Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: www.tandfonline.com/journals/thsj20

Comparison and validation of spatial reference evapotranspiration datasets over Africa

Suzan Dehati, Bich Ngoc Tran, Poolad Karimi & Marloes Mul

To cite this article: Suzan Dehati, Bich Ngoc Tran, Poolad Karimi & Marloes Mul (27 Jan 2026): Comparison and validation of spatial reference evapotranspiration datasets over Africa, Hydrological Sciences Journal, DOI: 10.1080/02626667.2025.2600684

To link to this article: https://doi.org/10.1080/02626667.2025.2600684

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2025.2600684

TWENTY-FIRST CENTURY HYDROLOGICAL CHALLENGES AND OPPORTUNITIES IN AFRICA OPEN ACCESS Check for updates

Comparison and validation of spatial reference evapotranspiration datasets over Africa

Suzan Dehati [a,b], Bich Ngoc Tran [a,c], Poolad Karimi [a] and Marloes Mul [a]

[a]Land and Water Management Department, IHE Delft Institute for Water Education, Delft, The Netherlands; [b]Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands; [c]Department of Water Management, Delft University of Technology, Delft, The Netherlands

ABSTRACT

Reference evapotranspiration (ET_0) is an important variable for water resources management and agricultural planning. Some regions, including Africa lack sufficient in-situ meteorological measurements to represent the climatic conditions. Open-access Global ET_0 data sets present a viable alternative that could potentially fill the gap. This study compares eight spatial ET_0 data sets against ET_0 estimated from 165 weather stations across Africa. Performance was assessed using statistical metrics, including R^2, Bias, RMSE, and RBias. Findings reveal that high-resolution data sets align better with in-situ data in temperate and tropical climates compared to low-resolution data sets. Results for arid regions appear to show low performance for all data sets, but results are less certain due to the availability of stations in this climate. This study also reveals that the input data contribute to 60–70% of the variability between data sets, with the remainder contributed by different model implementation, indicating the importance of good quality of input data.

ARTICLE HISTORY
Received 13 December 2024
Accepted 14 November 2025

EDITOR
S. Archfield

GUEST EDITOR
M. Dembélé

KEYWORDS
reference evapotranspiration; FAO-56 Penman–Monteith; Slob–de Bruin method; Africa; uncertainty; validation; TAHMO

1 Introduction

Efficient irrigation scheduling is essential for improving agricultural water productivity, particularly in regions where rainfall is highly variable and often insufficient (Singh *et al.* 2010, Greaves and Wang 2017, Attia *et al.* 2021). A key input of irrigation planning is the estimation of crop water requirement (CWR) (Allen *et al.* 1998, Aghdasi 2010, Ali 2010), which represents the amount of water needed to compensate for the evapotranspiration losses from crops. One widely accepted approach for calculating CWR is through the use of reference evapotranspiration (ET_0), which quantifies the atmospheric water demand from a well-watered reference surface, typically grass under optimal growing conditions (Allen *et al.* 1998). ET_0 forms the basis for deriving crop evapotranspiration (ETc) by applying crop coefficients that account for vegetation type, growth stage, and local conditions (Mehta and Pandey 2015, FAO 2020).

ET_0 is commonly estimated using the FAO-56 Penman–Monteith (PM) equation, which combines meteorological variables such as temperature, radiation, wind speed, and humidity (Allen *et al.* 1998). However, the effective use of this method depends on the availability of weather data. In many parts of Africa, the density of weather stations is below the minimum required by the World Meteorological Organization (WMO), which leads to data scarcity that severely limits the application of traditional ET_0 estimation techniques (Dinku 2019). Remote sensing or reanalysis meteorological data sets can be used to estimate ET_0 spatially.

There are many global open-access ET_0 data sets available including MSG (Trigo *et al.* 2018), Bristol (Singer *et al.* 2021), AQUASTAT (FAO 2021), WAPORv2 (FAO 2020), WaPORv3 (FAO 2024), and FEWSNET (Senay *et al.* 2008). However, the accuracy of these data sets varies significantly across different regions and climates in Africa. For example, a recent evaluation of the NOAA/MERRA-2 reanalysis-based ET_0 data set in Southern Africa found consistent overestimation compared to ground-based station data, with bias ranging from +9% up to +90% (average +36%) compared to ground estimates where the largest overestimations occurred in arid areas of Namibia and Botswana (Hobbins *et al.* 2023). Similarly, MSG-derived ET_0 has been shown to underestimate ET_0 in semi-arid environments due to localized advection effects, while performing well over larger irrigated or homogeneous areas (Trigo *et al.* 2018). This variability arises from differences in input data, model formulation, spatial resolution, and local environmental conditions. Some studies have evaluated the accuracy of ET_0 estimates from different reanalyses and remote sensing data (Paredes *et al.* 2021), but few have conducted comparative validation against in-situ observations at a continental scale particularly for Africa, where in-situ weather data are very limited and water demands are high (Pereira *et al.* 2015, Mahmoud and Gan 2019, Jovanovic *et al.* 2020).

CONTACT Suzan Dehati s.dehati@un-ihe.org Land and Water Management Department, IHE Delft Institute for Water Education, Westvest 7, Delft 2611 AX, The Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02626667.2025.2600684

Recent studies have proposed machine learning approaches to improve ET_0 estimation in data-scarce, heterogeneous environments. These methods capture nonlinear relationships in meteorological inputs and outperform traditional models in semi-arid climates. For example, Yonaba *et al.* (2024) tested 12 algorithms at nine stations in Burkina Faso (reporting R^2 = 0.99 and RMSE = 0.05 mm day^{-1} for XGBoost) and used SHAP to quantify each predictor's contribution to ET_0 prediction across different meteorological conditions. However, these station-trained models rely on local observations and have limited spatial transferability beyond the region represented by the training data. As a result, there is still a lack of clarity on which open access spatial ET_0 data sets are most suitable for specific climate zones, particularly in arid and semi-arid regions where local data are limited. This study aims to address these gaps by comparing and evaluating the performance of multiple ET_0 data sets across Africa against in-situ data from 165 stations. By analysing the influence of factors such as model choice, parameterization, spatial resolution, station characteristics, and input data, the study aims to provide a comprehensive overview of the factors that influence the ET_0 accuracy. Furthermore, the study examines the role of pixel size and grid resolution in the point to pixel validation process. Ultimately, the goal is to provide clear recommendations on the most suitable ET_0 data sets for different climates and elevations in Africa.

2 Data

2.1 ET_0 data sets

Eight spatial ET_0 data sets, with daily data for a five-year period from 01/01/2018 to 31/12/2022, were used for comparison with the in-situ ET_0 data. These include MSG (Trigo *et al.* 2018), Bristol (Singer *et al.* 2021), AQUASTAT (FAO 2021), WAPORv2 (FAO 2020), WaPORv3 (FAO 2024), and FEWSNET (Senay *et al.* 2008). Two additional reanalysis data sets, ET_0-ERA5 and ET_0-GEOS5, were produced as part of this research (Table 1). Although the majority of the ET_0 data sets provide global coverage, WaPORv2, ET_0-ERA5, and ET_0-GEOS5 are specifically available for Africa and the Near East.

Seven of the selected data sets are based on the FAO-56 PM method, while one, the MSG ET_0 data set, is based on the Slob–de Bruin (SdB) method (Trigo *et al.* 2018). The WaPORv2, WaPORv3, and AQUASTAT data sets produce daily ET_0 values representing in mm day^{-1}, based on meteorological and solar radiation data, following the FAO-56 procedure for a well-watered grass field with an albedo of 0.23. These data sets were generated using multiple sources of input data with different spatial resolution (Table 1 and see Supplementary material, Table S1). The key methodological differences are the empirical equations to estimate actual vapor pressure and calibration parameters for longwave radiation estimation (see Supplementary material, Table S2).

To evaluate the impact of input data alone, we calculated daily ET_0 using only the fifth-generation ECMWF reanalysis (ERA5) and the Goddard Earth Observing System version 5 (GEOS5) data sets. These data sets were selected because they have been used as input data for other ET_0 data sets; for instance, ERA5 is used in Bristol and AQUASTAT and GEOS5 in WaPORv2 and WaPORv3. The ET_0-ERA5 and ET_0-GEOS5 data sets were calculated from daily air temperature, atmospheric pressure, wind speed, vapor pressure, and solar radiation, following the same empirical equations and parameters of AQUASTAT (see Supplementary material, Table S2). Both ET_0-ERA5 and ET_0-GEOS5 are derived from reanalysis products that are known to have regional biases in near-surface radiation, temperature, and wind fields, which can propagate into ET_0 estimates. These limitations and their spatial patterns are discussed in detail in a companion study (Tran *et al.* 2025). Here, we include ET_0-ERA5 and ET_0-GEOS5 mainly as reference data sets to allow comparison with other open-access ET_0 products.

The Bristol Potential Evapotranspiration is derived from hourly climate variables from ERA5-Land, covering the period from 1981 to the present with a spatial resolution of 10 km over global land areas. Although termed potential evapotranspiration, the data set used the FAO-56 PM equation for grass as reference crop (Singer *et al.* 2021). Similarly, the FEWSNET Global Potential Evapotranspiration data sets calculated at 6-hour time steps (also available daily), with 100 km resolution, are labelled as potential evapotranspiration but use the same FAO-56 PM method using hypothetical grass with an albedo of 0.23. Therefore, both data sets were considered in the study.

Table 1. Characteristics of ET_0 data sets.

Data set	Temporal coverage	Spatial coverage	Temporal resolution	Spatial resolution	Input data source	Method	Reference
MSG	2017–present	−81° to 81°, −79° to 79°	Daily	5 km	MSG (radiation), ECMWF	SdB (de Bruin 1987)	Trigo *et al.* (2018)
Bristol	1981–present	Global	Hourly	10 km	ERA5-Land	FAO-56 (Allen *et al.* 1998)	Singer *et al.* (2021)
AQUASTAT	1979–present	Global	Daily	10 km	AgERA5, GMTED	FAO-56 (Allen *et al.* 1998)	FAO (2021)
WaPORv2	2009–present	Africa and Near East	Daily	20 km	GEOS5, SRTM, MSG (radiation)	FAO-56 (Allen *et al.* 1998)	FAO (2020)
ET_0-ERA5	2018–2022	Global	Daily	25 km	ERA5	FAO-56 (Allen *et al.* 1998)	This study
WaPORv3	2018–present	Global	Daily	30 km	GEOS5, GLO-90	FAO-56 (Allen *et al.* 1998)	FAO (2024)
ET_0-GEOS5	2018–2022	Africa and Near East	Daily	30 km	GEOS5	FAO-56 (Allen *et al.* 1998)	This study
FEWSNET	2008–present	Africa and Near East	6-hour	100 km	GDAS	FAO-56 (Allen *et al.* 1998)	Senay *et al.* (2008)

On the other hand, The MSG ET_0 data set is produced using a model developed by de Bruin *et al.* (2016), which estimates reference evapotranspiration from satellite-derived solar radiation using physical principles tailored to a well-watered grass surface. The method assumes that net radiation can be derived directly from shortwave radiation measured by SEVIRI/MSG, and then applies a simplified energy balance approach to estimate ET_0 for a reference crop defined by FAO-56 (Trigo *et al.* 2018).

2.2 In-situ data

The Trans-African Hydro-Meteorological Observatory (TAHMO) network, established in 2014, was designed to fill gaps in availability of weather data across sub-Saharan Africa, particularly in regions with limited or unavailable meteorological information. The network consists of ATMOS-41 weather stations manufactured by Meter Group, which record the variables required for estimating ET_0 including atmospheric pressure (kPa), radiation (W m^{-2}), relative humidity (-), temperature (degrees Celsius) and wind speed (m s^{-1}) (van de Giesen *et al.* 2014).

Meteorological data from 173 stations were collected at daily scale to calculate daily in-situ ET_0 as a benchmark for comparison against ET_0 data sets, with the locations of these stations shown in Fig. 1. The time series covers a 5-year period from 1 January 2018 to 31 December 2022, with approximately 1826 daily records per station. Data were selected from the stations that marked with the highest quality flags based on the TAHMO quality control procedure (van de Giesen *et al.* 2014), which is adapted from the Oklahoma Mesonetwork system (Shafer *et al.* 2000). This procedure combines automated and manual checks to ensure data reliability, including tests for data range, sensor performance, climatic consistency, sudden temporal changes, dips, spikes, and shifts in variance (Annor 2023). As an additional step, the percentage of missing daily records was calculated for each station, with an average of 27.8% missing data across all stations. Due to differences in intersection of grid coverage among data sets, 165 stations were used for the final evaluation to ensure that the same stations were applied consistently across all ET_0 data sets. To maintain consistency and avoid artificial alignment between data sets, no bias correction was applied to either the in-situ ET_0 or the ET_0 data sets; therefore, comparisons were based on raw daily ET_0 values.

2.3 Climate classification

Different climatic classes, like tropical, arid, or temperate areas, have unique climatic conditions that influence

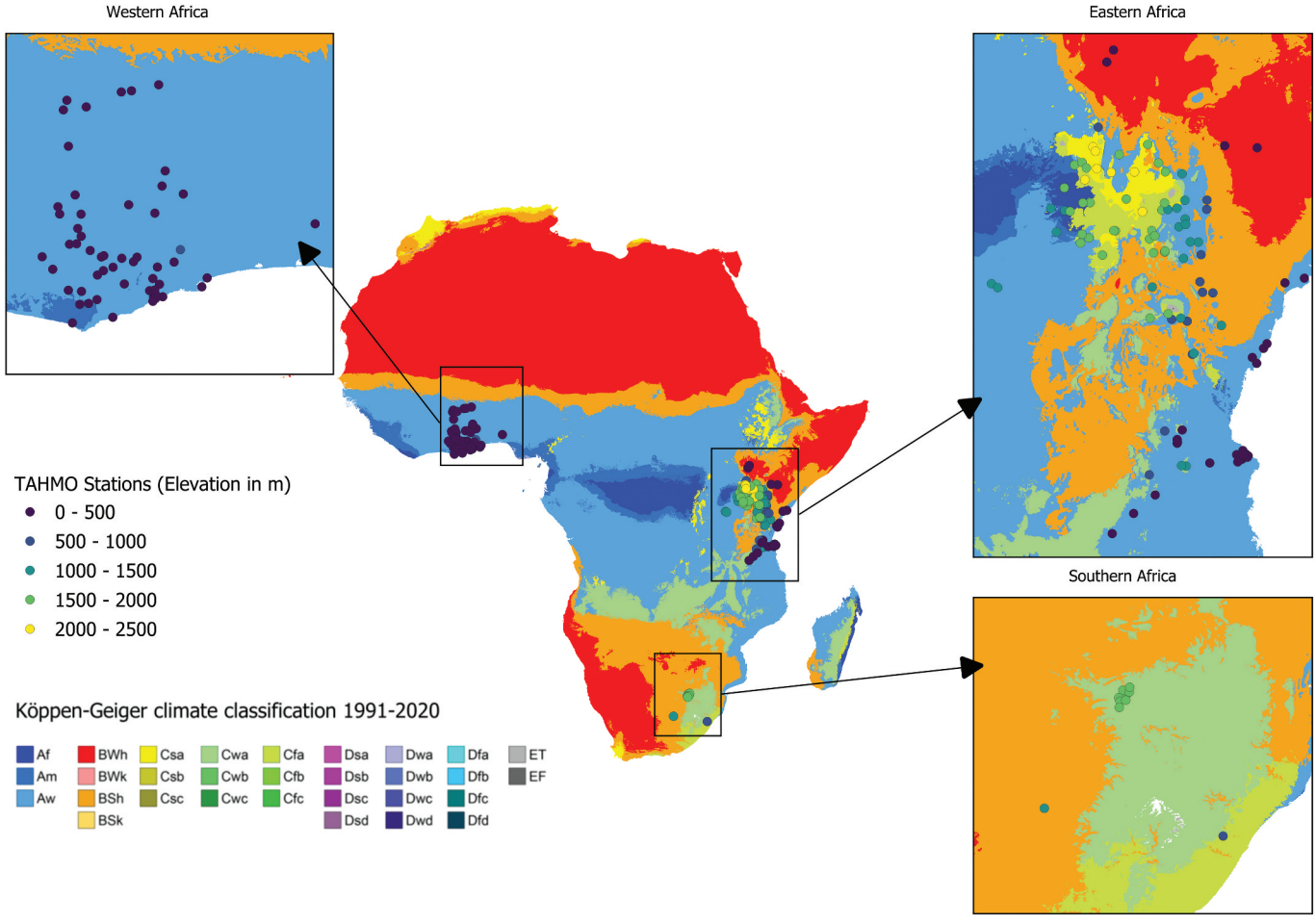

Figure 1. Climate classification map of study area and the locations of in-situ observations. Data source: TAHMO, Köppen–Geiger map (Beck *et al.* 2023).

evapotranspiration (Almorox *et al.* 2015). We therefore investigated how the different ET_0 data sets perform across different climate classes. To evaluate this, the Köppen–Geiger climate classification map with a 1-km resolution was used to identify the climate class of each station shown in Fig. 1. Table 2 shows the climate classes and the number of stations identified per climate class. Among 165 stations, 105 are located in (Aw) tropical savannah, 8 and 3 stations are in (BS) semi-arid, and (Bw) arid regions, respectively.

2.4 Elevation data

Elevation is another factor in ET_0 estimation as it directly affects atmospheric pressure, temperature, and solar radiation (Sun *et al.* 2020, Liu *et al.* 2021). Higher elevations lead to lower atmospheric pressure and cooler temperatures, both of which influence the rate at which water vaporizes. Additionally, solar radiation at higher elevations can vary, which may also impact ET_0 calculations. Therefore, the stations were also classified based on elevation data provided by the TAHMO network to assess the influence of elevation on performance of the data sets across different altitudes. Table 3 identifies the classes used based on elevation range and the number of stations in the corresponding class.

3 Methodology

3.1 Daily ET_0 calculation from in-situ data

In this study, daily meteorological data from 165 stations (1 January 2018 to 31 December 2022) were processed to calculate in-situ ET_0 with the FAO-56 PM equation (Allen *et al.* 1998; Equation 1). Following FAO-56 (1998; Chapter 2), we adopt the PM method with the fixed grass reference (h = 0.12 m, r_s = 70 s m^{-1}, a = 0.23), standardize all meteorological inputs to 2 m (with wind–height adjustment) and set daily G = 0. At this standard height, the aerodynamic is taken as $r_a = 208/u_2$ (s m^{-1}). The constants in Equation 1 follow FAO-56 daily conventions: 0.408 converts MJ m^{-2} day^{-1} to mm day^{-1}, 900 is the FAO-56 daily aerodynamic coefficient, 273 converts °C to K, and 0.34 derives from $(1+r_s/r_a)$ for the reference crop. The specific methodological implementations of the FAO-56 PM equation for the data sets are provided in Supplementary material Table S2.

Table 3. Elevation classes and number of stations per elevation class. The elevation classes were determined by the range of elevation (metres above sea level).

Elevation class	Number of station
0–500	83
500–1000	14
1000–1500	22
1500–2000	36
2000–2500	7

$$ET_0 = \frac{\left(0.408\Delta(R_n - G) + \gamma\frac{900}{(T+273)}\, u_2\,(e_s - e_a)\right)}{\Delta + \gamma\,(1 + 0.34 * u_2)} \quad (1)$$

where:

ET_0= vapotranspiration [mm day^{-1}]
Δ= slope of the saturation vapor pressure curve [kPa °C^{-1}]
R_n = net radiation [MJ m^{-2} day^{-1}]
G = soil heat flux density [MJ m^{-2} day^{-1}]
γ = psychrometric constant [kPa °C^{-1}]
T= mean daily air temperature at 2 m height [°C]
u_2= wind speed at 2 m height [m s^{-1}]
e_s= saturation vapor pressure [kPa]
e_a = actual vapor pressure [kPa]

3.2 Validation approach

All eight ET_0 data sets are available at daily scale; therefore, no temporal resampling was applied. Additionally, all the data sets have been pre-processed and corrected the windspeed from 10 m to 2 m before calculating the ET_0 (see Supplementary material, Table S2). For validation, a point-to-pixel comparison was employed by matching each station coordinate to the nearest pixel centroid of the corresponding data set. In cases where multiple stations fell within a single pixel, each station was independently compared to the same ET_0 value from that grid cell. While this approach is commonly used in the validation of gridded climate data, it can introduce uncertainties in areas with heterogeneous land cover or terrain. To assess its impact, a test has been done and the in-situ ET_0 data were spatially interpolated where multiple stations are located within one pixel to match the native resolution of each ET_0 data set using a weighted average of stations based on their distance from the grid centre. The results of this approach were compared with the point-to-pixel method and showed negligible differences in the performance of the data sets (see Supplementary material, Fig. S1). Therefore, the point-to-

Table 2. Climate classes, number of stations, and percentage area of the classes in Africa.

Class_ID	Class	Description	Number of Stations	Area (% of total)
2	Am	Tropical, monsoon	4	3.1
3	Aw	Tropical, savannah	105	27.8
4	Bw	Arid, desert, hot	3	44.3
6	BS	Arid, steppe, hot	8	14.7
9	Cs	Temperate, dry summer, warm summer	13	1.5
11	Cw	Temperate, dry winter, hot summer	14	5.4

pixel validation approach was retained. Despite its limitations, point-to-pixel comparisons remain a widely accepted approach particularly in regions where in-situ observations are limited (Li *et al.* 2018, Radmanesh *et al.* 2023, Ippolito *et al.* 2024, Lang *et al.* 2024).

3.3 Evaluation criteria

The ET_0 data sets were evaluated against daily in-situ ET_0 using four widely adopted statistical metrics; coefficient of determination (R^2), bias, root mean square error (RMSE), and relative bias (RBias). These metrics were selected based on their widespread use and direct relevance to applications in hydrology and climate studies (Mayr *et al.* 2019, Raziei and Parehkar 2021, Tran *et al.* 2023). The R^2 calculated as the square of the Pearson correlation coefficient is included because it measures how well the temporal variations of ET_0 from in-situ data correlate with those from spatial data sets (Mayr *et al.* 2019). Bias quantifies the mean difference and the systematic over or underestimation between in-situ and ET_0 data set values. RMSE shows the square root of the average of the squared differences between the data set and in-situ values and was chosen due to its sensitivity to large deviations which provides measure of data set accuracy by capturing significant discrepancies. Finally, RBias represents the bias as a percentage of the in-situ ET_0 value (Mayr *et al.* 2019, Raziei and Parehkar 2021).

The equations, range, unit, and best value for each metric are presented in Table 4. In this assessment, x is the spatial ET_0 data set and y is the in-situ ET_0. Although the literature (ASPRS 2015, Abdullah *et al.* 2024) often reports high performance thresholds for R^2 and low bias and RMSE values as indicative of good performance, these values are typically derived under controlled conditions. In contrast, spatial data sets like ET_0 have spatial grids and temporal intervals that do not perfectly align with in-situ data (Kohli *et al.* 2020, Wang *et al.* 2021). To ensure realistic classification, thresholds for R^2 and RBias were aligned with values commonly adopted in hydrologic model evaluation (Barbosa *et al.* 2019, Cardoso *et al.* 2019). In contrast, the magnitudes of bias and RMSE are known to be data-dependent (Dawson *et al.* 2007); therefore, their "Good/Acceptable" ranges were derived from the empirical distributions of all station–data set pairs and rounded to practical values, as reported in Table 4.

To determine the overall best-performing data set, the frequency of stations where each ET_0 data set was identified as the best performer was aggregated. The data set with the highest count of "best-performing" labels across all stations was considered the overall best-performing data set.

3.4 Performance of ET_0 data sets across climate and elevation classes

To assess the influence of environmental factors such as climate and elevation, the stations were grouped based on Köppen–Geiger climate classes and elevation ranges. The ET_0 performance metrics were then calculated for each group. This analysis is descriptive rather than a formal sensitivity analysis. The following section presents the results of this analysis, detailing the best-performing data set for each station, the overall best-performing data set, and variation in performance across different climate and elevation classes.

3.5 Evaluating relative importance of input data vs model implementation

Variances between ET_0 data sets can be derived from different sources, including different input data sets and different model implementations. To single out the variance caused by these two sources we calculated the variance between three different data sets using the same model implementation with different input data sets (AQUASTAT, ET_0-GEOS5, ET_0-ERA5) and the variance between three different products using the same input data (GEOS5), namely ET_0-GEOS5, WaPORv2 and WaPORv3. To quantity the relative contribution of input data and model parameterization to the total variance in ET_0 performance metrics, an analysis of variance (ANOVA)-based variance decomposition was applied (Giuntoli *et al.* 2015). For each station and metric (bias and RMSE), the between-group variance for each factor was calculated, and its proportion of the total variance was derived as the ratio of its variance component to the sum of both components. Station-level contributions were summarized by the median within each group (climate and elevation classes).

Table 4. Performance metrics used to validate the spatial with in-situ ET_0 data, including the thresholds for "Acceptable" and "Good" ranges for R^2, bias, RMSE, and RBias. For all equations, x represents the value from the ET_0 data set, y represents the value of in-situ ET_0, i represents the time step, and n represents the total number of stations.

Evaluation metrics	Formula	Unit	Value range	Best value	Acceptable	Good
Coefficient of determination (R^2)	$R^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}$	–	[0,1]	1	$0.5 \le R^2 < 0.7$	$0.7 \le R^2 < 0.9$
Bias	$\text{Bias} = \sum_{i=1}^{n}\frac{(x_i-y_i)}{n}$	mm day^{-1}	(−∞, +∞)	0	−0.6 ≤ Bias ≤ 0.6	−0.4 ≤ Bias ≤ 0.4
Root mean square error (RMSE)	$\text{RMSE} = \sqrt{\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{n}}$	mm day^{-1}	[0, +∞)	0	RMSE ≤ 1.0	RMSE ≤ 0.15
Relative bias (RBias)	$\text{RBias} = \frac{\text{Bias}\times 100}{\bar{y}}$	%	(−∞, +∞)	0	−15% ≤ RBias ≤ 15%	−10% ≤ RBias ≤ 10%

4 Results

4.1 Spatial and temporal variability of the ET_0 data sets

The average five-year ET_0 maps of the eight data sets are shown in Fig. 2, illustrating the spatial distribution across Africa. Overall, the spatial patterns are very similar although the magnitudes and spatial variability varies for the different data sets. The highest ET_0 values are observed in the Saharan zone (arid climate) with an average of 2540 mm year^{-1}, but with the lowest values for MSG 1556 mm year^{-1} followed by Bristol 2062 mm year^{-1} and FEWSNET 2244 mm year^{-1}.

Figure 3 shows the standard deviation and coefficient of variation (CoV) of average five-year ET_0 maps from eight data sets. The lowest CoV of ET_0 among the data sets is observed in the tropics (between 10°N and 20°S). The ET_0 data sets have the largest CoV in arid and semi-arid regions with CoV from 0.10 to 0.25.

The temporal variability of the ET_0 from the in-situ stations compared to the ET_0 data sets (averaged over all sites) is presented in Fig. 4. The observed seasonality is a combination of the seasonality of all the in-situ stations (West, East, and South Africa), the average ET_0 values range from 3 mm day^{-1} to 5.5 mm day^{-1}. All ET_0 data sets follow the same seasonal pattern; however, the magnitude of the values of the ET_0 data sets vary with the MSG following the in-situ observations the closest, followed by Bristol. AQUASTAT, WaPORv3, and ET_0-ERA5 show similar magnitudes, though higher than Bristol. ET_0-GEOS5 and WaPORv2 estimate even higher ET_0 with the largest overestimations by FEWSNET.

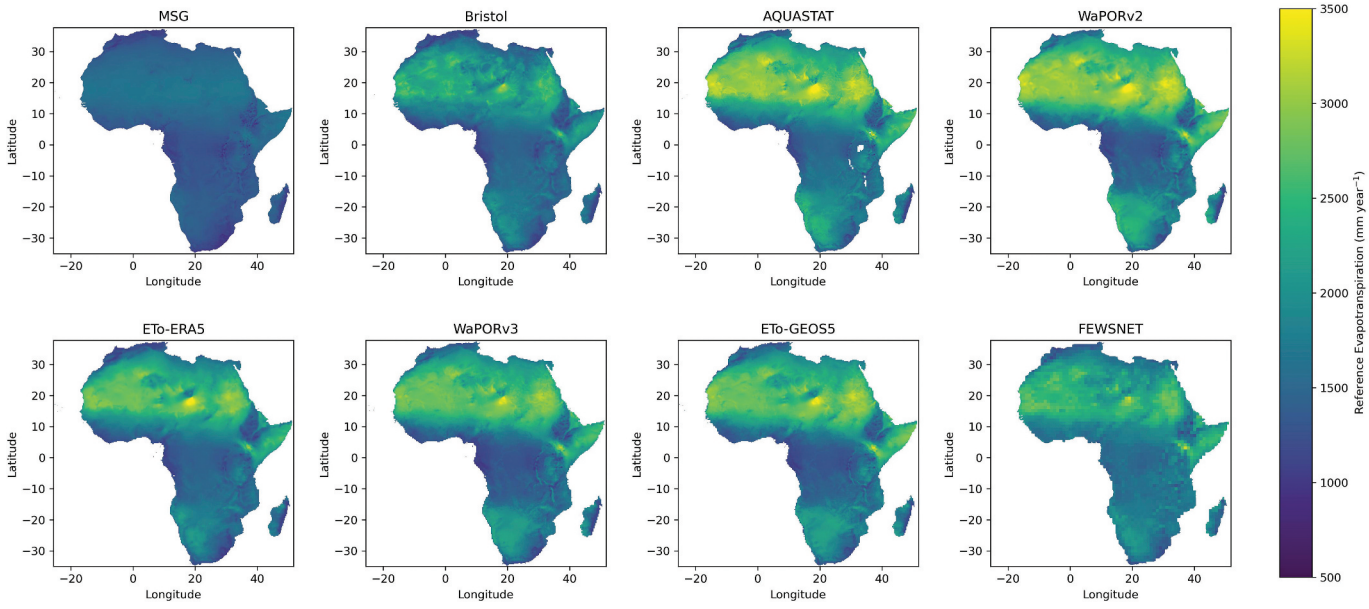

Figure 2. Five-year average (mm year^{-1}) of the eight ET_0 data sets from 2018 to 2022 across Africa, images are sorted based on high-resolution (MSG) to lower-resolution data sets (FEWSNET).

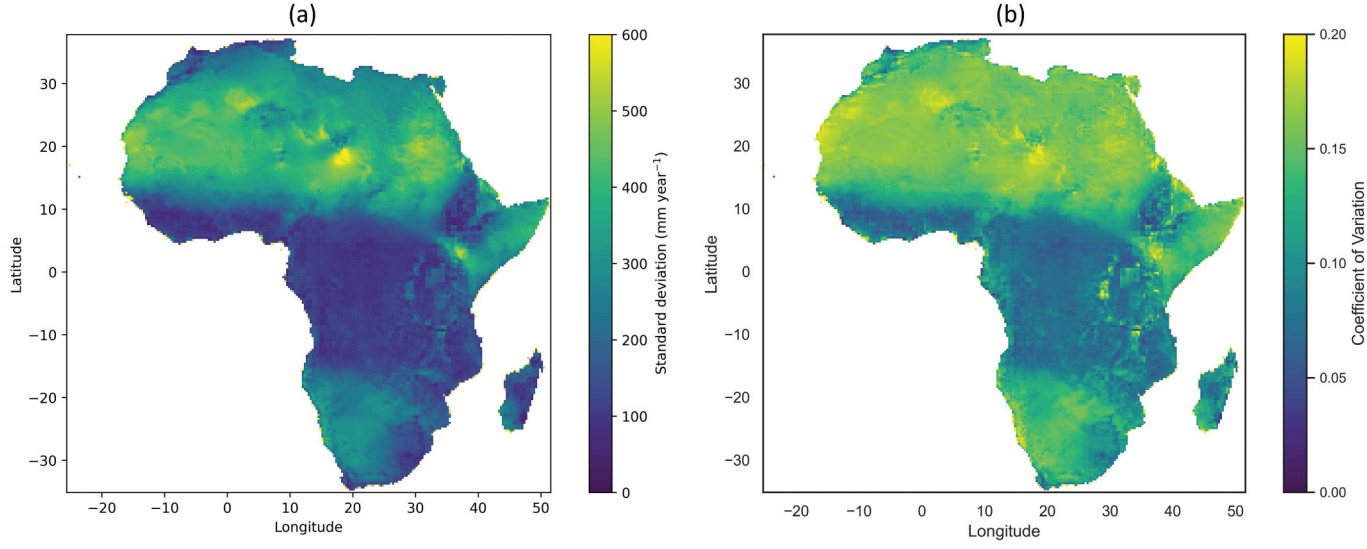

Figure 3. Map of standard deviation and coefficient of variation from eight ET_0 data sets.

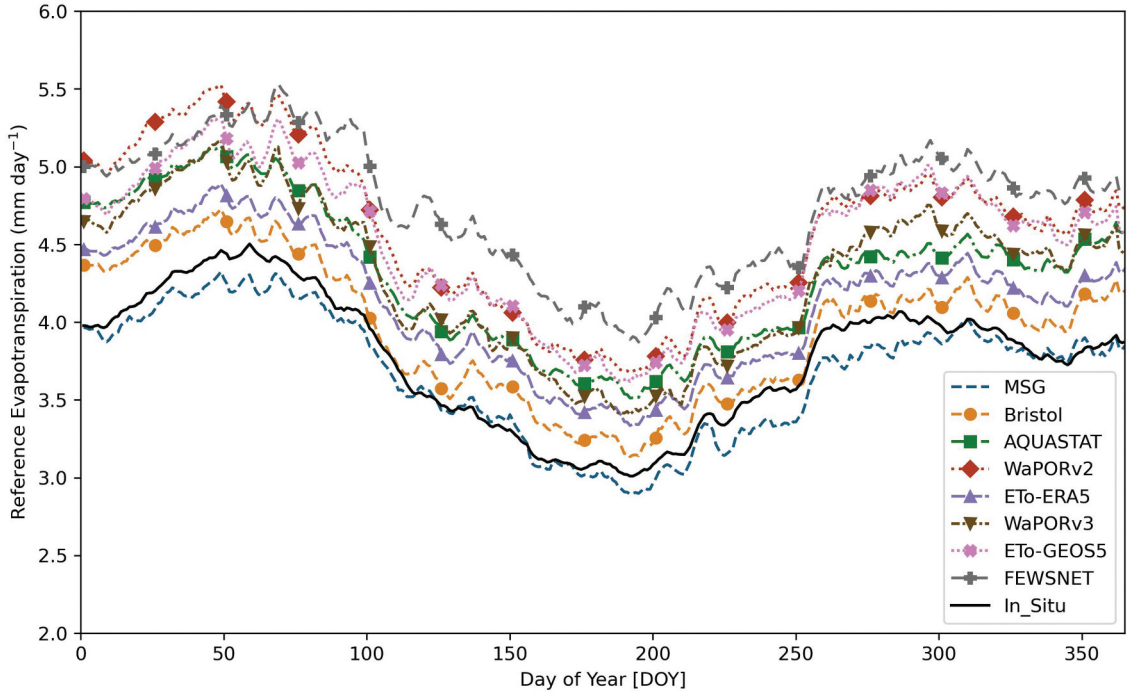

Figure 4. Comparison of seasonal variability of the spatial ET_0 data sets (mm day^{-1}) with in-situ data. Day-of-year (DOY) averages were calculated for each data set based on 5-year daily data from 2018 to 2022 from all stations, and a 7-day moving average was applied to smooth the variability.

4.2 Performance of ET_0 data sets

The results show that in general the data sets with higher spatial resolution yield better performance both in terms of higher correlation (R^2) and lower bias and RMSE when compared to the in-situ ET_0, with the exception of WaPORv2 and AQUASTAT (Fig. 5). WaPORv2 (20 km) has the highest median R^2 but at the same time its median bias and RMSE are the lowest and similar to FEWSNET (100 km). AQUASTAT (10 km) has the second highest median R^2 whereas its median bias and RMSE are close to WaPORv3 (30 km) and ET_0-GEOS5 (30 km). It is important to note that WaPORv2 was produced by combining input data sets with different spatial resolution, which are GEOS5 (30 km) climatic variables, SRTM elevation (90 m), and MSG transmissivity (5 km) (see Supplementary material, Table S1). The process of resampling these input data from native resolution to the final resolution of WaPORv2 (20 km) might have introduced more errors and bias at grid cell level. The final resolution of the data set may therefore not be the single indication of the resolution of the input data used to calculate ET_0. This result suggests that spatial resolution is a major factor but not the only one that determines the alignment of ET_0 data sets with the estimates from in-situ weather data.

The comparisons between the data sets using the same FAO-56 PM empirical equations (AQUASTAT, ET_0-ERA5, and ET_0-GEOS5) show that median R^2 of the ET_0-ERA5 is lower than AQUASTAT, but its median bias is lower and its median RMSE is comparable. Therefore, ET_0-ERA5 performs quite better when compared to in-situ ET_0 than AQUASTAT. The ET_0-GEOS5 has lower performance than ET_0-ERA5 and AQUASTAT, with remarkably lower R^2 and higher bias and RMSE. This is most likely due to the accuracy of input meteorological variables from GEOS5 reanalysis. Tran *et al.* (2025) showed that ET_0-GEOS5 perform worse than ET_0-ERA5 for all input variables of ET_0 when compared to measurements at the same stations used in this study. As a result, the ET_0 calculated using ERA5 also performs better than that from GEOS5. This highlights the importance of the quality of input data for estimating ET_0.

The R^2 results in that only WaPORv2 meets the "good" range ($0.7 \leq R^2 < 0.9$) for a strong correlation with the in-situ data (Fig. 5(a) and Table 5). This suggests that the WaPORv2 ET_0 data set captures the temporal patterns of the ET_0. Other data sets, including MSG, Bristol, AQUASTAT, ET_0-ERA5, and WaPORv3, fall within the "acceptable" range ($0.5 \leq R^2 < 0.7$), representing moderate agreement with in-situ ET_0 data, although not as strong as WaPORv2. Meanwhile, ET_0-GEOS5 and FEWSNET have R^2 values below the acceptable threshold, indicating lower performance as they are not able to capture the temporal variability of ET_0.

Looking at the bias (Fig. 5(b)), only MSG reaches a "good" bias score close to zero with minimal bias. Bristol with 0.25 and ET_0-ERA5 with 0.42 score in the "acceptable" range. In contrast, all other data sets exceed the acceptable bias range ($-0.6 \leq$ bias ≤ 0.6 mm day^{-1}), with notable examples such as ET_0-GEOS5, WaPORv2, and FEWSNET showing higher levels of deviation with 0.78, 0.88, and 1.02 mm day^{-1}, respectively (Table 6). This means that although MSG gives an unbiased estimate, other data sets overestimate the average daily ET_0 values.

The RMSE metric shows similar patterns (Fig. 5(c)). MSG stands out with the "good" RMSE criterion (RMSE $\leq$ 1.0 mm day^{-1}), showing a higher level of accuracy by producing smaller errors. Other data sets such as ET_0-GEOS5 and FEWSNET

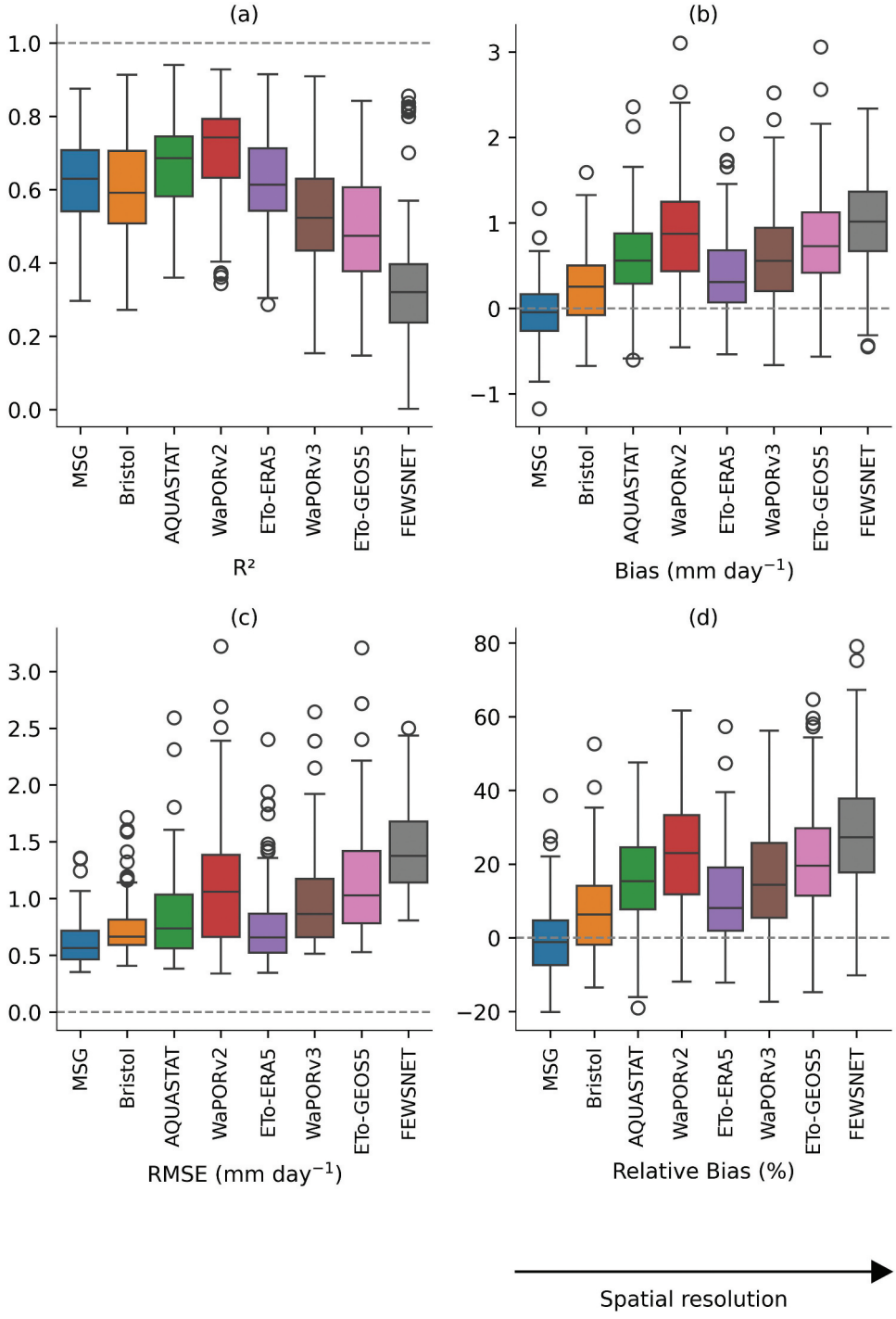

Figure 5. Distribution of performance metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}), RBias (%) for ET_0 data sets across 165 stations based on daily data for the period 2018–2022. The box plots illustrate the performance of each data set (MSG, Bristol, AQUASTAT, WaPORv2, ET_0-ERA5, WaPORv3, ET_0-GEOS5, FEWSNET). Outliers are shown as individual points, with boxes representing the interquartile range and the central line represents the median. The spatial resolution of data sets is indicated along the *x*-axis, with resolution decreasing from left to right.

Table 5. Mean values of the performance metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}) and RBias (%) of the ET_0 data sets for the 165 stations for the period 1 January 2018 to 31 December 2022.

	MSG	Bristol	AQUASTAT	WaPORv2	ET_0-ERA5	WaPORv3	ET_0-GEOS5	FEWSNET
R^2	0.62	0.60	0.67	0.71	0.63	0.53	0.49	0.33
Bias	−0.04	0.25	0.61	0.88	0.42	0.58	0.78	1.02
RMSE	0.61	0.73	0.84	1.11	0.76	0.98	1.15	1.43
RBias	−0.15	7.25	16.29	23.41	11.46	15.53	21.08	28.43

Table 6. Number of stations with the highest score based on the four metrics for each data set.

	MSG	Bristol	AQUASTAT	WaPORv2	ET_0-ERA5	WaPORv3	ET_0-GEOS5	FEWSNET
R^2	18		41	101	5			
Bias	80	28	6	3	27	13	4	4
RMSE	89	14	23	11	27	1		
RBias	80	28	6	3	27	13	4	4
	267	106	76	118	86	27	8	8

had the highest RMSE values of 1.15 and 1.43 mm day^{-1}, respectively (Table 5).

In terms of relative bias (Fig. 5(d)), MSG and Bristol show low bias with average −0.15%, 7.25%, respectively (Table 5), values which are within the "Good" range of −10% ≤ RBias ≤ 10%, whereas ET_0-ERA5 with 12.6% bias is in the "Acceptable" category. This minimal error level indicates that these data sets align closely with in-situ ET_0 values. However, ET_0-GEOS5, WaPORv2, and FEWSNET show highest levels of relative bias of 21.08%, 23.41%, and 28.43%, respectively.

Overall, MSG is the best performing data set across most of the metrics, with MSG showing the lowest bias and error. Bristol and ET_0-ERA5 also perform reasonably well, maintaining acceptable levels of accuracy and proportional error. Despite having the highest R^2, WaPORv2 has one of the largest biases, indicating difficulty in accurately estimating the absolute ET_0 values. Other data sets such as GEOS5 and FEWSNET also show consistently large errors and biases.

4.3 Best ET_0 data set per station

Figure 6 shows the best performing ET_0 data sets per station. For R^2, a large majority (61%) of the stations indicate WaPORv2 performs best, with the locations of the stations specifically located in the southern parts of West Africa (Ghana) and for stations in high elevation in East Africa. In South Africa, on the other hand, the AQUASTAT performs the best, which is also the case for the stations located in the drier (northern) parts of Ghana and the lower elevations in East Africa. WaPORv2 and AQUASTAT perform best in R^2 for 86% of all stations.

For the performance metrics that indicate deviation from the in-situ measurements (bias, RMSE and RBIAS), MSG has the highest number of stations that have the best performance (bias (80), RMSE (89) and RBias (80) (Table 6). There are three ET_0 data sets (Bristol, ET_0-ERA5, and AQUASTAT) that follow MSG in second place, but they are less widespread and there is no noticeable spatial pattern where certain ET_0 data sets outperform others. Only a few stations indicate that WaPORv2, WaPORv3, GEOS5, or FEWSNET ET_0 data sets are the best. Surprisingly, for two of the three stations located in the Bw climate class in East Africa (two most northern stations), FEWSNET ET_0 data show the best performance. For the stations in South Africa, the best performing data set is ET_0-ERA5 for the majority of the stations, with a noticeable difference for the station in the South West which is located in the BS climate class compared to the other stations located in the Cw climate class.

4.4 ET_0 performance by station characteristics

The performance of all the ET_0 data sets were further investigated by grouping the stations into the different climatic classes and elevation. The results of each ET_0 data set is provided in the Supplementary material, Fig. S2 and S3, with the MSG ET_0 data set used as an example in the main text.

4.4.1 Climate class

The results of the performance assessment per climate class are presented in Fig. 7. The analyses show some stark differences in the performance of the different ET_0 data sets in the different climate classes. MSG outperforms all other ET_0 data sets for the Aw and Cs climate class (except for R^2), which contains the majority of the stations, and could be responsible for the overall better performance as presented in the previous section. The performance of the different data sets also varies per climate class, with good performance for R^2 (>0.7) for almost all ET_0 data sets for the Cs and Cw climate classes (Fig. 7(a)), whereas for the dry climate classes (BS and Bw) R^2 is well below the acceptable range for most of ET_0 data sets. Similarly, the bias, RMSE, and RBias for these climate classes are well below the acceptable range.

In general, FEWSNET with low R^2 and higher RMSE and bias is not performing well compared to other data sets for most climate classes, except for the Bw climate class, where it performs quite well (and best in terms of bias and RBias), although these results are based on only three stations. MSG, Bristol, and ET_0-ERA5 are the ones performing the best in bias, RMSE, and RBias. MSG is the only ET_0 data set that has a negative bias for most climate classes, whereas all the other data sets overestimate ET_0 compared to station data.

A detailed evaluation of the performance metrics for MSG is presented in Fig. 8 (the analyses for the other data sets are presented in Supplementary material, Fig. S2 and S3). The analyses show that MSG performs best in the Cs and Cw (temperate climates), followed by the Am and Aw (tropical climates). These climate classes show higher R^2 values in Fig. 8 (a) and similar performance in bias, RSME, and RBias, with MSG underestimating ET_0 in Cs and Cw climate classes and overestimating in Am climate class (for Aw there seems to be equal number of stations over- as well as underestimating ET_0) . In contrast, BS and Bw climates show lower R^2 values and higher (negative) bias, RBias, indicating that ET_0 from MSG are less reliable in these drier regions. The bias values remain close to zero across most climate zones. However, a slightly larger negative bias with −0.3 and −0.8 mm day^{-1} is observed in the BS and Bw zones, respectively. ET_0 estimates tend to be high in drier climates (Fig. 2) with larger bias; this is countered in the RBias graph, where the Bw values are less extreme compared to the bias graph (Fig. 7).

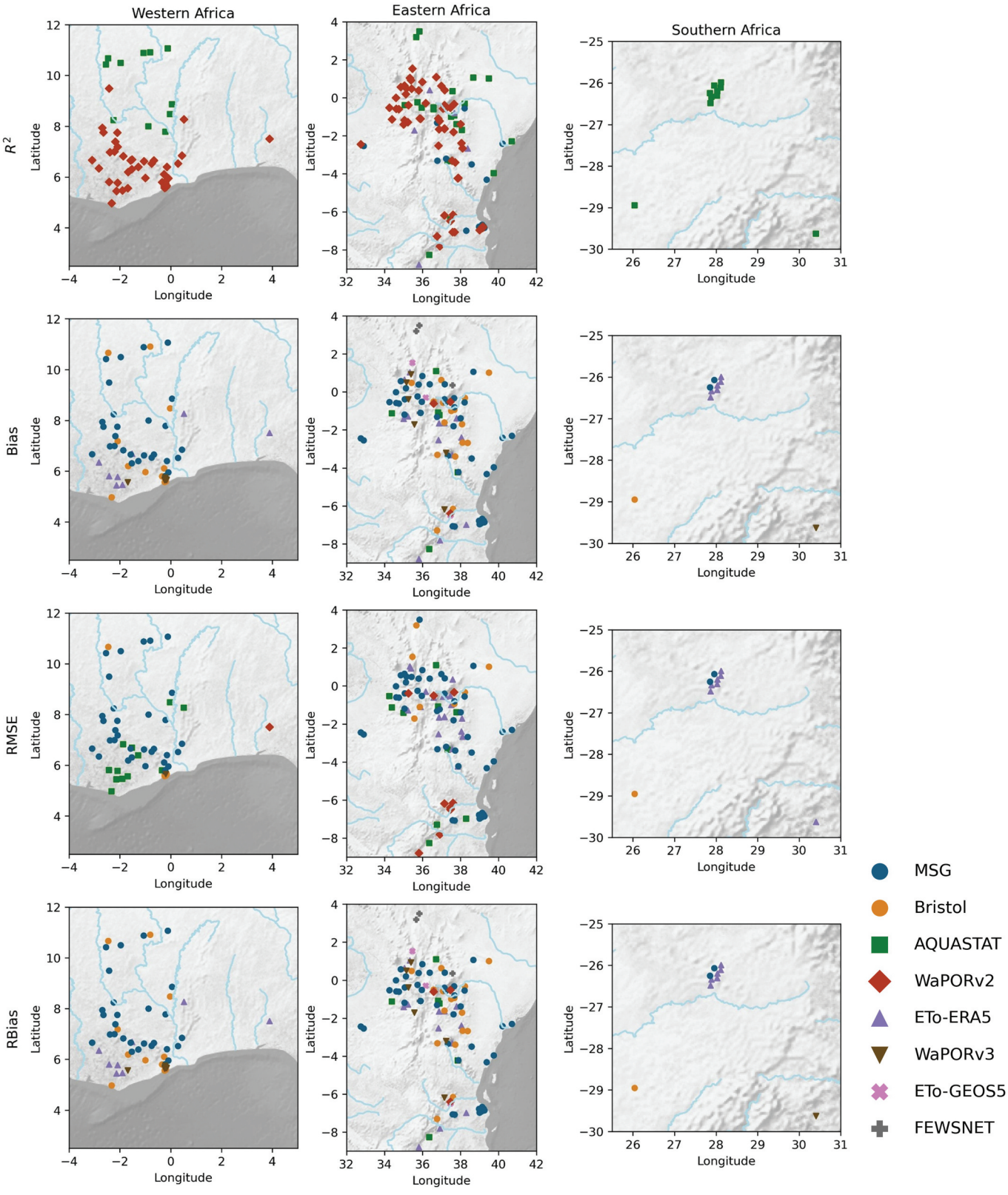

Figure 6. Spatial distribution of the best-performing ET_0 data sets based on the evaluation metrics for each station across Western (n = 59), Eastern (n = 104), and Southern Africa (n = 10). Each color corresponds to a data set, representing the highest-performing ET_0 data set at each station.

The high performance of MSG, particularly in tropical and temperate climates in terms of bias and RMSE can be explained by the way it calculates ET_0. Unlike the reanalysis-based data sets that rely on multiple meteorological inputs, MSG estimates ET_0 using solar radiation derived directly from satellite observations. As described by Trigo *et al.* (2018), this approach captures more accurately the daily and seasonal changes in radiation and avoids errors caused by inputs like

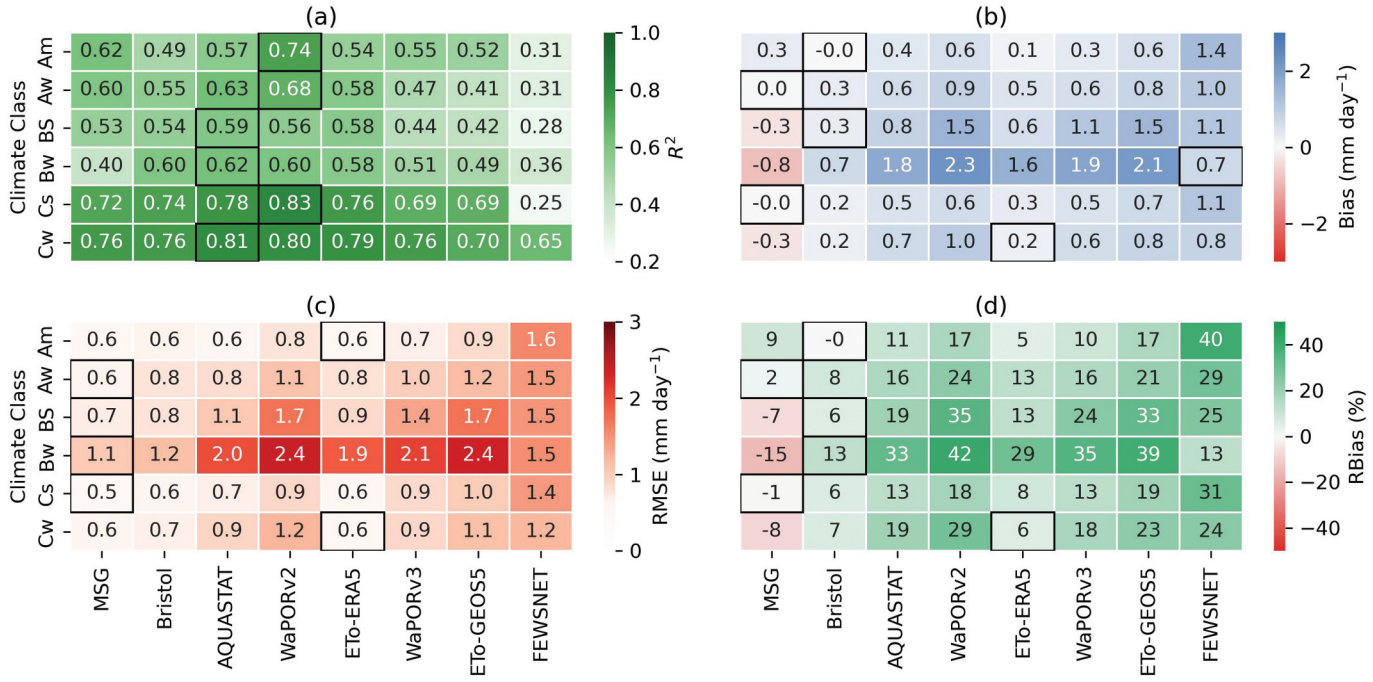

Figure 7. Heat maps of the performance metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}), and RBias (%) for ET_0 data sets across climate classes (Cw, Cs, Bw, BS, Aw, Am). The Black boxes in each subplot highlight the best performing ET_0 data set for the respective metric and climate class.

wind speed or humidity. In addition, MSG's approach aligns closely with the FAO reference conditions, which assume a flat, well-watered grass surface, which may also contribute to its strong performance in tropical and temperate climates. However, in arid and semi-arid regions where conditions differ from the reference surface, MSG's performance declines.

Although MSG shows the lowest RMSE and bias for most climate classes, it does not produce the highest R^2 values. This may be related to how ET_0 data sets respond to temporal variability versus absolute accuracy. R^2 measures the ability to capture temporal patterns, while RMSE reflects the magnitude of error. As MSG is based only on satellite-derived solar radiation, it performs well in capturing daily radiation, which reduces overall error. However, in some climates, particularly where other variables like wind speed and humidity play a stronger role in driving ET_0 (such as arid and semi-arid), MSG may not capture temporal fluctuations effectively. This results in even lower R^2, even when the absolute errors remain small. In contrast, data sets like WaPORv2 and AQUASTAT that use reanalysis inputs such as wind speed may capture variability better but introduce larger Biases due to uncertainties in those inputs, particularly windspeed, which has been shown to contribute the largest error among meteorological drivers in reanalysis-based ET_0 estimation (Tran *et al.* 2025).

A similar pattern is observed with WaPORv2, which shows high R^2 values in tropical regions. This could be due to its use of MSG-derived solar radiation, combined with meteorological data from GEOS-5 and elevation from SRTM (FAO 2020). This combination allows WaPORv2 to better respond to short-term variability in weather conditions, which helps it achieve higher R^2. In contrast, MSG ET_0 is driven primarily by radiation and does not account for wind speed, humidity, or other variables. As a result, MSG performs better in terms of average accuracy (bias and RMSE), but underperforms in temporal correlation, particularly in climates where non-radiative factors are more important. Despite the strong correlation of WaPORv2, it shows large bias and RMSE in most climate classes. This may be due to the resampling of its inputs, originally at different spatial resolutions, into a common 20 km grid. This step can introduce spatial mismatches and aggregation errors, especially in arid and semi-arid regions where environmental heterogeneity is high and more difficult to represent accurately (FAO 2020).

All ET_0 data sets (except FEWSNET) show the highest R^2 for the Cs and Cw climate classes and no clear pattern in ranking for the other climate classes (see Supplementary material, Fig. S2). Except for MSG all other ET_0 data sets overestimate ET_0 compared to the in-situ observations, indicated by a positive bias. The largest variance between data sets is observed in the Bw climate class with MSG showing a negative bias (<–1 mm day^{-1}), ET_0 data sets like Bristol and FEWSNET have similar but positive bias and the rest of the data set have significant larger positive bias (>1 mm day^{-1}).

4.4.2 Elevation

Figure 9 shows the heatmaps of the mean performance of the ET_0 data sets grouped per elevation class. In contrast to the comparison per climate class, the results are overall consistent across the different elevation classes, with WaPORv2 giving the highest R^2 values for each class and MSG giving the best results for almost all other performance metrics. For all the data sets (except FEWSNET), the performance metrics improve with increasing elevation.

A closer look at the MSG ET_0 performance shows that it performs better at higher elevations while lower elevations show greater variability (Fig. 10(a)). Figure 10(b,d) show that across most elevation ranges the mean of bias and

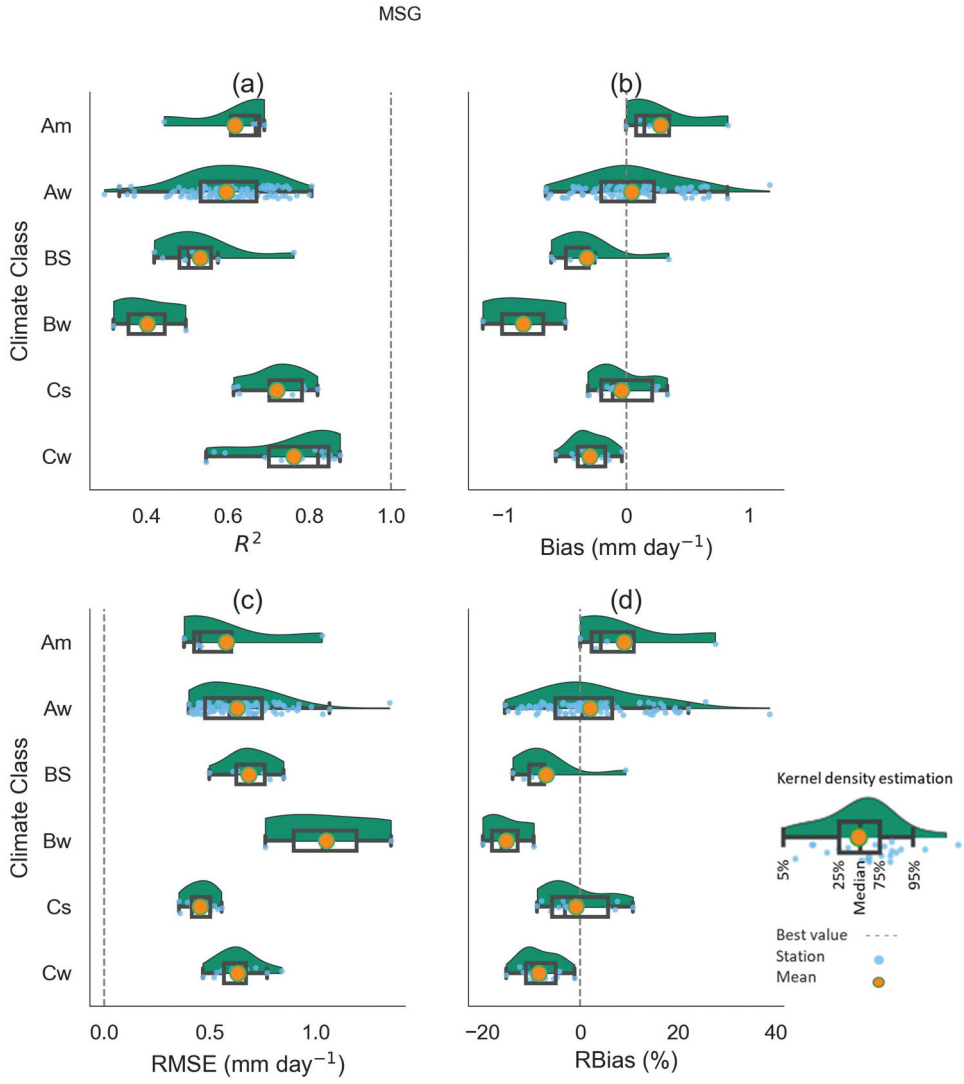

Figure 8. Distribution of performing metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}), RBias (%) for MSG ET_0 across climate classes; Am, Aw, BS, BW, Cs, Cw. The blue dots show the number of stations used for comparison. The green area under the curve represents the kernel density estimation of the distribution. The box plot represents the 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. The orange circle inside the box plot represents the mean value.

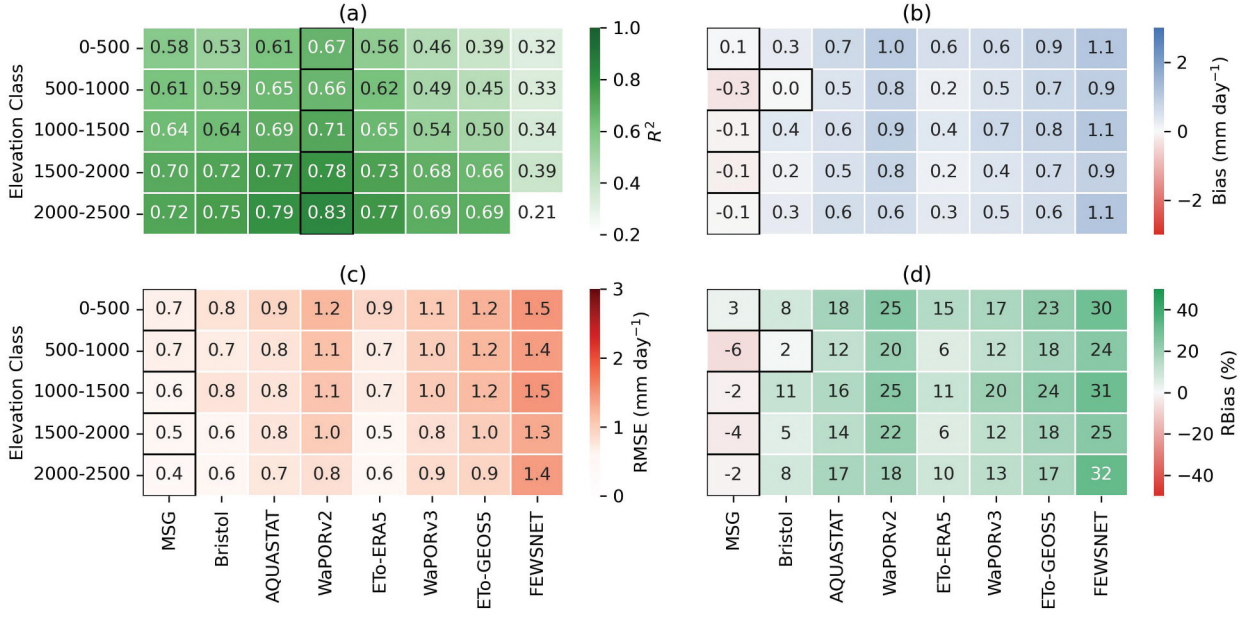

Figure 9. Heat maps of the performance metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}), and RBias (%), for ET_0 data sets across elevation classes (0–500 m, 500–1000 m, 1000–1500 m, 1500–2000 m, 2000–2500 m). The Black boxes in each subplot highlight the best performing ET_0 data set for the respective metric and elevation class.

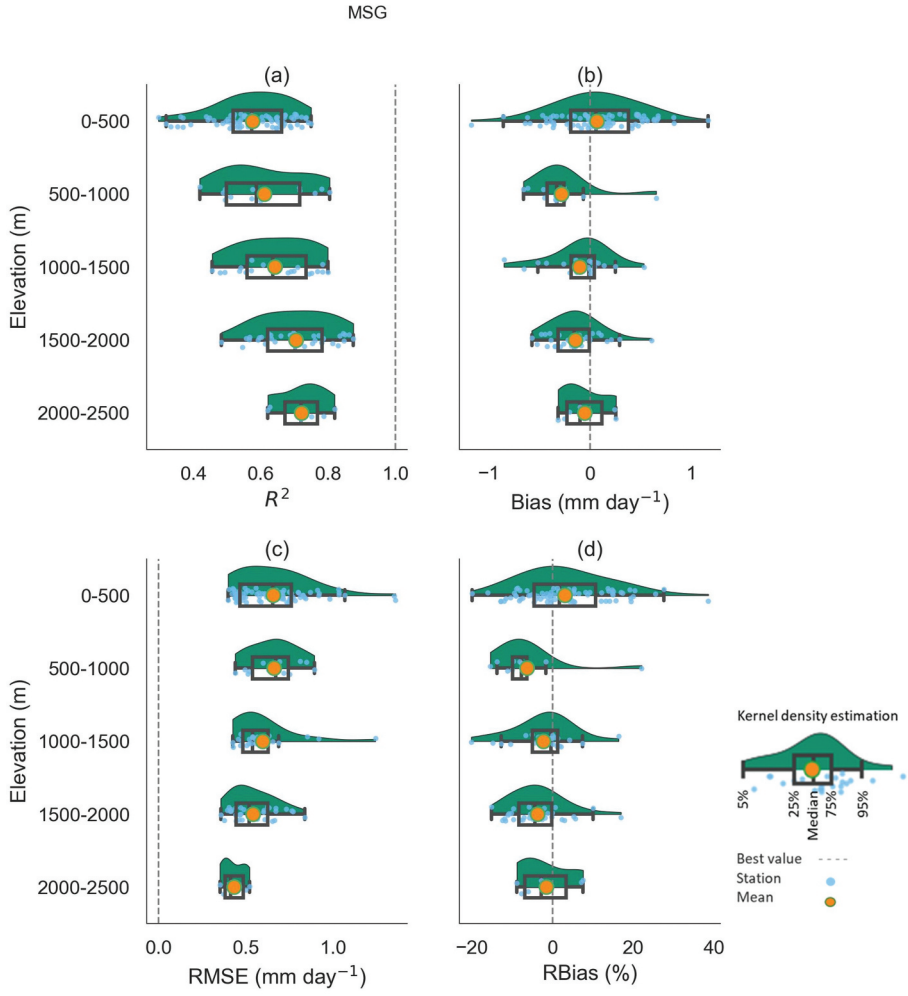

Figure 10. Distribution of performance metrics; R^2, bias (mm day^{-1}), RMSE (mm day^{-1}), RBias (%) for MSG ET_0 across elevation classes; 0–500, 500–1000, 1000–1500, 1500–2000, 2000–2500 m above area under the curve represents the kernel density estimation of the distribution. The box plot represents the 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. The orange circle inside the box plot represents the mean value.

RBias are close to zero. However, the bias and RBias at the 500–1000 m elevation range is slightly higher compared to the lower elevation, also for the bias and RBias the variability is the highest for the lowest elevation, which also contains most stations. The lowest elevation is where most stations show a positive bias (also illustrated with a positive mean bias and RBias); however, for the other elevation classes the majority of the stations have a negative bias and RBias. For the RMSE, the better performance at higher elevation continues (Fig. 10(c)).

All ET_0 data sets show a consistent higher R^2 with increasing elevation (except for FEWSNET). This general improvement with elevation can be attributed to both physical and data-related factors. Higher-elevation areas usually have lower humidity, fewer aerosols, and more stable atmospheric conditions, which improves the accuracy of radiation and vapor pressure inputs. In contrast, low-elevation and coastal regions tend to have more heterogeneous land–atmosphere interactions and local convection, which introduce noise in coarse-resolution data sets. Although the differences are not large for all metrics, the overall trend of better performance at higher elevations is consistent across most data sets.

4.5 Impact of input data versus parameterization

The impact of the uncertainty in meteorological input data on the ET_0 is illustrated by the difference between ET_0-GEOS5 and ET_0-ERA5 and AQUASTAT (Fig. 11(a)). These three data sets were calculated applying the same parameterization of FAO-56 PM with three different input data sets. In this case, the regions with higher ET_0-GEOS5 are regions where GEOS5 shows higher wind speed and lower solar radiation than ERA5 and AQUASTAT (Tran *et al.* 2025). The different input data sets lead to a difference of −104 mm year^{-1} in mean annual ET_0 between ET_0-GEOS5 and AQUASTAT, and 105 mm year^{-1} between ET_0-GEOS5 and ET_0-ERA5 (Fig. 12).

The impact of the choice of parameterization on ET_0 estimates is illustrated by comparing WaPORv2 and WaPORv3, which both use GEOS5 meteorological inputs but with different spatially calibrated parameters for net longwave radiation and different input elevation data (see Supplementary material, Table S1 and Table S2). This choice of parameters has a great impact on the spatial distribution of ET_0 (Fig. 11(b)). This difference in parameterization leads to a difference of −149 mm year^{-1} in mean annual ET_0 between ET_0-GEOS5 and WaPORv2, and 32 mm year^{-1} between ET_0-GEOS5 and WaPORv3 (Fig. 12).

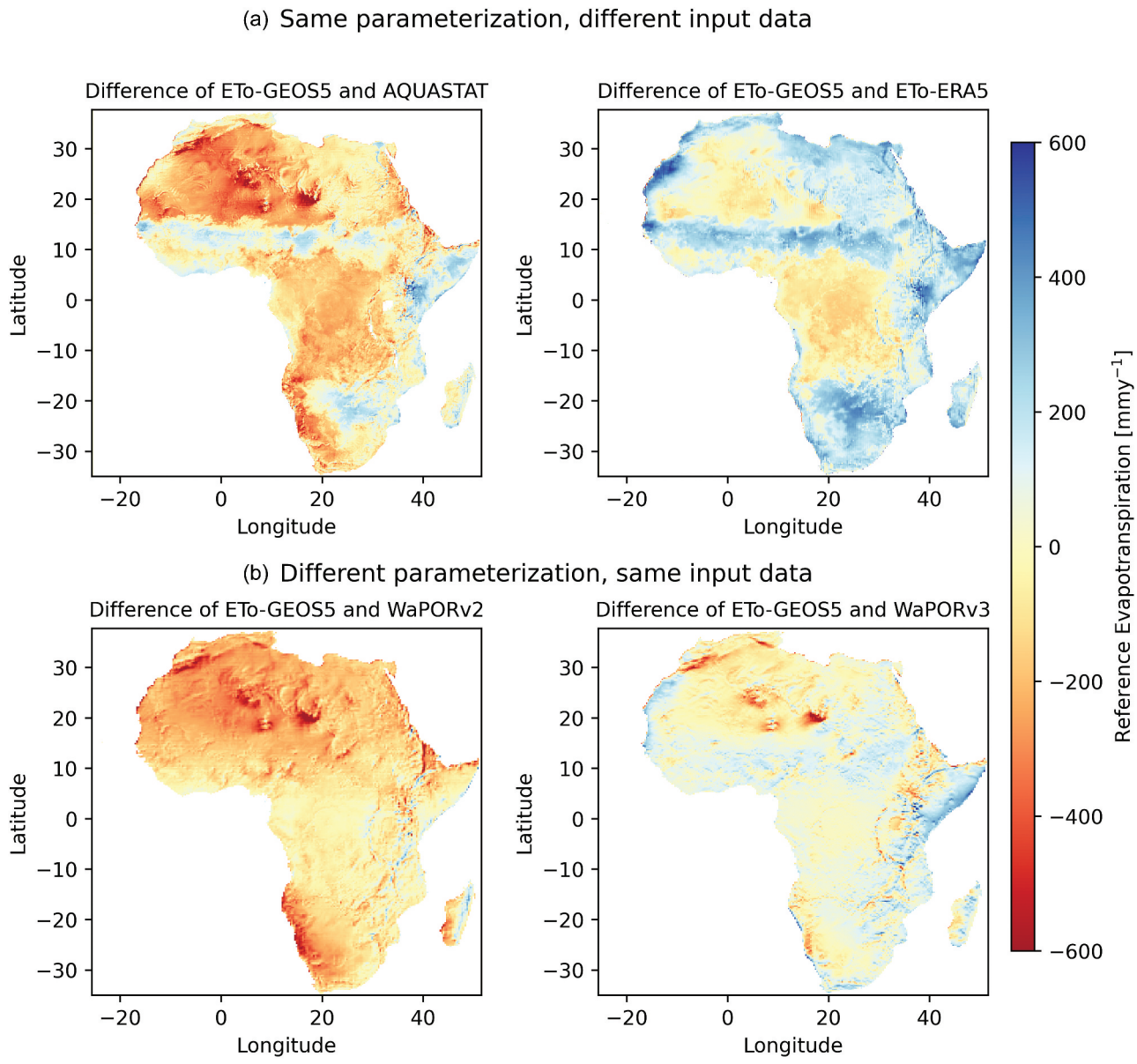

Figure 11. Comparison of five-year annual mean ET_0 between ET_0-GEOS5 and data sets with (a) the same parameterization and different input data (AQUASTAT and ERA5), and between ET_0-GEOS5 and data sets with (b) different parameterization and same input data (WaPORv2 and WaPORv3).

Although WaPORv3 has expanded the data set's coverage to global scale, our results show that WaPORv2 performs better in terms of temporal correlation with in-situ ET_0 (R^2 = 0.71 vs 0.53). However, WaPORv3 shows better performance in absolute accuracy, with lower bias (0.58 vs 0.88 mm day^{-1}), RMSE (0.98 vs 1.11 mm day^{-1}), and RBias (15.53% vs 23.41%). This might be due to the changes in radiation input sources and adjustment of the data set. WaPORv2 uses MSG-derived transmissivity for solar radiation, whereas WaPORv3 uses GEOS-5 inputs, which have coarser resolution. Additionally, WaPORv3 integrates multiple new data sources such as AgERA5 for temperature with the aim of global data production, whereas WaPORv2 was tailored more specifically for Africa and may thus perform better in this region.

Results of the ANOVA analyses show that differences in input data explain, on average, 60–70% of the total variance in bias and RMSE across stations, while model parameterization accounts for the remaining 30–40% (Fig. 13). The dominance of input data uncertainty was consistent across most climates and elevations, particularly in humid and mid-altitude regions, whereas parameterization effects became more relevant in high-elevation areas. These findings emphasize that, for operational applications such as irrigation scheduling, users should prioritize data sets that use high-quality or bias-corrected meteorological inputs (e.g. those based on ERA5). In contrast, in high-elevation or data-scarce areas, uncertainty from parameterization remains significant.

5 Discussion

Our findings align with previous research that highlights accuracy issues of spatial ET_0 data sets across Africa, particularly in

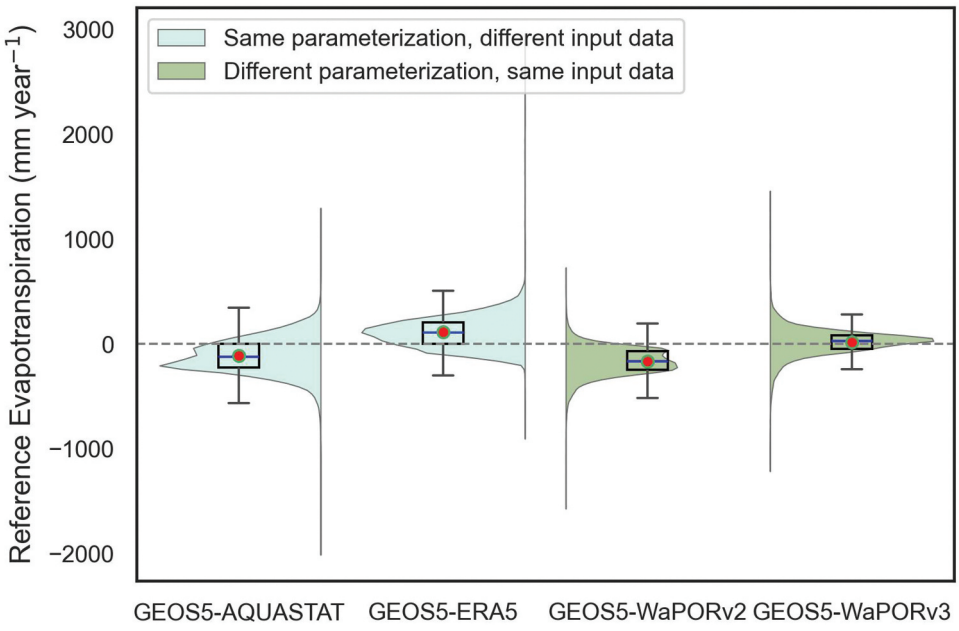

Figure 12. The distribution of the difference in mean annual ET_0 (mm year^{-1}) between ET_0-GEOS5 and products with the same parameterization and different input data (AQUASTAT and ET_0-ERA5) and products with the different parameterization and same input data (WaPORv2 and WaPORv3). Outliers are shown as individual points, with boxes representing the interquartile range and the central line represents the median.

arid and semi-arid environments. For example, Hobbins *et al.* (2023) reported significant positive biases in FAO-56 Penman–Monteith-based ET_0 estimates derived from MERRA2 data, which they validated against the SASSCAL station network. Their results showed significant positive bias (average +36%) in arid regions like Namibia. Similarly, Trigo *et al.* (2018) observed limitations in MSG-derived ET_0 due to local advection effects in semi-arid areas. Furthermore, Weerasinghe *et al.* (2020) demonstrated biases in satellite-based evapotranspiration products across Africa and emphasize the importance of input data quality, especially under water-stressed conditions. Zimba *et al.* (2023, 2024) found substantial discrepancies in satellite-based Actual Evapotranspiration (AETI) in miombo woodlands attributed to plant phenology and seasonal dynamics, which shows accurate input data are crucial for capturing evapotranspiration dynamics in ecosystems sensitive to climatic variability. Therefore, improving the quality of ET_0 estimates is important because ET_0 is a key input parameter in calculating AETI. Based on Weerasinghe *et al.* (2020), Zimba *et al.* (2023, 2024), better accuracy in ET_0 directly translates to improved AETI estimations, particularly in ecosystems sensitive to climatic variations such as miombo woodlands. Enhancing the accuracy of ET_0 thus improves the reliability of water balance analyses, irrigation scheduling, and agricultural water management.

To improve future ET_0 data sets, it is important to not only to increase the resolution, but also to combine different types of local station data. For instance, Rajput *et al.* (2024) and Ndiaye *et al.* (2020) suggest that blending satellite radiation data with better reanalysis weather data can help reduce bias and better represent local variation. Recent studies also suggest that machine learning models can improve ET_0 estimates, particularly in areas where station data are limited. Yonaba *et al.* (2024) showed that these models can capture complex patterns in weather data and perform better than traditional methods in semi-arid climates. Kiraga *et al.* (2024) and Rajput *et al.* (2024) also used AI tools like genetic algorithms and model combinations to improve ET_0 accuracy. These findings suggest that combining different data sources and using improved models can help create better ET_0.

Here we discuss two main challenges in this study, first the availability of stations and second the challenges of validating point data to pixel values.

5.1 Availability of in-situ stations

The availability of ground observations for the study period (2018–2022) in the area of interest is limited and biased, and the spatial distribution of the stations is uneven across Africa. We only managed to obtain stations in three regions of Africa (East Africa, West Africa, and South Africa) from the TAHMO network, with the majority of stations located in the tropical savannah climate class (Aw 64% of the stations), although this only covers 27.8% of the total area of the continent (see Table 2). The results are therefore most reliable for Aw climate class and conclusions for arid zones (BS/Bw) should be considered less certain. The underrepresentation of in-situ stations (three stations) located in the arid, desert, hot climate class (Bw), which covers 44.3% of the continent, is an example of this. This area specifically is where the major differences are observed between MSG and the other ET_0 data sets (Fig. 2). Even though a large part of this area is not inhabited, some of the largest irrigation schemes of the continent are located in this area, for example the Nile Delta in Egypt, Gezira scheme in Sudan, and the Office du Niger in Mali. The high variance between the different ET_0 data sets in these areas will therefore affect the crop water requirement estimations and irrigation scheduling recommendations.

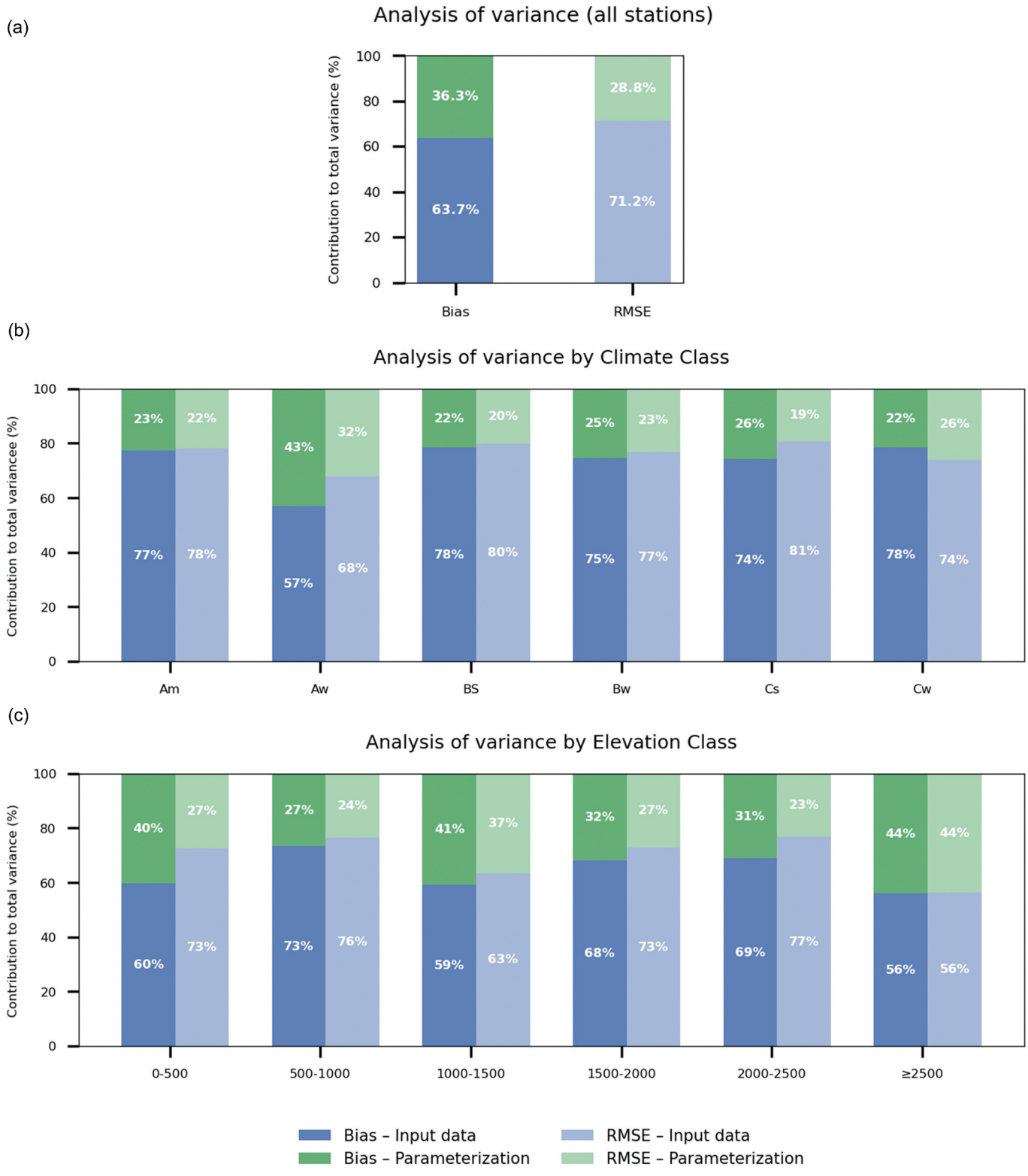

Figure 13. Distribution of the relative contribution to the total variance (%) in station-level ET_0 performance explained by two sources: input data and model parameterization. Results are shown (a) across all stations, (b) by climate classes, and (c) by elevation classes. Within each group, bias and RMSE are presented as adjacent bars. The blue segments represent the contribution of input data, and the green segments represent the contribution of parameterization, numeric labels inside the bars indicate the corresponding percentages.

The validation of the MSG ET_0 data set (Fig. 8) shows a low performance in this climate class, showing a significant negative bias and low R^2. On the other hand, the results of the other data sets (Fig. 7) all indicate positive biases, with Bristol and FEWSNET showing the smallest bias.

Despite the quality control applied to the in-situ data set, the performance of ATMOS-41 sensors has limitations that could influence the ET_0 validation results. For instance, temperature sensors may exhibit slight biases under strong solar radiation and low wind conditions, and the radiation sensor

may slightly underestimate high radiation levels (Mohit Anand and Molnar 2018, Dombrowski *et al.* 2021). Furthermore, an average of 27.8% of the data was missing across stations during the study period, which may affect the reliability of daily ET_0 estimates at some stations.

5.2 Validating point to pixel values

Due to the coarse resolution of the ET_0 data sets (highest resolution 5 km), a validation with a point measurement can be tricky. However, the comparison with applying a weighted average of all stations within one pixel showed negligible difference in the performance of the data sets compared to the pixel to point comparison (see Supplementary material, Fig. S1). For the higher-resolution data sets this can be explained as there are few pixels with multiple stations, for the ET_0 data set with the lowest resolution (FEWSNET), the differences show a slight improvement of the R^2 and bias.

A detailed analysis is provided in Fig. 14 which shows for two locations, one in East Africa and one in West Africa, a FEWSNET grid cell (100 km) with multiple ground stations within it (15 each). In our analyses the ground station data is compared to the value within the grid it falls, without considering the spatial variability of the ET_0 (the background shows this spatial variability by presenting the MSG ET_0 data). The comparison of the timeseries of the 15 stations compared to the FEWSNET grid value shows that all stations show a significant negative bias against the FEWSNET data throughout the time series, but with a similar temporal pattern. The stations also show the variability of observed ET_0 within the FEWSNET grid.

The performance results for the two grids are presented in Fig. 15. Overall, it shows large variability of the performance indicators from one station to another. Although this could be related to data quality of the stations it also shows the variability of ET_0 within the different grids. While expectations are that the variability of performance would be higher for the ET_0 data sets with lower resolution, this is not apparent in the analyses. The highest-resolution MSG shows the largest variability of R^2 among stations within the grids, which shows that higher resolution does not necessarily capture ET_0 variation at each site. Similar to Fig. 5, the performance indicators show that data sets with lower resolution show worse results compared to data sets with higher resolution. One exception is

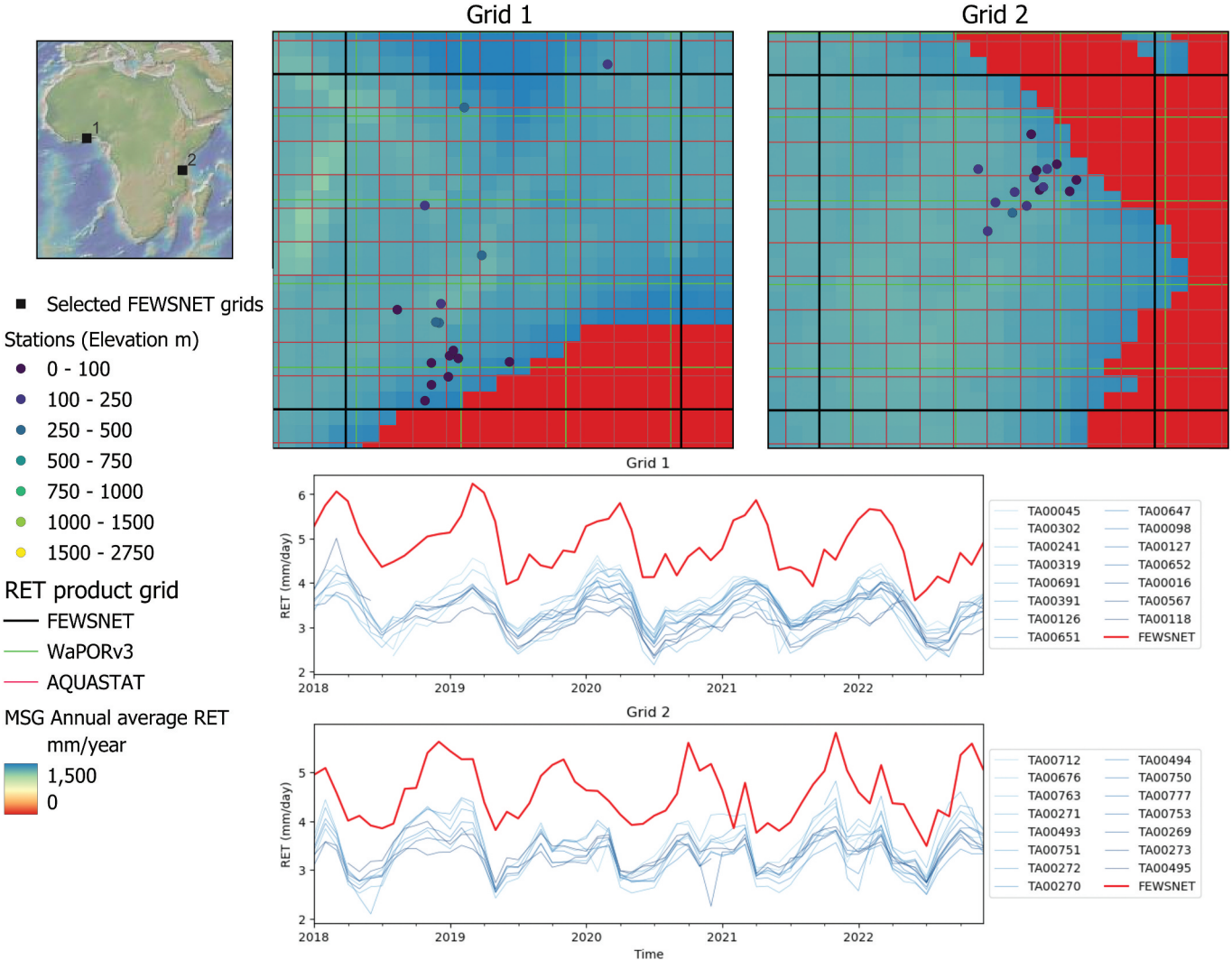

Figure 14. Spatial visualization of the grids from FEWSNET (Black), WaPORv3 (green), and AQUASTAT (red), along with the temporal variability of selected FEWSNET grids (Grid 1 and Grid 2) compared to station data within the corresponding grids from 2018 to 2022. The map (top left) highlights the location of the grids in Africa.

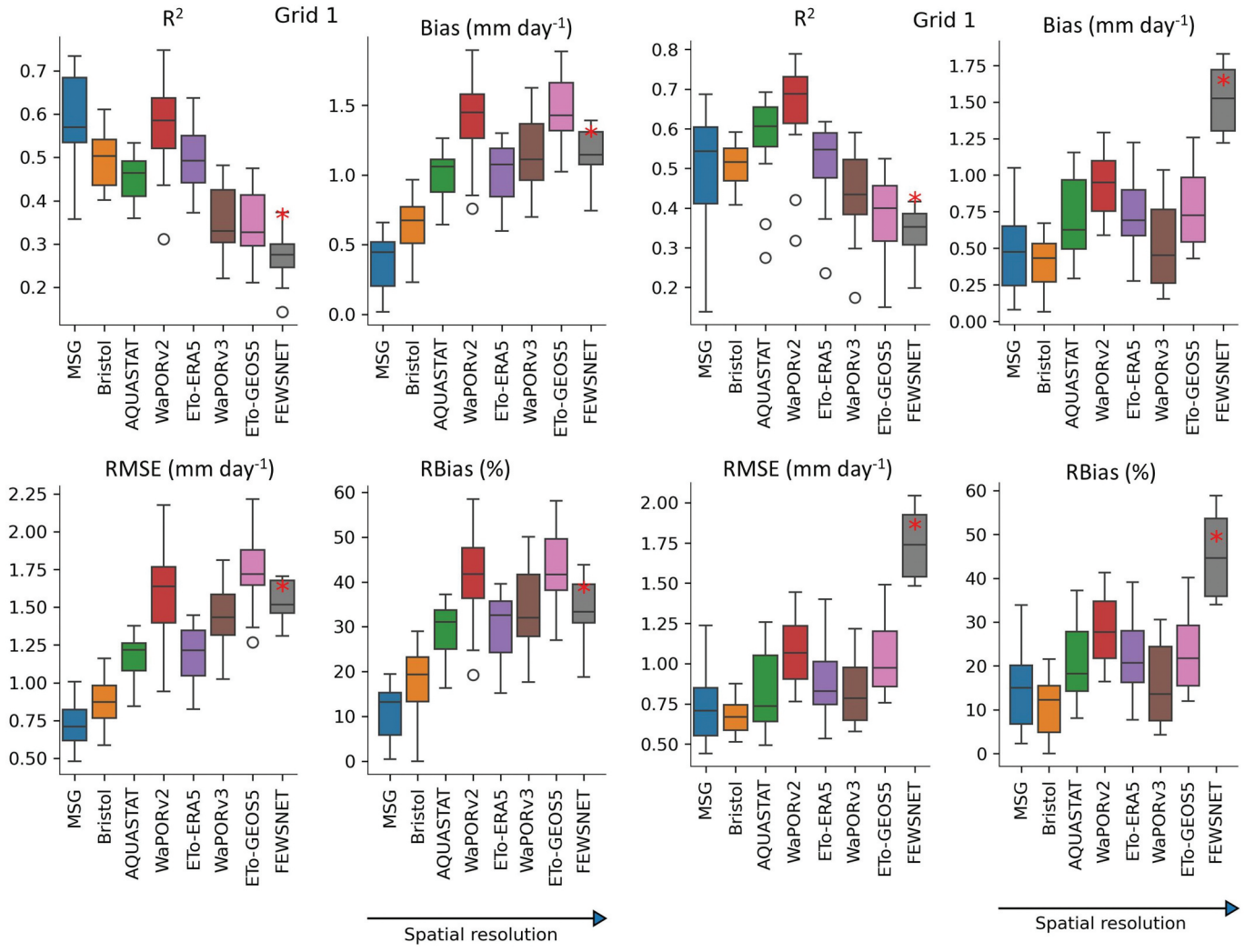

Figure 15. Validation metrics of stations within selected FEWSNET grids (Grid 1 and Grid 2). Outliers are shown as individual points, with boxes representing the interquartile range and the central line represents the median.

WaPORv2 data, which show good correlation with in-situ data, but with large bias and error.

The results show that the choice for the pixel to point validation approach did not influence the statistical results of this validation study. However, the high-resolution ET_0 data and station data show that the spatial variability for ET_0 is not captured well for the ET_0 data set with the lowest resolution (FEWSNET); however, higher-resolution data such as MSG do not necessarily mean better performance in each station.

6 Conclusion and recommendations

The primary aim of this study was to evaluate eight ET_0 data sets over Africa by comparing them against in-situ observations from 165 weather stations. The results reveal that the largest variance among the ET_0 data sets is observed in the drier parts of Africa (Bw climate) where some of the largest irrigation schemes of the continent are located. In particular the MSG estimates much lower ET_0 in those areas compared to the other ET_0 products. On the other hand, FEWSNET has the best performance in this climate class even though there is still a significant positive bias. The results are not conclusive as the number of stations (only three stations) in this climate class, which covers 44.3% of the continent, used for validation are not fully representative. The performance assessment shows that higher-resolution data sets (e.g. MSG, Bristol) overall perform better compared to low resolution ET_0 data sets (e.g. FEWSNET).

MSG may be suitable for applications where accurate average ET_0 values are important, particularly in tropical and temperate regions. Its low bias and RMSE across most climates support its use for magnitude-based analyses. In contrast, WaPORv2 shows the highest temporal correlation with in-situ ET_0 and may be more useful for tracking seasonal or short-term ET_0 dynamics. In arid and semi-arid regions, most data sets tend to overestimate ET_0, particularly ET_0-GEOS5 and FEWSNET, which show high bias. This can lead to excessive irrigation recommendations, resulting in inefficient water use. Where possible, local validation or bias correction of ET_0 data sets is recommended before using them in operational decision-making.

The results of this study show that the accuracy of ET_0 data sets depends not only on their spatial resolution, but also on the type and quality of the input data used. Our analysis comparing data sets that utilize identical models with varying input data versus data sets employing different models confirms that input data quality has large impact on ET_0 accuracy. Further investigation is required to precisely identify and address sources of positive Bias in FAO-56 Penman–Monteith-based data sets. Specifically, meteorological corrections, such as wind speed adjustment from 10 m to the standard 2 m height, may play a significant role in observed inaccuracies.

In summary, to effectively support irrigation planning and crop water requirement estimation in Africa, and to address the gaps in arid and semi-arid regions (BS/Bw), expanding weather-station coverage is essential, with priority given to irrigated areas where operational ET_0 accuracy is most critical. New sites should include high-quality radiation and wind measurements at standard heights to enable regional bias-correction of gridded inputs and support validation of spatial ET_0 data sets. Future versions of data sets such as WaPOR should use these observations for calibration and validation, integrating regional bias-correction of input meteorology. Additionally, improving the accuracy of input meteorological data is important, as this directly influences ET_0 data set performance and subsequently impacts agricultural productivity and water resource management across diverse African climates.

Acknowledgements

This research was supported by the Monitoring Land and Water Productivity by Remote Sensing (WaPOR Phase 2) project, funded by the Ministry of Foreign Affairs of the Netherlands (grant no. GCP/INT/729/NET). We are grateful for the support of the team from Trans-African Hydro-Meteorological Observatory (TAHMO) for making the meteorological data available for this study. We acknowledge the support of GPT-4, a language model created by OpenAI (version: GPT-4-turbo), which was used to detect programming errors and provide suggestions for improving the clarity, organisation, and structure of written content. All outputs generated by GPT-4 were carefully reviewed and verified by the authors to ensure that they aligned with the quality and standards expected of this work.

Data availability

ET_0-GEOS5 and ET_0-ERA5 data sets are available at https://doi.org/10.5281/zenodo.14418120. WaPORv2, WaPORv3, and AQUASTAT data sets are accessible at https://data.apps.fao.org/wapor/ (last access: 2024/12/13). The Bristol data set is accessible at https://doi.org/10.5523/bris.qb8ujazzda0s2aykkv0oq0ctp. The FEWSNET data set is accessible at https://earlywarning.usgs.gov/fews/product/81.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Ministerie van Buitenlandse Zaken [GCP/INT/729/NET].

ORCID

Suzan Dehati http://orcid.org/0009-0002-4361-4216
Bich Ngoc Tran http://orcid.org/0000-0001-6301-2699
Poolad Karimi http://orcid.org/0000-0002-5819-0981
Marloes Mul http://orcid.org/0000-0001-9469-3909

References

Abdullah, Q., 2024. Best practices in evaluating geospatial mapping accuracy according to the new ASPRS accuracy standards. *Photogrammetric Engineering & Remote Sensing*, 90 (5), 265–272. doi:10.14358/PERS.90.5.265

Aghdasi, F., 2010. *Crop water requirement assessment and annual planning of water allocation*. Master thesis. University of Twente. https://essay.utwente.nl/91580/

Ali, M.H., 2010. Crop water requirement and irrigation scheduling. *In:* M. H. Ali, ed. *Fundamentals of irrigation and on-farm water management: volume 1*. New York: Springer, 399–452. doi:10.1007/978-1-4419-6335-2_9

Allen, R., *et al.*, 1998. *Crop evapotranspiration: guidelines for computing crop water requirements*. FAO Irrigation and Drainage Paper No. 56. Rome: Food and Agriculture Organization of the United Nations (FAO).

Almorox, J., Quej, V.H., and Martí, P., 2015. Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, 528, 514–522. doi:10.1016/j.jhydrol.2015.06.057

Anand, M. and Molnar, P., 2018. Performance of TAHMO Zurich Weather Station: first tests with 6 months of data 2017/2018. doi:10.13140/RG.2.2.20185.16481

Annor, F.O., 2023. *Small reservoirs in Northern Ghana: monitoring, physical processes, and management*. Dissertation. TU Delft. doi:10.4233/uuid:81e5e8a8-2bee-4af5-b1bc-c7b210f9cb55

ASPRS Positional Accuracy Standards for Digital Geospatial Data, 2015. *Photogrammetric Engineering & Remote Sensing*, 81 (3), 1–26. doi:10.14358/PERS.81.3.A1-A26

Attia, A., *et al.*, 2021. Evaluating deficit irrigation scheduling strategies to improve yield and water productivity of maize in arid environment using simulation. *Agricultural Water Management*, 249, 106812. doi:10.1016/j.agwat.2021.106812

Barbosa, J.H.S., *et al.*, 2019. The influence of spatial discretization on HEC-HMS modelling: a case study. *International Journal of Hydrology*, 3 (5), 442–449. doi:10.15406/ijh.2019.03.00209

Beck, H.E., *et al.*, 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, 10 (1), 724. doi:10.1038/s41597-023-02549-6

Cardoso de Salis, H.H., *et al.*, 2019. Hydrologic modeling for sustainable water resources management in urbanized karst areas. *International Journal of Environmental Research & Public Health*, 16 (14), 2542. doi:10.3390/ijerph16142542

Dawson, C.W., Abrahart, R.J., and See, L.M., 2007. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, 22 (7), 1034–1052. doi:10.1016/j.envsoft.2006.06.008

De Bruin, H.A.R., 1987. From penman to makkink. *In*: *Evaporation and weather: technical meeting 44*, 25 March 1987 Ede, The Netherlands, 5–31.

de Bruin, H.D., *et al.*, 2016. A thermodynamically based model for actual evapotranspiration of an extensive grass field close to FAO reference, suitable for remote sensing application. *Journal of Hydrometeorology*, 17 (5), 1373–1382. doi:10.1175/JHM-D-15-0006.1

Dinku, T., 2019. Chapter 7—Challenges with availability and quality of climate data in Africa. *In*: A.M. Melesse, W. Abtew, and G. Senay, eds. *Extreme hydrology and climate variability*. Amsterdam, The Netherlands: Elsevier, 71–80. doi:10.1016/B978-0-12-815998-9.00007-5

Dombrowski, O., *et al.*, 2021. Performance of the ATMOS41 all-in-one weather station for weather monitoring. *Sensors*, 21 (3), 741. doi:10.3390/s21030741

FAO, 2020. *WaPOR database methodology*. Rome, Italy: FAO. doi:10.4060/ca9894en

FAO, 2021. *Reference evapotranspiration—AgERA5 derived (Global—Annual—~10km)—"FAO catalog"*. Available from: https://data.apps.fao.org/catalog//iso/0f594c49-58a2-4e65-85b9-1f23ce326001

FAO, 2024. *Reference evapotranspiration (Global—Annual—~30 km)—WaPOR v3—"FAO catalog"*. Available from: https://data.apps.fao.org/catalog/iso/7c0bc36b-9f92-49f9-9bd8-80272f250e0d

Giuntoli, I., *et al*., 2015. Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models. *Earth System Dynamics*, 6 (1), 267–285. doi:10.5194/esd-6-267-2015

Greaves, G.E. and Wang, Y.-M., 2017. Identifying irrigation Strategies for improved agricultural water productivity in irrigated maize production through crop simulation modelling. *Sustainability*, 9 (4), 630. doi:10.3390/su9040630

Hobbins, M., *et al*., 2023. A global long-term daily reanalysis of reference evapotranspiration for drought and food-security monitoring. *Scientific Data*, 10 (1), 746. doi:10.1038/s41597-023-02648-4

Ippolito, M., *et al*., 2024. Evaluation of daily crop reference evapotranspiration and sensitivity analysis of FAO Penman-Monteith equation using ERA5-Land reanalysis database in Sicily, Italy. *Agricultural Water Management*, 295, 108732. doi:10.1016/j.agwat.2024.108732

Jovanovic, N., *et al*., 2020. A review of strategies, methods and technologies to reduce non-beneficial consumptive water use on farms considering the FAO56 methods. *Agricultural Water Management*, 239, 106267. doi:10.1016/j.agwat.2020.106267

Kiraga, S., *et al*., 2024. Reference evapotranspiration estimation using genetic algorithm-optimized machine learning models and standardized Penman–Monteith equation in a highly advective environment. *Water*, 16 (1), 12. doi:10.3390/w16010012

Kohli, G., *et al*., 2020 *ECOSTRESS and CIMIS: A Comparison of Potential and Reference Evapotranspiration in Riverside County, California*. Available from: https://www.mdpi.com/2072-4292/12/24/4126 [Accessed 13 May 2025].

Lang, Q., *et al*., 2024. Effects of increasing spatial resolution on the spatial information content and accuracy of downward surface shortwave radiation. *International Journal of Applied Earth Observation and Geoinformation*, 133, 104128. doi:10.1016/j.jag.2024.104128

Li, X., *et al*., 2018. Intercomparison of six upscaling evapotranspiration methods: from site to the satellite pixel. *Journal of Geophysical Research: Atmospheres*, 123 (13), 6777–6803. doi:10.1029/2018JD028422

Liu, W. *et al*., 2021. Effect of elevation on variation in reference evapotranspiration under climate change in Northwest China. *Sustainability*, 13 (18). doi:10.3390/su131810151

Mahmoud, S.H. and Gan, T.Y., 2019. Irrigation water management in arid regions of Middle East: assessing spatio-temporal variation of actual evapotranspiration through remote sensing techniques and meteorological data. *Agricultural Water Management*, 212, 35–47. doi:10.1016/j.agwat.2018.08.040

Mayr, S., *et al*., 2019. Validation of earth observation time-series: a review for large-area and temporally dense land surface products. *Remote Sensing*, 11 (22), 2616. doi:10.3390/rs11222616

Mehta, R. and Pandey, V., 2015. Reference evapotranspiration (ETo) and crop water requirement (ETc) of wheat and maize in Gujarat. *Journal of Agrometeorology*, 17 (1). doi:10.54386/jam.v17i1.984

Ndiaye, P.M., *et al*., 2020. Evaluation and calibration of alternative methods for estimating reference evapotranspiration in the Senegal River Basin. *Hydrology*, 7 (2), 24. doi:10.3390/hydrology7020024

Paredes, P., *et al*., 2021. Daily grass reference evapotranspiration with Meteosat Second Generation shortwave radiation and reference ET products. *Agricultural Water Management*, 248, 106543. doi:10.1016/j.agwat.2020.106543

Pereira, L.S., *et al*., 2015. Crop evapotranspiration estimation with FAO56: past and future. *Agricultural Water Management*, 147, 4–20. doi:10.1016/j.agwat.2014.07.031

Radmanesh, Y., *et al*., 2023. Comparative evaluation of the accuracy of re-analysed and gauge-based climatic data in Iran. *Journal of Earth System Science*, 132 (4), 190. doi:10.1007/s12040-023-02202-1

Rajput, J., *et al*., 2024. Data-driven reference evapotranspiration (ET0) estimation: a comparative study of regression and machine learning techniques. *Environment, Development and Sustainability*, 26 (5), 12679–12706. doi:10.1007/s10668-023-03978-4

Raziei, T. and Parehkar, A., 2021. Performance evaluation of NCEP/NCAR reanalysis blended with observation-based datasets for estimating reference evapotranspiration across Iran. *Theoretical and Applied Climatology*, 144 (3), 885–903. doi:10.1007/s00704-021-03578-0

Senay, G.B., *et al*., 2008. Global Daily Reference Evapotranspiration Modeling and Evaluation1. *JAWRA Journal of the American Water Resources Association*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1752-1688.2008.00195.x

Shafer, M.A., *et al*., 2000. Quality assurance procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology*, 17 (4), 474–494. https://journals.ametsoc.org/view/journals/atot/17/4/1520-0426_2000_017_0474_qapito_2_0_co_2.xml

Singer, M.B., *et al*., 2021. Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present. *Scientific Data*, 8 (1), 224. doi:10.1038/s41597-021-01003-9

Singh, R., Kundu, D.K., and Bandyopadhyay, K.K., 2010. Enhancing agricultural productivity through enhanced water use efficiency. *Journal of Agricultural Physics*, 10 (2), 1–15.

Sun, J. *et al*., 2020. Elevation-dependent changes in reference evapotranspiration due to climate change. *Hydrological Processes*, 34 (26), 5580–5594. doi:10.1002/hyp.13978

Tran, B.N., *et al*., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. *Hydrology and Earth System Sciences*, 27 (24), 4505–4528. doi:10.5194/hess-27-4505-2023

Tran, B.N., *et al*., 2025. Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and the Southwest Asia. *Hydrological Sciences Journal*. doi:10.1080/02626667.2025.2600682

Trigo, I.F., *et al*., 2018. Validation of reference evapotranspiration from Meteosat Second Generation (MSG) observations. *Agricultural and Forest Meteorology*, 259, 271–285. doi:10.1016/j.agrformet.2018.05.008

van de Giesen, N., Selker, J., and Hut, R., 2014. The Trans-African Hydro-Meteorological Observatory (TAHMO). *WIREs Water*, 1 (4), 341–348. doi:10.1002/wat2.1034

Wang, L.W., *et al*., 2021. Synthesizing a regional territorial evapotranspiration dataset for Northern China. *Remote Sensing*, 13 (6), 1076. doi:10.3390/rs13061076

Weerasinghe, I., *et al*., 2020. Can we trust remote sensing evapotranspiration products over Africa? *Hydrology and Earth System Sciences*, 24 (3), 1565–1586. doi:10.5194/hess-24-1565-2020

Yonaba, R., *et al*., 2024. Accuracy and interpretability of machine learning-based approaches for daily ETo estimation under semi-arid climate in the West African Sahel. *Earth Science Informatics*, 18 (1), 87. doi:10.1007/s12145-024-01591-1

Zimba, H., *et al*., 2023. Phenophase-based comparison of field observations to satellite-based actual evaporation estimates of a natural woodland: miombo woodland, Southern Africa. *Hydrology and Earth System Sciences*, 27 (8), 1695–1722. doi:10.5194/hess-27-1695-2023

Zimba, H.M., *et al*., 2024. On the importance of plant phenology in the evaporative process of a semi-arid woodland: could it be why satellite-based evaporation estimates in the miombo differ? *Hydrology and Earth System Sciences*, 28 (15), 3633–3663. doi:10.5194/hess-28-3633-2024