Hydrological Sciences Journal

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

# Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and Southwest Asia

**Bich Ngoc Tran, Suzan Dehati, Solomon Seyoum, Johannes van der Kwast, Graham Jewitt, Remko Uijlenhoet & Marloes Mul**

**To cite this article:** Bich Ngoc Tran, Suzan Dehati, Solomon Seyoum, Johannes van der Kwast, Graham Jewitt, Remko Uijlenhoet & Marloes Mul (27 Jan 2026): Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and Southwest Asia, Hydrological Sciences Journal, DOI: 10.1080/02626667.2025.2600682

**To link to this article:** https://doi.org/10.1080/02626667.2025.2600682

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2025.2600682

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

# Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and Southwest Asia

Bich Ngoc Tran [a,b], Suzan Dehati [a,c], Solomon Seyoum [a], Johannes van der Kwast [a], Graham Jewitt [b,d], Remko Uijlenhoet [b] and Marloes Mul [a]

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

**ABSTRACT**

Recent developments of higher-resolution and lower-latency reanalysis data allow mapping reference evapotranspiration ($ET_o$) over large areas in a near real-time manner. This study evaluates the ERA5, AgERA5 and GEOS5 reanalysis datasets for meteorological input in Africa and Southwest Asia by comparing between data products and with 174 *in situ* sites. The inter-comparison reveals non-stationary differences between datasets and highlights temporal inconsistencies in the GEOS5 data. When evaluated against *in situ* measurements, GEOS5 demonstrates lower accuracy compared with ERA5 and AgERA5. Additionally, while all datasets accurately estimate air temperature and pressure, they overestimate windspeed and solar radiation, and underestimate vapour pressure. The propagation of uncertainty estimates of ERA5 through the FAO56 $ET_o$ equation shows particularly high uncertainty in the tropics. This study emphasizes the importance of applying multiple uncertainty assessment methods for better-informed use of reanalysis data, especially in data-scarce regions.

**ARTICLE HISTORY**
Received 1 November 2024
Accepted 29 September 2025

**EDITOR**
S. Archfield

**GUEST EDITOR**
J-M. Kileshye-Onema

**KEYWORDS**
reference evapotranspiration; reanalysis; Penman–Monteith; uncertainties; error propagation

## 1 Introduction

In many regions across Africa and Southwest Asia, a high percentage of the population experiences water shortage and water stress (Kummu *et al.* 2016). Climate change is projected to cause further water stress (Lelieveld *et al.* 2012, Kusangaya *et al.* 2014, Leal Filho *et al.* 2022), including regions with currently low physical water stress, such as Eastern and Central Africa (Adhikari *et al.* 2015, Abernethy *et al.* 2016). Agriculture, a key sector in Africa and Southwest Asia, is especially vulnerable to water stress, despite also being the largest consumer of water (de Pauw 2005, Hejazi *et al.* 2023). Thus, just and sustainable water management in agriculture is critically important for food and water security in these regions. This requires hydrological and meteorological information for estimating water demands and consumption, which is often limited and unevenly distributed over a vast land surface (Kusangaya *et al.* 2014).

In this context, estimating and mapping evaporation is instrumental in informing decisions to respond to water stress and ensure food security (Fisher *et al.* 2017). Evaporation is the transfer of water from liquid form to vapour in the atmosphere, which includes open water evaporation, soil water evaporation, plant transpiration and evaporation from canopy interception. This process is driven by the atmosphere's capacity to evaporate water when water is abundantly available, a concept introduced as potential evapotranspiration ($ET_p$) by Thornthwaite (1948). The term "potential" is sometimes confused with "reference" ET as described by Hargreaves and Samani (1982, 1985) (Xiang *et al.* 2020). However, these two terms need to be distinguished to avoid confusion and improper application (Xiang *et al.* 2020, Raza *et al.* 2022). The definition of PET lacks a strict definition of the evaporating surface. Because of its ambiguous definition, Allen *et al.* (1998) discouraged using the term PET when determining crop water requirement. Instead, reference ET ($ET_o$) is recommended, which is defined as the evapotranspiration rate of a hypothetical reference crop surface (e.g. grass or alfalfa) with uniform characteristics (i.e. surface albedo and crop height), without any water shortage or biophysical stress. The guideline to calculate $ET_o$ as reported in the Irrigation and drainage paper 56 published by the Food and Agriculture Organization (FAO) (Allen *et al.* 1998, Equation 6) is based on Penman–Monteith equation, hereafter called FAO56. The FAO56 report also provides a standard method to calculate crop water consumption or actual ET ($ET_a$) based on $ET_o$ values and crop coefficients. The FAO56 method has been popularized and applied extensively across multiple disciplines, especially for determining crop water requirement and consumption (Raza *et al.* 2022).

Almost all of the methods to calculate $ET_o$ (as well as $ET_a$), including the FAO56 method, require data of at least some meteorological variables that are often measured with standard

**CONTACT** Bich Ngoc Tran b.tran@un-ihe.org Land and Water Management Department, IHE Delft Institute for Water Education, PO Box 3015, Delft 2601 DA, the Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02626667.2025.2600682

weather stations, such as air temperature, pressure, humidity, windspeed and solar radiation. Since some variables are not always available, the FAO56 guideline provides alternative equations to estimate missing meteorological variables (Allen *et al.* 1998). Still, many regions lack weather stations, especially Africa and Southwest Asia (van de Giesen *et al.* 2014, Dinku 2019). Therefore, mapping $ET_o$ over large regions depends on either interpolated gridded weather datasets or climate reanalysis data (Abatzoglou *et al.* 2018, Singer *et al.* 2021). At the same time, the estimation of $ET_a$ aided by satellite observations is also dependent on models and forcing inputs (McCabe *et al.* 2017, Tran *et al.* 2023). Meteorological data are therefore essential forcing inputs for both $ET_o$ and $ET_a$ calculations. Despite the increasing development and use of spatial $ET_o$ data, there is limited understanding of the accuracy and precision of these data, and how they are related to the uncertainties of the meteorological input data.

Reanalysis is a compelling alternative to interpolated gridded weather datasets for mapping $ET_o$. Reanalysis is data generated using data assimilation techniques to couple numerical weather prediction (NWP) models with past observations. Recent developments in atmospheric reanalysis have greatly improved spatial resolution, notably the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) and its derived dataset AgERA5 (up to about 11 km resolution). In addition, the short latency of some NWP models and data assimilation systems facilitates operational near real-time monitoring of $ET_o$ and $ET_a$. For example, the FAO's portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) provides global daily $ET_o$ updated within 3 days calculated using the Goddard Earth Observing System version 5 (GEOS5) dataset, which has a latency of less than a day (FAO 2024). However, Parker (2016) forewarned that the reliability of reanalysis is unclear due to partial understanding of the errors and uncertainties in NWP models and thus advocated both quantitative and qualitative assessment of uncertainties in reanalysis data. The explicit and standardized quantification of uncertainties in NWP models and input parameters is often overlooked (Wang *et al.* 2024b). For example, Wang *et al.* (2024b) emphasized that uncertainties in cloud optical thickness, aerosol optical depth and ozone significantly impact solar radiation estimates, a key input for $ET_o$ and $ET_a$ calculation. Lang *et al.* (2024) demonstrated that coarse-resolution reanalysis data can introduce substantial errors in solar radiation estimates, due to mixed-pixel effects, especially under cloudy conditions (Wang *et al.* 2024a).

Since the sensitivity of $ET_o$ models to errors in meteorological forcing varies with different models, space and time (Fisher *et al.* 2017), it is important to analyse uncertainty in each $ET_o$ model when using reanalysis data. Furthermore, climate reanalysis and reanalysis-based evaporation estimates are increasingly used to study hydrological processes, which may lead to errors in these estimates being amplified and misinterpreted in hydrological studies. This necessitates a comprehensive review and description of both evaporation retrieval models and their forcing components (McCabe *et al.* 2017). The reliability of $ET_o$ calculated using particular reanalysis datasets has been investigated, mainly in southern Europe and China (e.g. Martins *et al.* 2017, Ippolito *et al.* 2024, Xu *et al.* 2024) where past weather observations for reanalysis are more available than in other parts of the world (Brönnimann *et al.* 2018, Soci *et al.* 2024). For instance, in Africa and Southwest Asia, where weather observations are scarce, the reanalysis data quality is largely unknown as well as the impact of the meteorological uncertainty on $ET_o$ calculations.

The objective of this article is therefore, to assess the uncertainty of meteorological forcing from reanalysis products (namely GEOS5, ERA5 and AgERA5) and the resulting uncertainty in $ET_o$ over Africa and Southwest Asia regions. In this study, we focused on $ET_o$ since the definition and calculation of $ET_o$ depends only on meteorological forcing. However, we also extend our discussion to the impact of uncertainty in meteorological forcing on $ET_a$ estimation where relevant.

## 2 Materials and methods

In this study, we assessed the uncertainty of three reanalysis data products (GEOS5, ERA5, AgERA5) for five meteorological inputs in the FAO56 $ET_o$ calculation: air temperature, atmospheric pressure, windspeed, vapour pressure and solar radiation for a five-year period (from 2018 to 2022). Our assessment entails three components: uncertainty between products, nominal accuracy and quantitative impact of uncertainty in inputs on $ET_o$ (Fig. 1). The uncertainty between products was assessed by spatial and temporal pair-wise comparison. The nominal accuracy[1] was assessed by comparison with time-series data from *in situ* measurements. Finally, the impact of uncertainty in inputs on $ET_o$ was assessed by two error propagation methods (Monte Carlo simulations and Taylor expansion).

### 2.1 Study area and in situ data

This study covers the land mass of Africa and Southwest Asia (30°S–60°N, 40°W–40°E) with a wide range of climates from arid to tropical (Fig. 2). About half of the study area is arid desert in the North Africa and Southwest Asia regions. There are a few climate monitoring networks that cover only fractions of the study area. The Trans-African Hydro-Meteorological Observatory (TAHMO) is an initiative that has successfully extended a network of meteorological and hydrological stations in sub-Saharan Africa (van de Giesen *et al.* 2014). Currently, TAHMO provides hydro-meteorological measurements from the largest number of stations in the region of interest.

*In situ* climate data were collected from 174 TAHMO stations (Fig. 2 and Table S3, see Supplementary material).

[1]The result from comparison with *in situ* measurement is considered nominal accuracy since *in situ* measurements also have errors and do not necessarily present true values of the grid cells due to spatial scale mismatch.

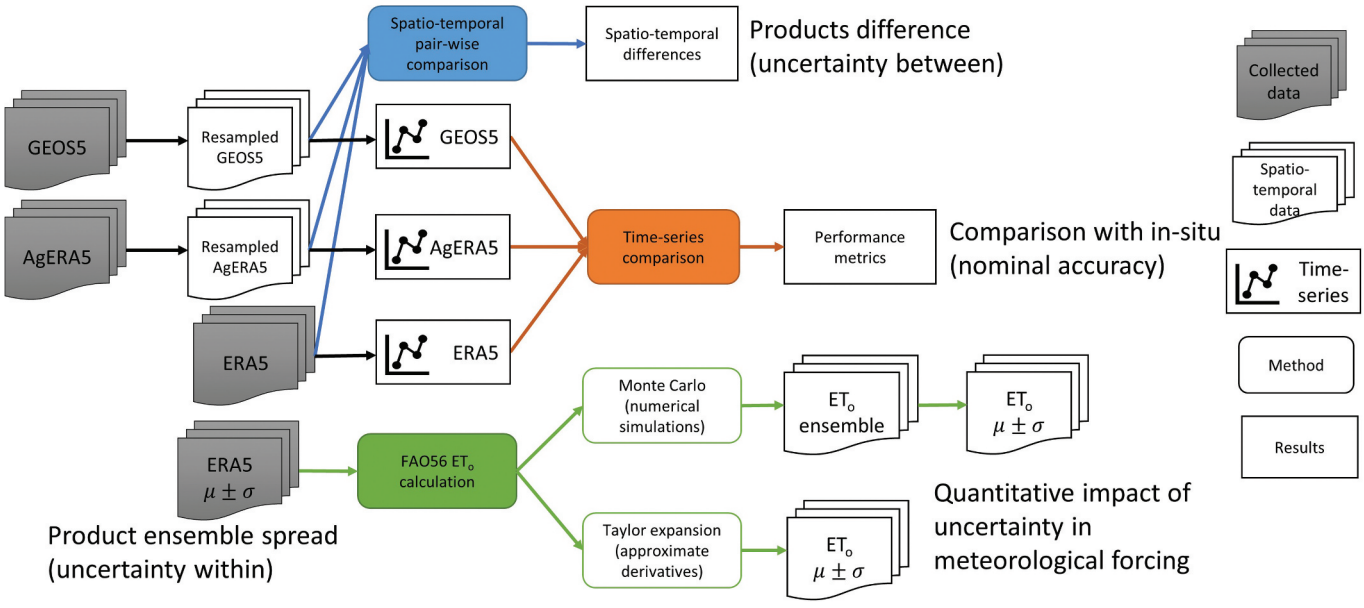


**Figure 1.** Schematization of the methodological framework. Collected reanalysis data (GEOS5, AgERA5 and ERA5) were resampled and processed for spatio-temporal pair-wise comparison to calculate difference between products, which represents uncertainty between products. Time-series were extracted at grid cells for comparison with *in situ* measurements, to calculate performance metrics, which represent nominal accuracy. The ensemble spread of ERA5, which represents uncertainty within ERA5 product, was used to propagate errors in FAO56 reference evapotranspiration ($ET_o$) calculation using Monte Carlo and Taylor expansion methods.

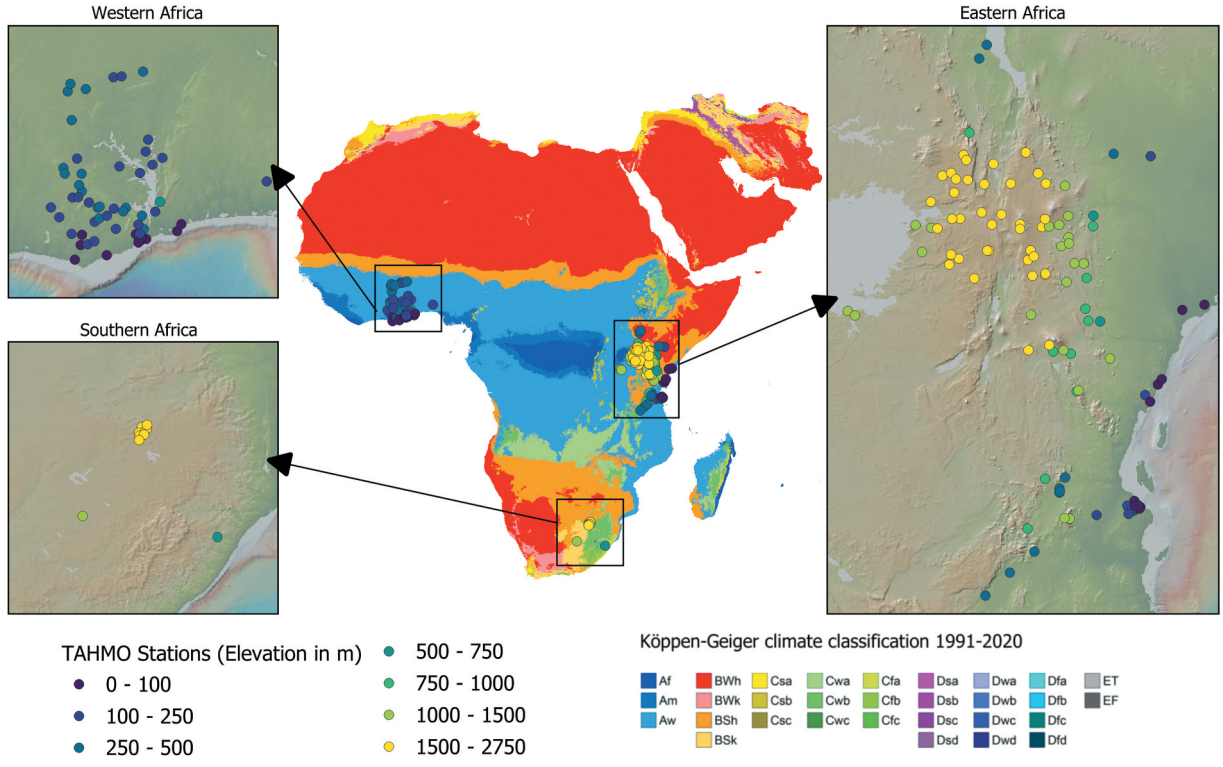


**Figure 2.** Climate classification map of study area and the locations of *in situ* observations. Data source: TAHMO, Köppen–Geiger map (Beck *et al.* 2023). Base map: Natural Earth NE1_50M_SR_W.

Table S1 (see Supplementary material) shows the distribution of TAHMO stations by climate classes. The dataset includes hourly measurement of standard meteorological variables (i.e. air temperature, relative humidity, wind speed and direction, solar radiation and atmospheric pressure) from ATMOS 41 sensors (METER 2023). Daily mean air temperature (°C), relative humidity (%), wind speed at 2 m (m $s^{-1}$), air pressure (mbar) and solar radiation (W $m^{-2}$) were computed by averaging hourly data (TAHMO 2023).

The quality of *in situ* data varies and depends on the accuracy specification of sensors (see Supplementary material,

Table S2), sensor performance and operation continuity. We acquired data from stations with the best quality flags provided by the TAHMO quality control procedure (van de Giesen *et al.* 2014). Quality control procedure for TAHMO stations is extended from the procedures in the Oklahoma Mesonetwork (Shafer *et al.* 2000). TAHMO employs both automated and manual methods to ensure data quality, including tests for range, sensor accuracy, climate condition, temporal changes, dips and spikes, changes in variance (Annor 2023, p. 125). In addition, we evaluated the quality and integrity of the acquired dataset by visually checking the timeseries of each climatic variable at each station.

## 2.2 Reanalysis data

### 2.2.1 Datasets description

We selected and retrieved data for a study period of 5 years from 1/1/2018 to 31/12/2022, which we considered a sufficient sample of daily values for analysing spatio-temporal pattern of data uncertainty (1826 data points per time-series). This is also the period for which we have access to TAHMO climate data. The hourly GEOS5 and ERA5 data and daily AgERA5 data were retrieved from the sources that are provided in Table 1.

GEOS5 stands for the Goddard Earth Observing System version 5, a global atmospheric model, developed by the NASA Global Modelling and Assimilation Office (GMAO) (Rienecker *et al.* 2008). The GEOS Data Assimilation System is the integration of GEOS5 Atmospheric General Circulation Model and the Gridpoint Statistical Interpolation Analysis (Rienecker *et al.* 2008). GMAO runs the GEOS Forward Processing stream, which generates both forecasts and assimilation products. The meteorological variables from the time-average 1-hourly data, 2-dimensional, single-level[2] atmospheric state variables (tavg1_2d_slv_Nx) and radiative fluxes (tavg1_2d_rad_Nx) data products were retrieved. The GEOS5 dataset includes the following variables: hourly air temperature at 2 m (K), northward and eastward components of wind at 10 m (m $s^{-1}$), specific humidity at 2 m (kg $kg^{-1}$), sea level pressure (Pa), surface pressure (Pa) and surface incoming shortwave flux (J $m^{-2}$ $h^{-1}$). The GEOS5 system derives air pressure at surface level from mean sea level pressure using the United States Geological Survey's 1-km Global Elevation (GTOPO30) raster data (Rienecker *et al.* 2008).

ERA5 is the fifth generation of the ECMWF atmospheric reanalysis of the global climate. ERA5 is generated by combining the model forecasts from the Integrated Forecasting System Cy41r2 with vast amounts of historical observations using the 4D-Var assimilation scheme (Hersbach *et al.* 2020). ERA5 provides data products for several climate variables at 137 pressure levels from the surface up to 80 km. In this study, the ERA5 hourly data on single levels (Hersbach *et al.* 2023) was retrieved from ECMWF's Climate Data Store (CDS). The ERA5 dataset includes the following variables: hourly 2 m temperature (K), 2 m dewpoint temperature (K), 10 m v-component and u-component of wind speed (m $s^{-1}$), mean sea level pressure (Pa), surface pressure (Pa) and surface solar radiation downwards (J $m^{-2}$ $h^{-1}$). ERA5 surface pressure was computed using surface elevation data interpolated from the Shuttle Radar Topography Mission Digital Elevation 30 m data (SRTM30) combined with other surface elevation datasets (ECMWF, 2023a).

AgERA5 comprises agrometeorological indicators derived from reanalysis, providing input needed for most crop growth models. Daily AgERA5 data is produced by aggregating ERA5 to daily time steps at the local time zone and downscaling towards a finer topography at a 0.1° spatial resolution. The ERA5 data was corrected using regression equations that were calibrated with the ECMWF's operational high-resolution atmospheric model (HRES) for each variable and grid (Boogaard *et al.* 2020). AgERA5 provides data products for 12 agro-meteorological variables at the surface level and daily timestep. The AgERA5 dataset was collected from ECMWF's CDS, which includes the following variables: daily 2 m temperature (K), 2 m dewpoint temperature (K), 10 m wind speed (m $s^{-1}$) and solar radiation flux (J $m^{-2}$ $d^{-1}$).

### 2.2.2 Reanalysis data pre-processing

Prior to our analyses, we applied a pre-processing procedure on the retrieved reanalysis data products to ensure consistency among variables and units for accurate comparisons. Simple linear temporal and spatial aggregation were applied to achieve the same resolution for comparison of different input datasets. For spatial and temporal

**Table 1.** Spatial and temporal resolution of the three reanalysis datasets used in the study. The specific data products were acquired from the data source given in parentheses.

| Dataset name | Spatial resolution | Temporal resolution | Product name (data source) |
|---|---|---|---|
| GEOS5 | 0.31° × 0.25° (~30 km at the equator) | hourly | GEOS-FP tavg1_2d_slv_Nx (https://opendap.nccs.nasa.gov/dods/GEOS-5/fp/0.25_deg/assim/tavg1_2d_slv_Nx.info) GEOS-FP tavg1_2d_rad_Nx (https://opendap.nccs.nasa.gov/dods/GEOS-5/fp/0.25_deg/assim/tavg1_2d_rad_Nx.info) |
| ERA5 | 0.25° × 0.25° (~28 km at the equator) | hourly | ERA5 hourly data on single levels from 1940 to present (https://doi.org/10.24381/cds.adbb2d47) (Hersbach *et al.* 2023) |
| AgERA5 | 0.1° × 0.1° (~11 km at the equator) | daily | Agrometeorological indicators from 1979 to present derived from reanalysis (https://doi.org/10.24381/cds.6c68c9bb) (Boogaard *et al.* 2020) |

[2]In the context of climate reanalysis, single level data includes variables measured or modelled close to the surface.

comparison, we analysed meteorological variables at the coarser resolution of the data products to avoid introducing errors due to spatial downscaling and elevation correction. Reanalysis data was converted to the same unit as *in situ* data.

The average windspeed at 10 m is derived from the windspeed components collected from ERA5 and GEOS5:

$$u_{10} = \sqrt{u_{10x}{}^2 + u_{10y}{}^2} \tag{1}$$

where $u_{10}$ is wind speed at 10 m (m $s^{-1}$), $u_{10x}$ is 10 m eastward wind or u-component (m $s^{-1}$) and $u_{10y}$ is 10 m northward wind or v-component (m $s^{-1}$).

The wind speed at 2 m was estimated from reanalysis windspeed at 10 m using the logarithmic wind speed profile (Allen *et al.* 1998, Eq. 47):

$$u_2 = u_{10}\frac{4.87}{\ln(67.8 \times z - 5.42)} \tag{2}$$

where $u_{10}$ wind speed at 10 m (m $s^{-1}$), $u_2$ wind speed at 2 m (m $s^{-1}$) and $z$ = 10 m is the height at which wind speed is calculated.

Notably, the reanalysis datasets do not have relative humidity except for AgERA5. Therefore, we compared vapour pressure derived from ERA5 and AgERA5 dew-point temperature and from GEOS5 specific humidity, which is required for the calculation of $ET_o$. The calculation of vapour pressure from specific humidity is:

$$e_a = \frac{q_v \cdot P}{\varepsilon} \tag{3}$$

where $e_a$ is vapour pressure (kPa), $q_v$ is specific humidity (kg $kg^{-1}$), $P$ is air pressure (kPa) and $\varepsilon$ is the ratio of molecular weight of water to dry air ($\varepsilon$ = 0.622).

The saturation vapour pressure at actual temperature was calculated following Allen *et al.* (1998, Eq. 11):

$$e_s(T) = 0.6108\ exp\left(\frac{17.27T}{T + 237.3}\right) \tag{4}$$

where $e_s(T)$ is saturated vapour pressure (kPa) at the actual air temperature (°C).

The actual vapour pressure $e_a$ (kPa) equals the saturated vapour pressure $e_s$ at the dewpoint temperature $T_d$ (°C): $e_a = e_s(T_d)$. Therefore, the saturated vapour pressure was calculated by substituting $T_d$ from reanalysis data for $T$ in Equation (4) following Allen *et al.* (1998, Eq. 14).

For *in situ* dataset, vapour pressure was derived from minimum and maximum relative humidity following Allen *et al.* (1998, Eq. 17) since dew-point temperature was not available:

$$e_a = \frac{e_s(T_{min}) \times \frac{RH_{min}}{100} + e_s(T_{max}) \times \frac{RH_{max}}{100}}{2} \tag{5}$$

where $RH_{min}$ is daily minimum relative humidity (%), $RH_{max}$ is daily maximum relative humidity (%) and $e_s(T)$ is the saturation vapour pressure (kPa) at the same temperature ($T$).

### 2.3 Spatial and temporal pair-wise comparison

The uncertainty between products was assessed by pair-wise comparison. Before that, we aggregated hourly reanalysis data to daily, daily to monthly, and monthly to yearly by arithmetic averaging. The higher-resolution datasets (GEOS5 and AgERA5) were resampled to the spatial resolution of ERA5 to ensure that all datasets represent the same level of detail, allowing for an unbiased comparison. For spatial comparison, we computed the yearly average of the differences between each pair for mean air temperature at 2 m, wind speed at 10 m, vapour pressure and solar radiation. For atmospheric pressure, we compared the pressure at sea level and the pressure at surface originally retrieved from GEOS5 and ERA5. Since the AgERA5 dataset does not include data of air pressure, it was excluded from air pressure comparison. For spatio-temporal comparison, we used the latitude-time Hovmöller diagrams of monthly average maps (Hovmöller 1949), which helps visually detect seasonal anomalies or any dynamics of the discrepancy between datasets.

### 2.4 Comparison with in situ data and performance metrics

The nominal accuracy of reanalysis data was evaluated by comparison with *in situ* measurements. Daily time-series of air temperature, air pressure, windspeed at 2 m, vapour pressure and solar radiation were extracted from the reanalysis datasets at the grids containing observation stations. The performance metrics we used to validate reanalysis data against *in situ* data include the coefficient of determination ($r^2$); the root mean square error (RMSE), the bias and the relative bias (PBIAS) (Table 2). The $r^2$ metric is the square of Pearson correlation coefficient, which measures how well the variables from *in situ* data are correlated to the temporal variation of variables derived from reanalysis

**Table 2.** Performance metrics used to validate the reanalysis data with *in situ* observations. For all equations, $x$ represents the value from reanalysis product, $y$ represents the value from *in situ* data and $i$ represents the time step.

| Evaluation metrics | Formula | Unit | Value range | Best value |
|---|---|---|---|---|
| Coefficient of determination | $r^2 = \frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}$ | – | [0,1] | 1 |
| BIAS | $BIAS = \sum_{i=1}^{n}\frac{(x_i-y_i)}{n}$ | Same unit as $x$ and $y$ | (-∞, +∞) | 0 |
| PBIAS | $PBIAS = \frac{BIAS \times 100}{\bar{x}}$ | % | (-∞, +∞) | 0 |
| Root Mean Square Error (RMSE) | $RMSE = \sqrt{\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{n}}$ | Same unit as $x$ and $y$ | [0, +∞) | 0 |

products. The bias measures mean residuals, while the RMSE measures the root mean square difference between reanalysis and *in situ* data. These metrics are widely used in *in situ* validation of earth observation data (Mayr *et al.* 2019, Tran *et al.* 2023).

### 2.5 Error propagation methods

We analysed the impact of uncertainty in $ET_o$ propagated from the meteorological inputs from reanalysis on $ET_o$ by applying error propagation methods in the FAO56 calculation of $ET_o$. We applied and compared the Monte Carlo (MC) method with the Taylor method.

The MC method is a statistical approach to estimate the uncertainty in a complex mathematical model by performing random sampling. It entails randomly simulating inputs based on known or assumed probability distributions, applying these inputs to the model, and deriving the uncertainty and variability from the resulting outputs. The MC method is well-suited for non-linear functions and models with multiple variables. However, it can be computationally expensive and time-consuming when a large number of simulations are needed to obtain accurate estimates of the probability distribution.

The Taylor method is based on the theory of error propagation, which applies the Taylor expansion for linear approximation of non-linear functions (Taylor 1997). The main advantages of using the Taylor method are efficient computation and the analytical form of the variance of the output error. However, when the operation $\boldsymbol{g}(\cdot)$ is strongly non-linear or involves many inputs, like the $ET_o$ calculation, the approximation error may increase and computational efficiency may decrease (Heuvelink 1998, p. 43). Therefore, comparing the Taylor method and the MC method can provide more insights to guide future applications of $ET_o$ error propagation. Our intention was to evaluate whether the Taylor method can be an alternative to the MC method for operational $ET_o$ uncertainty estimation.

#### 2.5.1 Calculation of FAO56 reference evapotranspiration

The daily $ET_o$ (mm $d^{-1}$) was calculated following the FAO56 Penman–Monteith equation for reference crop following the procedure described by Allen *et al.* (1998):

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_{mean}+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (6)$$

where $\Delta$ is the slope of saturation vapour pressure curve (kPa $^oC^{-1}$), $R_n$ is the net radiation at the reference crop surface (MJ $m^{-2}$ $d^{-1}$), $G$ is the soil heat flux density (MJ $m^{-2}$ $d^{-1}$) assumed to be zero for day period, $T_{mean}$ is the daily mean air temperature ($^oC$), $\gamma$ is the psychrometric constant (kPa $^oC^{-1}$), $e_s$ is saturation vapour pressure (kPa), $u_2$ is daily average wind speed at 2 m (m $s^{-1}$) and $e_a$ is actual vapour pressure (kPa). $T_{mean}$ was calculated as the average of minimum and maximum air temperature ($T_{min}$ and $T_{max}$):

$$T_{mean} = \frac{T_{min} + T_{max}}{2} \quad (7)$$

The slope of saturation vapour pressure curve ($\Delta$) was calculated following Allen *et al.* (1998, Eq. 13). The psychrometric constant ($\gamma$) was calculated following Allen *et al.* (1998, Eq. 8).

The net radiation ($R_n$) was calculated by subtracting the net longwave radiation ($R_{nl}$) from the net shortwave solar radiation ($R_{ns}$):

$$R_n = R_{ns} - R_{nl} = (1 - \alpha)R_s - R_{nl} \quad (8)$$

where $\alpha = 0.23$ is surface albedo for the hypothetical grass reference and $R_s$ is the incoming solar radiation from reanalysis data (MJ $m^{-2}$ $d^{-1}$). was calculated following Allen *et al.* (1998, Eq. 39):

$$R_{nl} = \sigma \left[ \frac{(T_{min} + 273.16)^4 + (T_{max} + 273.16)^4}{2} \right] \times (0.34 - 0.14\sqrt{e_a}) \times \left(1.35 \frac{R_s}{R_{so}} - 0.35\right) \quad (9)$$

where $\sigma = 4.903 \cdot 10^{-9}$ MJ $K^{-4}$ $m^{-2}$ $d^{-1}$ is Stefan–Boltzmann constant, $\left(0.34 - 0.14\sqrt{e_a}\right)$ and $\left(1.35 \frac{R_s}{R_{so}} - 0.35\right)$ are correction terms for air humidity and cloudiness respectively and $R_{so}$ is the clear-sky radiation (MJ $m^{-2}$ $d^{-1}$) calculated from the extraterrestrial radiation $R_a$, which was, in turn, calculated following Allen *et al.* (1998, Eq. 21).

#### 2.5.2 Uncertainty estimation of reanalysis datasets

Since only the ERA5 data product comes with uncertainty estimation, we focused on the propagation of ERA5 data uncertainty estimates in the FAO56 $ET_o$ calculation. The uncertainty estimates of the ERA5 data product was produced by ECMWF by sampling from 10 underlying ensemble members every 3 hours (ECMWF 2023b). This uncertainty estimation addresses mostly random errors in the observations and sea surface temperature model parameterization. The estimates are closely related to the uncertainty of the assimilated observations, which have evolved considerably over time. The ERA5 uncertainty estimates vary in different zones due to the uneven distribution of measurements that are used to correct the forecasting model through data assimilation. However, the systematic errors were not addressed and not correlated to the computed uncertainty estimates (ECMWF, 2023b).

The mean and spread (standard deviation) of the ERA5 ensemble (see Supplementary material, Fig. S1) were retrieved from ECMWF's CDS (Hersbach *et al.* 2023). We considered each meteorological variable as a quantitative spatial attribute $\boldsymbol{A}(\cdot) = \{\boldsymbol{A}(\boldsymbol{x}) | \boldsymbol{x} \in D\}$ where D is the domain of interest in a 3-dimensional space (longitude-latitude-time). Then, $\boldsymbol{A}(x)$ is the value of $\boldsymbol{A}(\cdot)$ at a certain point in time and space. An error model of that variable was assumed: $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{V}(\boldsymbol{x})$, where $\boldsymbol{b}(\boldsymbol{x})$ is a deterministic function of $x$, and $\boldsymbol{A}(\boldsymbol{x})$ and $\boldsymbol{V}(\boldsymbol{x})$ are random variables (Heuvelink 1998, p. 10). We assumed that $\boldsymbol{V}(\boldsymbol{x})$ were the random errors that follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$, where $\mu$ is the mean (equal to zero) and $\sigma^2$ is the variance. Instead of assuming an error model for $\sigma(\boldsymbol{x})$, we utilized the hourly ensemble spread $\sigma(\boldsymbol{x})$ at each ERA5 grid,

which presents the temporal and spatial variability of errors, the covariance of meteorological variables and autocorrelation.

#### 2.5.3 Monte Carlo method

The Monte Carlo (MC) method was applied to propagate the random errors from the aforementioned five meteorological variables to $ET_o$. Since each daily output $ET_o$ map is computed from the five input maps $\boldsymbol{A_i}(\cdot)(i = 1, \ldots, 5)$, the output $ET_o$ maps are also random functions: $\boldsymbol{U}(\cdot) = \boldsymbol{g}(\boldsymbol{A}_1(\cdot), \ldots, \boldsymbol{A}_5(\cdot))$, where $\boldsymbol{g}(\cdot)$ is the FAO56 $ET_o$ calculation. The MC method randomly samples $N$ sets of realizations $\boldsymbol{a_{i,j}}(\boldsymbol{x})(j = 1, \ldots, N)$ from the distribution of $\boldsymbol{A_i}(\boldsymbol{x})$ described by the daily ERA5 ensemble mean and standard deviation maps. For each set of realizations, an $ET_o$ map $\boldsymbol{u_j}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{a}_{1,j}(\boldsymbol{x}), \ldots, \boldsymbol{a}_{5,j}(\boldsymbol{x}))$ was calculated. The error of $ET_o$ was estimated by calculating the statistics of $N$ outputs $\boldsymbol{u_j}(\boldsymbol{x})$, including standard deviation normalized with mean values ($\sigma_{norm}$), and percentiles.

Although the MC method can generate the entire distribution of $\boldsymbol{u_j}(\boldsymbol{x})$, the level of accuracy may be arbitrary depending on the method of the random generator and number of simulations. Since the efficiency of MC method is proportional to $\sqrt{N}$ (Heuvelink 1998, p. 106), we experimented with $N$ between 100 to 10,000 (10 times less efficient) to assess the impact of $N$ on the results. Due to the computational burden of random sampling, we employed Latin-Hypercube sampling (LHS) (Stein 1987, Lee *et al.* 2014) to improve computational efficiency.

#### 2.5.4 Taylor method

For implementing the Taylor method, we utilized the Python package *uncertainties* (Lebigot 2017) to facilitate the calculation of derivatives. For that, we defined the daily ensemble mean and ensemble spread of ERA5 reanalysis products as the nominal mean and standard deviation of uncertain arrays (*uncertainties.unumpy.uarray*) for each meteorological variable. The *uarray* objects of each meteorological variable were used directly as inputs in each equation in the FAO56 $ET_o$ calculation procedure to estimate the uncertain arrays of daily $ET_o$ over the study area. The advantage of this step-wise calculation is that we can analyse also the propagated errors of intermediate variables, such as $\gamma$ and $R_n$.

## 3 Results and discussion

### 3.1 Spatio-temporal comparison of reanalysis datasets

The scatterplots of meteorological variables show that all products are well correlated grid by grid (Fig. 3). The correlation between paired datasets for air pressure, air temperature and vapour pressure are very high (0.97 to above 0.99). Meanwhile, windspeed (0.85–0.94) and solar radiation (0.82–0.98) show slightly lower correlation, especially between GEOS5 and other two products. The data points in the scatterplot exhibit the largest spread for solar radiation, followed by windspeed, vapour pressure, air temperature and air pressure. The highest correlation in the three pairs is between AgERA5 and ERA5 (Fig. 3(a, d, g, j)), which is expected since AgERA5 is derived from ERA5. However, the comparison between AgERA5 and ERA5 shows the largest mean bias of air temperature (−0.41°C) and windspeed (0.14 m $s^{-1}$) among three pairs (Fig. 3(a, d)).

The difference between datasets varies with location (Fig. 4). In case of air temperature, the largest difference between AgERA5 and ERA5 can be seen in areas with variable elevation (e.g. the Great Rift Valley and the Iranian highlands) (Fig. 4(a) and Fig. S3a, see Supplementary material). This can be explained by the fact that an elevation correction using vertical lapse rate is applied to derive AgERA5 from ERA5 (Boogaard *et al.* 2023). GEOS5 air temperature is lower near the equator, while it is higher at mid-latitude when compared with both ERA5 and AgERA5 (Fig. 4(b, c)).

The difference between AgERA5 and ERA5 windspeed is close to zero in the majority of the grid cells, and slightly negative near the coastline and mountainous areas (Fig. 4 (d)). This is likely due to the topographical correction algorithm of AgERA5. Compared with GEOS5, both AgERA5 and ERA5 show greater windspeed below the Sahel, in Iran and in Southern Africa and lower windspeed in the Sahara and Congo rainforest, with differences up to 3 m $s^{-1}$ (Fig. 4(e, f)). Moreover, the difference between the products over the Congo rainforest increases below the equator. Given that GEOS5 is prone to a larger error (Rienecker *et al.* 2008, p. 25) and ERA5 exhibits lower uncertainty (see Supplementary material, Fig. S1) in the southern hemisphere, the larger difference between the two products is likely due to the errors in GEOS5.

In the case of vapour pressure, the difference between products is greater in the middle range (between 20 and 25 mbar) (Fig. 3(g–i)). The difference between AgERA5 and ERA5 is slightly positive over the majority of the area (lower than 2 mbar) and greater in Western Africa and the coastline of the Red Sea (Fig. 4(g)). The difference in vapour pressure between GEOS5 and ERA5 is highest among the three pairs. GEOS5 vapour pressure is found to be higher than ERA5 and AgERA5 in the Sahel and Congo rainforest, but lower in the Ethiopian highlands (Fig. 4(h, i)).

GEOS5 estimates higher solar radiation values than do ERA5 and AgERA5 in the humid tropical parts of Western and Central Africa (10°S–20°N, 0°–30°E), and lower over the rest of the continent (Fig. 4(k, l)). Although AgERA5 is derived from ERA5, it shows slightly lower solar radiation values than ERA5 in most areas, except for the mountainous areas around Lake Victoria and Ethiopia (Fig. 4(j)). This is likely due to the topographic correction used for AgERA5.

When comparing air pressure at the surface, ERA5 and GEOS5 exhibit greater similarity than for other variables (Fig. 3(m)). Differences between the two products are primarily observed in regions with high elevation and coastlines (Fig. 4(m)). However, these differences appear to be random and only present in a small number of grid cells, which does not affect the overall correlation. The difference between the two datasets for air pressure at the surface is much smaller than the difference in air pressure at sea level (see Supplementary material, Fig. S2), which underlines the effect of topographical correction of air pressure for both datasets.

The differences between the datasets vary seasonally, especially for air temperature, vapour pressure and solar radiation

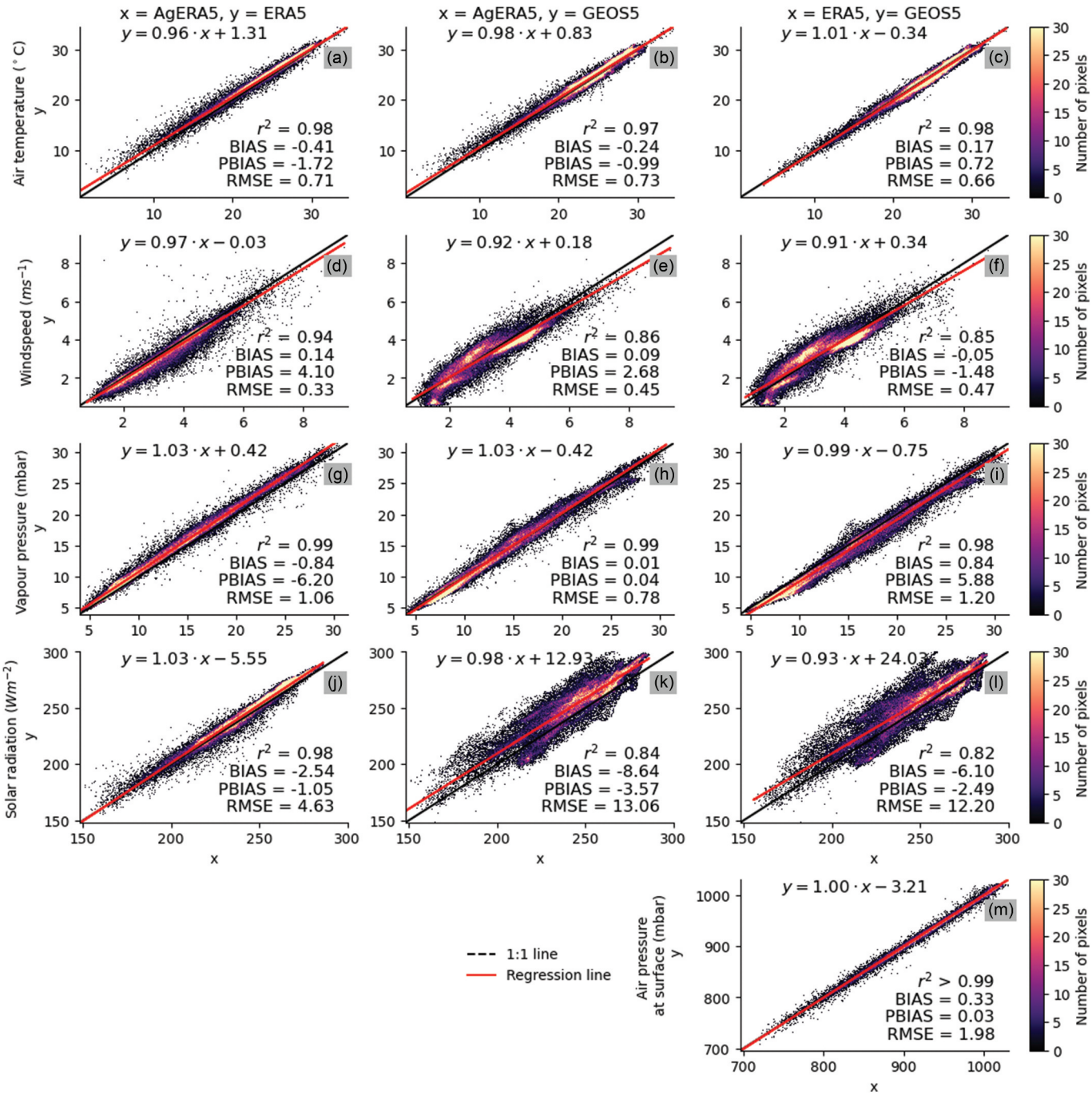


**Figure 3.** Comparison of mean annual air temperature, windspeed, vapour pressure, solar radiation and air pressure at the surface from ERA5, AgERA5 and GEOS5 for the period 2018–2022. The scatterplots show the correlation between each pair of two datasets (column) for each meteorological variable (row). The performance metrics ($r^2$, RMSE, BIAS, PBIAS) and linear regression coefficients were calculated grid-wise by reshaping mean annual 2-dimensional arrays into 1-dimensional series, showing the spatial correlation between two products.

(Fig. 5). The highest discrepancy in air temperature between the GEOS5 and ERA5 or AgERA5 products occurs between May and July (up to 2°C). Moreover, during these months, the differences of air temperature near the equator are larger than the other months. In the case of vapour pressure, the largest difference between GEOS5 and the other two products occurs between May and July below the equator and between July and September above the equator. The vapour pressure difference shows a seasonal pattern aligning with air temperature differences, but it has an inverse relationship – positive air temperature differences correspond to negative vapour pressure differences, and vice versa. The difference in solar radiation between GEOS5 and the other two products also varies seasonally. From June to October, GEOS5 provides higher solar radiation values, whereas for the rest of the year, GEOS5 values of solar radiation are lower below the equator. In the case of windspeed, the latitudinal pattern of differences between the products is mostly consistent throughout the year. The difference in air pressure at the surface is also very low (<5 mbar) and not seasonally variable.

The time-series of air temperature, windspeed and vapour pressure difference between GEOS5 and the other two datasets

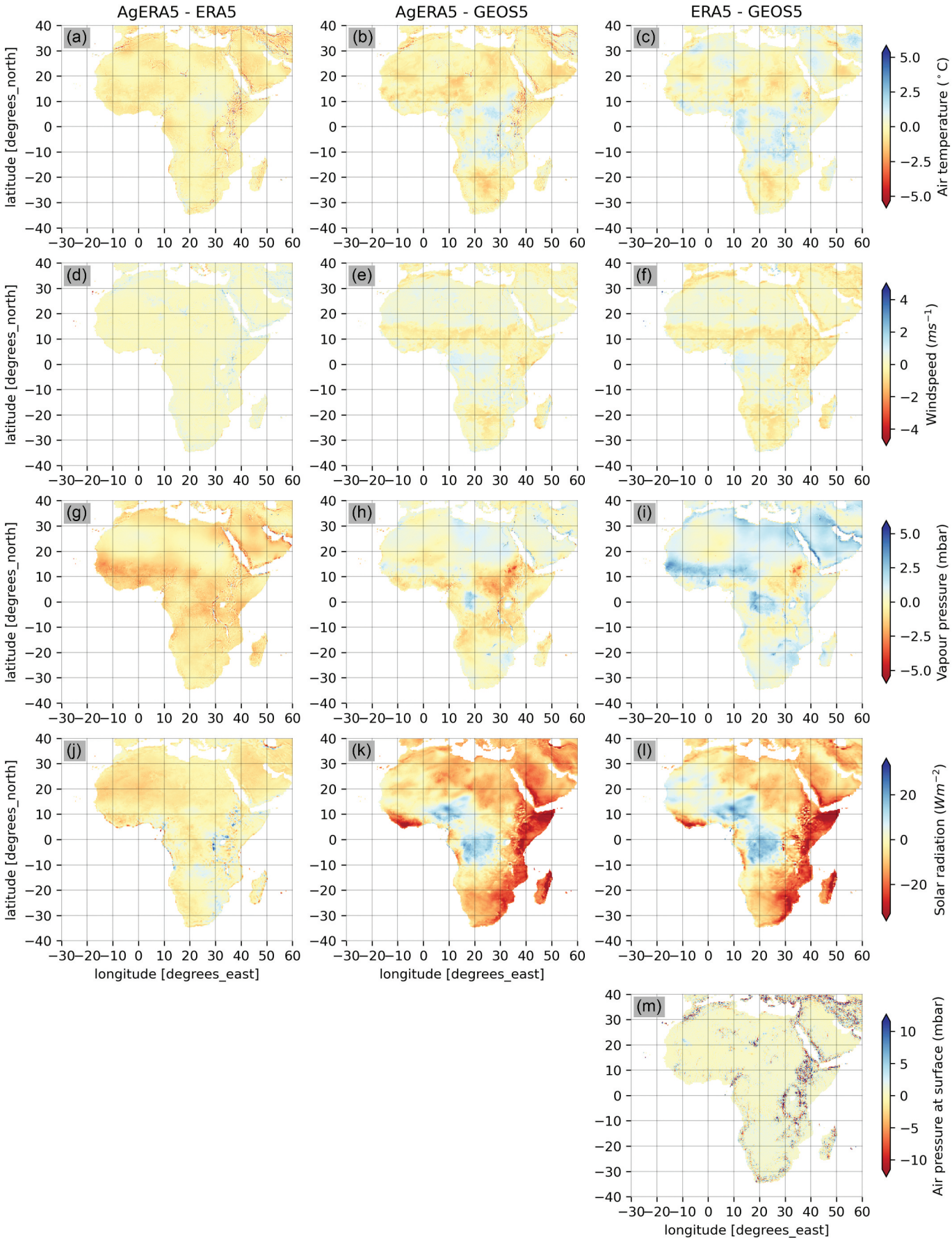


**Figure 4.** Mean annual difference of air temperature, windspeed, vapour pressure, solar radiation and air pressure at the surface from ERA5, AgERA5 and GEOS5 for the period 2018–2022.

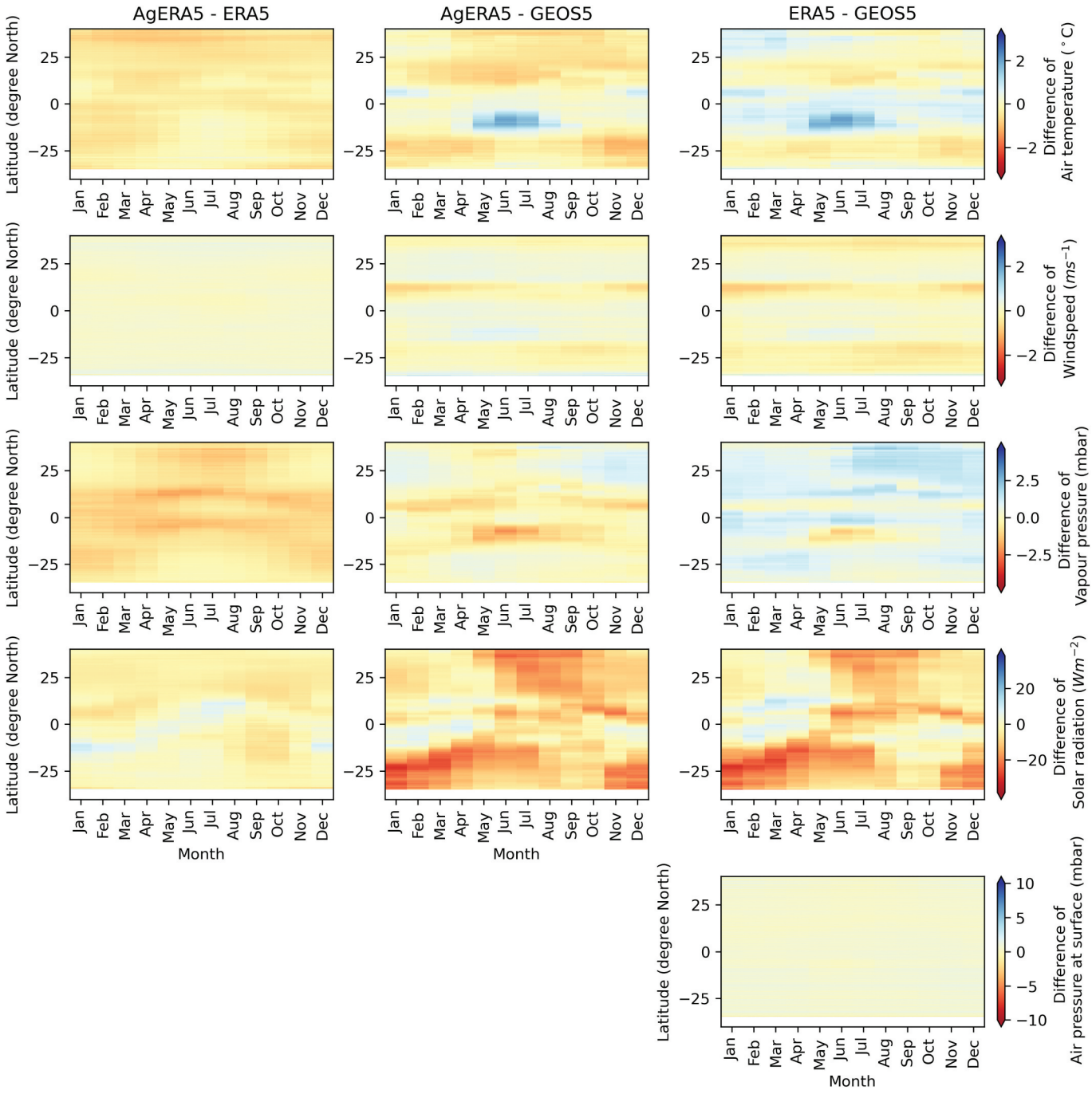


**Figure 5.** Hovmöller diagrams showing the monthly and latitudinal variation of mean difference in air temperature, windspeed, vapour pressure, solar radiation and air pressure at the surface between GEOS5, AgERA5 and ERA5 for the period 2018–2022.

show a shift in pattern from March 2020 (Fig. 6). This shift also coincides with an increase in the spatial difference between the data products, where GEOS5 air temperature becomes even lower near the equator and higher at mid-latitude compared with ERA5 and AgERA5. After the shift, the difference in windspeed between the data products becomes more positive in the tropics (between 20°N and 20°S), where the values of GEOS5 become smaller than ERA5 and AgERA5. In the case of vapour pressure, this shift causes GEOS5 values to become lower than ERA5 and AgERA5 between 10°S and 10°N. For other latitudes, the shift has less impact on vapour pressure differences.

According to the notices from Global Modelling and Assimilation Office (GMAO), the developer of GEOS5, the GEOS-FP model was updated on 7 April 2020. This upgrade introduced two changes that addressed (1) a bias in the heating tendency within the stratosphere and (2) errors in the diagnostics of convective mass flux (GMAO, 2020). Although it was not clearly mentioned how this upgrade affected each variable in the final data product, the shift observed in Fig. 6 suggests that these changes might have caused the temporal inconsistency of the reanalysis datasets. The Hovmöller plots for solar radiation and air pressure at the surface do not show any discernible shift between GEOS5 and the other two

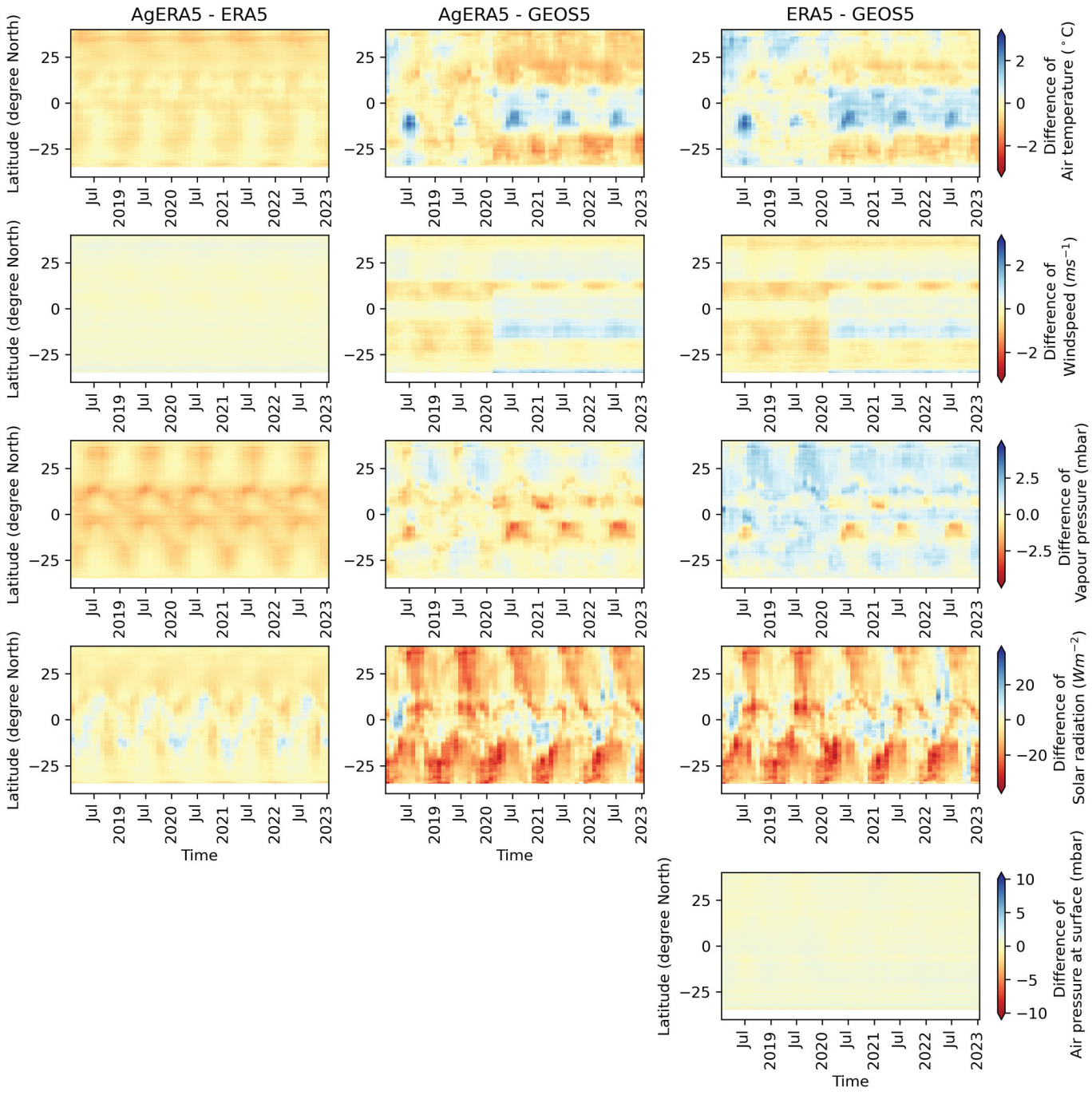


**Figure 6.** Hovmöller diagrams showing the temporal and latitudinal variation of mean difference in air temperature, windspeed, vapour pressure, solar radiation and air pressure at the surface between GEOS5, AgERA5 and ERA5 for the period 2018–2022.

products. Since GEOS5 derives solar radiation data from the tavg1_2d_rad_Nx product and other variables originate from tavg1_2d_slv_Nx, the discrepancy suggests a systematic change within the GEOS5 model that specifically affects the tavg1_2d_slv_Nx product.

### 3.2 Comparison with in situ measurements at TAHMO sites

In terms of nominal accuracy, the comparison with TAHMO stations shows that all datasets exhibit comparable performance (Fig. 7). ERA5 and AgERA5 show slightly better performance metrics than GEOS5 (Fig. 7 and Table S4, see Supplementary material). AgERA5 performs slightly better than ERA5 for all variables, except for vapour pressure. Among the five variables, reanalysis datasets perform best for air pressure at surface and air temperature, showing bias close to zero and high correlation. Meanwhile, the nominal accuracy for windspeed, vapour pressure and solar radiation is much lower. In general, all three datasets overestimate windspeed and solar radiation, and underestimate vapour pressure. In the case of windspeed and solar radiation, the reanalysis datasets

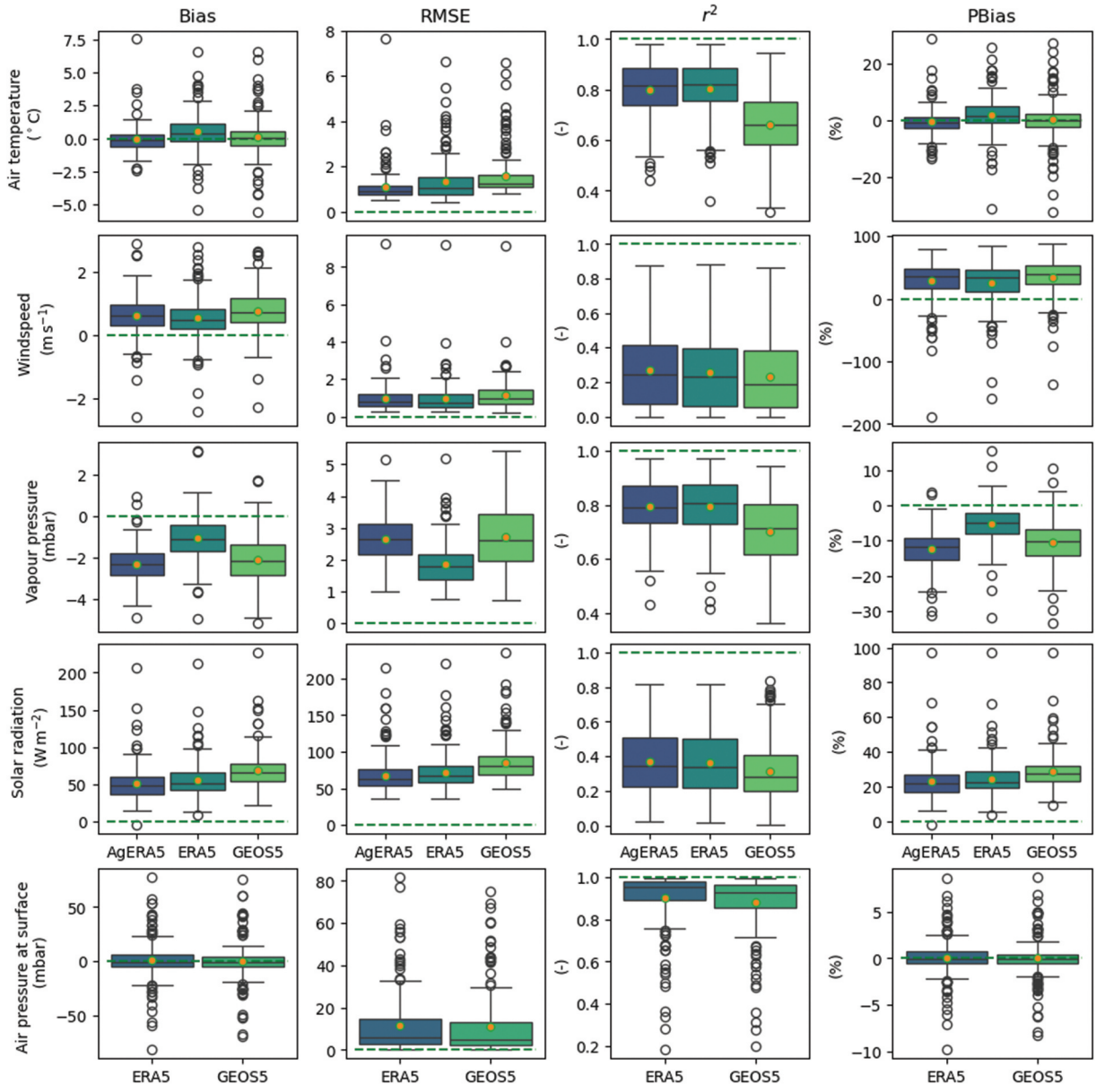


**Figure 7.** Performance metrics of meteorological variables from reanalysis datasets (indicated by different colours) compared with measurements at 174 TAHMO sites. The box-and-whiskers plots represent the 25th (Q1), 50th (median), 75th (Q3) percentiles of the probability distribution. The orange circles inside the box-and-whisker plots represent the mean value. The White circles represent outliers, which exceed the range [Q1 − 1.5 × IQR, Q3 + 1.5 × IQR], where IQR = Q3 − Q1 is the interquartile range.

show very low correlation ($r^2 < 0.5$) at almost all stations. Some outlier stations have windspeed PBias up to −200%, while solar radiation PBIAS is up to +100%. For vapour pressure, the correlation between reanalysis and *in situ* data is similar to air temperature ($r^2 > 0.6$), although PBIAS is generally from −5% to −10%. The overestimation of windspeed and solar radiation, and the underestimation of vapour pressure derived from dew-point temperature is also observed in validation studies of other reanalysis data in Iran (Radmanesh *et al.* 2023), Spain and Portugal (Martins *et al.* 2017) and China (Xu *et al.* 2024).

Figure 8 shows the spatial variability of the average $r^2$ at the TAHMO sites, averaged over the three reanalysis datasets. The maps for RMSE, BIAS and PBIAS are shown in Figs. S4–6 (see Supplementary material). Overall, the area with the best performance is Southern Africa, with fewer stations covering a smaller surface area. The other two areas have a lower performance, due to their complex topography (Eastern Africa) and proximity to the coast (Western Africa). These observations are aligned with findings from validation of reanalysis datasets in other regions (Pelosi *et al.* 2020, Pelosi and Chirico 2021). In the case of air temperature and vapour pressure, the average $r^2$ of the reanalysis datasets shows a decreasing pattern towards the coastline, which is not observed for solar radiation, air pressure and windspeed.

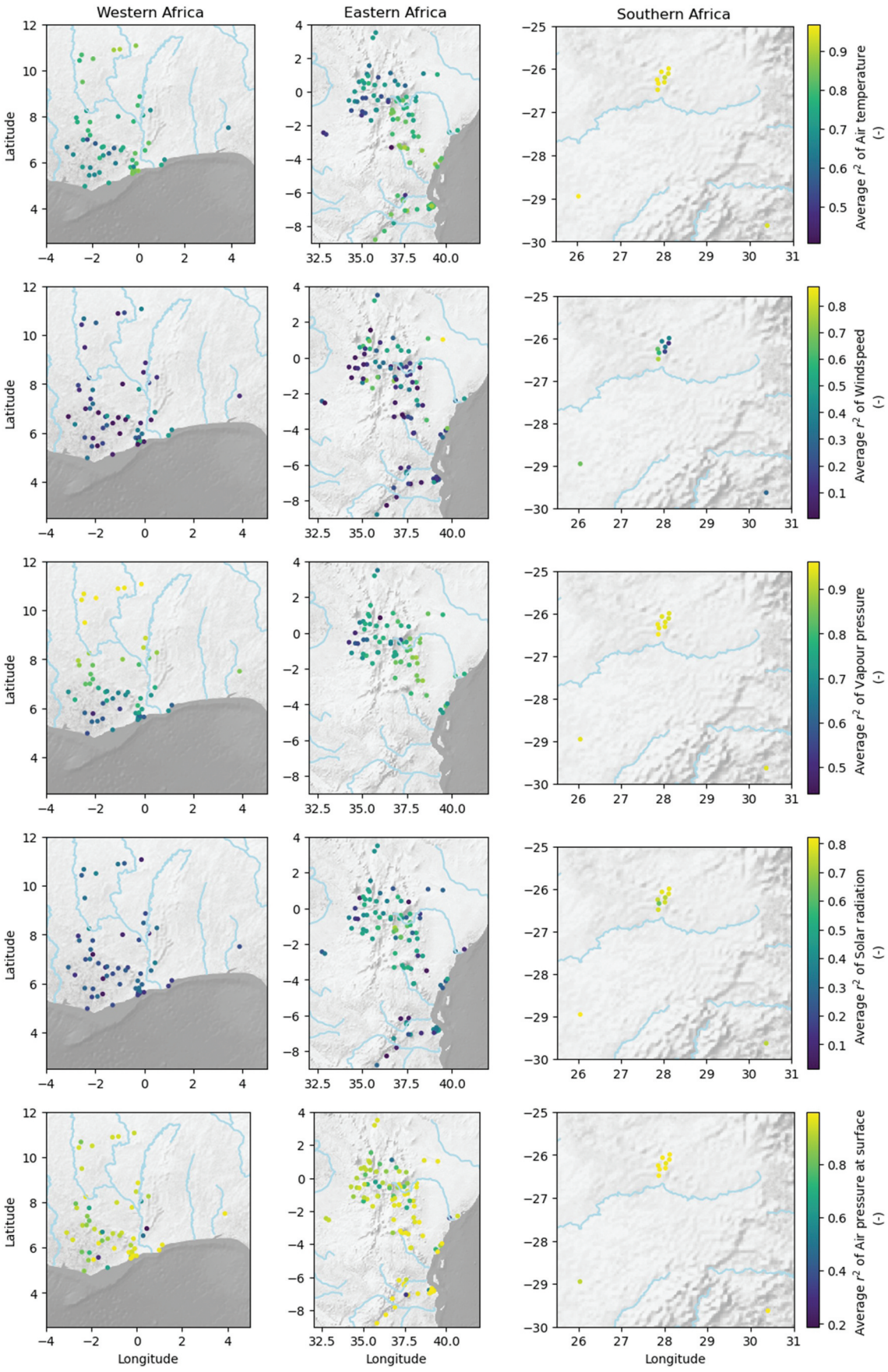


**Figure 8.** Spatial variability of the average $r^2$ computed at the TAHMO sites, averaged over the three reanalysis datasets, AgERA5, ERA5 and GEOS5. Base map: Natural Earth Shaded Relief and Rivers.

### 3.3 Comparing the uncertainty of reanalysis from multiple assessments

The uncertainties associated with the meteorological variables from the reanalysis datasets were assessed by different approaches. The uncertainty estimates from the ERA5 ensemble represent the random errors within ECMWF's NWP model and its data assimilation system. Meanwhile, the discrepancy between different reanalysis datasets indicates the relative errors between the different NWP models. The performance metrics calculated using *in situ* measurements as reference indicate the nominal accuracy of the reanalysis data. Here, we compare and discuss the findings from the three approaches to identify commonalities.

The uncertainty estimates for the ERA5 reanalysis show that all variables have the largest ensemble spread between May and September, especially around 20°N (see Supplementary material, Fig. S1). The seasonal variability of the discrepancy between reanalysis datasets is also higher between May and September, but only for air temperature and between the equator and 20°S (Fig. 5). While ERA5 has a larger ensemble spread in the northern hemisphere than in the southern hemisphere in general, the differences between the reanalysis products do not show the same pattern. This spatial variability between products is more aligned with the reported errors from the GEOS5 NWP initial estimates and assimilated observations for wind and humidity, which is larger in the southern hemisphere than in the northern hemisphere (Rienecker *et al.* 2008, p. 25). Unfortunately, the GEOS5 data products do not include uncertainty estimates or quality indicators for quantitative analysis.

The spatial variability of the ERA5 ensemble spread is different for each variable (see Supplementary material, Fig. S1). Air temperature has the largest yearly average uncertainty estimates in the tropics. Dew point temperature and wind component speed are most uncertain in the Sahara and semi-arid Southern Africa and air pressure in Central Africa (see Supplementary material, Fig. S1). The discrepancy between the datasets also tends to be higher in the tropics for air temperature (Fig. 4). However, for vapour pressure and windspeed, the discrepancy is higher in the Congo rainforest and below the Sahel.

The spatial difference between ERA5 and AgERA5 is mainly observed in mountainous and coastal areas. This is likely because AgERA5 is derived by calibrating ERA5 data with ECMWF's operational high-resolution atmospheric model (HRES). According to Boogaard *et al.* (2023), the greatest enhancements of AgERA5 compared with HRES are observed at grid points situated in mountainous regions or along coasts and lakes, improved particularly for the variables of temperature, humidity and windspeed. However, AgERA5 data is still limited by the accuracy of the HRES operational

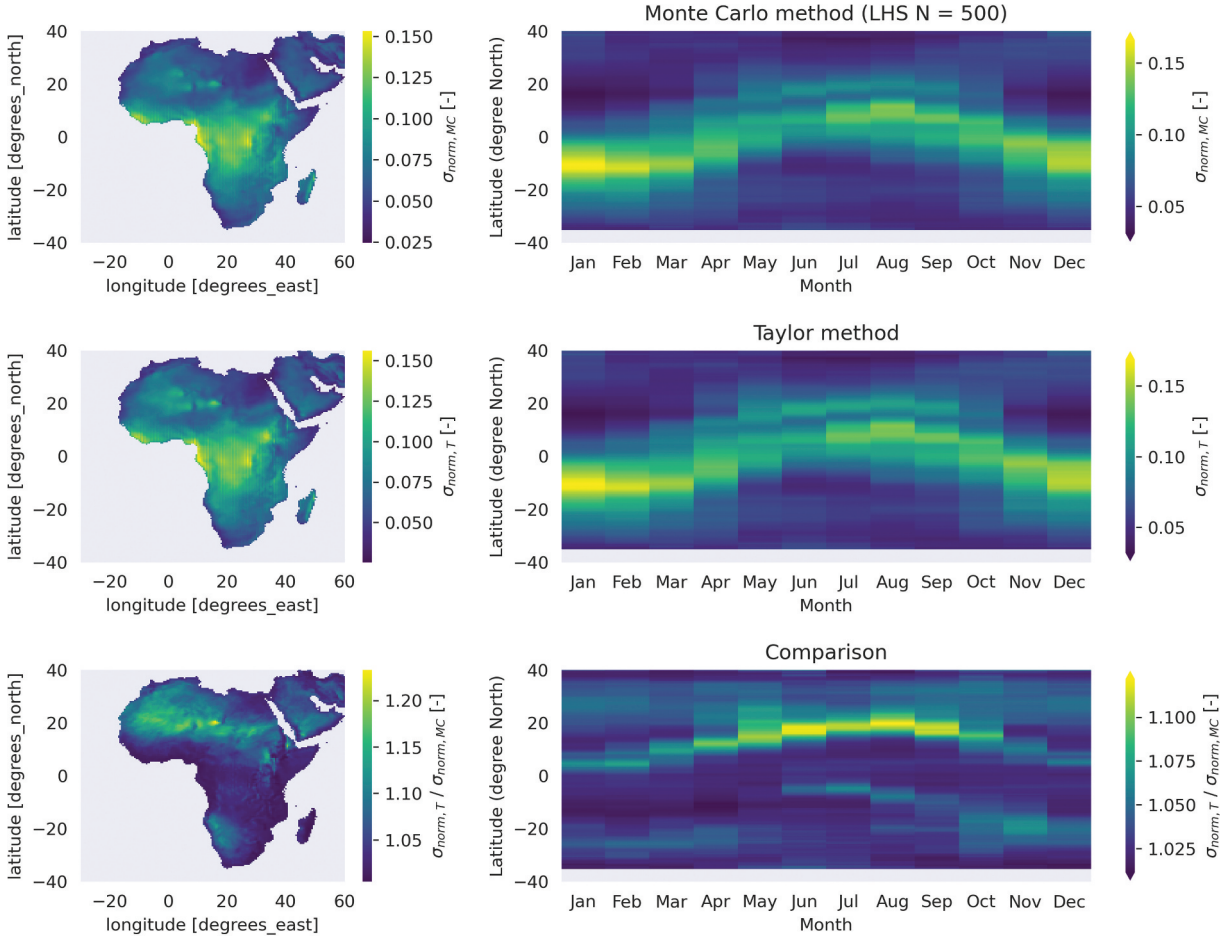


Figure 9. Yearly average and monthly-latitudinal variation of normalized standard deviation $\sigma_{norm}$ of daily reference evapotranspiration estimated by Monte Carlo method (first row), Taylor method (second row) and their ratio (third row).

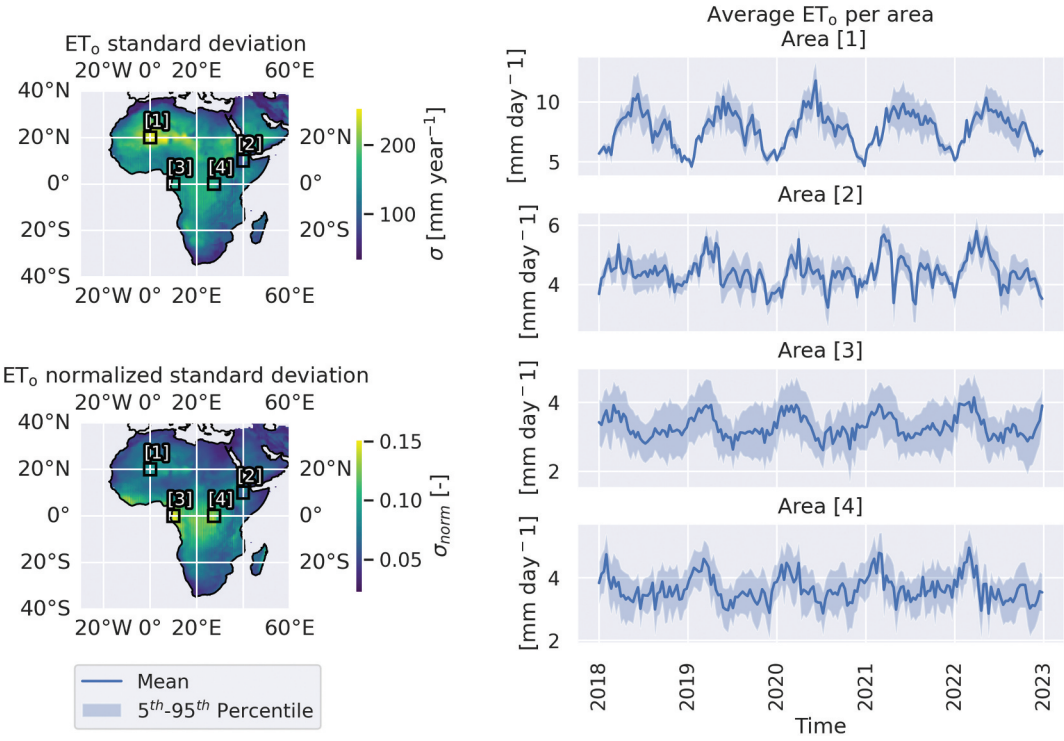


**Figure 10.** Standard deviation of daily reference evapotranspiration ($ET_o$) resulted from 500 Monte Carlo simulations, averaged for 5 years (top left) and normalized with daily mean $ET_o$ (bottom left), and the 10-day ETo in four specific areas indicated in the maps, with 90% confidence interval (right column).

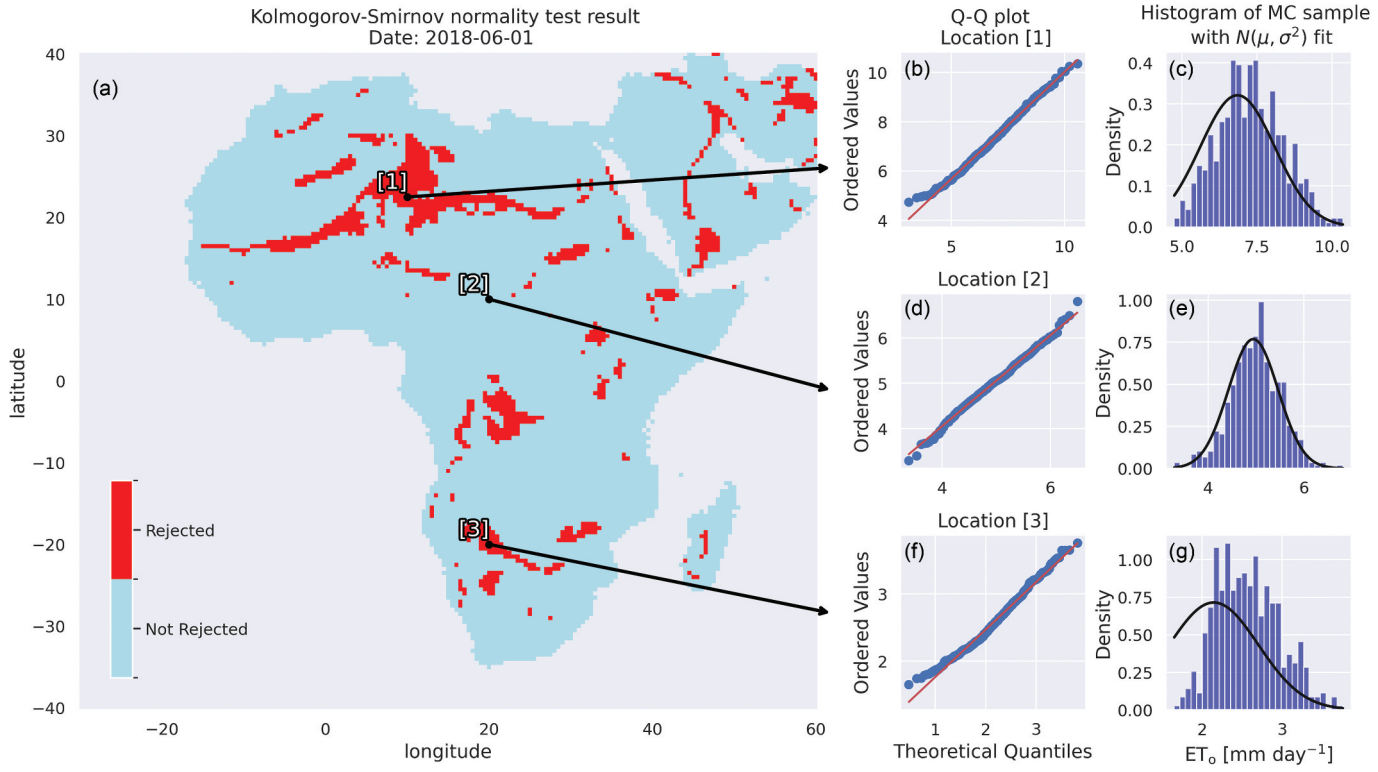


**Figure 11.** Kolmogorov–Smirnov normality test comparing Monte Carlo sample of calculated $ET_o$ ($N = 500$) to the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu$ and $\sigma$ obtained from the Taylor method (significance level of $\alpha = 0.05$). The *Q-Q* plot and histogram at location [1] and [3] (normal distribution is rejected) and location [2] (not rejected) are on the right panel.

model, since it is assumed that HRES represents actual conditions the best. The performance metrics at coastal stations (see Supplementary material, Fig. S7) show AgERA5 performs better for windspeed, but worse for air temperature, and the same for solar radiation and vapour pressure in comparison with all stations (Fig. 7). In case of stations at elevations higher than 1000 m (see Supplementary material, Fig. S8), the performance metrics of AgERA5 relative to ERA5 are not much improved compared with all stations.

### 3.4 Propagation of reanalysis error in reference evapotranspiration calculation

The $ET_o$ uncertainty estimated using the Taylor method shows the exact same spatial and temporal pattern compared with the MC method (Fig. 9). However, it overestimates the normalized standard deviation up to about 20% in the Sahara, especially in the summer months (Fig. 9). In areas with a high uncertainty of $ET_o$, such as the west coast of Central Africa, the Taylor method shows better agreement with the MC method ($\sigma_{\text{norm,MC}}/\sigma_{\text{norm,T}} \approx 1$). The MC simulation result with $N$ = 500 realizations were found to be optimal, as they provided similar outcomes to simulations with larger sample size, while significantly reducing computational time, which is slightly more than using the Taylor method (see Supplementary material, Fig. S9).

The advantage of the MC method is that a probability density function (PDF) and confidence interval can be estimated from the sampled realizations. Daily $ET_o$ was calculated for 500 MC simulations of ERA5 meteorological inputs by Latin-Hypercube sampling (LHS). From the ensemble of $ET_o$, the standard deviation () and 90% confidence interval of the $ET_o$ values were estimated for each grid cell. Figure 10 shows the spatial and temporal variation of the estimated $ET_o$ uncertainty. The Sahara shows the highest standard deviation (mm $d^{-1}$) but a relatively lower normalized standard deviation ($\sigma_{\text{norm}}$) compared with other regions. This is likely due to higher $ET_o$ values in the Sahara (Fig. 10 Area [1]). The tropics shows a high $\sigma_{\text{norm}}$, especially at the west coast of Central Africa (Fig. 10 Area [3]) and the Congo rainforest (Fig. 10 Area [4]). The time-series of 90% confidence intervals shows that this high uncertainty is consistent throughout the years for these regions (Fig. 10). Meanwhile, for other regions, the estimated uncertainty varies with season (Fig. 10 Area [1] and Area [2]).

Since the month of June is where the MC method and the Taylor method show the largest difference in estimated $\sigma_{\text{norm}}$ (Fig. 9), we compared the PDF of the MC samples to a normal distribution $\mathcal{N}(\mu, \sigma^2)$ characterized by the $\mu$ and $\sigma$ obtained from the Taylor method on one day of June (Fig. 11). In general, the PDF obtained from the MC method fits quite well with the normal distribution obtained from the Taylor method, even when the $ET_o$ model involves a non-linear combination of input variables. At locations where the normal distribution is rejected by the Kolmogorov–Smirnov test, the $Q$-$Q$ plot and histogram show that the largest discrepancy is observed in the tail quantiles, where $ET_o$ is extremely low (Fig. 11 location [1] and [3]). This suggests that while the Taylor method provides a reasonable estimate of the PDF for the central part of the $ET_o$ values range, it overestimates the occurrence of extremely low $ET_o$ values. Therefore, in cases where accurate estimation of tail quantiles is critical, the MC method is preferred over the Taylor method. Since the MC results are not in analytical form, it is not feasible to analyse the effect of changing the input error on the output without running the entire simulation again (Heuvelink 1998, pp. 44–45). For this reason, where the Taylor method matches well with the MC result in the central quantiles, it can be applied to update uncertainty estimates when more information about input errors becomes available.

### 3.5 Impact of uncertainties in meteorological data from reanalysis

The uncertainties associated with meteorological variables from reanalysis datasets can affect the accuracy and precision of $ET_o$ estimation. The propagation of ERA5 errors in $ET_o$ estimates shows that random errors are particularly high in the tropics and the Sahel (Fig. 9). The estimated uncertainty in $ET_o$ also varies seasonally, following similar patterns to the errors in input variables (see Supplementary material, Fig. S1). This indicates that non-stationary uncertainty in the reanalysis data also leads to non-stationary uncertainty in $ET_o$ calculations.

Since $ET_o$ estimates are used to calculate crop water requirement, high uncertainty in $ET_o$ can affect irrigation recommendations. Given that $ET_a$ in arid and semi-arid climates is predicted to be primarily controlled by water supply (rainfall or irrigation) (Zhang *et al.* 2016), effective irrigation management is essential to limit agricultural water consumption. Therefore, precise and accurate $ET_o$ estimates are crucial in these regions. The estimated uncertainty of $ET_o$ propagated from input reanalysis data is generally less than 15% of $ET_o$, based on $\sigma_{\text{norm}}$ (Fig. 9). In terms of mm $year^{-1}$, this uncertainty is highest in the Sahel (Fig. 10). Therefore, application of $ET_o$ for irrigation in this region could be improved considerably with more accurate input meteorological data.

Furthermore, the differences between reanalysis datasets indicate that the choice of input dataset may influence the estimated $ET_o$, particularly for solar radiation, vapour pressure and wind speed (Fig. 3). Since the nominal accuracy of ERA5 and AgERA5 at TAHMO sites is generally better than that of GEOS5 (Fig. 7), these datasets are more recommended for field-scale estimation. If GEOS5 is needed for high-latency applications, the error propagation methods employed in this study can be applied to GEOS5, assuming its uncertainty quantification is feasible. This will provide more robust results with a clearer understanding of the confidence or reliability of the estimates.

### 3.6 Limitations of uncertainty assessment methods

The multiple uncertainty assessment methods used in this study have certain limitations. Firstly, the inter-comparison approach requires all datasets to have equivalent spatial resolution (grid size), temporal resolution and physical quantity.

We resampled a higher-resolution dataset to coarser resolution to avoid bias caused by the loss of detail in the coarser-resolution dataset. However, it is difficult to separate the impact of spatial resampling from the differences between datasets. For example, AgERA5 data is derived from ERA5 by applying both spatial downscaling and bias correction for topographical condition and coastal areas (Boogaard *et al.* 2024). As a result, the differences between AgERA5 and ERA5 represent both these corrections and the errors introduced by downscaling and subsequently upscaling processes.

Secondly, comparing reanalysis data with *in situ* measurements only provides nominal accuracy at a few clusters of TAHMO sites, mainly in tropical savannah climate. These clusters do not cover the tropics and the Sahara, which are the areas with the largest uncertainties indicated by spatio-temporal comparison and ERA5 ensemble spread. Furthermore, the accuracy of *in situ* measurements is also a factor influencing the resulting performance metrics. For example, *in situ* windspeed measurement at 2 m can be affected by local terrain effects at some sites (Pelosi and Chirico 2021). For equivalent comparison, the reanalysis windspeed at 10 m was used to derive windspeed at 2 m using a logarithmic windspeed profile (Equation 2), however, this cannot account for local terrain effects.

Thirdly, the error propagation method requires an assumption for the error models of input variables. In this study, we assumed that the errors in meteorological variables from reanalysis data follow a normal distribution. For ERA5, we characterized the error distribution using the mean and standard deviation provided by its ensemble uncertainty quantification. However, in the case of AgERA5 and GEOS5, where such uncertainty quantification is not available, making assumptions about the mean and standard deviation of errors would be inherently incomplete, as they vary for every grid cell and every time step. Additionally, none of these approaches account for systematic error and uncertainty due to different NWP models. Lastly, it is important to note that the methodological uncertainties associated with the FAO56 $ET_o$ equation (Equation (6)) were not evaluated in this study. Given that the FAO56 definition of $ET_o$ is based on idealized conditions, ambiguity regarding the consideration of local advection effects may introduce biases in FAO56 $ET_o$ estimates when applied to arid and semiarid regions with non-ideal or smaller fields (de Bruin *et al.* 2016, Pereira *et al.* 2021). Therefore, updates on the FAO56 $ET_o$ equation still need to be considered for its use in estimating water demand.

## 4 Conclusion

The present study aimed to evaluate the uncertainty of meteorological data inputs for FAO56 $ET_o$ calculation from GEOS5, ERA5 and AgERA5 reanalysis data products through spatio-temporal inter-comparison (between-product uncertainty), comparison with *in situ* measurements (nominal accuracy) and ensemble spread (within-product uncertainty). The study also analyses the error propagation in the FAO56 $ET_o$ calculation using reanalysis as meteorological forcing. The spatio-temporal inter-comparison of all data products shows that the differences between products are non-stationary. The major differences between climatic input datasets are in Central and Southern Africa, and Southwest Asia. This uncertainty between reanalysis datasets is due to the model uncertainty of the employed numerical weather prediction model and estimated errors of observations in data assimilation systems. All reanalysis datasets predict air temperature and air pressure well, but overestimate windspeed and solar radiation and underestimate vapour pressure at the reference sites in the TAHMO network. Comparison with *in situ* data shows that all reanalysis datasets have comparable performance, but ERA5 and AgERA5 perform slightly better. Although having better latency, GEOS5 has lower nominal accuracy and some temporal inconsistency due to changes in the data assimilation system. Therefore, near-real time applications that depend on the GEOS5 dataset are subject to more errors and not recommended to be used for trend analysis. The error propagation results show that the uncertainty in $ET_o$ estimates propagated from the estimated uncertainty in the ERA5 reanalysis dataset is consistently higher in the tropics. The Taylor method showed a consistent spatial and temporal pattern of uncertainty and adequate accuracy compared with the Monte Carlo method. Since every uncertainty assessment method has its limitations, applying multiple approaches and comparing their results could help identify the limitation in reanalysis data and better inform the application of reanalysis data, especially in data-scarce regions like Africa and Southwest Asia.

## Acknowledgements

We gratefully acknowledge all researchers, technicians and supporters at the Trans-African Hydro-Meteorological Observatory (TAHMO) for the effort of making this dataset available. We especially thank Frank Annor and Kwame Duah for providing the TAHMO dataset. We also thank Annemarie Klaassen (former) and Henk Pelgrum at eLEAF, and Bert Coerver at FAO for sharing information about the access and processing of reanalysis datasets. We also acknowledged the use of ChatGPT, a language model developed by OpenAI (version: GPT-4) to identify errors in programming code and suggest improvements to enhance the clarity and structure of written text. The output generated by GPT-4 is thoroughly reviewed and quality-controlled by the authors to ensure that the final outputs meet the required standards of accuracy, clarity and relevance.

## Disclosure statement

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

## Funding

This research has been supported by the monitoring land and water productivity by remote sensing (WaPOR phase 2) project, which is funded by the Ministry of Foreign Affairs of the Netherlands [grant no. GCP/INT/729/NET]; Ministerie van Buitenlandse Zaken.

## ORCID

Bich Ngoc Tran http://orcid.org/0000-0001-6301-2699
Marloes Mul http://orcid.org/0000-0001-9469-3909

## Code and data availability

The computational notebooks for analyses and result visualization are available at https://doi.org/10.5281/zenodo.13970799. ERA5 data are accessible at https://doi.org/10.24381/cds.adbb2d47. AgERA5 data are accessible at https://doi.org/10.24381/cds.6c68c9bb. GEOS5 data are accessible at https://opendap.nccs.nasa.gov/dods/GEOS-5/fp/0.25_deg/assim. TAHMO dataset is available on request to the data provider at https://tahmo.org/climate-data/.

## References


Abatzoglou, J.T., *et al*., 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191. doi:10.1038/sdata.2017.191

Abernethy, K., Maisels, F., and White, L.J.T., 2016. Environmental issues in Central Africa. *Annual Review of Environment and Resources*, 41 (1), 1–33. doi:10.1146/annurev-environ-110615-085415

Adhikari, U., Nejadhashemi, A.P., and Herman, M.R., 2015. A review of climate change impacts on water resources in East Africa. *Transactions of the ASABE*, 58, 1493–1507. doi:10.13031/trans.58.10907

Allen, R.G., *et al*., 1998. *Crop evapotranspiration - guidelines for computing crop water requirements, FAO irrigation and drainage paper*. FAO - Food and Agriculture Organization of the United Nations, Rome.

Annor, F.O., 2023. *Small reservoirs in northern Ghana: monitoring, physical processes, and management*. Delft, The Netherlands: Delft University of Technology. doi:10.4233/UUID:81E5E8A8-2BEE-4AF5-B1BC-C7B210F9CB55

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

Boogaard, H., *et al*., 2020. Agrometeorological indicators from 1979 to present derived from reanalysis. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. doi:10.24381/cds.6c68c9bb

Boogaard, H., *et al*., 2024. *Data stream 2: AgERA5 historic and near real time forcing data: algorithm theoretical basis (ATBD)* [WWW document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. Available from: https://confluence.ecmwf.int/pages/viewpage.action?pageId=278550984 [Accessed 12 August 2024].

Boogaard, H., Van der Grijn, G., and Schubert, J., 2023. *Global agriculture downscaling and bias correction [WWW document]*. Copernic. Knowl. Base - ECMWF Conflu. Wiki. Available from: https://confluence.ecmwf.int/display/CKB/Global+Agriculture+Downscaling+and+bias+correction [Accessed 6 August 2024].

Brönnimann, S., *et al*., 2018. Observations for reanalyses. *Bulletin of the American Meteorological Society*, 99 (9), 1851–1866. doi:10.1175/BAMS-D-17-0229.1

de Bruin, H.A.R., *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

de Pauw, E., 2005. *Monitoring agricultural drought in the Near East, in: Monitoring and predicting agricultural drought: a global study*. Oxford, New York: Oxford University Press.

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

European Centre for Medium-Range Weather Forecasts (ECMWF), 2023a. *ERA5 data documentation* [WWW document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. Available from: https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation [Accessed 29 July 2024].

European Centre for Medium-Range Weather Forecasts (ECMWF), 2023b. *ERA5 uncertainty estimation* [WWW document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. Available from: https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation [Accessed 29 July 24].

FAO, 2024. *Reference evapotranspiration (global - daily - approximately 30km) - WaPOR v3 - "FAO catalog"* [WWW document]. Available from: https://data.apps.fao.org/catalog//iso/6f487301-9731-4317-b620-faeccc976443 [Accessed 18 July 2024].

Fisher, J.B., *et al*., 2017. The future of evapotranspiration: global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, 53 (4), 2618–2626. doi:10.1002/2016WR020175

Global Modeling and Assimilation Office (GMAO), 2020. *GEOS FP to be updated on April 7, 2020* [WWW document]. GMAO - Glob. Model. Assim. Off. Res. Site. Available from: https://gmao.gsfc.nasa.gov/news/geos_system_news/2020/GEOS_FP_upgrade_5_25_1p5.php [Accessed 6 August 2024].

Hargreaves, G.H. and Samani, Z.A., 1982. Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108, 225–230. doi:10.1061/JRCEA4.0001390

Hargreaves, G.H. and Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1 (2), 96–99. doi:10.13031/2013.26773

Hejazi, M., *et al*., 2023. Impacts of water scarcity on agricultural production and electricity generation in the Middle East and North Africa. *Frontiers in Environmental Science*. 11. 10.3389/fenvs.2023.1082930

Hersbach, H., *et al*., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146 (730), 1999–2049. doi:10.1002/qj.3803

Hersbach, H., *et al*., 2023. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) *ERA5 hourly data on single levels from 1940 to present*. doi:10.24381/cds.adbb2d47

Heuvelink, G.B.M., 1998. *Error propagation in environmental modelling with GIS*. London: CRC Press. doi:10.4324/9780203016114

Hovmöller, E., 1949. The trough-and-ridge diagram. *Tellus*, 1 (2), 62–66. doi:10.1111/j.2153-3490.1949.tb01260.x

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

Kummu, M., *et al*., 2016. The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability. *Scientific Reports*, 6 (1), 38495. doi:10.1038/srep38495

Kusangaya, S., *et al*., 2014. Impacts of climate change on water resources in Southern Africa: a review. *Physics and Chemistry of the Earth, Parts A/B/C*, 67, 47–54. doi:10.1016/j.pce.2013.09.014

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

Leal Filho, W., *et al*., 2022. Understanding responses to climate-related water scarcity in Africa. *Science of the Total Environment*, 806, 150420. doi:10.1016/j.scitotenv.2021.150420

Lebigot, E.O., 2017. *Uncertainties: a Python package for calculations with uncertainties* [WWW document]. Uncertainties Python package 301. Available from: https://pythonhosted.org/uncertainties/ [Accessed 30 July 2024].

Lee, A., *et al*., 2014. *Randomized designs — pyDOE 0.3.6 documentation* [WWW document]. PyDOE Exp. Des. Package Python. Available from: https://pythonhosted.org/pyDOE/randomized.html#latin-hypercube [Accessed 30 July 2024].

Lelieveld, J., *et al*., 2012. Climate change and impacts in the eastern Mediterranean and the Middle East. *Climatic Change*, 114 (3–4), 667–687. doi:10.1007/s10584-012-0418-4

Martins, D.S., *et al*., 2017. Assessing reference evapotranspiration estimation from reanalysis weather products. An application to the Iberian Peninsula. *International Journal of Climatology*, 37 (5), 2378–2397. doi:10.1002/joc.4852

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

McCabe, M.F., *et al*., 2017. The future of earth observation in hydrology. *Hydrology and Earth System Sciences*, 21 (7), 3879–3914. doi:10.5194/hess-21-3879-2017

METER, 2023. *ATMOS41 manual* [WWW document]. Available from: https://library.metergroup.com/Manuals/20635_ATMOS41_Manual_Web.pdf [Accessed 29 July 2024].

Parker, W.S., 2016. Reanalyses and observations: what's the difference? *Bulletin of the American Meteorological Society*, 97 (9), 1565–1572. doi:10.1175/BAMS-D-14-00226.1

Pelosi, A., *et al.*, 2020. Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water*, 12 (6), 1669. doi:10.3390/w12061669

Pelosi, A. and Chirico, G.B., 2021. Regional assessment of daily reference evapotranspiration: can ground observations be replaced by blending ERA5-Land meteorological reanalysis and CM-SAF satellite-based radiation data? *Agricultural Water Management*, 258, 107169. doi:10.1016/j.agwat.2021.107169

Pereira, L.S., *et al.*, 2021. Updates and advances to the FAO56 crop water requirements method. *Agricultural Water Management*, 248, 106697. doi:10.1016/j.agwat.2020.106697

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

Raza, A., *et al.*, 2022. Misconceptions of reference and potential evapotranspiration: a PRISMA-guided comprehensive review. *Hydrology*, 9 (9), 153. doi:10.3390/hydrology9090153

Rienecker, M.M., *et al.*, 2008. The GEOS-5 data assimilation system-documentation of versions 5.0.1. 5.1.0, and 5.2.0 (No. NASA/TM-2008-104606-VOL-27).

Shafer, M.A., *et al.*, 2000. Quality assurance procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology*, 17 (4), 474–494. doi:10.1175/1520-0426(2000)017<0474:QAPITO>2.0.CO;2

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

Soci, C., *et al.*, 2024. The ERA5 global reanalysis from 1940 to 2022. *Quarterly Journal of the Royal Meteorological Society*, 150 (764), 4014–4048. doi:10.1002/qj.4803

Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29 (2), 143–151. doi:10.1080/00401706.1987.10488205

TAHMO, 2023. *TAHMO weather station manual* [WWW document]. Available from: https://usercontent.one/wp/tahmo.org/wp-content/uploads/2023/06/TAHMOSTATIONmanual2023.pdf [Accessed 29 July 2024].

Taylor, J., 1997. *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements*. 2nd ed. Sausalito, California: Published by University Science Books.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. *Geographical Review*, 38 (1), 55–94. doi:10.2307/210739

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

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

Wang, L., *et al.*, 2024a. Quantifying and mitigating errors in estimating downward surface shortwave radiation caused by cloud mask data. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–15. doi:10.1109/TGRS.2024.3438879

Wang, L., *et al.*, 2024b. Hourly solar radiation estimation and uncertainty quantification using hybrid models. *Renewable and Sustainable Energy Reviews*, 202, 114727. doi:10.1016/j.rser.2024.114727

Xiang, K., *et al.*, 2020. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. *Agricultural Water Management*, 232, 106043. doi:10.1016/j.agwat.2020.106043

Xu, C., *et al.*, 2024. Evaluation of ERA5, ERA5-Land, GLDAS-2.1, and GLEAM potential evapotranspiration data over mainland China. *Journal of Hydrology: Regional Studies*, 51, 101651. doi:10.1016/j.ejrh.2023.101651

Zhang, K., Kimball, J.S., and Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. *Wires Water*, 3 (6), 834–853. doi:10.1002/wat2.1168