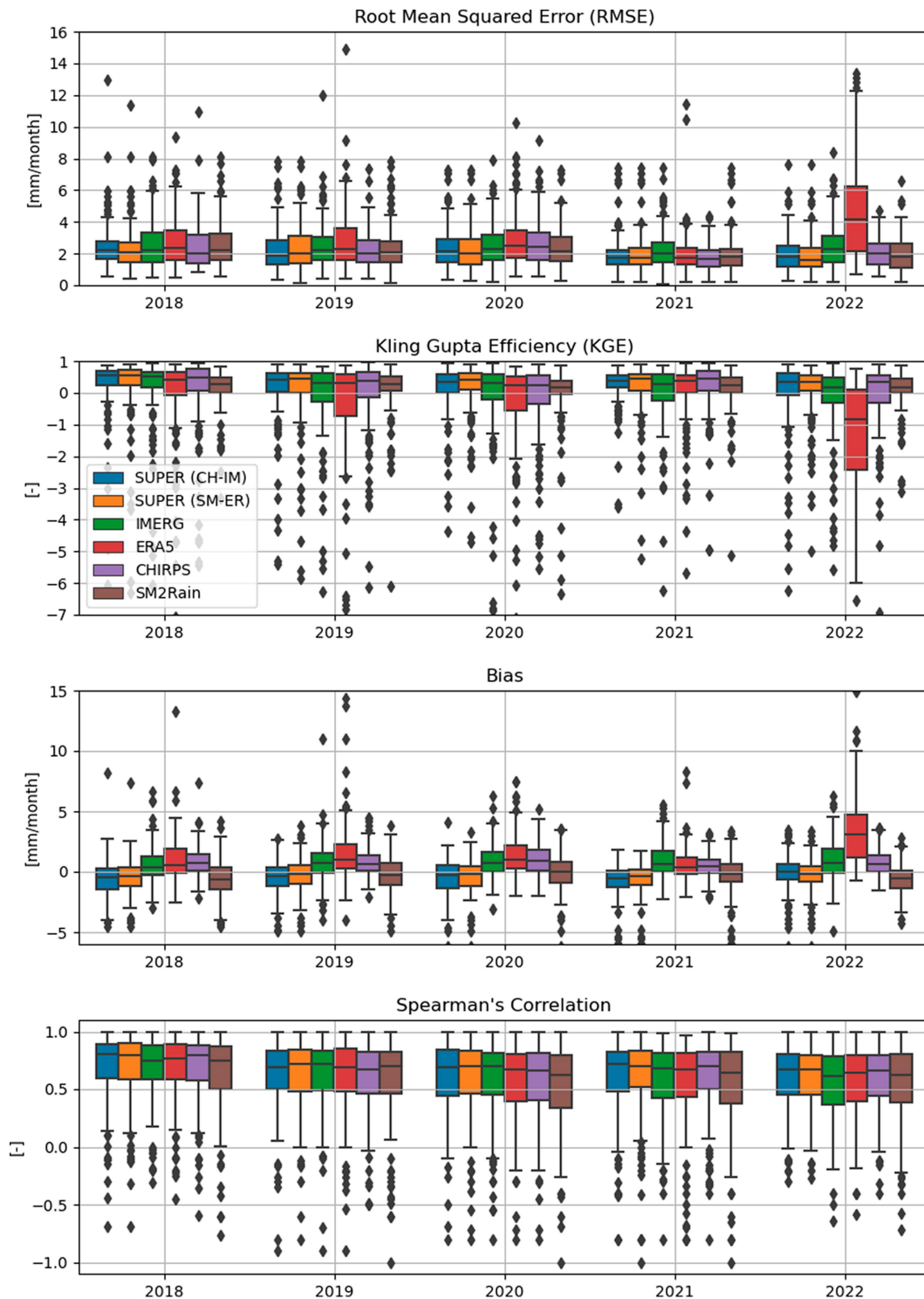
FIG. 9. Yearly performance metrics of SUPER, SM2Rain, and other rainfall products on a point-to-pixel basis on a monthly resolution, spread across all stations considered per year.

TABLE 6. Performance metrics of SUPER, SM2Rain, and other rainfall products on a monthly resolution, averaged over all stations and the entire 5-yr period. The ρ is the Spearman correlation.

	RMSE (mm day^{-1})	KGE (—)	Bias (mm day^{-1})	ρ
SUPER (IM-CH)	2.23	0.42	−0.43	0.64
SUPER (SM-ER)	2.21	0.44	−0.34	0.65
SM2Rain	2.33	0.30	−0.24	0.57
ERA5	3.38	−0.05	1.62	0.59
IMERG (late)	2.55	0.31	0.81	0.62
CHIRPS	2.30	0.36	0.754	0.63

located in dry, medium, and wet areas and verified data availability for each station. Although the case study period and stations are used in the calibration of SM2Rain, the main purpose is to compare rainfall products with TAHMO time series rather than validate SM2Rain or SUPER. Since the number of stations did not significantly impact SM2Rain's performance, excluding these stations from calibration would likely not have significantly influenced the results in this section.

The daily time series from TA00023 (dry), TA00679 (medium), TA00317 (wet), and the three satellite-based rainfall products are presented in Fig. 12. Based on the RMSE and KGE, we observe that SUPER performs best for the dry and medium stations, while IMERG has the best performance for the wet station. Additionally, SM2Rain outperforms IMERG for the dry station, and its RMSE is lower for the medium station and comparable to IMERG for the wet station. However, by examining the time series itself, it can be seen that SM2Rain clearly fails to capture the intermittent nature of rainfall during wet conditions, as it more behaves like a soil moisture product itself. This also relates to the poor performance of satellite-derived soil moisture in wet conditions. Under wet conditions, there tend to be more vegetation and saturated soil, leading to challenges in measuring changes in

soil moisture caused by rainfall (Barrett et al. 2009; Hahn et al. 2021). According to this, SM2Rain seems to be more reliable in drier conditions.

For the monthly time series (Fig. 13), we selected TA00453 (dry), TA00569 (medium), and TA00274 (wet) as ground stations. Considering the performance metrics, we see that SUPER outperforms both SM2Rain and IMERG for all categories. Based on the time series analysis, it is evident that SM2Rain is more capable of capturing the monthly time series compared to the daily time series, even at the wet station. In the dry season, we observe a small peak of rainfall in November and December 2020 that the TAHMO station does not detect. Upon examining the daily time series for this specific station during these months, the station recorded only very small rainfall intensities for several days, whereas the satellite products reported around 3 days with high-intensity events. This discrepancy could be due to several reasons, such as the TAHMO station not working properly during this period or just missing these high-intensity events due to point–pixel comparison.

Note that these results are just examples at six different stations. Different stations were selected for daily and monthly data because we wanted to include more than just three
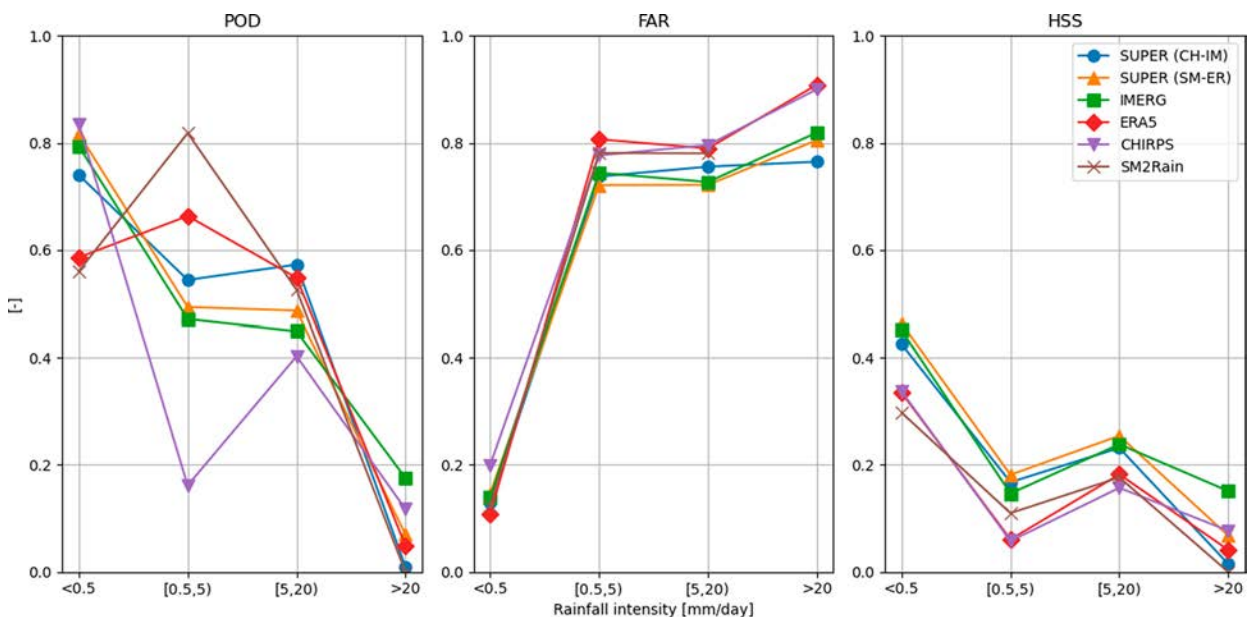

FIG. 10. Detection capabilities of all rainfall products across various rainfall intensities, averaged over all stations considered.
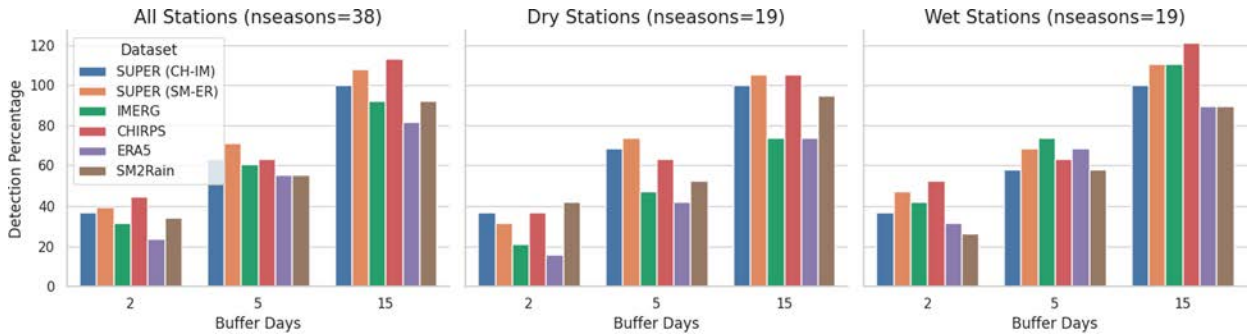
FIG. 11. Detection percentage of rainfall season onset across different datasets for all stations (8), dry stations (4), and wet stations (4) for 2, 5, and 15 days of buffer. >100% detection indicates that during the rainy season, measured by TAHMO, a drop below the threshold was recorded by the rainfall product, followed by five consecutive pentads with threshold exceedance.

stations in the analysis. Results may vary at different locations because of a large number of factors, such as biased stations, or certain landscapes influencing satellite measurements. Nevertheless, these findings emphasize the importance of not relying solely on performance metrics for validating satellite rainfall products and, more crucially, for their implementation in operational settings.

4. Discussion

The accuracy of SM2Rain is likely to be influenced by the choice of soil moisture product as input. Therefore, it might be worthwhile to test SM2Rain with other available satellite soil moisture products. We did a small comparison between the SM2Rain version of this study with SM2Rain V1.5 from

Brocca et al. (2019). However, the actual influence of implementing TAHMO data can be better assessed when using the same ASCAT product in both SM2Rain versions. Validation results of SM2Rain simulations with a comparable ASCAT product can be found in appendix B. These results suggested no large differences between the use of both products. Other examples of soil moisture products that can be used in SM2Rain include soil moisture data derived from the Soil Moisture Ocean Salinity (SMOS) mission, which uses SAR data, and the Soil Moisture Active Passive (SMAP) mission. However, these products have lower spatiotemporal resolutions compared to ASCAT. Additionally, high-resolution soil moisture data over agricultural crops derived from *Sentinel-1* might be interesting to consider in the development of SM2Rain (Satalino et al. 2014).
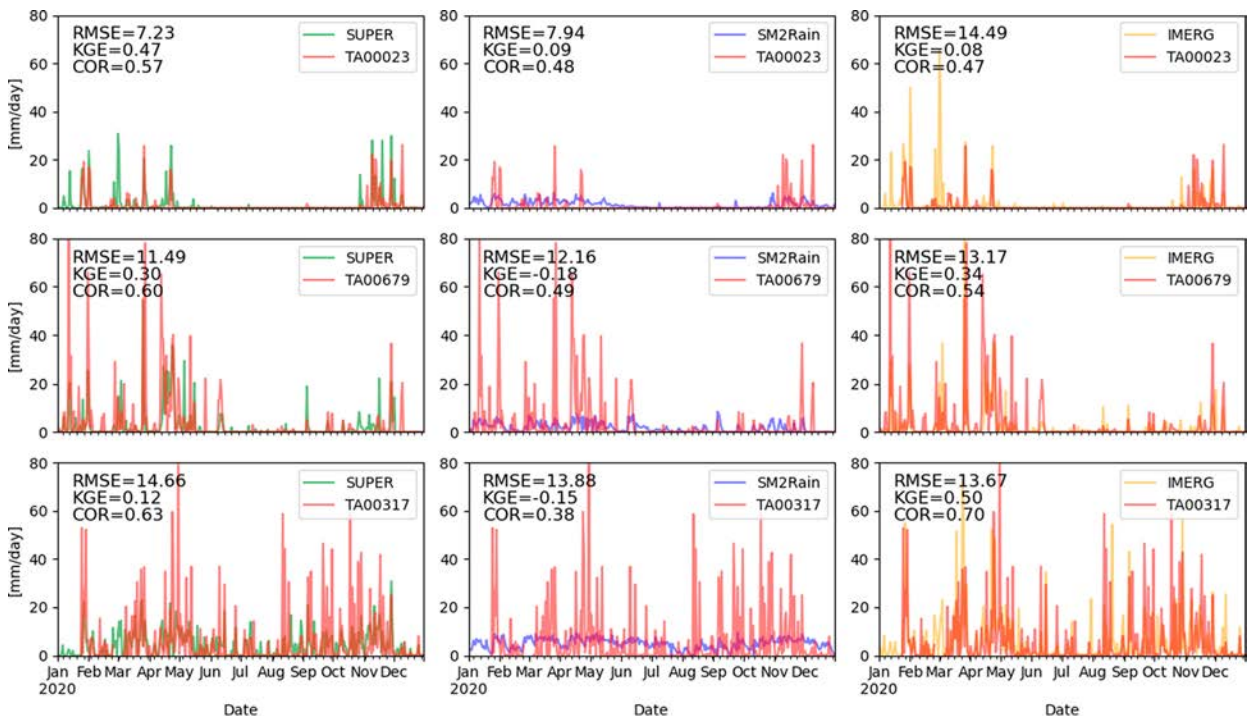

FIG. 12. Daily time series of SUPER (SM-ER), SM2Rain, and IMERG vs three TAHMO stations.
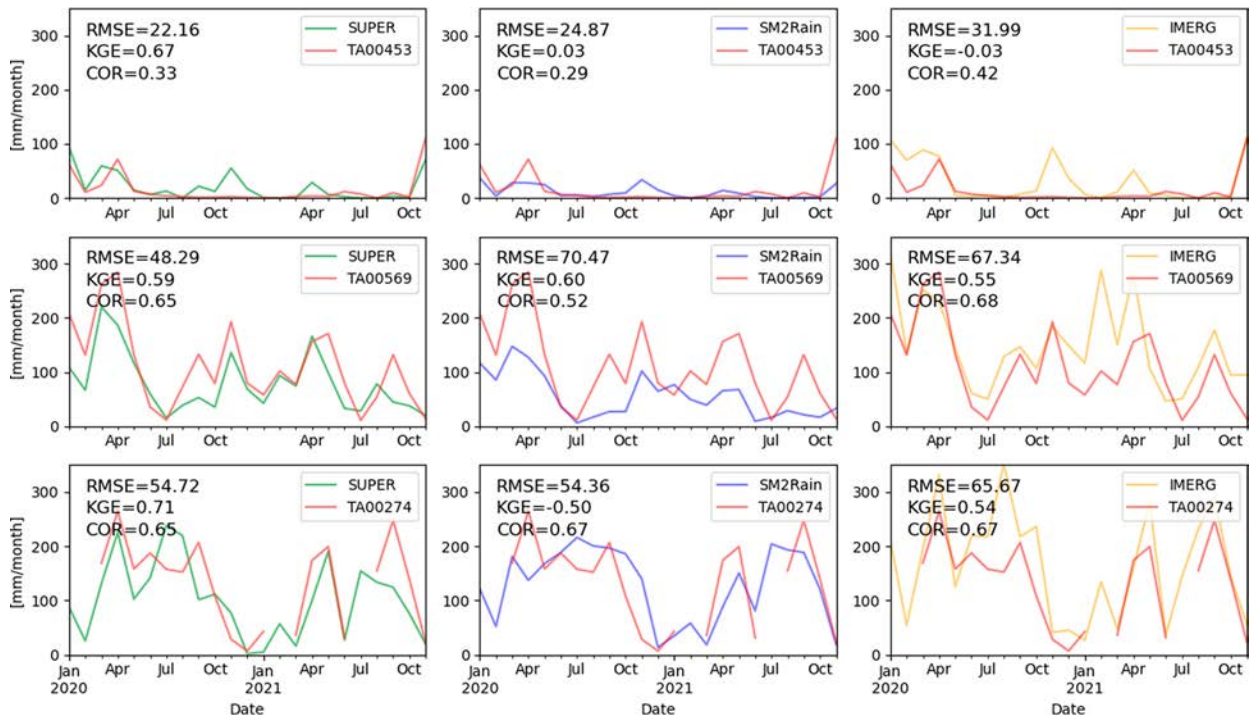
FIG. 13. Monthly time series of SUPER, SM2Rain, and IMERG vs three TAHMO stations.

Soil moisture data from 2021 onward used in SM2Rain are derived from the H26 product, while data before 2021 are generated from H141/142 products. The primary difference between the data before and after 2021 lies in the type of data assimilation used, and the historical H141/142 products uses some offline Land Data Assimilation System (LDAS) data for configuration, while H26 uses stand-alone surface analysis for this (Fairbairn et al. 2019). The choice of using this in H26, compared to previous operational versions, is mainly because of its computing speed (EUMETSAT H SAF 2021c). The exact impact of the difference in data assimilation methods on SM2Rain performance is not clear, however.

Although the original SM2Rain dataset is globally available, its performance has hardly been tested in East Africa. As of writing, the study by Boluwade (2020) is the only evaluation of SM2Rain in East Africa that focuses on daily rainfall validation. They performed a similar evaluation for CHIRPS and SM2Rain in Uganda and Ghana using TAHMO data as the reference dataset. Their validation in Uganda used 40 TAHMO stations for the year 2018. They found that SM2Rain (RMSE = 7.78 mm day^{-1}) performs relatively better than CHIRPS in Uganda (RMSE = 8.28 mm day^{-1}). We observed an RMSE of 8.16 mm day^{-1} for the same country and year (57 stations), indicating a more substantial improvement compared to the original product. Besides this, the most significant strength of this regional SM2Rain implementation lies in its potential for operational applications and robustness, primarily attributed to the integration of local rain gauges.

As mentioned, the primary reason for using SM2Rain is to effectively implement TAHMO stations in the development of the regional rainfall product, thereby creating an error-independent rainfall product for the merging algorithm. The choice of using the H26 soil moisture product rather than the products used in the original SM2Rain (V1.5) was because of its potential for operational implementation. Although the use of H26 in SM2Rain potentially introduces a bias in the QC due to error dependency between H26 and ERA5 from the use of the HTESSEL land surface model in both datasets, we assume that this does not lead to a large bias in the QC-based merging process. This assumption is further supported by the additional analysis conducted with H119/H120 in appendix B. Moreover, as noted, previous studies have shown that the SM2Rain (V1.5) product is relatively independent of reanalysis and IR/MW-based rainfall products, whereas this version of SM2Rain is calibrated with ERA5 for the generation of a global dataset. Our current analysis also demonstrates that assuming some other error cross-dependency, by assuming ERA5–CHIRPS–IMERG independency, resulted in realistic product-truth correlations. Therefore, we conclude that the impact of some error cross-dependency is likely not substantial in both SUPER configurations (CH-IM) and (SM-ER). In contrast, we found that certain QC combinations led to unrealistic error cross-dependencies, caused by a biased QC analysis due to the incorrect assumption of interproduct error independence in the selected combinations.

Based on the case studies and the rainy season analysis, our findings indicate that SM2Rain is less reliable under wet conditions but tends to be more useful for determining the timing of the onset of the rainy season in drier climates. Although the number of stations and the time period used

in the analysis were relatively limited, the product demonstrated a reasonable ability to capture the timing of the rainy season offset for the dry stations. Also, this product is good at capturing small rainfall events. In East Africa and many other areas, the beginning of the rainy season is a critical period that determines the success of agricultural activities (Ngetich et al. 2014).

In addition to the specific capabilities of SM2Rain, the other considered rainfall products also tend to be more useful in particular settings. For example, we found that IMERG is the best performing product among the individual products, specifically in capturing high rainfall events, while CHIRPS performs best on a monthly scale and in the overall detection of rainy season offsets. In the best performing SUPER configuration, IMERG also has the highest contribution, which correlates with the strong performance of this product among the individual products. With the SUPER merging, we chose a static merging approach, as the merging weights remained constant over the considered time period. However, as a next step, it might be interesting to explore a more dynamic approach based on seasonality, climate, and other geophysical properties of a pixel or region. As mentioned and confirmed in the literature, every satellite rainfall product has strengths and weaknesses in certain applications (Le Coz and van de Giesen 2020). Therefore, it may be worthwhile to develop a more dynamic merging approach that assigns higher merging weights to products in specific settings or merges in different smaller time periods. However, when using SUPER merging in these variable periods, it is important to keep in mind that TC analysis (and QC) requires a minimum number of sample points to make reliable error variance estimates (Tsamalis 2022).

Several other studies investigated the performance of satellite-based rainfall products over Africa, focusing on IMERG, ERA5, and CHIRPS. For instance, Macharia et al. (2022) did a performance validation of CHIRPS against gauge data from TAHMO. They found that CHIRPS had the lowest performance in terms of RMSE and bias compared to the other investigated products (TAMSAT, GSMaP, and GSMaP_gauge) in East Africa. However, when aggregating to a monthly scale, CHIRPS performed best after GSMaP_gauge. Dezfuli et al. (2017) validated IMERG (V04A) data against its predecessor (TMPA), as well as CHIRPS and GPCC datasets, using rainfall observations from one station in Kenya and two stations in Ghana for 2015. They found that IMERG performed best in capturing the distribution of daily rainfall intensity, while CHIRPS showed the largest differences with gauge observations for daily rainfall at the Kenyan station. The recent study of Mekonnen et al. (2023) conducted a multiscale assessment of eight rainfall products across the African continent from 2001 to 2020, including CHIRPS, IMERG, and ERA5. They found that CHIRPS has the lowest RMSE on a monthly scale (RMSE = 57.54 mm month^{-1}, or 1.89 mm day^{-1}), followed by IMERG (RMSE = 58.03 mm month^{-1}, or 1.91 mm day^{-1}) and ERA5 (RMSE = 62.77 mm month^{-1}, or 2.06 mm day^{-1}). However, this study used a dataset other than TAHMO for evaluation, and the number of stations used for evaluation was relatively limited, despite spanning a longer time frame than

possible with TAHMO data. Overall, the findings of these studies generally align with this study.

In the results, we found a number of outliers among the stations for the performance metrics. For example, in Fig. 8, some stations show very low performance for specific years. Despite applying additional quality checks and filtering out certain data periods, some TAHMO data may still have issues due to various factors. For SM2Rain, we observed a relatively high number of stations with poor performance. This is largely due to the product's reduced reliability in wetter climates, as highlighted by the case studies. When investigating other existing outliers, we found cases where rainfall products measured several peaks while TAHMO indicated a relatively dry period. In particular, the sharp drop in ERA5 performance in 2022 was caused by numerous peaks recorded by ERA5, whereas TAHMO indicated drought. Other products also detected some peaks during this period, but to a lesser extent. While ground data are often considered the reference for validation, quality issues can arise in specific periods. Given the large number of stations analyzed, identifying and systematically filtering out all quality issues remains challenging due to the uncertainties in the underlying causes of discrepancies. Additionally, differences in scale between point-based (TAHMO) and pixel-based (satellite) measurements can contribute to missing events, leading to lower performance metrics.

Similar to ERA5, SM2Rain, and IMERG, the SUPER rainfall product provides rainfall estimates at a spatial resolution of 0.1°, approximately corresponding to 11 km in the region of interest. However, convective clouds, the predominant mechanism generating rainfall in East Africa, are characterized by extremely high spatial variability (Palmer et al. 2023). This variability is further amplified by the complex topography of the region (Indeje et al. 2000). As the validation of the results is done on a pixel-to-point basis, this notable subpixel variability may significantly contribute to the imprecision of satellite rainfall estimates. Some research on small-scale rainfall variability using a dense network of rain gauges has been conducted in Niger (Flitcroft et al. 1989; Taupin 1997; Lebel et al. 1992). However, more research is needed to understand the small-scale rainfall variability at a subpixel level and how this relates to the accuracy of modern satellite rainfall products in sub-Saharan Africa.

Next to the high spatial variability of convective rains, this rainfall type is also characterized by its high temporal variability. Temporal mismatches, particularly concerning convective rains, impact the performance of satellite rainfall products. Measurements on a subdaily scale are necessary to capture extreme events (Maggioni et al. 2022). The temporal resolution of rainfall estimates by SUPER is 1 day, constrained by the temporal resolutions of the selected rainfall products: 1 day for both CHIRPS and SM2Rain, 1 h for ERA5, and 30 min for IMERG. Therefore, subdaily rainfall measurements are to some extent incorporated into the merged product. The study of Freitas et al. (2020) suggests that IMERG provides relatively useful subdaily scale rainfall information in Brazil. However, the evaluation reveals significant over- and underestimations of IMERG in terms of rainfall duration and intensity properties. Additionally, the study by Hu and Yuan (2021)

identified timing and intensity issues with ERA5 precipitation in the Tibetan Plateau. While aggregating these higher-resolution datasets partially mitigates these problems, it is noteworthy that these challenges persist even at higher temporal resolutions. SM2Rain and CHIRPS are likely to have even more temporal mismatches and therefore miss extreme events. The study of Xia et al. (2021), for example, assessed the performance of these two products over the Pearl River basin in China and found that these products perform poorly in capturing extreme events. Nevertheless, it is important to note that errors from temporal mismatches remain poorly investigated on the African continent.

Using the selected operational rainfall products and the regional implementation of SM2Rain, SUPER could be operationally implemented in the future. However, when considering the same individual rainfall products used in this study, this product would have a minimum latency of 3 weeks, aligning with the latency of CHIRPS. This would mean that the current setup will not be feasible for near-real-time applications. However, the workflow is designed in a way that other products are easy to be implemented. Some possible operational IR/MW products for reducing the latency that match the current spatial resolution of SUPER or having a higher resolution are TAMSAT (5-day latency) or GSMaP_MVK (3-day latency) (Tarnavsky et al. 2014; Kubota et al. 2017). This implies that, given the dependency on ERA5, the minimum latency of SUPER would be 5 days when replacing CHIRPS with either of these products. While SUPER may not be suitable for very near-real-time (NRT) applications, it remains valuable for agricultural management, for example.

5. Summary and conclusions

In this study, we used in situ rainfall data from TAHMO stations to develop a regional rainfall product for East Africa through the implementation of the SM2Rain algorithm. The generated regional product was subsequently merged with ERA5, IMERG (late), and CHIRPS using the Statistical Uncertainty Analysis-Based Precipitation Merging framework (SUPER). The development of the regional SM2Rain product was essential for three primary reasons: 1) to meet the prerequisite of SUPER, which requires three independent products for the quadruple collocation; 2) to potentially operationalize the workflow designed in this study, as the original SM2Rain product is not operationally implemented yet; and 3) to effectively integrate ground data into the merging algorithm.

The impact of TAHMO station data on SM2Rain's performance was investigated by calibrating the model with varying numbers of stations (10, 25, 50, and 75). It was found that calibrating with a higher number of stations did not lead to a large increase in performance, although the spread in performance among the stations was marginally reduced when using more stations. A possible reason for the limited impact of increasing the number of stations in calibration is the consistency of parameters across the grid in the different rainfall classes.

The merging algorithm of SUPER is based on the error variances of rainfall products derived from quadruple collocation.

We tested all possible SUPER configurations by assuming different interproduct error dependencies. It was found that both CHIRPS–IMERG (CH–IM) and ERA5–SM2Rain (ER–SM) have possible interproduct error dependencies, and therefore, both configurations are considered in the merging algorithm. Rainfall products with a lower error variance in a pixel receive a higher weight in the final merging. In the SM-ER configuration, IMERG has the highest contribution, followed by CHIRPS, ERA5, and SM2Rain. In CH-IM, ERA5 has the highest contribution, followed by SM2Rain, IMERG, and CHIRPS. Another aspect of the SUPER framework is the rain/no-rain filter, which is based on a categorical triple collocation.

The performance of the satellite rainfall products was validated with TAHMO data on a pixel-to-point basis, both considering daily and monthly resolutions. In general, SUPER (SM-ER) outperformed both SUPER (CH-IM) and the individual products in terms of the considered performance metrics. Comparing the individual products, SM2Rain performs best in terms of RMSE and bias, while IMERG exhibits the highest Kling–Gupta efficiency and Spearman's correlation with daily rainfall data from TAHMO. When aggregating to a monthly resolution, however, CHIRPS has the highest mean performance among the individual products, followed by both IMERG and SM2Rain, and ERA5. Furthermore, the detection of rainy season onset was validated, with CHIRPS and SUPER (ER–SM) showing the highest overall detection. SM2Rain is more useful for detecting the rainy season offset timing in dry areas.

The ultimate goal of this study was to develop a workflow to enhance the accuracy of satellite rainfall measurements in East Africa for potential operational applications. The results of this study showed an improvement in rainfall estimates compared to existing state-of-the-art products, providing a potential for improved satellite rainfall estimates in East Africa. In addition, important insights were obtained regarding the value of standard metrics. By comparing the time series from satellite and TAHMO data, we demonstrated that it is crucial not only to validate using performance metrics directly with the stations but also to understand how the satellite products relate to ground-level rainfall dynamics.

Data availability statement. TAHMO data are accessed through the TAHMO portal (https://portal.tahmo.org/), with data requests available via info@tahmo.org. ASCAT data were retrieved from the FTP server at ftphsaf.meteoam.it. IMERG GPM data were accessed from https://gpm1.gesdisc.eosdis.nasa.gov/opendap/. CHIRPS and CHPClim datasets were downloaded from https://data.chc.ucsb.edu/products/.

ERA5 data were accessed using the Climate Data Store (CDS) API (https://cds.climate.copernicus.eu/how-to-api). Python scripts and Jupyter notebooks used for generating SM2Rain and SUPER rainfall estimates, along with the analysis, are available in GitHub (Hoogelander 2024).

APPENDIX A

Derivation Least Squares Solution

In SUPER, the error variances of the rainfall products are estimated by using the following linear system:

$$\overline{y} = \begin{bmatrix} \sigma_a^2 \\ \sigma_b^2 \\ \sigma_c^2 \\ \sigma_d^2 \\ \sigma_{ab} \\ \sigma_{ac}\sigma_{ad}/\sigma_{cd} \\ \sigma_{bc}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ac}\sigma_{cd}/\sigma_{ad} \\ \sigma_{bc}\sigma_{cd}/\sigma_{bd} \\ \sigma_{ad}\sigma_{cd}/\sigma_{ac} \\ \sigma_{bd}\sigma_{cd}/\sigma_{bc} \\ \sigma_{ac}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ad}\sigma_{bc}/\sigma_{cd} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \overline{x} = \begin{bmatrix} \beta_a^2\sigma_\Theta^2 \\ \beta_b^2\sigma_\Theta^2 \\ \beta_c^2\sigma_\Theta^2 \\ \beta_d^2\sigma_\Theta^2 \\ \beta_a\beta_b\sigma_\Theta^2 \\ \sigma_{\varepsilon a}^2 \\ \sigma_{\varepsilon b}^2 \\ \sigma_{\varepsilon c}^2 \\ \sigma_{\varepsilon d}^2 \\ \sigma_{\varepsilon a}\sigma_{\varepsilon b} \end{bmatrix}, \quad \overline{y} = \mathbf{A}\overline{x}.$$

In this system, σ_i^2 is the dataset variance of product i, σ_{ij}^2 is the dataset covariance between products i and j, $\beta_i^2\sigma_\Theta$ is the signal variance of product i, $\sigma_{\epsilon i}^2$ is the error variance of product i, and $\sigma_{\epsilon i}\sigma_{\epsilon j}^2$ is the interproduct error covariance between i and j.

To get the error variances and interproduct covariance, the system can be solved using the common least squares solution:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\overline{y}, \tag{A1}$$

which gives

$$\begin{bmatrix} \beta_a^2\sigma_\Theta^2 \\ \beta_b^2\sigma_\Theta^2 \\ \beta_c^2\sigma_\Theta^2 \\ \beta_d^2\sigma_\Theta^2 \\ \beta_a\beta_b\sigma_\Theta^2 \\ \sigma_{\varepsilon a}^2 \\ \sigma_{\varepsilon b}^2 \\ \sigma_{\varepsilon c}^2 \\ \sigma_{\varepsilon d}^2 \\ \sigma_{\varepsilon a}\sigma_{\varepsilon b} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -0.5 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & -0.5 \end{bmatrix} \begin{bmatrix} \sigma_a^2 \\ \sigma_b^2 \\ \sigma_c^2 \\ \sigma_d^2 \\ \sigma_{ab} \\ \sigma_{ac}\sigma_{ad}/\sigma_{cd} \\ \sigma_{bc}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ac}\sigma_{cd}/\sigma_{ad} \\ \sigma_{bc}\sigma_{cd}/\sigma_{bd} \\ \sigma_{ad}\sigma_{cd}/\sigma_{ac} \\ \sigma_{bd}\sigma_{cd}/\sigma_{bc} \\ \sigma_{ac}\sigma_{bd}/\sigma_{cd} \\ \sigma_{ad}\sigma_{bc}/\sigma_{cd} \end{bmatrix}$$

APPENDIX B

Effect of Different ASCAT Datasets on SM2Rain Performance and SUPER Analysis

To further assess the robustness of the SUPER framework with the selected rainfall products, an additional evaluation was conducted using the Metop ASCAT Surface Soil Moisture Climate Data Record v7 (H119/H120) product provided by H SAF. By using H120, the analysis minimizes potential dependencies between soil moisture and precipitation datasets, ensuring a theoretically more independent error characterization compared to using SM2Rain derived from H26 in SUPER.
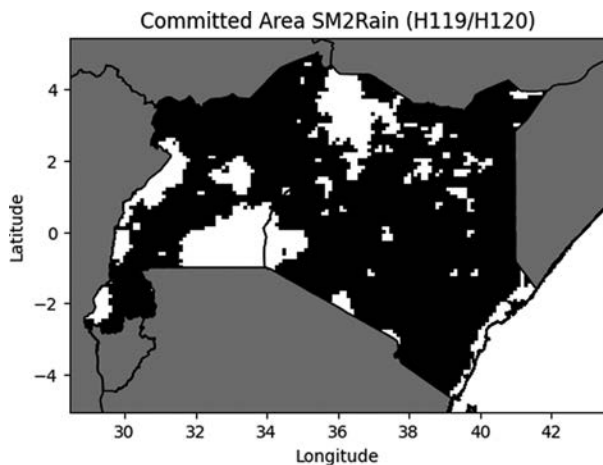
FIG. B1. Committed area of ASCAT H119/H120 products. Next to masking out water bodies, low-confidence areas are also excluded, mainly located in areas close to water bodies, dense vegetation, and complex topography.

a. ASCAT datasets

The H119/H120 dataset provides a long-term soil moisture time series derived from ASCAT observations with a 12.5-km sampling resolution. Unlike the H26 dataset, H119/H120 is not influenced by the HTESSEL land surface model, making it less error-dependent with ERA5 and therefore theoretically more suitable for QC-merging in SUPER. The dataset extends from 2007 to the present, but its irregular latency limits its suitability for operational rainfall products.

Unlike H26, which is organized as a time series on a regularly spaced grid, H119/H120 data are structured in grid cells with varying observation locations. Due to the ASCAT swath path, these locations are irregularly distributed in both space and time. To ensure a consistent spatial and temporal resolution, we applied linear interpolation to generate a regular grid with evenly spaced observations. In some cases, gaps in the data required interpolation over more than 1 day in larger areas, which may introduce uncertainties and affect data quality. In addition to the water body mask, we applied the committed area mask, an extra variable in the H119/H120 dataset, to the soil moisture data. The committed area defines a restricted geographical region with high confidence in soil moisture retrieval. We also applied the same masking approach to other products used in SUPER to ensure consistent merging and validation. Figure B1 shows the regions

considered for the generation of SM2Rain with H119/H120, and further analysis.

After the preprocessing steps of H119/H120, we generated SM2Rain maps using the same steps as described in section 3a and merged this with the other products using the SUPER framework.

b. Product performances

We validated SM2Rain using both H119/H120 and H26 soil moisture products. SUPER was generated assuming both IMERG-CHIRPS and SM2Rain-ERA5 error cross correlations, similar to the approach outlined in section 3b. For a rough comparison, we only evaluate the validation metrics over the entire period from 2018 to 2022, considering only stations with more than 80% of data available during this time. However, we solely used the stations located within the committed area of H119/H120. The validation results of all different SM2Rain and SUPER configurations for both the daily and monthly resolutions can be found in Tables B1 and B2.

We found that SM2Rain (H120) shows slightly lower performance compared to SM2Rain (H26) at the daily scale, with higher RMSE and lower KGE values. The bias in SM2Rain (H120) is also higher than in H26 at this resolution, indicating some differences in the accuracy of the rainfall estimation. Despite these differences, the Spearman correlation ρ for both datasets is comparable, suggesting that the relationship between observed and predicted rainfall is similar for both H120 and H26 at the daily level.

At the monthly scale, the performance differences between SM2Rain (H120) and SM2Rain (H26) are negligible, with both configurations showing similar results in terms of RMSE, KGE, and bias. These findings indicate that, at coarser temporal resolutions, the two datasets produce nearly identical performance in rainfall estimation.

c. SUPER analysis

For the SUPER analysis, we again derived the QC-based product-truth Pearson correlations for both SUPER configurations and compared these with the correlations with TAHMO stations. The results of this analysis can be found in Table B3. When comparing the configurations in Table B3 with Table 4, we see that the QC-derived product-truth correlations using the H120 configuration generally align more closely with the correlations found with TAHMO. Between configurations (IM-CH) and (SM-ER) with H120, (CH-IM) shows a relatively

TABLE B1. Performance metrics of SM2Rain and SUPER using H26 and H119/H120 in SM2Rain on a daily resolution, averaged over all stations in the committed area and the entire 5-yr period. The ρ is the Spearman correlation.

	RMSE (mm day^{-1})	KGE (—)	Bias (mm day^{-1})	ρ
SM2Rain (H26)	10.25	−0.05	−0.26	0.47
SM2Rain (H120)	11.91	−0.14	1.60	0.51
SUPER (IM-CH) (H26)	11.19	0.16	−0.23	0.54
SUPER (IM-CH) (H120)	11.40	0.20	−0.40	0.55
SUPER (SM-ER) (H26)	11.58	0.26	−0.16	0.55
SUPER (SM-ER) (H120)	12.07	0.27	−0.20	0.55

TABLE B2. Performance metrics of SM2Rain and SUPER using H26 and H119/H120 in SM2Rain on a monthly resolution, averaged over all stations in the committed area and the entire 5-yr period. The ρ is the Spearman correlation.

	RMSE (mm day^{-1})	KGE (—)	Bias (mm day^{-1})	ρ
SM2Rain (H26)	2.12	0.31	−0.23	0.59
SM2Rain (H120)	2.17	0.34	−0.04	0.56
SUPER (IM-CH) (H26)	2.03	0.42	−0.19	0.65
SUPER (IM-CH) (H120)	2.03	0.42	−0.35	0.66
SUPER (SM-ER) (H26)	2.05	0.43	−0.11	0.66
SUPER (SM-ER) (H120)	2.05	0.43	−0.15	0.66

TABLE B3. Mean QC-based and traditionally based product-truth Pearson's correlations derived from QC analysis, with ASCAT products H119/H120 used in SM2Rain.

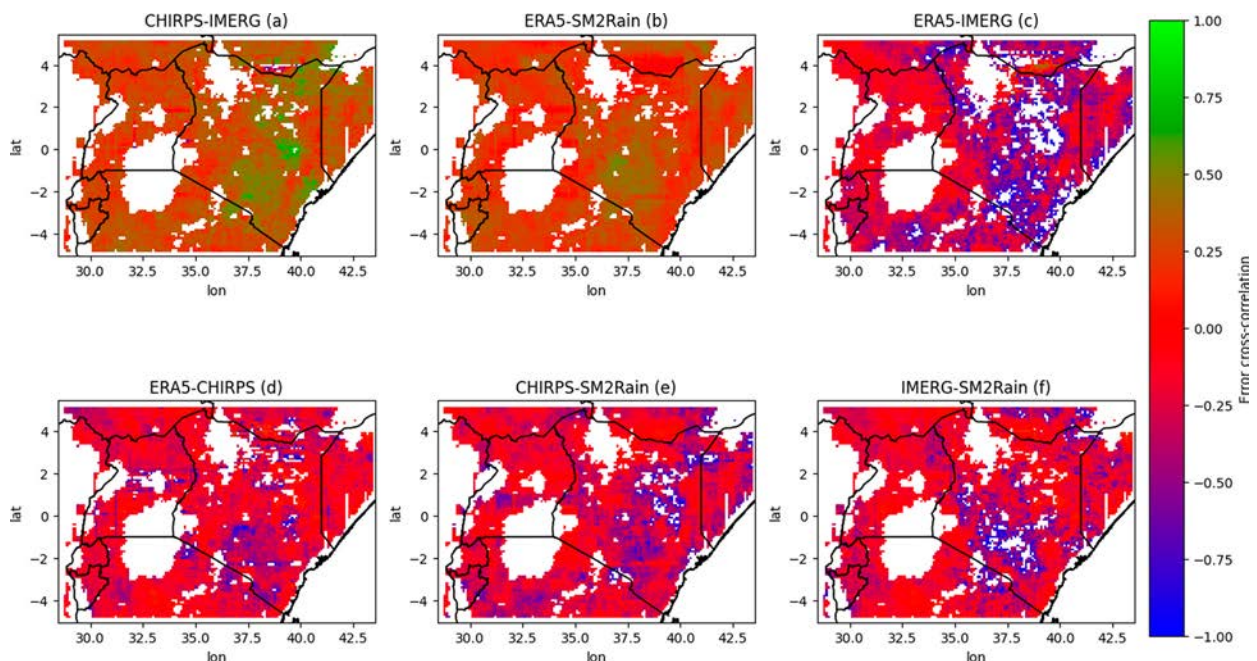|	CHIRPS–IMERG (H120)	ERA5–SM2Rain (H120)	TAHMO
SM2Rain (H120)	0.68	0.70	0.65
ERA5	0.70	0.77	0.65
IMERG	0.61	0.62	0.73
CHIRPS	0.55	0.61	0.67

FIG. B2. QC-based ECCs between products using different SUPER configurations, with ASCAT products H119/H120 used in SM2Rain. Pixels having |ECC| > 1 are excluded from the maps.

closer match with the TAHMO correlations, although a significant discrepancy remains with IMERG and CHIRPS. However, this discrepancy is now also present in configuration (SM-ER).

Figure B2 shows the QC-based ECC of the different products on a pixel-to-pixel basis for all possible QC configurations. Similar to Fig. 6, the spatial distribution of ECC values varies across different product comparisons, with only configurations (Figs. 6a and b) exhibiting realistic ECCs across the entire area. However, combinations (Figs. 6c–f) now retain

relatively more realistic data, as fewer pixels are filtered out due to physically impossible ECC values. The highest ECCs in (CH-IM) and (SM-ER) are concentrated in southeast Kenya, with (CH-IM) showing slightly larger values than (SM-ER). While these results align more closely with the expected ECC relationship between IMERG and CHIRPS, the overall difference between (CH-IM) and (SM-ER) remains small. By directly comparing configuration (CH-IM) with SM2Rain across both H26 (Fig. 6) and H119/H120 (Fig. B2), the differences
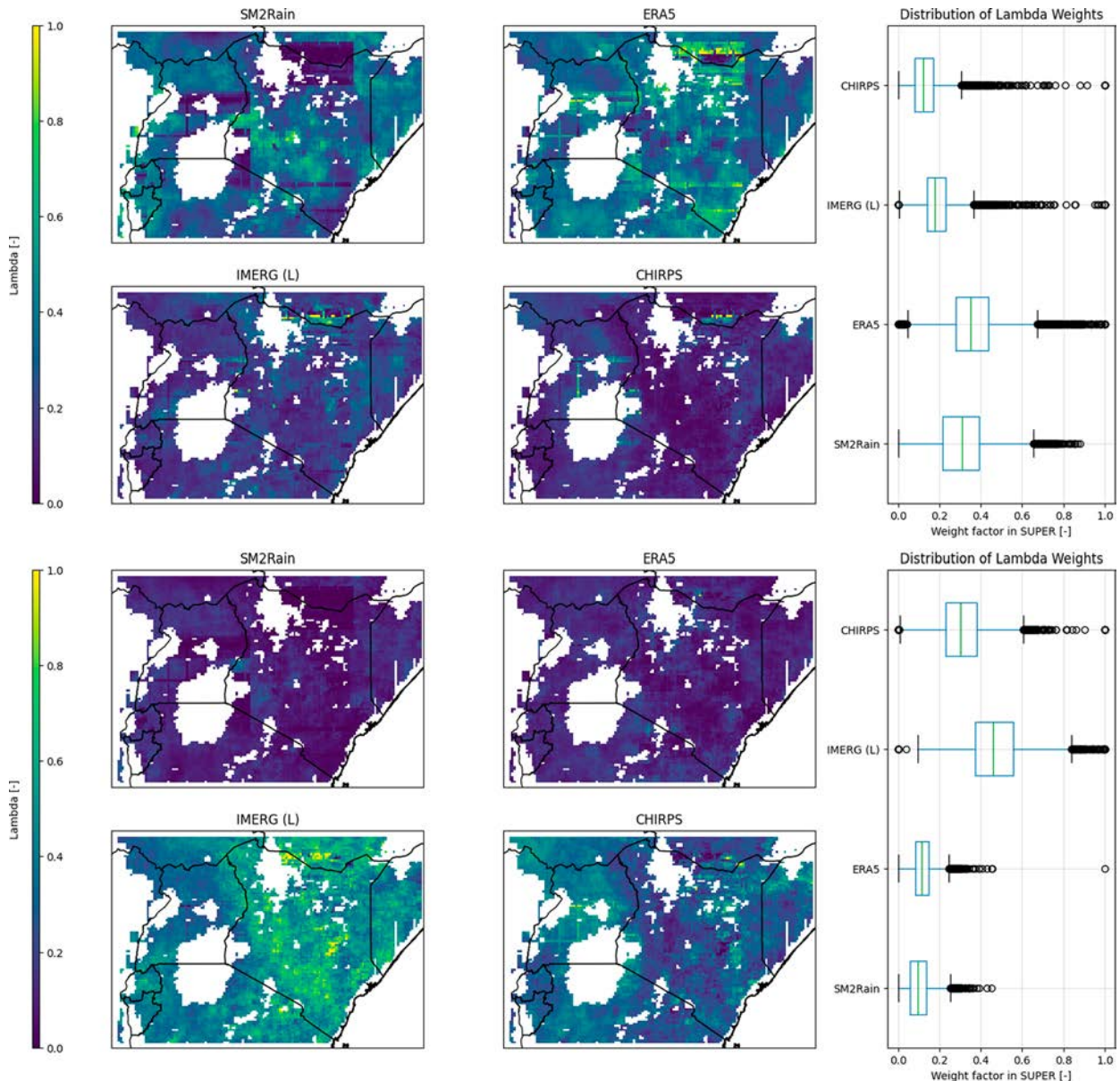
FIG. B3. QC-based ECCs between products using different SUPER configurations, with ASCAT products H119/H120 used in SM2Rain. Pixels having |ECC| > 1 are excluded from the maps.

remain small. However, for configuration (SM-ER), the differences are more pronounced. Specifically, the ECC values in Fig. 6 for configuration (SM-ER) are generally higher than those in Fig. B2. This suggests that the results in Fig. 6 are influenced by the choice of using H26 in SM2Rain, potentially introducing a bias in the ECC estimates.

Last, the lambda weights derived from the QC analysis with both (CH-IM) and (SM-ER) configurations are presented in Fig. B3. In general, the distribution of lambda weights among the products follows similar patterns as found in Fig. 7, although the contributions per product are less pronounced. For example, in configuration (CH-IM), ERA5 and SM2Rain have the highest contribution, but their lambda weights are generally lower

than in Fig. 7, and the same trend is observed for configuration (SM-ER). Additionally, in some areas, there is a low spatial coherence in the lambda weights, particularly for SM2Rain, which may be due to the interpolation of H119/H120.

APPENDIX C

Rainy Season Analysis

For the rainy season analysis, we used the flexible rainy season definition as proposed in the study of Seregina et al. (2019). Here, we outline the steps taken from this study to apply this method to the data used in this study.
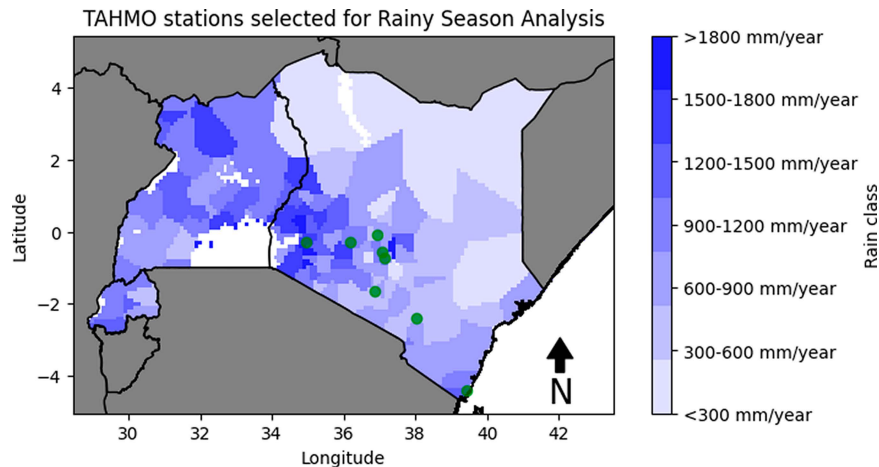
FIG. C1. Locations of the eight selected stations used for the rainy season analysis. Three stations fall under rain class 2, one station under rain class 3, and four stations under rain class 4.

Since a long-term climatology is not available for TAHMO stations, and the period 2018–22 may be too short, we included the years 2016 and 2017 and considered stations with more than 80% data availability between 2016 and 2022 to derive the most reliable approximation of the threshold possible with TAHMO data. This resulted in a total of eight stations, whose locations are indicated in the map in Fig. C1. Among these, three stations fall under rain class 2, one station under rain class 3, and four stations

under rain class 4. For an additional wet/dry analysis, we categorized stations with a rain class below 4 as "dry" and those with rain class 4 as "wet." Prior to computing the long-term pentad average over the years, a Lanczos low-pass filter is applied to the pentad rainfall average of every station data, with a cutoff period of six pentads. From this climatology, the threshold of the start of the rainy season is determined using the standard normal homogeneity test (SNHT) from Alexandersson (1986). The offset of the rainy
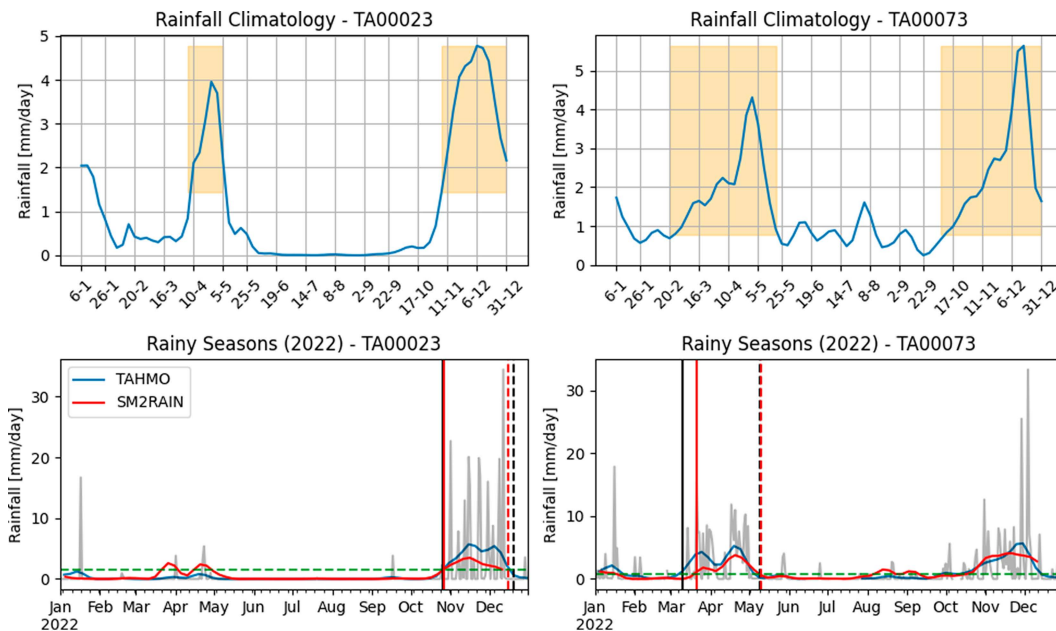

FIG. C2. Rainy season analysis for TAHMO stations TA00023 and TA00073. The top two graphs show the climatologies based on data from 2016 to 2022, with yellow boxes indicating the rainy season periods. The bottom two graphs display the filtered pentad rainfall data for both SM2Rain and TAHMO, along with daily TAHMO rainfall. Solid vertical lines represent the offset of the rainfall season, with black for TAHMO and red for SM2Rain, while the vertical dashed line indicates the cessation of the rainy season. The horizontal dashed line indicates the rainy season threshold.

season is defined as the first pentad of five consecutive pentads in which the filtered rainfall data exceed the threshold. The cessation is identified as the first pentad in which rainfall falls below the threshold. An example of the climatology, the climatological rainy season, and 2 years of rainfall data with identified rainy seasons of TAHMO stations TA00023 and TA00073 and SM2Rain is presented in Fig. C2.

For validation, we focus again on the period 2018–22. During this time, the stations detected a total of 38 rainy seasons. We assess whether the onset of the rainy season, as identified in different rainfall products, aligns with the onset determined by TAHMO, using buffers of 2, 5, and 15 days. Additionally, we analyze the results separately for the four drier stations and four wetter stations by categorizing them based on their rainfall class.

REFERENCES

Ageet, S., A. H. Fink, M. Maranan, J. E. Diem, J. Hartter, A. L. Ssali, and P. Ayabagabo, 2022: Validation of satellite rainfall estimates over equatorial East Africa. *J. Hydrometeor.*, **23**, 129–151, https://doi.org/10.1175/JHM-D-21-0145.1.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, **6**, 661–675, https://doi.org/10.1002/joc.3370060607.

Barrett, B. W., E. Dwyer, and P. Whelan, 2009: Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sens.*, **1**, 210–242, https://doi.org/10.3390/rs1030210.

Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

Boluwade, A., 2020: Remote sensed-based rainfall estimations over the East and West Africa regions for disaster risk management. *ISPRS J. Photogramm. Remote Sens.*, **167**, 305–320, https://doi.org/10.1016/j.isprsjprs.2020.07.015.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, and C. Albergel, 2013: Scaling and filtering approaches for the use of satellite soil moisture observations. *Remote Sensing of Energy Fluxes and Soil Moisture Content*, G. P. Petropoulos, Ed., CRC Press, 411–426.

——, and Coauthors, 2019: SM2RAIN–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations. *Earth Syst. Sci. Data*, **11**, 1583–1601, https://doi.org/10.5194/essd-11-1583-2019.

Chen, F., W. T. Crow, L. Ciabatta, P. Filippucci, G. Panegrossi, A. C. Marra, S. Puca, and C. Massari, 2021: Enhanced large-scale validation of satellite-based land rainfall products. *J. Hydrometeor.*, **22**, 245–257, https://doi.org/10.1175/JHM-D-20-0056.1.

Dembo, R. S., and T. Steihaug, 1983: Truncated-Newton algorithms for large-scale unconstrained optimization. *Math. Program.*, **26**, 190–212, https://doi.org/10.1007/BF02592055.

Dezfuli, A. K., C. M. Ichoku, G. J. Huffman, K. I. Mohr, J. S. Selker, N. van de Giesen, R. Hochreutener, and F. O. Annor, 2017: Validation of IMERG precipitation in Africa. *J. Hydrometeor.*, **18**, 2817–2825, https://doi.org/10.1175/JHM-D-17-0139.1.

Dinku, T., 2019: Challenges with availability and quality of climate data in Africa. *Extreme Hydrology and Climate Variability*, A. M. Melesse, W. Abtew, and G. Senay, Eds., Elsevier, 71–80, https://doi.org/10.1016/B978-0-12-815998-9.00007-5.

Dong, J., W. T. Crow, and R. Reichle, 2020: Improving rain/no-rain detection skill by merging precipitation estimates from different sources. *J. Hydrometeor.*, **21**, 2419–2429, https://doi.org/10.1175/JHM-D-20-0097.1.

——, and Coauthors, 2022: Statistical Uncertainty Analysis-Based Precipitation Merging (SUPER): A new framework for improved global precipitation estimation. *Remote Sens. Environ.*, **283**, 113299, https://doi.org/10.1016/j.rse.2022.113299.

EUMETSAT H SAF, 2021a: Product User Manual (PUM) Metop ASCAT surface soil moisture climate data record v7 12.5 km sampling (H119) and Extension (H120), v0.3. Eumetsat H SAF Tech. Doc. HSAF/CDOP3/PUM/, 33 pp., https://hsaf.meteoam.it/Products/Detail?prod=H119.

——, 2021b: Product User Manual (PUM) of H26 soil wetness index in the roots region by ASCAT soil moisture assimilation. Eumetsat H SAF Tech. Doc. SAF/HSAF/CDOP3/PUM/, 25 pp., https://hsaf.meteoam.it/Products/Detail?prod=H26.

——, 2021c: Algorithm Theoretical Baseline Document (ATBD) H26. Eumetsat H SAF Tech. Doc. SAF/HSAF/CDOP3/ATBD/, 27 pp., https://hsaf.meteoam.it/CaseStudy/GetDocumentUserDocument?fileName=h26_atbd_v2.pdf&tipo=ATBD.

Fairbairn, D., P. de Rosnay, and P. A. Browne, 2019: The new stand-alone surface analysis at ECMWF: Implications for land–atmosphere DA coupling. *J. Hydrometeor.*, **20**, 2023–2042, https://doi.org/10.1175/JHM-D-19-0074.1.

Flitcroft, I. D., J. R. Milford, and G. Dugdale, 1989: Relating point to area average rainfall in semiarid West Africa and the implications for rainfall estimates derived from satellite data. *J. Appl. Meteor.*, **28**, 252–266, https://doi.org/10.1175/1520-0450(1989)028<0252:RPTAAR>2.0.CO;2.

Freitas, E. D. S., and Coauthors, 2020: The performance of the IMERG satellite-based product in identifying sub-daily rainfall events and their properties. *J. Hydrol.*, **589**, 125128, https://doi.org/10.1016/j.jhydrol.2020.125128.

Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak, 2015a: A global satellite-assisted precipitation climatology. *Earth Syst. Sci. Data*, **7**, 275–287, https://doi.org/10.5194/essd-7-275-2015.

——, and Coauthors, 2015b: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10.1038/sdata.2015.66.

Gruber, A., C.-H. Su, W. T. Crow, S. Zwieback, W. A. Dorigo, and W. Wagner, 2016: Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. *J. Geophys. Res. Atmos.*, **121**, 1208–1219, https://doi.org/10.1002/2015JD024027.

Hahn, S., W. Wagner, S. C. Steele-Dunne, M. Vreugdenhil, and T. Melzer, 2021: Improving ASCAT soil moisture retrievals with an enhanced spatially variable vegetation parameterization. *IEEE Trans. Geosci. Remote Sens.*, **59**, 8241–8256, https://doi.org/10.1109/TGRS.2020.3041340.

Hoogelander, V., 2024: vhoogelander-SM2Rain_TAHMO_SRP merging_EastAfrica. Zenodo, accessed 4 November 2024, https://doi.org/10.5281/zenodo.14035125.

Hu, X., and W. Yuan, 2021: Evaluation of ERA5 precipitation over the eastern periphery of the Tibetan Plateau from the perspective of regional rainfall events. *Int. J. Climatol.*, **41**, 2625–2637, https://doi.org/10.1002/joc.6980.

Huffman, G. J., D. T. Bolvin, R. Joyce, O. A. Kelley, E. J. Nelkin, J. Tan, D. C. Watters, and B. J. West, 2019: Integrated Multi-satellitE Retrievals For GPM (IMERG) technical documentation. IMERG Tech. Doc., 97 pp., https://gpm.nasa.gov/sites/default/files/2023-07/IMERG_TechnicalDocumentation_final_230713.pdf.

Indeje, M., F. H. M. Semazzi, and L. J. Ogallo, 2000: ENSO signals in East African rainfall seasons. *Int. J. Climatol.*, **20**, 19–46, https://doi.org/10.1002/(SICI)1097-0088(200001)20:1<19::AID-JOC449>3.0.CO;2-0.

Kilavi, M., and Coauthors, 2018: Extreme rainfall and flooding over central Kenya including Nairobi City during the long-rains season 2018: Causes, predictability, and potential for early warning and actions. *Atmosphere*, **9**, 472, https://doi.org/10.3390/atmos9120472.

Kimani, M. W., J. C. B. Hoedjes, and Z. Su, 2017: An assessment of satellite-derived rainfall products relative to ground observations over East Africa. *Remote Sens.*, **9**, 430, https://doi.org/10.3390/rs9050430.

Kubota, T., and Coauthors, 2017: Recent progress in Global Satellite Mapping of Precipitation (GSMAP) product. *2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Fort Worth, TX, Institute of Electrical and Electronics Engineers, 2712–2715, https://doi.org/10.1109/IGARSS.2017.8127556.

Lebel, T., H. Sauvageot, M. Hoepffner, M. Desbois, B. Guillot, and P. Hubert, 1992: Rainfall estimation in the Sahel: The EPSAT-NIGER experiment. *Hydrol. Sci. J.*, **37**, 201–215, https://doi.org/10.1080/02626669209492582.

Le Coz, C., and N. van de Giesen, 2020: Comparison of rainfall products over sub-Saharan Africa. *J. Hydrometeor.*, **21**, 553–596, https://doi.org/10.1175/JHM-D-18-0256.1.

Liu, R., X. Zhang, W. Wang, Y. Wang, H. Liu, M. Ma, and G. Tang, 2024: Global-scale ERA5 product precipitation and temperature evaluation. *Ecol. Indic.*, **166**, 112481, https://doi.org/10.1016/j.ecolind.2024.112481.

Macharia, D., K. Fankhauser, J. S. Selker, J. C. Neff, and E. A. Thomas, 2022: Validation and intercomparison of satellite-based rainfall products over Africa with TAHMO in situ rainfall observations. *J. Hydrometeor.*, **23**, 1131–1154, https://doi.org/10.1175/JHM-D-21-0161.1.

MacLeod, D. A., and Coauthors, 2021: Drivers and subseasonal predictability of heavy rainfall in equatorial East Africa and relationship with flood risk. *J. Hydrometeor.*, **22**, 887–903, https://doi.org/10.1175/JHM-D-20-0211.1.

Maggioni, V., C. Massari, and C. Kidd, 2022: Errors and uncertainties associated with quasiglobal satellite precipitation products. *Precipitation Science*, S. Michaelides, Ed., Elsevier, 377–390, https://doi.org/10.1016/B978-0-12-822973-6.00023-8.

Maidment, R. I., and Coauthors, 2017: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Sci. Data*, **4**, 170063, https://doi.org/10.1038/sdata.2017.63.

Mekonnen, K., and Coauthors, 2023: Accuracy of satellite and reanalysis rainfall estimates over Africa: A multi-scale assessment of eight products for continental applications. *J. Hydrol.*, **49**, 101514, https://doi.org/10.1016/j.ejrh.2023.101514.

Muita, R., and Coauthors, 2021: Towards increasing data availability for meteorological services: Inter-comparison of meteorological data from a synoptic weather station and two automatic weather stations in Kenya. *Amer. J. Climate Change*, **10**, 300–316, https://doi.org/10.4236/ajcc.2021.103014.

Muñoz Sabater, J., and Coauthors, 2021: Era5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, **13**, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021.

Ngetich, K. F., M. Mucheru-Muna, J. N. Mugwe, C. A. Shisanya, J. Diels, and D. N. Mugendi, 2014: Length of growing season, rainfall temporal distribution, onset and cessation dates in the Kenyan highlands. *Agric. For. Meteor.*, **188**, 24–32, https://doi.org/10.1016/j.agrformet.2013.12.011.

Palmer, P. I., and Coauthors, 2023: Drivers and impacts of eastern African rainfall variability. *Nat. Rev. Earth Environ.*, **4**, 254–270, https://doi.org/10.1038/s43017-023-00397-x.

Satalino, G., A. Balenzano, F. Mattia, and M. W. J. Davidson, 2014: C-band SAR data for mapping crops dominated by surface or volume scattering. *IEEE Geosci. Remote Sens. Lett.*, **11**, 384–388, https://doi.org/10.1109/LGRS.2013.2263034.

Schunke, J., P. Laux, J. Bliefernicht, M. Waongo, W. Sawadogo, and H. Kunstmann, 2021: Exploring the potential of the cost-efficient TAHMO observation data for hydro-meteorological applications in sub-Saharan Africa. *Water*, **13**, 3308, https://doi.org/10.3390/w13223308.

Seregina, L. S., A. H. Fink, R. van der Linden, N. A. Elagib, and J. G. Pinto, 2019: A new and flexible rainy season definition: Validation for the Greater Horn of Africa and application to rainfall trends. *Int. J. Climatol.*, **39**, 989–1012, https://doi.org/10.1002/joc.5856.

Skaugen, T., 1997: Classification of rainfall into small- and large-scale events by statistical pattern recognition. *J. Hydrol.*, **200**, 40–57, https://doi.org/10.1016/S0022-1694(97)00003-6.

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, **103**, 7755–7766, https://doi.org/10.1029/97JC03180.

Tarnavsky, E., D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and F. Kayitakire, 2014: Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *J. Appl. Meteor. Climatol.*, **53**, 2805–2822, https://doi.org/10.1175/JAMC-D-14-0016.1.

Taupin, J.-D., 1997: Caractérisation de la variabilité spatiale des pluies aux échelles inférieures au kilomètre en région semi-aride (région de Niamey, Niger). *C. R. Acad. Sci.*, **325IIA**, 251–256, https://doi.org/10.1016/S1251-8050(97)88297-3.

Tsamalis, C., 2022: Clarifications on the equations and the sample number in triple collocation analysis using SST observations. *Remote Sens. Environ.*, **272**, 112936, https://doi.org/10.1016/j.rse.2022.112936.

van de Giesen, N., R. Hut, and J. Selker, 2014: The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdiscip. Rev.: Water*, **1**, 341–348, https://doi.org/10.1002/wat2.1034.

Xia, X., Y. Liu, W. Jing, and L. Yao, 2021: Assessment of four satellite-based precipitation products over the Pearl River basin, China. *IEEE Access*, **9**, 97 729–97 746, https://doi.org/10.1109/ACCESS.2021.3095239.