frontiers in Environmental Science

TYPE Original Research
PUBLISHED 27 August 2026
DOI 10.3389/fenvs.2026.1859873

OPEN ACCESS

EDITED BY
Dericks Shukla,
Indian Institute of Technology Mandi,
India

REVIEWED BY
Panfeng Zhang,
Jilin Normal University, China
Ipshita Pradhan,
DExtER Lab, IIT Mandi, India

*CORRESPONDENCE
Fabian Mitze,
fabian.mitze@zeu.uni-giessen.de

RECEIVED 19 April 2026
REVISED 10 July 2026
ACCEPTED 31 July 2026
PUBLISHED 27 August 2026

CITATION
Mitze F, Jacobs SR, Breuer L, Campos Zeballos J, Dowling TPF and Weeser B (2026) Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data.
Front. Environ. Sci. 14:1859873.
doi: 10.3389/fenvs.2026.1859873

Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data

Fabian Mitze[1,2]*, Suzanne Robin Jacobs[1,2], Lutz Breuer[1,2], Jazmin Campos Zeballos[1,3], Thomas P. F. Dowling[4] and Björn Weeser[1]

[1]Center for International Development and Environmental Research, Justus Liebig University Giessen, Giessen, Germany, [2]Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, Giessen, Germany, [3]Faculty of Agriculture, University of Bonn, Bonn, Germany, [4]School of Environment, Faculty of Science, University of Auckland, Auckland, New Zealand

Many remote tropical mountainous regions lack precise, spatially distributed air temperature data. This limits both scientific research and applied solutions in fields such as hydrology, meteorology, health and agriculture. As regular monitoring networks do not provide sufficient spatial coverage in these regions, alternatives are required. To address this issue, this study proposes a novel approach using participatory monitoring (PM) data for point-to-point bias correction of ERA5-Land air temperature data. ERA5-Land data was trained on PM data from different remote mountainous regions in Ecuador (n = 67), Honduras (n = 23) and Tanzania (n = 275) using simple linear regression models. Validation was conducted using automatically measured air temperature, applying different metrics, including mean absolute error (MAE) and coefficient of determination (R^2). Validation demonstrated an improvement across all stations, with the most significant reduction in MAE, from 5.49 °C to 1.76 °C in Ecuador. Concurrently, R^2 increased across all stations up to 0.83. The decrease in deviation was also statistically significant for all stations. The study demonstrated that incorporating PM data into simple linear regression bias correction can reduce the regional bias of ERA5-Land air temperature data. This can be regarded as a pragmatic solution for remote regions, where the absence of weather stations is a common issue.

KEYWORDS
citizen science, correction, hydrometeorological data, regression, remote sensing

1 Introduction

The World Meteorological Organization (WMO) has confirmed that 2024 was the warmest year on record with the global air temperature 1.55 °C above the pre-industrial level (WMO, 2025a). This serves as yet another reminder that climate change is a matter of pressing concern. It may have far-reaching consequences for human health, the global economy, and agriculture, necessitating significant adaptation measures (Lee et al., 2023). Countries belonging to the “Global South” might lack resources for climate change adaptation and mitigation which is why effects might be felt sooner and more intensely (Abraham, 2018; Sen Roy, 2018; United Nations, 2022; Ngcamu, 2023; Cavazos et al., 2024). Given that all these challenges are directly or indirectly linked to rising air temperature (Lee et al., 2023), reliable data on this parameter can be considered critical for many kinds of climate change research.

Traditionally, hydrometeorological data such as air temperature has been collected using professional manual weather stations managed by meteorological authorities (WMO, 2017). Today, many of these are replaced by automatic weather stations which are considered the state of the art for this type of data collection (WMO, 2017). Common problems with these stations are their complexity of maintenance, costs and their highly uneven distribution around the globe. For instance, the distribution of the approximately 80,000 weather stations included in the Global Historical Climatology Network (GHCN) daily database (Menne et al., 2012), highlight this issue: Relatively few stations are located in countries like Honduras (n = 7), Ecuador (n = 15) or Tanzania (n = 14) compared to Western European countries like the Netherlands (n = 353) or Germany (n = 1,069) (NOAA, 2025). This becomes more evident when considering the weather station density, i.e., the number of GHCN stations per unit area of the country. For instance, Honduras has a density of 0.062 stations * 1,000 km^{-2}, whereas the Netherlands has a density of 8.497 stations * 1,000 km^{-2}.

In recent decades, remote sensing-based monitoring has become a powerful alternative to such ground-based data collection, for instancefor the estimation of large-scale air or land surface temperature (Prihodko and Goward, 1997; Sun et al., 2005; Hooker et al., 2018). Possible limitations of this type of data can be, however, the often coarse spatial or temporal resolution. A current state-of-the-art global dataset that offers air temperature reanalysis values is ERA5-Land, which has a coverage with 9 km spatial- and a one hour temporal resolution (Muñoz-Sabater et al., 2021). Nevertheless, this spatial resolution may still be too coarse for specific applications, for instance the calculation of evapotranspiration in mountainous areas where air temperature varies greatly over short distances at different elevations.

To overcome these limitations, there have been multiple attempts to adapt large-grid cell model outputs to make them more useful for the local scale (Zhu et al., 2021; Niazkar et al., 2023; Sebbar et al., 2023; Dhawan et al., 2024; Xu et al., 2024; Zhang et al., 2024). Removing a systematic error or deviation in modelled data, such as ERA5-Land, to match a specific location is commonly known as bias correction (Maraun, 2016; Muñoz-Sabater et al., 2021). Some studies explicitly focused on bias correction for air or land surface temperature in mountainous regions in different parts of the globe (Zhu et al., 2021; Sebbar et al., 2023; Dhawan et al., 2024; Satyapragyan et al., 2026). Different methods can be used for this, such as statistical methods (Lompar et al., 2019; Dhawan et al., 2024; Zhang et al., 2025), machine learning (Niazkar et al., 2023; Sebbar et al., 2023; Dhawan et al., 2024; Zhang et al., 2024) or deep learning (Niazkar et al., 2023; Xu et al., 2024). However, the data used for the bias correction of the remotely sensed dataset is typically obtained from conventional sources, such as regular weather stations. As the distribution of these stations is spatially limited, especially in remote areas, data for bias correction is not available everywhere and alternative data sources may be required. This raises the question whether data from other, non-traditional sources in remote mountainous regions can also be used for such bias correction approaches.

A noteworthy alternative source for hydrometeorological data could be participatory monitoring (PM), where volunteers participate in the data collection process (Danielsen et al., 2009; Rajagopalan et al., 2020; Beele et al., 2022; Campos Zeballos et al., 2025; Mitze et al., 2026). Measuring instruments provided to participants often have low costs and are usually easy to use by lay people to use (Buytaert et al., 2014; Davids et al., 2019). This makes it a particularly viable alternative for regions with limited financial resources (Buytaert et al., 2014; Fankhauser and McDermott, 2014; Mitze et al., 2026). Therefore, this kind of approach could improve spatial resolution compared to traditional monitoring networks, providing data from areas that have previously been poorly monitored. A recent example by Mitze et al. (2026) showed that the accuracy of air temperature data measured by participants with analog sensors is of considerable accuracy compared to regular automatic sensors, despite construction-related outliers, which could be excluded. However, participants do not measure at a regular interval, for example in hourly intervals, as automatic sensors do. This irregularity of measurements and the uncertainty regarding a long-term perspective are likely the most significant challenges for integrating PM into existing measurement networks (Njue et al., 2019). Because of that, a 1:1 replacement of automatic sensor measurements with PM could present new challenges, especially for further applications or long-term use of the data. Assuming a high correlation between these PM measurements and other data sources such as remote sensing data with regular time intervals, this issue could be addressed by applying regression models to predict missing timestamps. At the same time, such a combination could address the spatial accuracy of the remote sensing data, thus compensating for the inherent limitations of each data type. While there are multiple bias correction approaches, only a few studies are available that apply such corrections to an hourly timescale on ERA5-Land data (Niazkar et al., 2023; Dhawan et al., 2024; Zhang et al., 2025). Furthermore, these studies all used automatically measured data from regular weather stations. To date no research has evaluated the use of PM data for the purpose of bias correction of continuous datasets such as ERA5-Land. For this reason, this study utilizes PM air temperature data for a simple, point-to-point bias correction. The corresponding PM data was collected in different tropical remote mountainous regions in Ecuador, Honduras and Tanzania as part of the hydrometeorological PM project "HydroCrowd" (Campos Zeballos et al., 2025; Mitze et al., 2026). In HydroCrowd, participatory monitoring is examined as an alternative data collection method in remote tropical regions (https://ihp-wins.unesco.org/citizens4water/project/18). The three countries were chosen as study sites for HydroCrowd due to various strategic and practical reasons, such as existing connections with local partners, as well as different management and socio-cultural contexts, which allowed testing different approaches regarding target groups of PM. Because of the limited availability of hydrometeorological data in the study regions investigated, these are considered well-suited for the approach pursued in this study.

The objective of this study is to analyze whether such PM air temperature data can be used for bias correction to improve the accuracy of ERA5-Land air temperature. As there are no other sources of continuous hydrometeorological data available here, this could contribute to a better understanding of the complex microclimates in these regions (Sebbar et al., 2023), which, in

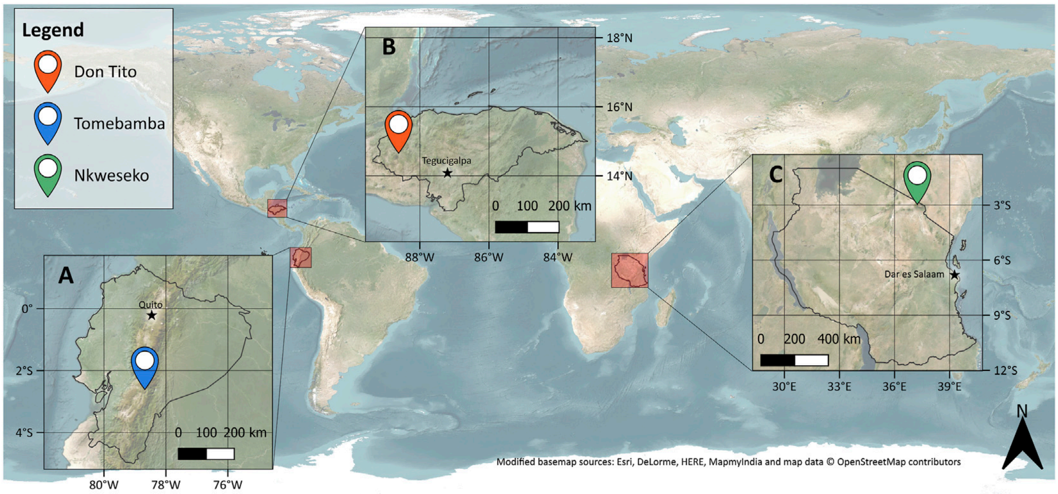

FIGURE 1
Locations of the stations used for the analysis in the study regions in **(A)** Southern Ecuador, **(B)** Western Honduras and **(C)** Northern Tanzania.

turn, can be vital for understanding of local climate change impacts, adaptation to climate change, agricultural planning or health protection. Furthermore, future PM projects could benefit from insights gained from the application of collected data for even more targeted PM campaigns.

2 Materials and methods

2.1 Study region and climate

In this study, air temperature data collected between May 2023 and August 2025 from the PM project HydroCrowd was used. The study regions are located in different tropical mountain settings predominantly in and around national parks in Ecuador, Honduras and Tanzania (Campos Zeballos et al., 2025; Mitze et al., 2026). An automatic sensor was used to monitor the air temperature at one station in each country where high participation was expected, for validation purposes (Figure 1).

Further information on the participatory approach and the data collection itself can be found in prior publications by Campos Zeballos et al. (2025) and Mitze et al. (2026). The characteristics of the PM data are presented in more detail in the Section 2.2.1.

2.1.1 Ecuador

From the study region in Ecuador the station "Tomebamba" was selected for this analysis. The station is located in the municipality of the same name, province Azuay, Southern Ecuador (Figure 1A). It was installed in the center of the town, around 37 km north-east of the city of Cuenca, at an elevation of approx. 2,380 m a.s.l. (Mitze et al., 2026). Due to the absence of a climatological station in the immediate vicinity, data from the nearest WMO station in Cuenca is used to describe the prevailing climatic conditions. This station is at a comparable elevation (2,530 m a.s.l.) and according to the World Bank, it has the same temperate Andean climate (temperate oceanic climate) (World Bank, 2025). The annual average temperature is 16.3 °C, with low seasonal variation (WMO, 2025b), with November having the highest (22.2 °C) and July having the lowest average air temperature (18.6 °C) (WMO, 2025b). Annual rainfall amounts to approximately 876 mm, with peak rainfalls in April, with over 120 mm (WMO, 2025b).

2.1.2 Honduras

The station "Don Tito" is located on a pasture in the vicinity of Celaque National Park in Western Honduras (Figure 1B). It was installed at an elevation of approx. 1,182 m a.s.l. About 5 km southwest of the city of Gracias. The region has a tropical savannah climate with an average temperature of 24 °C and annual average precipitation of 1,600 mm (Valdez et al., 2017). The area has a distinct rainy season from May to October, while the rest of the year is dry (Aguilar, 2005).

2.1.3 Tanzania

From the study region in Tanzania, the station "Nkweseko" was selected. It is located at the Nkweseko research station close to the border of Kilimanjaro National Park at an elevation of approx. 1,700 m a.s.l. (Figure 1C). The area has a typical equatorial daytime climate with varying temperature and rainfall depending on elevation (Hemp, 2005). Rainfall has a bimodal pattern with two rainy seasons from October to December and March to May (Shagega et al., 2025) resulting in approx. 2,616 mm per year (Gütlein et al., 2018). According to Gütlein et al. (2018) and estimations of Hemp (2005), mean annual air temperature is around 18.4 °C –18.8 °C.

2.2 Data sources

2.2.1 Participatory monitoring and automatic sensor data

For the participatory monitoring approach, analog thermometers (TFA-Dostmann, K1.100273, Wertheim, Germany;

FIGURE 2
The three sensors mounted on a weather station at "Don Tito" in Honduras: a hygrometer for relative humidity, a thermometer for air temperature and a rain gauge for rainfall measurements.

TABLE 1 Stations with participatory monitoring (pm) and automatically (as) measured data and the corresponding measurement periods for the stations (n = number of T_{pm} and m = number of T_{as} and coordinates in WGS 1984).

Station name	Country	Latitude	Longitude	n	m	Period monitored
Nkweseko	Tanzania	−3.184554	37.240911	275	12,912	26.08.2023–14.02.2025
Tomebamba	Ecuador	−2.742007	−78.682567	67	4,680	07.12.2023–19.06.2024
Don tito	Honduras	14.532804	−88.61398	23	10,080	08.05.2024–02.07.2025

accuracy: ±1 °C) and other sensors were mounted on standardized, roofed wooden signpost, called weather stations, which were installed at different locations in the three study regions (Campos Zeballos et al., 2025; Mitze et al., 2026) (Figure 2).

Measurements were always taken at the exact location of the weather stations by reading the analog sensor and submitted via a smartphone application or a web interface to the HydroCrowd database (Campos Zeballos et al., 2025; Mitze et al., 2026). As the roofing proved to be inadequate, outliers due to exposure to intense sunlight during certain hours of the day, particularly during early morning or late afternoon, were detected in the measurements (Mitze et al., 2026). To exclude these outliers from further use, the same interquartile range outlier elimination approach as presented by Mitze et al. (2026) was used. This resulted in a moderate loss in available PM measurements (Tomebamba: 8.2%, Don Tito: 17.9% and Nkweseko: 12.4%).

At these three stations, automatic stand-alone loggers (Lascar Electronics, EL-USB-2+, Whiteparish, United Kingdom; accuracy: ±0.3 °C) for continuous measurements of air temperature and relative humidity were installed (Mitze et al., 2026). Table 1 provides an overview of the period and number of PM and automatic measurements monitored. On average, much less than one measurement per day was submitted, which limits potential autocorrelation in the measurements at each location. Please note that PM temperature data will be referred to as T_{pm} and the data from the automatic sensors as T_{as}, in the following. Due to repeated theft and vandalism, it was not possible to provide continuous data for the full measurement period for all stations (Mitze et al., 2026). In particular, this has significantly limited the data available for the analysis at the stations "Don Tito" and "Tomebamba". The substantially smaller sample size of Don Tito compared to the other stations, must be taken into account as a potential limitation for the subsequent modelling. Nevertheless, data from all stations will be used to determine whether a small number of measurements might be sufficient for bias correction.

2.2.2 ERA5-land data

The land component of the fifth generation of European ReAnalysis (ERA5-Land) (Muñoz-Sabater et al., 2021) was used as the continuous dataset, to which bias correction should be applied. It was developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and offers a global dataset from 1950 to the present with 50 variables describing the water and energy cycles over land (Muñoz-Sabater et al., 2021). It is based on a numerical integration of the ECMWF land surface model using the ERA5 climate reanalysis dataset (Muñoz-Sabater et al., 2021). This dataset relies on various data sources like meteorological station, weather ballons, satellites and other remote sensing data sources data (Muñoz-Sabater et al., 2021; Copernicus, 2025). As the same spatial and temporal resolution of data is not available for all regions in the world, interpolation was used to complete the reanalysis dataset (Muñoz-Sabater et al., 2021).

The global ERA5-Land dataset is available at a native 9 by 9 km spatial and 1-h temporal resolution (Copernicus Climate Change Service, 2019; Muñoz-Sabater et al., 2021). This product was preferred over other remote sensing datasets, due to its gapless continuous temperature availability, at global range. The 2 m air temperature of this dataset (referred to as T_{rs} in the following) was retrieved and processed via the Python earthengine-api of Google Earth Engine (Gorelick et al., 2017).

2.3 Bias correction

In this study, the bias of T_{rs} is considered as the deviation from T_{as} at the exact location of the three PM weather stations. T_{pm} was utilized for the training of point-to-point bias correction models, as the availability of automatic measurements is limited in the three study regions.

2.3.1 Data preprocessing

In a first step, timestamps of T_{pm} were assigned to the nearest full hour to match those of T_{as} and T_{rs}, for instance a measurement from 8:42 was assigned 9:00, etc. (compare Mitze et al., 2026). As T_{rs} is available in coordinated universal time (UTC), time stamps were corrected to the local time zones of the study regions (Ecuador: +5, Honduras: +6 and Tanzania: −3).

In the next step, timestamps of T_{pm} were used to retrieve the corresponding T_{rs} values for the three stations (training datasets). In the same way, T_{rs} values were also retrieved for the entire hourly T_{as} dataset (see Table 1) in order to validate the bias correction models (validation datasets). From each T_{rs} grid pixel a point value was extracted using the coordinates of the PM station (Table 1). Only the pixel value in which each station is located were used because ERA5-Land grid pixel values already represent a spatial average of the 9 by 9 km grid pixel. Incorporating surrounding pixels, for instance, via buffering or neighborhood averaging would introduce unwarranted spatial assumptions and could obscure meaningful gradients in the data, which we were trying to avoid.

2.3.2 Model training

Empirical quantile mapping is currently one of the most widely used methods for bias correction in climate research, e.g., for air temperature data (Copernicus Climate Change Service, 2025). However, the small dataset size, especially for Tomebamba (n = 67) and Don Tito (n = 23), was considered insufficient to determine representative quantiles. As the analysis of Mitze et al. (2026) showed high spearman correlation (Spearman, 1961) between analog and automatically measured air temperature data, it was assumed that there is also a high correlation between T_{pm} and T_{rs}. Therefore, simple linear regression was applied to carry out bias correction. For this, the corresponding linear regression function from the machine learning Python package Scikit-learn (Pedregosa et al., 2011a) was used to train a model for each of the three locations. The T_{rs} of the training datasets of each location were used as predictor variables for the regression models, while the corresponding T_{pm} were used as the target variables. Nevertheless, correlation between T_{pm} and T_{rs} will be considered in the statistical analysis of the data (see Section 2.4).

Due to the relatively small amount of PM data (see Table 1), we did not perform a train-test-validation split of the PM data unlike in traditional machine learning approaches, i.e., all T_{pm} measurements from each station were used only for model training. The trained models were then tested using only the validation datasets. Subsequently, the trained models were applied to predict T_{as} (i.e., bias correct T_{rs}) of the validation datasets by using T_{rs} as an input. The bias corrected values are referred to as T_{bc} in the following.

2.4 Validation and statistical analysis

2.4.1 Descriptive analysis

A descriptive analysis was conducted for the training (T_{pm} and T_{rs}) datasets, to provide a more extensive foundation for the discussion of bias correction. For that, different summary statistics (range, median, mean and standard deviation) were calculated for T_{pm} and T_{rs} of the training datasets. Boxplots were created for validation datasets to facilitate comparison of the different distributions.

To further analyze the correlation and differences between T_{rs} and T_{pm} of the training data, the normal distribution was tested using a Shapiro-Wilk test (Shapiro and Wilk, 1965), which is appropriate for smaller datasets. As normal distribution could not be confirmed ($p < 0.001$), Spearman Rank correlation ((Spearman, 1961), r_{spear}, Equation 1) was used to determine the correlation. Furthermore, 95% confidence intervals (95%-CI) were calculated for each Spearman's rho value.

$$r_{spear} = \frac{\operatorname{cov}[R[X], R[Y]]}{\sigma_{R[X]}\sigma_{R[Y]}} \quad (1)$$

where cov = covariance of the rank variables, σ_X, σ_Y = standard deviations of the rank variables.

Potential statistical differences between T_{pm} and T_{rs} were tested using a two-sided Wilcoxon signed-rank test (Wilcoxon, 1945) as normal distribution could not be confirmed in the previous step. Additionally, the temporal patterns of T_{pm} (i.e., the distribution of the actual T_{pm} values) were visually analyzed by examining the times of the day at which PM data were collected. The underlying assumption was that participants were more likely to have taken measurements during the daytime.

2.4.2 Model validation

To validate the bias correction approach, T_{rs} and T_{bc} of the validation datasets, were compared to the corresponding T_{as} for each station, whereby the comparison to T_{rs} is considered as a baseline. The mean absolute error (MAE, Equation 2), the root mean squared error (RMSE, Equation 3) and the coefficient of determination (R^2, Equation 4) were used to determine the bias correction performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{2}$$

where n = number of observations, y_i is T_{as} value and $\hat{y}_i$ is the T_{rs} or T_{bc} value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{3}$$

where n = number of observations, y_i is T_{as} value and x_i is the T_{rs} or T_{bc} value.

$$R^2 = 1 - \frac{SSres}{SS_{tot}} \tag{4}$$

where SS_{res} = residual sum of squares and SS_{tot} = total sum of squares. Please note, that the general residual-based formulation was used for R^2 which can yield negative values when predictions are worse than a simple mean predictor (Pedregosa et al., 2011b).

2.4.3 Statistical and data structure analysis

The bias correction was further analyzed by providing statistical supporting evidence. In a first step, normal distribution of the validation data (T_{rs}, T_{as} and T_{bc}) was checked. Here, the Anderson-Darling-test (Anderson and Darling, 1952) was used as it is considered more robust for larger datasets (see Table 1). As normal distribution could not be confirmed for all stations ($p < 0.001$), a non-parametric test was conducted in a next step. Since the study aims to analyze whether the distribution of the residuals of T_{bc} is significantly smaller (i.e., better) than the residuals of T_{rs}, a one-sided Wilcoxon signed-rank test (Wilcoxon, 1945) was performed. The following hypotheses were formulated:

- H_0 (Null Hypothesis): The bias of T_{bc} is not significantly smaller than that of T_{rs}.
- H_1 (Alternative Hypothesis): The bias of T_{bc} is significantly smaller than that of T_{rs}.

To account for the large sample size and minimize the probability of type 1 errors, an alpha level of 0.01 was used for this test to ensure more conservative statistical inference.

Further analysis was then conducted to understand the strengths and limitations of the bias correction using PM data. For this, samples of the validation datasets were plotted as time series to analyze the temporal pattern of the bias corrections in comparison to the ERA5-Land and automatic sensor data. By comparing this to the temporal distribution analysis of the training datasets (see Section 2.4.1), the influence of the temporal patterns in training data collection should be investigated. In order to proceed with this analysis, the bias (i.e., deviation from T_{as}) of the validation data was grouped for each hour of the day of each dataset. Thereafter, the Anderson-Darling-test was applied to all groups to verify normal distribution. As it could not be confirmed for all groups, the Friedman test (Friedman, 1937) was then conducted to compare the distribution of all bias samples against each other.

If this test was positive, indicating statistically significant differences between the individual hours of the day, a two-sided Wilcoxon signed-rank test (Wilcoxon, 1945) was conducted as a *post hoc* test in order to ascertain whether there are statistically significant differences from zero (i.e., significant deviation from T_{as}) for each hour (compared values for each hour of the day: n = 538 (Nkweseko), n = 195 (Tomebamba) and n = 420 (Don Tito)), with an alpha level of 0.01.

3 Results

3.1 Descriptive analysis

The training data (T_{pm} and T_{rs}) from the three stations are characterized by substantial differences which are highlighted by the summary statistics in Table 2.

An unbalanced number of measurements was available, leading to Nkweseko having more than 10 times more measurements available for training than Don Tito. In addition, substantial differences in the mean between T_{pm} and T_{rs} can be observed. A statistical analysis revealed significant differences in the distribution for all stations ($p < 0.001$). While for Nkweseko the range of the measured T_{pm} was slightly wider than for the corresponding T_{rs} data, range and mean values for Tomebamba and Don Tito differed more clearly from one another. For instance, the upper boundary of the temperature range for T_{pm} and T_{rs} at Tomebamba differed by almost 13 °C (Nkweseko: 3 °C and Don Tito: 6 °C). Nevertheless, all stations showed a comparable high degree of correlation: Tomebamba (r_{spear} = 0.83, 95%-CI = [0.74, 0.89]), Don Tito (r_{spear} = 0.85, 95%-CI = [0.65, 0.93]) and Nkweseko (r_{spear} = 0.86, 95%-CI = [0.82, 0.89]). Note that due to the small sample size for Don Tito, the given correlation holds greater uncertainty than Tomebamba and Nkweseko which is reflected in the wider 95% confidence interval.

The temporal pattern (i.e., number of measurements collected by participants within each hour of the day) of T_{pm} shows large differences between Nkweseko and Tomebamba on the one hand and Don Tito on the other hand (Figure 3).

Although Nkweseko and Tomebamba have high temporal coverage, no measurements were conducted in the early night until morning. In sharp contrast to that, at Don Tito measurements were only taken during the late morning and predominantly during the afternoon. Figure 4 displays the distribution of the validation datasets (T_{as} and T_{rs}) and the bias corrected air temperature (T_{bc}) using boxplots.

A comparison of the ranges of T_{bc} and T_{as} reveals that the bias correction adjusts the entire range predominantly good to the automatically measured data (T_{as}) for Nkweseko and Tomebamba. While the range for Nkweseko was only shifted slightly, it was extended for Don Tito at the lower and upper boundary. However, the small range for T_{rs} was drastically improved for Tomebamba. A visual comparison of medians and means indicates that for Nkweseko the bias correction adjusted air temperature downwards while for Tomebamba and Don Tito it was adjusted upwards. Note that a few high-end outliers (above 30 °C) are present in T_{as} for all stations, especially for Nkweseko (Figure 4A), that were not fully matched by the bias correction.

TABLE 2 Summary statistics of the training data (where $_{pm}$ refers to T_{pm} and $_{rs}$ to T_{rs} data) for the three stations used in this study (n = number of T_{pm} measurements used, Std = Standard deviation).

Station name	n	Range$_{pm}$	Range$_{rs}$	Median$_{pm}$	Median$_{rs}$	Mean$_{pm}$	Mean$_{rs}$	Std$_{pm}$	Std$_{rs}$
Nkweseko	275	11.0–31.0	12.3–28.2	17.0	20.2	17.8	20.1	3.8	3.1
Tomebamba	67	14.0–30.0	9.4–17.1	18.0	12.7	20.7	13.1	4.7	1.8
Don tito	23	24.0–33.0	18.6–27.3	28.0	21.9	28.3	22.4	2.8	2.5

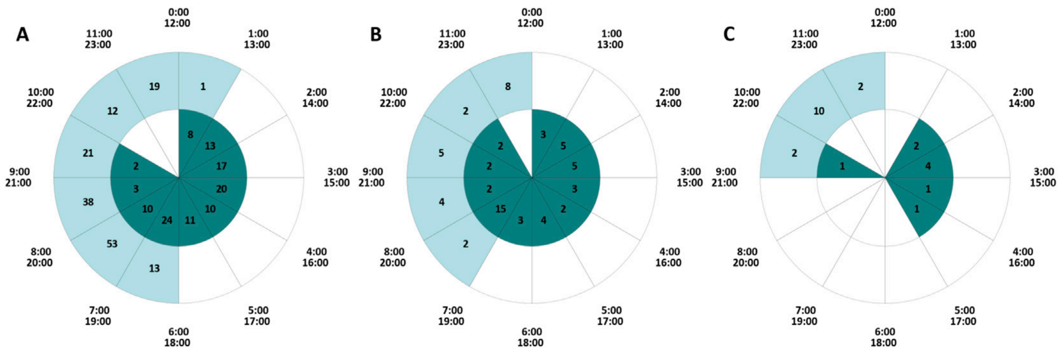

FIGURE 3
Availability of participatory monitoring data for the three stations **(A)** Nkweseko, **(B)** Tomebamba and **(C)** Don Tito in the morning (light blue shading) and afternoon (petrol shading), with black numbers indicating how many measurements are available for each hour.

3.2 Model validation

For a first overview of the validation of the bias correction the distribution of T_{rs} and T_{bc} is plotted against T_{as} (Figure 5).

Clearly, the bias-corrected data for Tomebamba and Don Tito align more closely with the automatically measured data than the original ERA5-Land (T_{rs}). For Nkweseko this is not so clear as the difference is substantially smaller than for the other two stations. Nevertheless, the data from Nkweseko and Tomebamba are also not evenly distributed along the 1:1 line, which would indicate a perfect fit. Especially higher temperatures are more likely underestimated by both T_{rs} and T_{bc}. The data from Don Tito on the other hand shows a very good fit to the 1:1 line with slightly more scattering between 25 ˚C and 30 ˚C. These impressions are confirmed by the validation metric results (Table 3).

For all three stations MAE and RMSE were reduced, where the reduction for Nkweseko was the smallest with 0.29 ˚C, while the value at Don Tito nearly halved with 1.41 ˚C. The most substantial improvement was achieved for Tomebamba, where MAE was the highest and was also reduced the most, by almost 4 ˚C. Simultaneously, R^2 was improved for all stations, where the improvement was the smallest for Nkweseko (+0.07) and the highest for Tomebamba (+1.71). Please note, that for the baseline (T_{rs}) at Tomebamba a very low R^2 of −0.98 was observed, which indicates a very poor fit with the automatically measured data, i.e., the model performs worse than a simple mean predictor. Overall, the mean represented by the MAE deviation at all stations could be reduced to less than 2 ˚C.

3.3 Statistical and data structure analysis

The previously presented results are supported by statistical evidence. A one-sided Wilcoxon signed-rank test was performed on the test data from all stations, which was significant in all cases ($p < 0.001$), rejecting the null hypothesis for all stations tested. This indicates that the distributions of the residuals of T_{bc} (bias correction) are significantly smaller (i.e. better) than those of T_{rs} (baseline).

To better understand and further analyze the structure of the bias corrected data, the first month of each station from the validation dataset (n = 720 hourly values) including the bias corrected values is displayed as an example in a timeseries plot in Figure 6.

Examining the time series for the first month of each station in Figure 6 gives an impression of which times of the day are corrected well and which are not. While daily maximum temperatures are often overestimated by ERA5-Land at Nkweseko, the bias correction corrected this for multiple days. Lower temperatures on the other hand became slightly worse after the bias correction. Rising temperatures in the morning and falling temperatures in the evening, on the other hand are corrected better. At Tomebamba (Figure 6B) both maximum and minimum temperatures were adjusted and fit better to the actual temperatures (T_{as}), while some maximum temperatures are still underestimated. Similar results could be observed for air temperature at Don Tito, where especially underestimated daily maximum temperatures were corrected in most cases (Figure 6C).

Considering these differences, especially for daily maximum temperatures, which are predominantly recorded at noon or early

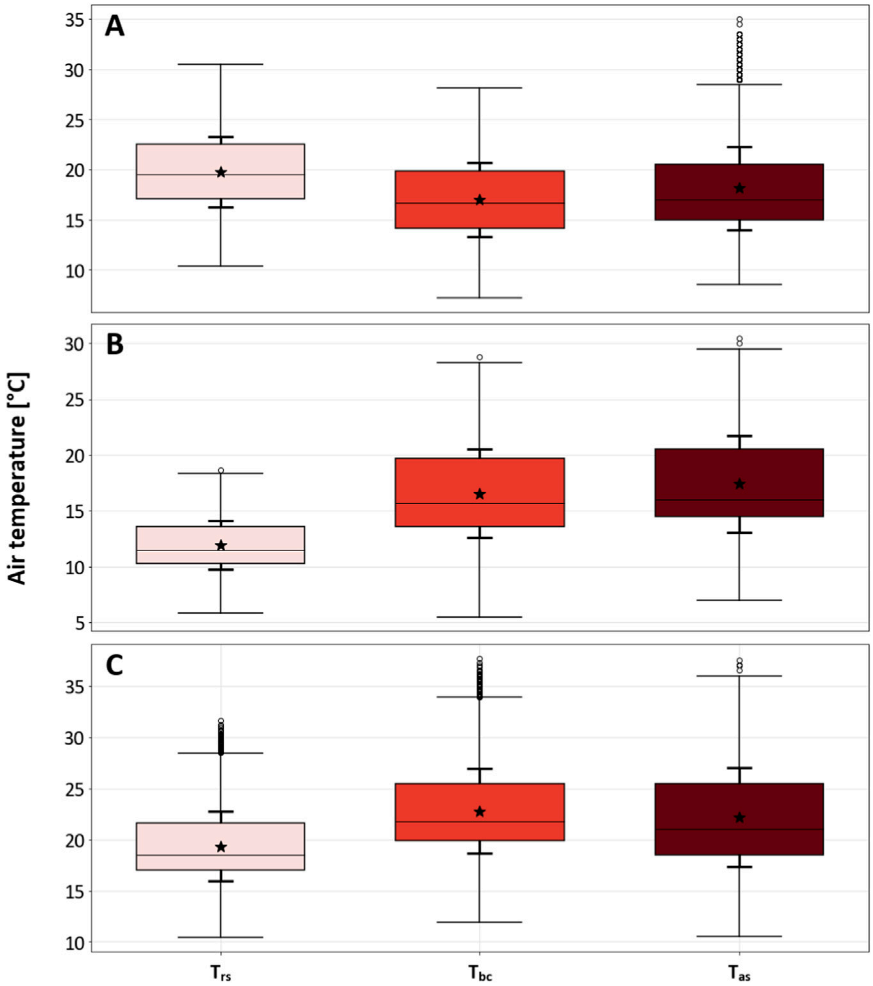

FIGURE 4
Boxplots for the validation air temperature (°C) data at the three stations **(A)** Nkweseko, **(B)** Tomebamba and **(C)** Don Tito (T_{rs} = ERA5-Land 2m air temperature, T_{bc} = bias corrected ERA5-Land 2m air temperature, T_{as} = automatically measured air temperature). The black star indicates the mean value and the thick smaller whiskers standard deviation.

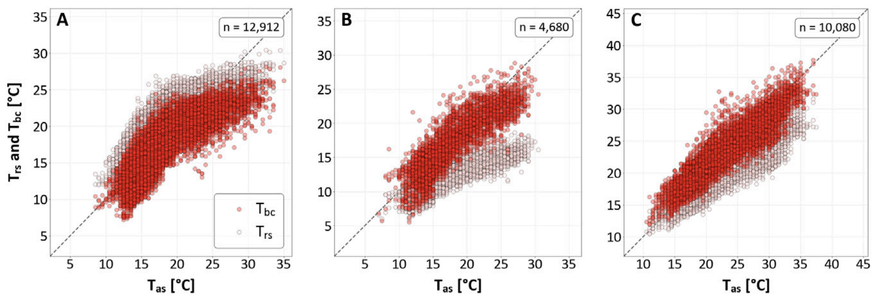

FIGURE 5
Scatterplots comparing the validation air temperature (°C) data distribution of T_{rs} (ERA5-Land 2m air temperature; pale red) and T_{bc} (bias corrected ERA5-Land 2m air temperature; red) at the three stations **(A)** Nkweseko, **(B)** Tomebamba and **(C)** Don Tito.

afternoon, boxplots averaging air temperature deviation (from T_{as}) for each hour of the day for all stations are presented in Figure 7.

While for Nkweseko (Figure 6A) T_{rs} overestimated T_{as} for most of the day, except for the late afternoon where it was quite similar, it turns to the opposite for the bias correction. Here air temperature is underestimated at most hours of the day, with only low deviations during the morning from 8–10 a.m., while the highest are observed in the late afternoon from 4 to 6 p.m. For Tomebamba (Figure 6B)

TABLE 3 Validation metric results for the three stations, with n = number of compared values, R^2_{rs} = baseline results using T_{rs} and R^2_{bc} = bias correction results using T_{bc}.

Station	n	R^2_{rs}	MAE_{rs} (°C)	$RMSE_{rs}$ (°C)	R^2_{bc}	MAE_{bc} (°C)	$RMSE_{bc}$ (°C)
Nkweseko	12,934	0.55	2.27	2.77	0.62	1.98	2.54
Tomebamba	4,682	−0.98	5.49	6.09	0.73	1.76	2.27
Don tito	10,080	0.46	2.95	3.57	0.83	1.54	1.97

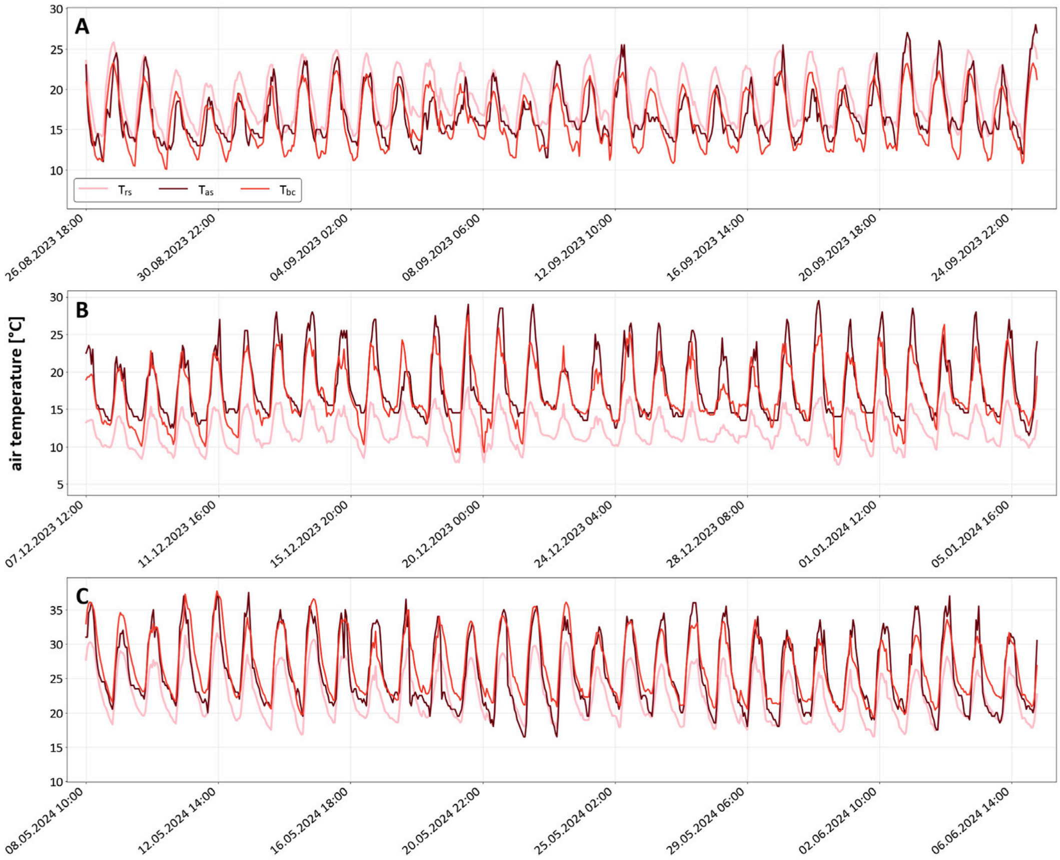

FIGURE 6
Comparison of the first month (n = 720 hourly values) of the validation air temperature (°C) data (T_{rs} = ERA5-Land 2m air temperature, light pink; T_{bc} = bias corrected ERA5-Land 2m air temperature, orange; T_{as} = automatically measured air temperature, dark red) for **(A)** Nkweseko, **(B)** Tomebamba and **(C)** Don Tito.

the most substantial improvement can be recognized, similar to prior results. Deviation for all hours of the day is substantially reduced, only leaving higher deviations at noon to late afternoon (1–5 p.m.). At Don Tito, however, the bias correction increases overall air temperature values excessively. This causes the slight underestimation to turn into a slight overestimation at night. Meanwhile, values before noon show low deviation, but values in the afternoon are still more underestimated. Overall, the bias corrected values of all stations show a common characteristic: air temperature values from the early midday until the late afternoon have slightly higher deviations than the rest of the day. Air temperature also has a much higher value range during the daytime (approximately from 8 a.m. to 6 p.m.) compared to nighttime. A statistical analysis using the Friedman test comparing deviation for all hours of the day was significant ($p < 0.001$) indicating statistically significant differences between the different times of the day. To follow this up, a *post hoc* Wilcoxon signed-rank test was used to compare each hour of the day with 0 (i.e., no deviation). Full statistical results can be found in Supplementary Material. For Nkweseko the test indicated statistically significant deviations from zero for all hours of the day except 8 and 11 a.m., while for Tomebamba only for 8 and 10 a.m. no significant differences were determined. On the other hand, for all hours of the day except between 9 and

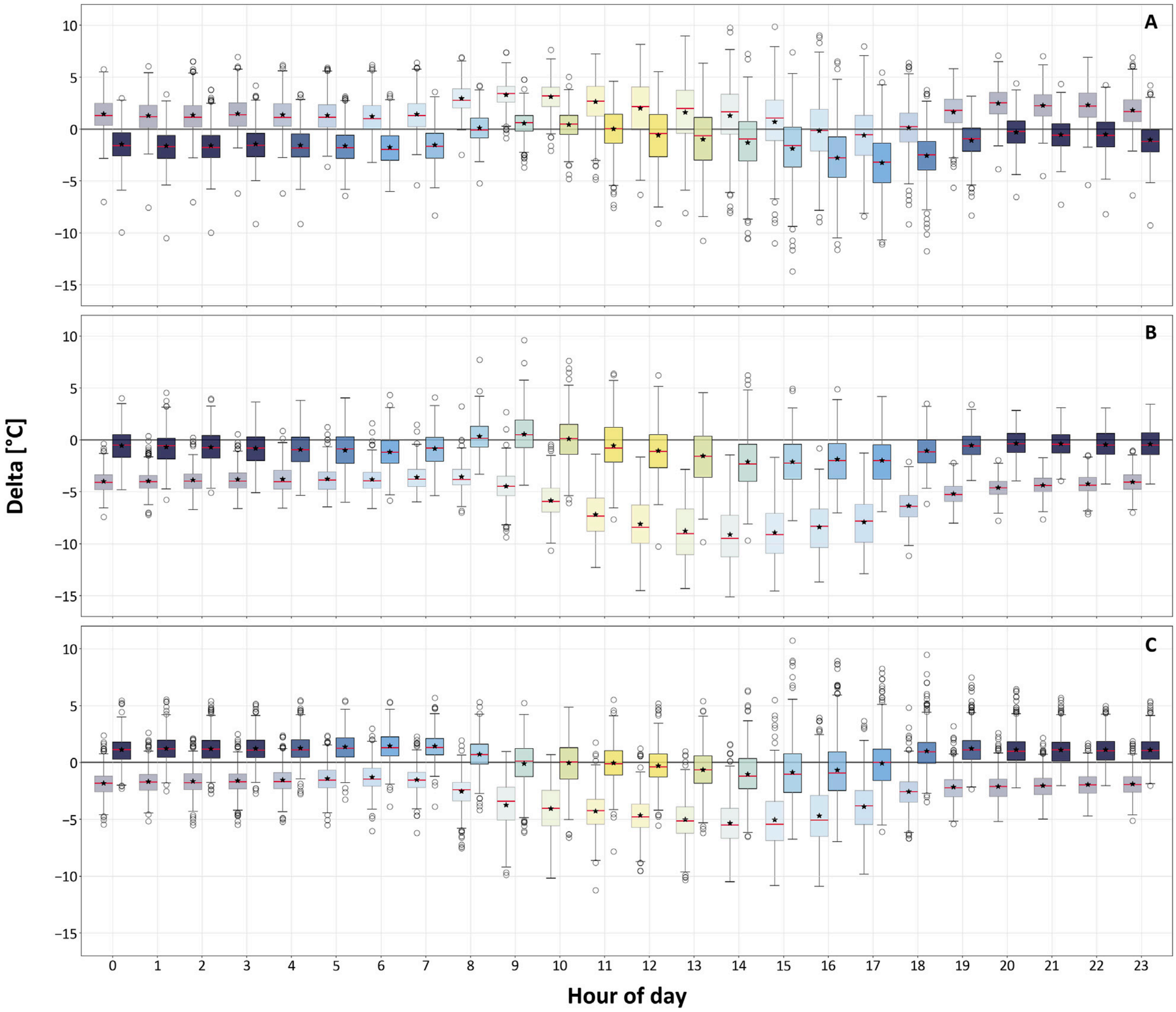

FIGURE 7
Boxplots (vibrant colors) comparing delta of the air temperature (°C) bias correction (T_{bc} - T_{as}) for every hour of the day at **(A)** Nkweseko, **(B)** Tomebamba and **(C)** Don Tito. The pale boxplots in the background display delta for the original ERA5-Land air temperature data (T_{rs} - T_{as}). The red line indicates the median, the black star the mean value of each hour while the colors represent the time of the day.

11 a.m. and 5 p.m. significant differences were determined for Don Tito.

4 Discussion

The utilization of participatory monitoring air temperature data for bias correction enhanced the accuracy of ERA5-Land 2 m air temperature in various tropical remote locations. Simple linear regression was carried out for bias correction. However, it should be noted that the degree of correction exhibited by the stations was not uniform, with an average deviation of up to 2 °C being observed. The performance of the stations was found to vary, thus providing a foundation for critical analysis of the data and approach utilized.

4.1 Using participatory monitoring data for bias correction

Several previous studies analyzed bias correction or downscaling on ERA5-Land data in different study regions with varying spatiotemporal scales. Study regions covered a wide range of topography from the Yellow river (Xu et al., 2024; Zhang et al., 2024) and lower Jinsha River Basin (Zhang et al., 2025) to mountainous topography in the Higher Central Himalaya (Satyapragyan et al., 2026), the High Atlas Mountains (Sebbar et al., 2023) or the Southern Alps (Niazkar et al., 2023; Dhawan et al., 2024). None of them covered tropical mountainous regions. Among different metrics for validation, MAE, RMSE and R^2 were used most frequently, similar to our case (Niazkar et al., 2023; Sebbar et al., 2023; Dhawan et al., 2024; Xu et al., 2024; Zhang et al., 2024;

Satyapragyan et al., 2026). While in our study only linear regression model was applied to correct T_{rs}, the methods varied in most investigations using different statistical methods, machine learning or machine learning ensembles (Niazkar et al., 2023; Sebbar et al., 2023; Dhawan et al., 2024; Xu et al., 2024; Zhang et al., 2024). Regularly, multiple weather stations were used as the main source of correction data (Zhu et al., 2021; Niazkar et al., 2023; Sebbar et al., 2023; Dhawan et al., 2024; Xu et al., 2024; Zhang et al., 2024; Satyapragyan et al., 2026) while for our analysis only one station for each location was used, what might be a limiting factor for our approach. In some cases additional data like digital elevation models, land surface temperature, land cover data or vegetation indices were used as well (Sebbar et al., 2023; Xu et al., 2024; Zhang et al., 2024). As our approach aimed to keep the bias correction as simple as possible, this was not further considered in this study. However, a conceivable but less obvious option for integrating more predictor variables in our approach, would be to use additional ERA5-Land variables for the regression, such as land surface temperature or soil temperature. Apart from that, integrating other additional time series remote sensing data in our analysis could be complex, as most products provide data at intervals of days or multiple days. Using PM data with irregular measurement frequency requires higher temporal resolution of training data, which only ERA5-Land provided at all locations investigated. Nevertheless, there is a potential possibility to integrate, for instance MODIS land surface temperature (Hulley and Hook, 2017), as it is possible to fill gaps in the dataset and to calculate hourly values (Shiff et al., 2021; Yu et al., 2022; Zare Khormizi et al., 2025). However, this is only possible under clear sky conditions. This limits its application in tropical mountainous regions, as the cloud cover might be more frequent there (Tropical Montane Cloud Forests, 2013; Bassiouni et al., 2017).

In general, the use of analog participatory monitoring data for bias correction in ERA5-Land data remains unexplored prior to this work. PM datasets have characteristics like irregular measurement frequency or unequal coverage of different time periods and therefore do not replicate a full temporal cycle, as our analysis has shown. Furthermore, the PM temperature data used in our study is of good quality but still deviates from automatically measured air temperature (MAE: Nkweseko = 0.74, Tomebamba = 0.95 and Don Tito = 1.54 °C), even though no statistically significant differences between analog and automatic measurements were found (see Mitze et al., 2026). The deviation at Don Tito also roughly corresponds to the improvement achieved by the bias correction (1.44 °C). As the bias correction was benchmarked with automatically measured data, this could be another source of distortion.

These characteristics are potential weaknesses of PM data in comparison to automatically measured data and should be taken into account when comparing our results with those of previous studies. Therefore, it is logical that the accurafy of our approach does not fully match that of approaches using regular weather station data. Hence, this might be the greatest limitation of our data. For instance, Niazkar et al. (2023) and Dhawan et al. (2024) achieved similar or even better results for MAE (0.7°– 2.0 °C), RMSE (1.0 °C – 2.5 °C) and R^2 (up to 0.98) but also used more complex machine learning models and high-resolution data for air temperature correction.

Another aspect of using PM data for bias correction, is the reliability of the approach under the impact of climate change with rising air temperatures. This is a well-known phenomenon called ‘concept drift’, whereby the statistical properties of the target variable - air temperature, in this case - change over time, causing the model to become inaccurate (Widmer and Kubat, 1996). As there is regularly no continuous time series of PM data available for training the bias correction model, it may become outdated at a certain point in time. Usually, there is no exact point of time for when this could happen. For this reason, there are several approaches to deal with this issue, for instance recalibrating the model periodically or when a drift threshold is exceeded, i.e., the accuracy is dropping (Suárez-Cetrulo et al., 2023; Najafi et al., 2026). As air temperature is expected to significantly increase over the next decades, especially in tropical areas a decadal recalibration might be an appropriate way for PM data (Diffenbaugh and Scherer, 2011; Po-Chedley et al., 2022; Ying et al., 2022). This poses a further challenge for the use of PM data, as long-term perspectives remain one of the key challenge of successful PM campaigns (Buytaert et al., 2014; Cunha et al., 2017; Pandeya et al., 2021).

Overall, using PM data for bias correction proved suitable for enhancing the accuracy of air temperature data for specific locations in our study regions. However, this approach may not achieve the same level of accuracy as other, more complex bias correction methods using more extensive and more accurate input data obtained from automatic sensors. Also, recalibration might be necessary after a certain time due to climate change.

4.2 The role of input data and bias correction method

While annual, monthly or daily temporal scales were used for most climate related bias corrections, the two aforementioned examples by Niazkar et al. (2023) and Dhawan et al. (2024) also showed the applicability of bias correction of hourly data, which has not yet been investigated in depth (Dhawan et al., 2024). These studies included several years of data for bias correction model training, while in our case only 23 to 275 irregular analog measurements were used, as not more data were available from the PM project. The results showed quite different performances for the three stations. The lowest deviation from automatic sensors were achieved at Don Tito with MAE = 1.54 °C, while this station also provided the lowest number of training samples (n = 23). On the other hand, the station with the most samples, Nkweseko (n = 275) performed worst (MAE = 1.98 °C). Furthermore, the automatically measured data at Nkweseko (T_{as}) contained more high-end outliers which were not correctly reproduced by the regression (see Figure 4A). This suggests that more measurements, especially during the peak temperatures could improve the accuracy of the approach. Also, Tomebamba and Nkweseko show a similar distribution of measurements over the day (see Figure 3), despite having different total amounts of measurements. While this suggests that more measurements might not necessarily ensure better results, the small sample size (n = 23) at Don Tito results in statistically unstable MAE and R^2 values, which are strongly dependent on the subsample of measurements. As such, one has to be careful with generalization of the results from Don Tito, as the validation rather

TABLE 4 Stations elevations (meters above sea level), the mean elevation of their corresponding ERA5-Land pixels and the elevation difference.

Station name	Station elevation (m a.s.l.)	ERA5-land pixel elevation (m a.s.l.)	Difference (m a.s.l.)
Don tito	1,182	1,199	17
Tomebamba	2,380	3,332	952
Nkweseko	1,700	1,483	−217

reflects how well the model has captured the sample's characteristics rather than the generalizability (Pradhan et al., 2024).

Nevertheless, the very low number at Don Tito prompts the question of whether there is a minimum number of measurements required when using PM data. This should be further explored in future investigations, for instance by sampling different subsets of PM data from the same location. However, it could be more important that the available PM measurements cover as wide a range of air temperatures as possible, regardless of how many measurements are available. None of the stations provided measurements for every hour of the day: during the night between 0 and 6 a.m. no measurements were available at all. Therefore, specifically with regard to nighttime and peak temperatures, we can speak of an extrapolation of the model. Because of this, further PM projects could investigate the added value of measurements that are spread more evenly throughout the day. As there is relatively low fluctuation in air temperature throughout the year in tropical climates, short PM campaigns in such regions covering a wide range of day and night air temperatures might be sufficient, and avoid the problem of ensuring long-term participation, which is often a key challenge in PM projects (Buytaert et al., 2014; Cunha et al., 2017; Pandeya et al., 2021). Moreover, the analysis of the temporal distribution of T_{pm} and the results for each hour of the day do not indicate that the incorrect prediction of T_{bc} can be linked to the incomplete diurnal coverage, as higher deviations were observed predominantly during the hours for which most data were collected. However, it cannot be entirely ruled out that the correction could have been even better with a data set covering all times of day evenly. Also, the deviation of T_{pm} from the automatic sensor data (Mitze et al., 2026) can only partially explain the performance, as Nkweseko (MAE = 0.74 °C) has the lowest and Don Tito (MAE = 1.54 °C) has the highest deviation.

As the varying performance and limitation of this approach cannot be entirely attributed to structure of the participatory monitoring data, the cause might also be linked to the underlying structure of the ERA5-Land data, the correspondence of the 9 km grid value with the location of the station and the linear regression bias correction used.

The spatial reproduction of temperature variability, which the 9 km pixel of ERA5-Land offers for mountainous areas might be another possible source of distortion. There is likely a correlation between the ERA5-Land temperature values and ground-based measurements within a given pixel. However, a perfect fit may not be ensured for every location within a pixel because the ERA5-Land value is only an average for the entire area. To analyze elevation differences, we calculated the mean elevation for each ERA5-Land pixel, where the stations are located. For that we used the NASA 30m Digital Elevation Model (NASA JPL, 2020) and averaged the elevation for each ERA5-Land pixels area (see Table 4).

While the difference is low for Don Tito with only 17 m, the pixel that contains Tomebamba has roughly 952 m higher mean elevation and Nkweseko' pixel is even lower than the station (−217 m). The huge differences between the stations highlight the challenges of using ERA5 land data for mountainous terrain. This is especially the case for Nkweseko and Tomebamba, which are both located in terrain, where the elevation increases sharply within a short distance. The 9 km grid pixel can therefore only offer a rough estimation for the area, which can be seen in the example of Nkweseko displayed in Figure 8. Due to large elevation differences over short distances, values of adjacent pixels can differ considerably: the station Nkweseko is located within a pixel with a reanalysis temperature around 15 °C, the pixels diagonally above it around the Kibo summit (blue pixels in the right upper corner) show values substantially below 10 °C. Furthermore, the value of the pixel might be considerably lower or higher than the actual temperature at any given location within the pixel. Including a lapse-rate adjustment prior to bias-correction could partially correct this, as it would adjust the temperature value based on the orography elevation of the ERA5-Land pixel and the elevation of the site. While this could reduce the systematic error of ERA5-Land temperature values, it would not adjust the temporal variation in deviation induced by other sources. Furthermore, such a systematic offset is also partially addressed through the intercept of the linear regression model. For both Tomebamba and Nkweseko, the deviation was reduced by the linear regression model from 5.49 °C (Tomebamba) and 2.27 °C (Nkweseko) to 1.76 °C and 1.98 °C, respectively. For these reasons, lapse-rate adjustment was not considered in this study. It would, however, be worth investigating whether including a local lapse-rate adjustment would further reduce the deviation to enhance the reliability of the approach.

Furthermore, the model based on linear correlation can only work sufficiently with high linear correlation between the predictor (T_{rs}) and the target variable (T_{pm}). Although this is quite high in our study with r_{spear} ranging from 0.83 to 0.86, it does not establish a perfect monotonic relationship between the variables. There might be several explanations for that. Firstly, our PM data is of good quality but not a precise replication of automatically measured data which might affect the correlation with ERA5-Land data, which is partly based on automatically measured data, to some degree. Another reason might be the linear regression model used for the bias correction, which can only transmit a linear relationship. Using more complex models that catch other characteristics apart from just linear relationships between variables could possibly lead to different results. This limitation can be seen in the hourly based comparison of the deviation in Figure 7. The bias correction using linear regression improved overall accuracy of the air temperature by only shifting the overall dataset into a specific direction, with ERA5-Land 2m air temperature at Nkweseko being decreased, while being increased for Tomebamba and Don Tito. The original ERA5-Land trend used for prediction can still be recognized. As linear regression cannot be expected to considerably change the distribution of the data, other

10.3389/fenvs.2026.1859873

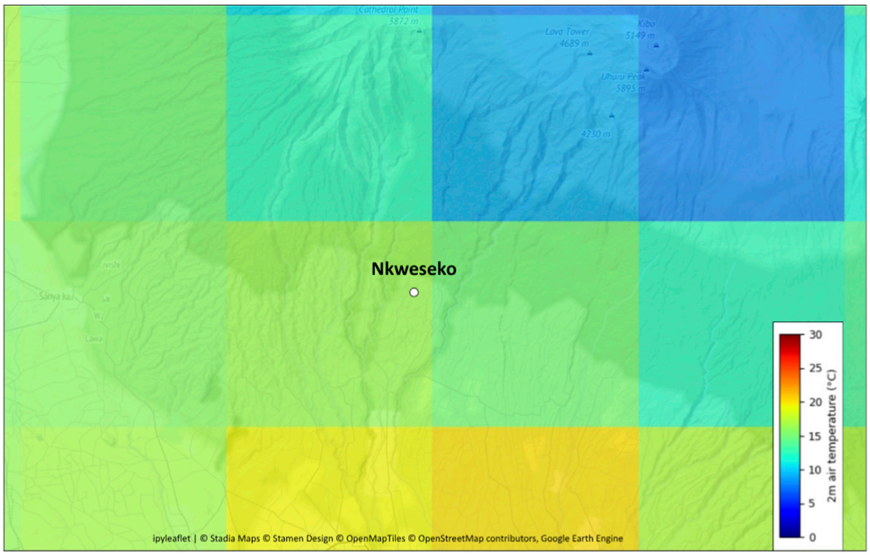

FIGURE 8
Example of ERA5-Land 2 m air temperature pixels at the southern slopes of Kilimanjaro around the Nkweseko station (white point) from the 30th of April 2024.

methods of bias correction besides linear regression should be tested in further research using PM data.

In summary, the limitations of our approach, which applies simple linear regression and participatory monitoring data, can most likely be linked to the original structure of the ERA5-Land air temperature data and the regression method used. The impact of restricted availability of PM data also remains a possible limitation, albeit not definitively established.

5 Conclusion

This analysis investigated how participatory monitoring could be used for the bias correction of ERA5-Land air temperature data in diverse remote tropical mountainous regions in Ecuador, Honduras and Tanzania where limited automatically measured hydrometeorological data are available. For bias correction, linear regression was used and for validation purpose, ERA5-Land and the corrected data were compared to automatically measured air temperature using MAE, RMSE and R^2.

Statistically significant smaller residuals were detected for the bias-corrected air temperature at all locations. Validation results with different metrics showed varying degree of success of the bias correction with the lowest MAE at Don Tito (Honduras) with 1.54 °C and the highest for Nkweseko (Tanzania) with 1.98 °C. Nevertheless, the most significant improvement has been observed for Tomebamba with MAE reduced from 5.49 °C to 1.76 °C. A thorough examination of the test data revealed that the automatically measured data does not perfectly align with the bias-corrected data with only a few hours of the day having no significant deviations from zero.

It can be concluded that with small participatory monitoring datasets it is possible to significantly enhance ERA5-Land air temperature accuracy at remote tropical mountainous locations. Asymmetric distribution of the training data did not seem to result in distorted prediction results, although this was not specifically evaluated and should be addressed in future research. Rather, the ERA5-Land data and the bias correction method used might influence the results to a higher degree. Therefore, participatory monitoring in combination with ERA5-Land might be one solution for remote regions with insufficient automatic weather station density. As this data is required for a variety of applications, such as agriculture, health protection, hydrology or in general climate change related issues, it might be a crucial contribution to an improved data situation in such regions.

In the next step, the approach should be tested in other geographic regions using analog temperature measurements. Another possible next step could be an extended analysis of other ERA5-Land parameters that might have higher correlation with PM air temperature data. Here, for instance, land surface temperature or soil temperature could be used to test if further accuracy-improvements are possible. Also, regressions for other parameters like relative humidity could be investigated, if such measurements are also available from PM projects.

To check if other bias correction methods are better suited for PM data, we suggest testing this approach with other techniques. This could be traditional machine learning methods, like random forest regression or even deep learning models like Multilayer Perceptrons. As the location of the PM station within the ERA5-Land pixel seems to have an influence on the correction accuracy, other PM data at different locations and varying elevations within the same pixel should be tested as well to further analyze the influence of the location of the station.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.5281/zenodo.17712676.

Author contributions

FM: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. SJ: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing. LB: Funding acquisition, Supervision, Writing – review and editing. JC: Data curation, Writing – review and editing. TD: Supervision, Writing – review and editing. BW: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. LB and SJ acknowledge funding of the DFG project "Biodiversity and the supply of water-related NCPs" (BR 2238/35–1/-2 and JA 3059/4–2) as part of the DFG Research Unit 5064 "The role of nature for human wellbeing in the Kilimanjaro Social-Ecological System" (Kili-SES). FM, JZ, SJ, and BW would like to thank the Kurt Eberhard Bode Foundation for funding the junior research group "HydroCrowd" (grant number 0122/40195/2022).

Acknowledgments

The authors would like to thank all our project partners for their help in setting up the hydrometeorological monitoring network in Honduras, Ecuador and Tanzania. Special thanks go to Mancomunidad de Municipios del Parque Nacional Montaña de Celaque (Mapance) and Alliance of Bioversity International and CIAT in Honduras; Empresa Pública Municipal de Telecomunicaciones, Agua Potable, Saneamiento y Gestión Ambiental del cantón Cuenca (ETAPA) and Universidad de Cuenca in Ecuador; the Kilimanjaro National Park Authority (KINAPA) and the DFG Research Unit in Tanzania, for their support of the HydroCrowd project. Finally, we would like to say thank you to all volunteers for their invaluable contribution, without their dedication and commitment, the success of this project would not have been attainable.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. The authors declare that generative AI was used during the preparation of the manuscript in order to correct grammar and spelling of the text using the online version of DeepL Write (DeepL SE, Cologne, Germany). After using this tool, the authors reviewed and edited the content as needed.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2026.1859873/full#supplementary-material

References

Abraham, E. M. (2018). Understanding climate change adaptation in Africa: key considerations. *Manag. Environ. Qual. Int. J.* 29, 165–179. doi: 10.1108/MEQ-01-2017-0001

Aguilar, A. (2005). Remote sensing of forest regeneration in highland tropical forests. *GIScience and Remote Sens.* 42, 66–79. doi: 10.2747/1548-1603.42.1.66

Anderson, T. W., and Darling, D. A. (1952). Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *Ann. Math. Statistics* 23, 193–212. doi: 10.1214/aoms/1177729437

Bassiouni, M., Scholl, M. A., Torres-Sanchez, A. J., and Murphy, S. F. (2017). A method for quantifying cloud immersion in a tropical Mountain forest using time-lapse photography. *Agric. For. Meteorology* 243, 100–112. doi: 10.1016/j.agrformet.2017.04.010

Beele, E., Reyniers, M., Aerts, R., and Somers, B. (2022). Quality control and correction method for air temperature data from a citizen science weather station network in leuven, Belgium. *Earth Syst. Sci. Data* 14, 4681–4717. doi: 10.5194/essd-14-4681-2022

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., et al. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Front. Earth Sci.* 2. doi: 10.3389/feart.2014.00026

Campos Zeballos, J., Valencia, J., Codalli, F., Mitze, F., Shagega, F., Weeser, B., et al. (2025). Evaluating participatory monitoring in mountainous tourist regions. *Front. Environ. Sci.* 13, 1537278. doi: 10.3389/fenvs.2025.1537278

Cavazos, T., Bettolli, M. L., Campbell, D., Sánchez Rodríguez, R. A., Mycoo, M., Arias, P. A., et al. (2024). Challenges for climate change adaptation in Latin America and the Caribbean region. *Front. Clim.* 6, 1392033. doi: 10.3389/fclim.2024.1392033

Copernicus (2025). Climate reanalysis | copernicus. Available online at: https://climate.copernicus.eu/climate-reanalysis (Accessed November 11, 2025).

Copernicus Climate Change Service (2019). *ERA5-Land Hourly Data from 1950 to Present*. doi: 10.24381/CDS.E2161BAC

Copernicus Climate Change Service (2025). Copernicus climate change service - global impacts. *What Is Bias Correction?*. Available online at: https://climate.copernicus.eu/sites/default/files/2021-01/infosheet7.pdf (November 11, 2025).

Cunha, D. G. F., Marques, J. F., Resende, J. C. D., Falco, P. B. D., Souza, C. M. D., and Loiselle, S. A. (2017). Citizen science participation in research in the environmental sciences: key factors related to projects' success and longevity. *An. Acad. Bras. Ciênc.* 89, 2229–2245. doi: 10.1590/0001-3765201720160548

Danielsen, F., Burgess, N. D., Balmford, A., Donald, P. F., Funder, M., Jones, J. P. G., et al. (2009). Local participation in natural resource monitoring: a characterization of approaches. *Conserv. Biol.* 23, 31–42. doi: 10.1111/j.1523-1739.2008.01063.x

Davids, J. C., Devkota, N., Pandey, A., Prajapati, R., Ertis, B. A., Rutten, M. M., et al. (2019). Soda bottle science—citizen science monsoon precipitation monitoring in Nepal. *Front. Earth Sci.* 7, 46. doi: 10.3389/feart.2019.00046

Dhawan, P., Torre, D. D., Niazkar, M., Kaffas, K., Larcher, M., Righetti, M., et al. (2024). A comprehensive comparison of bias correction methods in climate model simulations: application on ERA5-Land across different temporal resolutions. *Heliyon* 10, e40352. doi: 10.1016/j.heliyon.2024.e40352

Diffenbaugh, N. S., and Scherer, M. (2011). Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Clim. Change* 107, 615–624. doi: 10.1007/s10584-011-0112-y

Fankhauser, S., and McDermott, T. K. J. (2014). Understanding the adaptation deficit: why are poor countries more vulnerable to climate events than rich countries? *Glob. Environ. Change* 27, 9–18. doi: 10.1016/j.gloenvcha.2014.04.014

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* doi: 10.1080/01621459.1937.10503522

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth engine: planetary-Scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Gütlein, A., Gerschlauer, F., Kikoti, I., and Kiese, R. (2018). Impacts of climate and land use on N2O and CH4 fluxes from tropical ecosystems in the mt. Kilimanjaro region, Tanzania. *Glob. Change Biol.* 24, 1239–1255. doi: 10.1111/gcb.13944

Hemp, A. (2005). Climate change-driven forest fires marginalize the impact of ice cap wasting on Kilimanjaro. *Glob. Change Biol.* 11, 1013–1023. doi: 10.1111/j.1365-2486.2005.00968.x

Hooker, J., Duveiller, G., and Cescatti, A. (2018). A global dataset of air temperature derived from satellite remote sensing and weather stations. *Sci. Data* 5, 180246. doi: 10.1038/sdata.2018.246

Hulley, G., and Hook, S. (2017). *MOD21 MODIS/terra Land Surface Temperature/3-Band Emissivity 5-Min L2 1km V006*. doi: 10.5067/MODIS/MOD21.006

Lee, H., Calvin, K., Dasgupta, D., Krinmer, G., Mukherji, A., Thorne, P., et al. (2023). *Synthesis Report of the IPCC Sixth Assessment Report (AR6), Geneva, Switzerland: Longer Report*. IPCC. Available online at: https://mural.maynoothuniversity.ie/id/eprint/17733/ (Accessed May 17, 2026).

Lompar, M., Lalić, B., Dekić, L., and Petrić, M. (2019). Filling gaps in hourly air temperature data using debiased ERA5 data. *Atmosphere* 10, 13. doi: 10.3390/atmos10010013

Maraun, D. (2016). Bias correcting climate change simulations - a critical review. *Curr. Clim. Change Rep.* 2, 211–220. doi: 10.1007/s40641-016-0050-x

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* 29, 897–910. doi: 10.1175/JTECH-D-11-00103.1

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Codalli, F., Shagega, F. P., et al. (2026). Validation of analog sensor measurements in hydrometeorological participatory monitoring in various tropical countries. *Front. Earth Sci.* 14, 1721642. doi: 10.3389/feart.2026.1721642

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13, 4349–4383. doi: 10.5194/essd-13-4349-2021

Najafi, H., Lagerwall, G. L., Obeysekera, J., and Liu, J. (2026). Machine learning in climate downscaling: a critical review of methodologies, physical consistency, and operational applications. *Water* 18, 271. doi: 10.3390/w18020271

NASA JPL (2020). *NASADEM Merged DEM Global 1 Arc Second V001*. doi: 10.5067/MEASURES/NASADEM/NASADEM_HGT.001

Ngcamu, B. S. (2023). Climate change effects on vulnerable populations in the global south: a systematic review. *Nat. Hazards* 118, 977–991. doi: 10.1007/s11069-023-06070-2

Niazkar, M., Piraei, R., Menapace, A., Dhawan, P., Torre, D. D., Larcher, M., et al. (2023). Bias correction of ERA5-Land temperature data using standalone and ensemble machine learning models: a case of northern Italy. *J. Water Clim. Change* 15, 271–283. doi: 10.2166/wcc.2023.669

Njue, N., Stenfert Kroese, J., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L., et al. (2019). Citizen science in hydrological monitoring and ecosystem services management: state of the art and future prospects. *Sci. Total Environ.* 693, 133531. doi: 10.1016/j.scitotenv.2019.07.337

NOAA (2025). Search | climate data online (CDO) | national climatic data center (NCDC). *Clim. Data Online Search*. Available online at: https://www.ncei.noaa.gov/cdo-web/search (Accessed November 4, 2025).

Pandeya, B., Buytaert, W., and Potter, C. (2021). Designing citizen science for water and ecosystem services management in data-poor regions: challenges and opportunities. *Curr. Res. Environ. Sustain.* 3, 100059. doi: 10.1016/j.crsust.2021.100059

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011a). Scikit-learn: machine learning in python. *J. Machine Learn. Research* 12, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011b). Scikit-learn: machine learning in Python, sklearn.metrics/r2_score. *J. Mach. Learn. Res.* 12, 2825–2830.

Po-Chedley, S., Fasullo, J. T., Siler, N., Labe, Z. M., Barnes, E. A., Bonfils, C. J. W., et al. (2022). Internal variability and forcing influence model-satellite differences in the rate of tropical tropospheric warming. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2209431119. doi: 10.1073/pnas.2209431119

Pradhan, I. P., Mahanta, K. K., Liou, Y.-A., Chauhan, A., and Shukla, D. P. (2024). Machine learning based high-resolution air temperature modelling from landsat-8, MODIS, and *in-situ* measurements with ERA-5 inter-comparison in the data sparse regions of Himachal Pradesh. *Bull. Atmos. Sci.& Technol.* 5, 22. doi: 10.1007/s42865-024-00085-8

Prihodko, L., and Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sens. Environ.* 60, 335–346. doi: 10.1016/S0034-4257(96)00216-7

Rajagopalan, P., Andamon, M. M., and Paolini, R. (2020). Investigating thermal comfort and energy impact through microclimate monitoring-a citizen science approach. *Energy Build.* 229, 110526. doi: 10.1016/j.enbuild.2020.110526

Satyapragyan, S., Yadav, J. S., and Bhambri, R. (2026). Bias corrections of ERA5 and ERA5-land temperature using automatic weather station data in the higher central himalaya: implications for hydro-meteorological and glaciological research. *J. Hydrology Regional Stud.* 63, 103079. doi: 10.1016/j.ejrh.2025.103079

Sebbar, B., Khabba, S., Merlin, O., Simonneaux, V., Hachimi, C. E., Kharrou, M. H., et al. (2023). Machine-learning-based downscaling of hourly ERA5-Land air temperature over mountainous regions. *Atmosphere* 14, 610. doi: 10.3390/atmos14040610

Sen Roy, S. (2018). "Climate change in the global south: trends and spatial patterns," in *Linking Gender to Climate Change Impacts in the Global South*. Editor S. Sen Roy (Cham: Springer International Publishing), 1–25. doi: 10.1007/978-3-319-75777-3_1

Shagega, F. P., Codalli, F., Jacobs, S., Munishi, S. E., Windhorst, D., and Breuer, L. (2025). Quantifying preferential flow occurrence in dependence of land cover on the southern slopes of Mount Kilimanjaro, Tanzania. *J. Hydrology Regional Stud.* 58, 102215. doi: 10.1016/j.ejrh.2025.102215

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. doi: 10.2307/2333709

Shiff, S., Helman, D., and Lensky, I. M. (2021). Worldwide continuous gap-filled MODIS land surface temperature dataset. *Sci. Data* 8, 74. doi: 10.1038/s41597-021-00861-7

Spearman, C. (1961). The proof and measurement of association between two things. *East Norwalk, CT, U. S. Applet.*, 45–58. doi: 10.1037/11491-005

Suárez-Cetrulo, A. L., Quintana, D., and Cervantes, A. (2023). A survey on machine learning for recurring concept drifting data streams. *Expert Syst. Appl.* 213, 118934. doi: 10.1016/j.eswa.2022.118934

Sun, Y.-J., Wang, J.-F., Zhang, R.-H., Gillies, R. R., Xue, Y., and Bo, Y.-C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. *Theor. Appl. Climatol.* 80, 37–48. doi: 10.1007/s00704-004-0079-y

Tropical Montane Cloud Forests (2013). 79–85. doi: 10.1016/B978-0-12-384703-4.00519-0

United Nations (2022). UNCTAD handbook of statistics 2022. Available online at: https://www.un-ilibrary.org/content/books/9789210021784 (November 11, 2025).

Valdez, M. C., Chang, K.-T., Chen, C.-F., Chiang, S.-H., and Santos, J. L. (2017). Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. *Geomatics, Nat. Hazards Risk* 8, 876–892. doi: 10.1080/19475705.2016.1278404

Widmer, G., and Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Mach. Learn* 23, 69–101. doi: 10.1007/BF00116900

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biom. Bull.* 1, 80–83. doi: 10.2307/3001968

WMO (2017). *Challenges in the Transition from Conventional to Automatic Meteorological Observing Networks for long-term Climate Records*. Available online at: https://library.wmo.int/idurl/4/55812 (Accessed November 14, 2025).

WMO, (2025a). *WMO Confirms 2024 as Warmest Year on Record at About 1.55°C Above Pre-industrial Level*. Geneva, Switzerland: Press release. Available online at: https://wmo.int/news/media-centre/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level (Accessed November 4, 2025).

WMO (2025b). *World Weather Information Service - Climate Report for Cuenca, Ecuador for the Period 1981 - 2010*. Geneva, Switzerland: World Weather Information Service. Available online at: https://worldweather.wmo.int/en/city.html?cityId=2047 (Accessed October 31, 2025).

World Bank (2025). *Ecuadorecu - Country Overview | Climate Change Knowledge Portal - Province Azuay*. Available online at: https://climateknowledgeportal.worldbank.org/country/ecuador (Accessed October 31, 2025).

Xu, L., Du, J., Ren, J., Hu, Q., Qin, F., Mu, W., et al. (2024). Improving the ERA5-Land temperature product through a deep spatiotemporal model that uses fused multi-source remote sensing data. *Remote Sens.* 16, 3510. doi: 10.3390/rs16183510

Ying, J., Collins, M., Cai, W., Timmermann, A., Huang, P., Chen, D., et al. (2022). Emergence of climate change in the tropical Pacific. *Nat. Clim. Chang.* 12, 356–364. doi: 10.1038/s41558-022-01301-z

Yu, P., Zhao, T., Shi, J., Ran, Y., Jia, L., Ji, D., et al. (2022). Global spatiotemporally continuous MODIS land surface temperature dataset. *Sci. Data* 9, 143. doi: 10.1038/s41597-022-01214-8

Zare Khormizi, H., Jafari, M., Ghafarian Malamiri, H., Tavili, A., and Keshtkar, H. (2025). Generating MODIS hourly land surface temperature under clear sky conditions using fourier series analysis. *Int. J. Appl. Earth Observation Geoinformation* 136, 104341. doi: 10.1016/j.jag.2024.104341

Zhang, Y., Li, J., and Liu, D. (2024). Spatial downscaling of ERA5 reanalysis air temperature data based on stacking ensemble learning. *Sustainability* 16, 1934. doi: 10.3390/su16051934

Zhang, L., Yan, Z., Huang, K., and Zhang, W. (2025). Evaluation and statistical bias correction of ERA5-Land meteorological variables for a humid river basin in southwest China. *Sci. Rep.* 15, 41101. doi: 10.1038/s41598-025-24942-4

Zhu, J., Zhu, S., Yu, F., Zhang, G., and Xu, Y. (2021). A downscaling method for ERA5 reanalysis land surface temperature over urban and Mountain areas. *Zggx* 25, 1778–1791. doi: 10.11834/jrs.20211257