Reports on Geodesy and Geoinformatics, 2025, Vol. 119, pp. 85–98

DOI: 10.2478/rgg-2025-0009
Received: 11 August 2024 / Accepted: 19 March 2025
Published online: 25 April 2025

ORIGINAL ARTICLE

Improving TerraClimate hydroclimatic data accuracy with XGBoost for regions with sparse gauge networks: A case study of the Meknes plateau and the Middle Atlas Causse, Morocco

Yassine Hammoud [1*], Youssef Allali [1] and Abderrahim Saadane [1]

[1]Centre of Water, Natural Resources, Environment, and Sustainable Development (CERNE2D), Faculty of Sciences, Mohammed V University, 4 Avenue Ibn Batouta BP 1014 RP, Rabat, Morocco

[*]yassine.hammoud@um5r.ac.ma

Abstract

Access to reliable hydroclimatic data, including precipitation, temperature, evapotranspiration, and runoff is crucial for effective water resource management, especially in water-stressed regions like Morocco. However, the scarcity of meteorological stations makes data collection difficult. Satellite products offer a promising alternative to these stations for monitoring and forecasting hydroclimatic trends. This study focuses on the Meknes Plateau and the Middle Atlas Causse to assess the reliability of TerraClimate data and explore their optimization using the XGBoost Machine Learning algorithm. Comparative evaluation between measured data and raw TerraClimate data reveals a satisfactory correlation, though data accuracy imperfections persist. Applying the XGBoost algorithm significantly improves the raw TerraClimate data, reducing the average Mean Absolute Error (MAE) across all parameters from 3.08 to 0.29, and the average Root Mean Square Error (RMSE) from 4.84 to 0.46, and increasing the average Nash-Sutcliffe Efficiency (NSE) from 0.82 to 0.99. These improvements validate this approach in enhancing hydroclimatic data quality in the studied region. In conclusion, this study highlights the potential of satellite products, especially TerraClimate, combined with optimization techniques, for example, the XGBoost algorithm, to address hydroclimatic data shortages in water-stressed regions. The results constitute a robust foundation for future initiatives aimed at improving water resource management and resilience to water challenges in Morocco.

Key words: hydroclimatic data, Middle Atlas Causse, Meknes plateau, TerraClimate, XGBoost, water resource management

1 Introduction

Since the late 20th century, Morocco has been experiencing a water crisis, affected by climate change, which has manifested itself as constant temperature increases and decreased precipitation (Cherif et al., 2023), which has placed Morocco under water stress (World Bank Group, 2022a), thus threatening both food and economic security (World Bank Group, 2022b). Despite substantial investments in hydraulic infrastructure, many regions still suffer direct consequences of water deficits, with droughts becoming more frequent and severe (Cherif et al., 2023). In this concerning context, it has become imperative for Morocco to adopt effective adaptation measures to ensure a sustainable water management.

Hydroclimatic parameters such as precipitation, temperature, evapotranspiration, and runoff play a crucial role in water resource planning and management (Chowdhury et al., 2016). However, despite the existence of a long-established meteorological network, Morocco faces challenges in terms of spatial data coverage, especially in mountainous areas like the Middle Atlas and the Meknes Plateau (Hanchane et al., 2023). This data gap hampers water-related risk monitoring and complicates climate trend analysis, thus necessitating the adoption of alternative solutions. It is in this context that the use of satellite data emerges as an intriguing solution.

Recent advances in remote sensing for hydroclimatic data in Morocco

Recent studies have demonstrated the potential of satellite-based hydroclimatic data to address data scarcity in regions with sparse meteorological networks. For instance, Hanchane et al. (2023) evaluated the performance of TerraClimate rainfall data in the Fes-Meknes region of Morocco, highlighting its ability to capture spatial and temporal variability despite some biases. Similarly, Rachdane et al. (2023) assessed the accuracy of GPM-IMERG precipitation data in a semi-arid basin in Morocco, finding that bias correction significantly improved the data's reliability. These studies underscore the importance of preprocessing satellite data to enhance their accuracy before application in hydrological modeling.

Preprocessing methods for satellite data

Preprocessing satellite data is a critical step in ensuring their reliability for hydroclimatic studies. Common preprocessing steps include: bias correction, gap filling, and spatial interpolation. For example, Karmouda et al. (2022) applied geo-statistical methods to correct biases in Tropical Rainfall Measuring Mission (TRMM) precipitation data over the Ouergha basin in northern Morocco, achieving significant improvements in data accuracy. Similarly, Ouatiki et al. (2021) used machine learning techniques to preprocess PERSIANN-CDR data, demonstrating their effectiveness in reducing errors and improving the correlation with ground-based observations. These preprocessing methods are essential for ensuring that satellite data can be used as a reliable alternative to ground-based measurements, especially in data-scarce regions.

The role of machine learning in enhancing satellite data

Machine learning algorithms, particularly XGBoost, have gained popularity in recent years owing to their ability to process complex datasets and improve the accuracy of satellite-derived hydroclimatic data. For instance, Niazkar et al. (2024) reviewed the applications of XGBoost in water resources engineering, demonstrating its effectiveness in optimizing hydrological models and enhancing the accuracy of satellite data. In Morocco, the use of machine learning for preprocessing satellite data remains underexplored, presenting a significant opportunity for advancing hydroclimatic research in the region.

Objectives of the study

This study aims to address the gap in the literature by: (1) evaluating the accuracy of raw TerraClimate data compared to ground-based observations in the Meknes Plateau and Middle Atlas Causse, (2) improving the quality of TerraClimate data through the application of the XGBoost machine learning algorithm, and (3) analyzing hydroclimatic trends in the studied region. We aim to contribute to the growing body of research of the use of satellite data and machine learning for water resource management in data-scarce regions.

Table 1 summarizes all the literature cited.

2 Methodology and materials

2.1 Study area

The study area encompasses the Middle Atlas Causse and the Meknes Plateau, two neighboring regions located between the Southern Rif rift and the Middle Atlas range (Figure 1) – areas characterized by distinct geological and hydrological features that affect their hydroclimatic behavior. The Middle Atlas Causse is a karstic limestone plateau with high permeability, while the Meknes Plateau consists of lacustrine formations with lower permeability (Bentayeb and Leclerc, 1977; Daguin, 1927; Amraoui, 2005).

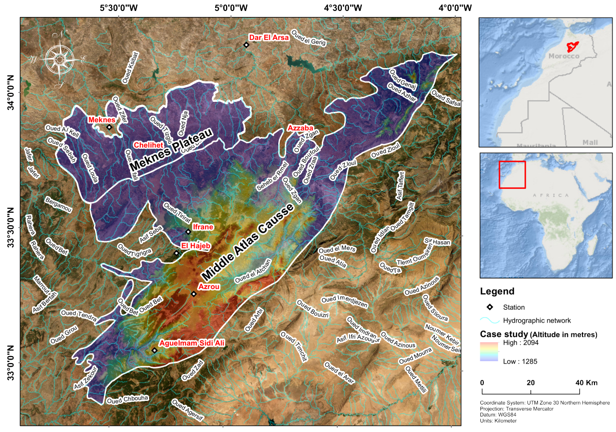


Figure 1. Geographical location of the study area

The region experiences a sub-humid mountainous climate with Mediterranean influences, as classified by Köppen (Koppen, 1936). Average annual precipitation ranges from 300 to 600 mm, with higher amounts in the mountainous areas of the Middle Atlas Causse, where snowfall is frequent in winter, and average annual temperatures vary between 12 and 16°C, with hot, dry summers and mild, humid winters.

The studied area is part of the Sebou watershed (Sebou Hydraulic Basin Agency, 2024), which is the largest one in Morocco in terms of area and flow. The Middle Atlas Causse is drained by several wadis that feed either the Sebou River to the north or the Oum Er-Rbia Basin to the south. The Meknes Plateau is primarily drained by tributaries of the Beht wadi, which flows into the El Kansera reservoir. These hydrological features, combined with the region's karstic and lacustrine aquifers, contribute to its significant water potential, which is used for irrigation, generating hydroelectricity, and as a drinking water supply (Faraj et al., 1966; Eddif et al., 2018; Amraoui, 2005).

2.2 Data collection

The data used consist of hydroclimatic variables spanning a sufficiently long period to capture hydroclimatic variations and trends accurately. The precipitation, temperature, evapotranspiration, and runoff data were collected and computed for the period between 1971 and 2001 (Tabyaoui, 2005). The choice of this period is based on its quality and the need to reserve the remaining data for subsequent analyses, particularly, for forecast applications within the scope of this research.

These data were sourced at eight meteorological stations in the study region: Aguelmam Sidi Ali, Azrou, Azzaba, Chelihet, Dar El Arsa, El Hajeb, Ifrane, and Meknes, which were chosen because they offer all the measurements for the entire study period. Table 2 summarizes the eight stations used in this study.

The precipitation and temperature data were retrieved from two providers: the Sebou Hydraulic Basin Agency (ABHS) and the Provincial Directorate of Agriculture (DPA) of Meknes. The evapotranspiration and runoff were deduced from the precipitation and temperature values in addition to other coefficients such as the crop coefficient (Kc).

Table 1. Summary of recent studies on satellite data evaluation and preprocessing methods in Morocco and similar regions

Study	Satellite Product	Region	Preprocessing Methods	Key Findings
Hanchane et al. (2023)	TerraClimate	Fes-Meknes, Morocco	Bias correction, spatial interpolation	TerraClimate data showed good spatial and temporal variability after correction.
Rachdane et al. (2023)	GPM-IMERG	Semi-arid basin, Morocco	Bias correction, machine learning	GPM-IMERG data improved significantly after bias correction.
Karmouda et al. (2022)	TRMM	Ouergha basin, Morocco	Geo-statistical methods	TRMM data accuracy improved with geo-statistical preprocessing.
Ouatiki et al. (2021)	PERSIANN-CDR	Semi-arid watershed, Morocco	Machine learning (SVM, Random Forest)	PERSIANN-CDR data showed high correlation with ground observations after ML.
Niazkar et al. (2024)	Various	Global (review)	XGBoost, other ML algorithms	XGBoost is highly effective for optimizing satellite data in hydrological models.

Table 2. The eight stations used in this study

Station	Longitude (WGS84)	Latitude (WGS84)	Altitude (m)	Basin
Aguelmam Sidi Ali	5° 8' 24" W	33° 4' 48.72" N	1800	Sebou
Azrou	4° 38' 44.52" W	33° 16' 12" N	1250	Sebou
Azzaba	5° 19' 0.12" W	33° 49' 56.28" N	480	Sebou
Chelihet	4° 55' 42.6" W	33° 48' 40.32" N	720	Sebou
Dar El Arsa	5° 13' 12" W	34° 11' 44.16" N	140	Sebou
El Hajeb	5° 7' 56.64" W	33° 25' 12" N	1050	Sebou
Ifrane	5° 31' 48" W	33° 30' 22.68" N	1650	Sebou
Meknes	5° 52' 48" W	33° 52' 48" N	575	Sebou

Meanwhile, the satellite data pertaining to the same parametres and for the same study period and region were extracted from the TerraClimate dataset (Abatzoglou et al., 2018), which is uses monthly temporal resolution and spatial resolution of approximately 4 km (1/24th degree) and covers the period 1958–2020. Table 3 summarizes the characteristics of the TerraClimate dataset.

2.3 Data preparation

During the data preparation phase, assessing the quality of the precipitation and temperature data is crucial, especially for calculating evapotranspiration and runoff. The objective of this phase was to obtain a comprehensive assessment of the hydroclimatic parameters compared to TerraClimate satellite-acquired data in the region of the Meknes Plateau and the Middle Atlas Causse.

To address the challenge of spatial matching between gridded satellite data and point-based station data, we used a spatial interpolation technique to align the TerraClimate gridded data with the locations of the meteorological stations. Specifically, we employed the Inverse Distance Weighting (IDW) method, which is widely used (Shepard, 1968). This ensured that the satellite data accurately represented the conditions at the station locations. The interpolation method was validated against station data to ensure accuracy, and the results showed a high degree of correlation ($R^2 > 0.9$). This approach is consistent with best practices in remote sensing and hydroclimatic data analysis, where spatial interpolation is often used to bridge the gap between gridded and point data (Li et al., 2015).

Numerous satellite-based hydroclimatic data sources are available, such as, among others: GPM, TRMM, PERSIANN, ERA5, and MODIS. This study relies on TerraClimate data due to their high quality and careful preprocessing, ensuring their reliability. Additionally, the fine spatial resolution of these data (4 km × 4 km) and their monthly temporal resolution make them particularly suitable for hydroclimatic studies. These data provide a valuable source for the four essential parameters under analysis: precipitation, temperature, evapotranspiration (both actual and potential), and runoff. The use of TerraClimate satellite data allows for a robust scientific approach, providing a reliable basis for an in-depth analysis of hydroclimatic interactions, while contributing to a deeper understanding of environmental dynamics at the studied scale. Table 4 summarizes the characteristics of each parameter.

2.4 Data repair

Filling data gaps

In hydroclimatic studies, it is common for datasets to have gaps that affect the quality of the results (Aieb et al., 2019). Since any statistical study must rely on a reliable and controlled dataset, checking the selected series is a preliminary step to any analysis (Castellani, 1986).

In this study gaps have been identified in the time series of precipitation and temperature, which may be due to technical errors or data corruption (Beaulieu et al., 2007).

Lack of data must be compensated for using one of the following specific methods: (1) duplicating the rainfall value from a neighboring station to fill the gap, (2) analyzing the annual trend of observed precipitation at the regional scale

Table 3. Characteristics of the TerraClimate dataset

Feature	Description
Type	Dataset of monthly climate and climatic water balance
Temporal Coverage	1958–2024
Spatial Coverage	Global terrestrial surfaces
Spatial Resolution	~4 km (1/24th degree grid)
Source Data	Combined from WorldClim (spatial climatology) and CRU TS4.0 (time-varying data)
Interpolation Method	Climatically aided interpolation
Variables	Precipitation, maximum and minimum temperature, wind speed, vapor pressure, solar radiation, reference evapotranspiration, vapor pressure deficit, PDSI (Palmer Drought Severity Index), Runoff
Indices	Standardized Precipitation- Evapotranspiration Index (SPEI), Palmer Drought Severity Index (PDSI), and more (depending on user-defined parameters)
Format	NetCDF files
Access	Freely available for download
Applications	Ecological and hydrological studies, climate change analysis, crop modeling, resource management, drought monitoring
Resources	https://www.climatologylab.org/terraclimate.html https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE

to estimate the missing precipitation (Abdullah and Al-Ansari, 2022), (3) applying the IDWA (Inverse Distance Weighted Averaging) method (based on linear regression) to calculate missing precipitation (Brandsma and Buishand, 1998), or (4) replacing missing rainfall with the average rainfall (Li et al., 2015).

Method (4) was chosen for this study for the following reasons:

- *Simplicity and reliability*: The average rainfall method is easy to implement, as it provides a reliable estimate when data from neighboring stations are not available, or when regional trends are inconsistent.
- *Consistency*: Using the average rainfall ensures that the filled data are consistent with the overall climatological patterns of the study area, minimizing the risk of introducing biases.
- *Applicability*: This method is particularly suitable for the studied region, where precipitation patterns are relatively stable over time, and the interannual variability is well captured by the average values.

The gaps were filled using the following formula:

$$M = \frac{1}{n} \sum Xi, \tag{1}$$

where: Xi – monthly values, n – number of months, M – interannual monthly averages.

Gaps were primarily found in the precipitation and temperature time series, particularly for the months of: January, February, and December, where data were missing or incomplete due to technical issues at some meteorological stations. These gaps were filled using the average monthly values calculated from the available data for the same months across the study period (1971–2001).

Homogenization testing

Homogenization is a critical step in ensuring the consistency and reliability of hydroclimatic data, as it detects and corrects non-natural inhomogeneities resulting from modifications in observation networks. It involves comparing data from one station with those from neighboring stations to ensure that climate changes are not interpreted as inhomogeneities (Beaulieu et al., 2007).

Various homogenization techniques have been developed to take into account several influencing factors, including the specific variable being homogenized, the spatial and temporal variability of the data (which depends on the location of stations), the length of the time series, the amount of missing data, the availability of metadata, and the density of the observation network (Aguilar et al., 2003; Beaulieu et al., 2007).

In this study, we conducted homogenization. The double mass analysis (Kohler, 1949) was used to evaluate the homogeneity of data series. This method involves simple linear regression on cumulative precipitation values from one station compared to those from another station, followed by plotting the regression line. The homogeneity of the series from the studied station is determined by the random dispersion of values around this line. However, a change in slope in the regression line indicates a change in one or both of the stations, suggesting a possible inhomogeneity. To assess temperature homogeneity, the r_{xy} correlation estimator between two random variables was applied, allowing for the detection of any inconsistencies in temperature series (Aguilar et al., 2005):

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}, \tag{2}$$

with:

s_x^2 – estimator of the variance of the random variable x,
s_y^2 – estimator of the variance of the random variable y,
s_{xy} – estimator of the covariance of random variables x and y:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right). \tag{3}$$

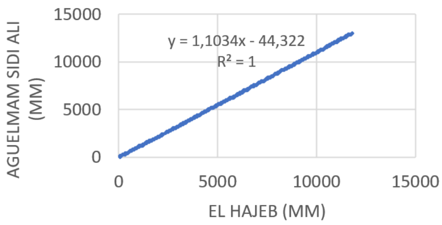
The results indicate that most of the precipitation and temperature series are homogeneous (see Figure 2 and 3). The double mass analysis performed on precipitation data revealed a stable linear relationship between the analyzed stations, suggesting no significant breakpoints or artificial shifts in the data. Similarly, the interstation correlation for temperature yielded coefficients above 0.9, confirming the consistency of the thermal records. The impact of homogenization is evident in the improved predictive performance of the XGBoost algorithm.

Analysis of spatial variability

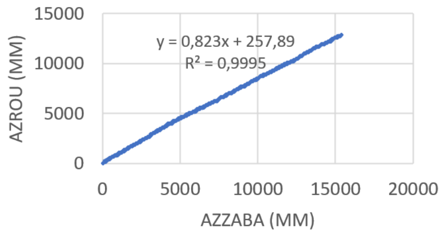
To account for the spatial variability of the data, a detailed analysis of their spatial distribution was conducted. For each month and each parameter, the differences in values between individual weather stations and the TerraClimate data were examined using spatial analysis techniques (Kessabi et al., 2023), such as: anomaly mapping and geographic variability analysis (Amiri and Mesgari, 2017). This allowed for the assessment of the consistency of TerraClimate data with ground measurements on a spatial basis and determined to what extent the monthly averages accurately represented the spatial variability of the weather conditions in the studied area.

Table 4. Characteristics of hydroclimatic factors provided by TerraClimate

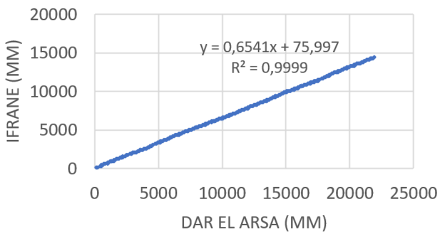
Variable	Spatial resolution	Temporal resolution	Source
Precipitation	1/24° (~4 km)	Monthly	CRU TS 3.22, interpolated
Temperature	1/24° (~4 km)	Monthly	CRU TS 3.22, interpolated
Potential Evapotranspiration	1/24° (~4 km)	Monthly	Penman-Monteith, using CRU TS 3.22 as inputs
Actual Evapotranspiration	1/24° (~4 km)	Monthly	VIC, forced with TerraClimate precipitation and temperature data
Runoff	1/24° (~4 km)	Monthly	VIC, forced with TerraClimate precipitation and temperature data



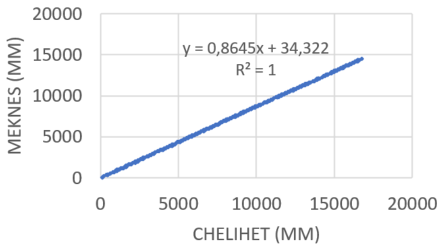
(a) Double accumulation between Aguelmame and El Hajeb



(b) Double accumulation between Azrou and Azzaba

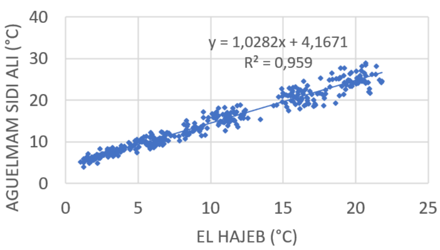


(c) Double accumulation between Ifrane and Dar El Arsa

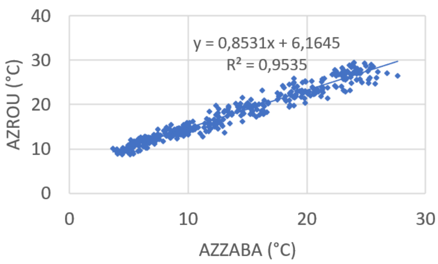


(d) Double accumulation between Chelihet and Meknes

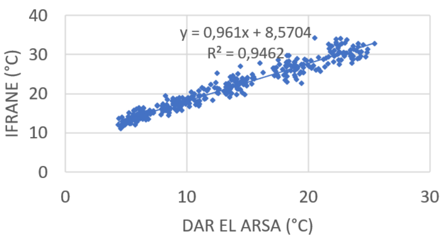
Figure 2. Precipitation homogenization test



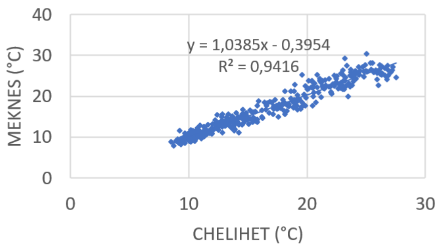
(a) Correlation between Aguelmame and El Hajeb



(b) Correlation between Azrou and Azzaba



(c) Correlation between Ifrane and Dar El Arsa



(d) Correlation between Chelihet and Meknes

Figure 3. Temperature homogenization test

2.5 Estimation of evapotranspiration and runoff

Data on evapotranspiration and runoff are not included in the initial dataset due to difficulties in accessing measured data. However, it is possible to estimate these values from known variables using equations such as Thornthwaite (1948) for potential evapotranspiration, Allen et al. (1998) for actual evapotranspiration, and the modified Tixeron-Berkaloff equation by Romantchouk for runoff.

Estimation of evapotranspiration

Evapotranspiration, the sum of evaporation and transpiration processes, is a fundamental variable in understanding land surface management (Vörösmarty et al., 1998). Evaporation is defined as the direct transfer of water to the atmosphere from various sources, including soil and bodies of water. In parallel, transpiration represents the movement of water from the root system through a plant, followed by its release into the air as water vapor. Two main components characterize evapotranspiration: potential evapotranspiration (pET), which represents the amount of water that would be evaporated and transpired by a crop, soil, or ecosystem given adequate water availability; and actual evapotranspiration (aET), which corresponds to the amount of water actually evaporated and transpired under real conditions (Allen et al., 1998).

Due to the complexity and time required to directly measure atmospheric vapor flux, evapotranspiration is generally estimated using various alternative methods that do not require direct measurements (Alexandris, 2013).

In this study, potential evapotranspiration was assessed using the Thornthwaite's (1948) equation. This approach allows for the calculation of monthly and annual water balance at a station level, based on monthly precipitation values and average monthly temperatures. It is noteworthy that the calculation of monthly potential evapotranspiration is conditioned by an average monthly temperature not exceeding 38°C, a favorable condition in the study area (Bonnet, 1970):

$$pET = 16\left(\frac{10t}{I}\right)^a \cdot K, \tag{4}$$

where:

pET : Corrected monthly potential evapotranspiration [mm],
t : Average monthly temperature [°C],
I : Annual thermal index: sum of monthly indices calculated from mean monthly temperatures according to the formula:
$i = \left(\frac{t}{5}\right)^{1.514}$ where t is the mean monthly temperature in °C; the annual thermal index I is: $I = \sum_{i=1}^{12} i$,
a : Coefficient calculated by the expression:
$a = 6.75 \cdot 10^{-7} \cdot I^3 - 7.71 \cdot 10^{-5} \cdot I^2 + 1.79 \cdot 10^{-2} \cdot I + 0.49$,
K : Latitude- and time-dependent correction coefficient presented in the relevant table. The annual potential evapotranspiration value is the sum of the monthly values.

Estimating actual evapotranspiration presents challenges, particularly due to its dependence on the water stock at the given moment., Allen's FAO-56 model, simple and operational, since it relates the plant's phenological data to the measured meteorological data (Mjejra et al., 2014; Er-Raki et al., 2007) is commonly used to overcome this difficulty. It is based on potential evapotranspiration along with the use of a crop coefficient (Kc), aiming to adjust the potential evapotranspiration to actual conditions.

The crop coefficient describes particular types of crops at particular stages of development. The coefficient is frequently used to estimate evapotranspiration (ET): as the crop reaches later vegetative development stages, the crop coefficient tends to approach the value 1, and consequently, actual evapotranspiration (aET) is equivalent to potential evapotranspiration (Herrera-Puebla et al., 2020):

$$aET = pET \cdot Kc. \tag{5}$$

Kc is one of the most challenging parameters to measure in the field (Runtunuwu, 2007), hence the need for a reliable approach to its estimation. Several studies have used multispectral vegetation indices to estimate Kc values. In a study based on 957 stations across various regions, Runtunuwu (2007) found a close correlation between Kc and the Normalized Difference Vegetation Index (NDVI) and represented it by the following equation:

$$Kc = 0.08 + 1.83 \cdot \text{NDVI} \tag{6}$$

Subsequently, this equation was integrated into the calculation of aET.

The NDVI calculation was based on Sentinel-2A multispectral images according to the equation:

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R}, \tag{7}$$

where: NIR means Near-Infrared, which is the reflectance in the near-infrared region of the electromagnetic spectrum; R stands for Red, which is the reflectance in the red region of the electromagnetic spectrum.

Estimation of runoff

Runoff is the flow of water on the soil surface and is a significant component of the hydrological cycle. It refers to the water layer that could not infiltrate or evaporate and is expressed by the ratio of the runoff volume (m^3) to the basin area (m^2). It is calculated using the Tixeron-Berkaloff formula as follows:

$$Ra = \frac{P^3}{3 \cdot pET^2} \quad \text{if} \quad P < 600 \text{ mm}, \tag{8}$$

$$Ra = \frac{P^2}{3} \quad \text{if} \quad P \geq 600 \text{ mm}, \tag{9}$$

where:

Ra : Annual runoff [mm]
P : Annual precipitation [mm].

2.6 Evaluation and boosting of data

Following the adjustment of data a regression analysis was implemented to compare the measured values with the TerraClimate captured values to assess their reliability as alternatives to the measured data.

For each hydroclimatic parameter studied, a quantitative comparison of the time series was conducted using five statistical metrics: Pearson Correlation Coefficient (PCC), Nash-Sutcliffe Efficiency Coefficient (NSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Bias (BIAS) (Rachidi et al., 2023; Ebert, 2007; Laurent et al., 1998; Lettenmaier and Wood, 1993; Rachdane et al., 2022; Beck et al., 2017; Ghozat et al., 2021; Saeedi et al., 2022). The respective equations and optimal values for these metrics are summarized in Table 5.

To optimize the performance of TerraClimate data related to the studied variables and to improve their suitability as alternatives to measured data, various regression models are applied. These models include Support Vector Machine (SVM) (Shrestha and Shukla, 2015; Adnan et al., 2024; Ezzaouini et al., 2022),

Table 5. Evaluation metrics

Indicator	Description	Equation	Optimal values
Pearson Correlation Coefficient (PCC)	Statistical measure that evaluates the strength and direction of the linear relationship between two data sets.	$\mathrm{PCC} = \frac{\sum_{i=1}^{n}(G_i-\bar{G})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(G_i-\bar{G})^2}\sqrt{\sum_{i=1}^{n}(S_i-\bar{S})^2}}$	1
Bias	Measure of the systematic tendency of the prediction to deviate from reality. A positive bias indicates a tendency to overestimate, while a negative bias indicates a tendency to underestimate.	$\mathrm{Bias} = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} G_i}$	0
Root Mean Square Error (RMSE)	Measure of the average deviation between predicted and observed values. RMSE is sensitive to outliers and gives more weight to significant errors.	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)^2}$	0
Nash-Sutcliffe Efficiency Coefficients (NSE)	Indicator of model efficiency. A value of 1 indicates perfect performance, 0 indicates performance equivalent to a constant prediction, and negative values indicate performance inferior to constant prediction.	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(S_i-G_i)^2}{\sum_{i=1}^{n}(G_i-\bar{G})^2}$	1
Mean Absolute Error (MAE)	Measure of the average absolute deviation between predicted and observed values. It is less sensitive to outliers than RMSE, as it does not consider the magnitude of errors.	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert S_i - G_i\rvert$	0

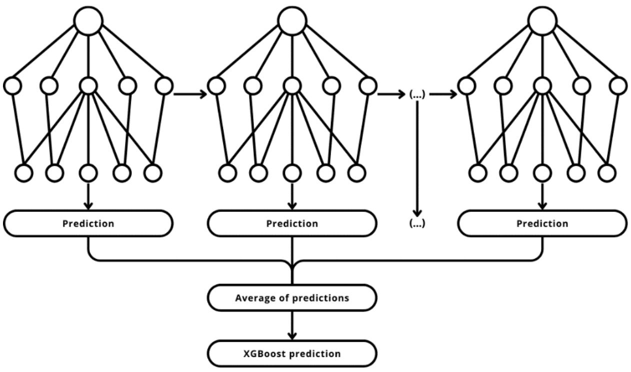


Figure 4. XGBoost process diagram

Random Forest (Xu et al., 2019; Chen et al., 2020), Artificial Neural Networks (ANN) (Tan et al., 2017; Kinouchi et al., 2021), Convolutional Neural Networks (CNN) (Wang and Xuan, 2022), Gradient Boosting Decision Trees (GBDT) (Kofidou et al., 2023; Han et al., 2018), and Extreme Gradient Boosting (XGBoost) (Niazkar et al., 2024; Ali et al., 2022), among others.

In this study, the XGBoost model (Chen and Guestrin, 2016) was used to optimize performance, ensuring fast and accurate learning, and confirming it as the machine learning model of choice in the field of data science, particularly for hydroclimatic simulations (Hao and Bai, 2023).

XGBoost is a set of machine learning algorithms based on the structure of GBDT (Figure 4). XGBoost is used in regression and classification cases (Chen and Guestrin, 2016). It has been widely applied in many fields due to its high prediction accuracy and low computational costs (Fan et al., 2018; Yu et al., 2020). XGBoost is defined by the following equation:

$$y = \sum_{k=1}^{K} f_k(x)\,, \tag{10}$$

where f_k is tree structure with K leaves and leaf weights.

Data splitting for training, validation and testing

The dataset was divided into training (70%) and testing (30%) sets to ensure the model's generalizability. This split ratio is consistent with common practices in machine learning, where a larger portion of the data is used for training to capture the underlying patterns, while a smaller portion is reserved for testing to evaluate the model's performance on unseen data (Hastie et al., 2009). The training set was used to train the XGBoost model, while the testing set was used to evaluate its performance on unseen data. This approach ensures that the model's effectiveness is not limited to the training period but can also generalize to new data. Additionally, we performed cross-validation during the training phase to further validate the model's robustness and avoid overfitting (Kohavi, 1995).

Model structure and hyperparameters

The XGBoost model was implemented with default hyperparameters, including a learning rate of 0.1, a maximum depth of 6, and 100 estimators. These parameters were chosen based on their proven effectiveness in similar hydroclimatic studies (Chen and Guestrin, 2016). The learning rate controls the step size at each iteration, while the maximum depth limits the complexity of the trees to prevent overfitting. The number of estimators determines the number of boosting rounds, which was set to 100 to balance model performance and computational efficiency. The input data structure consisted of monthly hydroclimatic variables (precipitation, temperature, evapotranspiration, and runoff), and the output was the optimized values of these variables. This structure aligns with best practices in machine learning for hydrological applications (Niazkar et al., 2024).

Next, a trend analysis of the parameters over time was undertaken to evaluate, by observation, the divergences and correlations between the measured data and the raw TerraClimate data initially, and between the measured data and those enhanced by the XGBoost model subsequently.

3 Results and discussion

3.1 Comparison of measured data and raw Terraclimate data

The evaluation of raw TerraClimate satellite data involved their comparison with weather station observations using the statistical metrics defined in Table 5.

The analysis revealed significant spatial variability in the performance of TerraClimate data, driven by the geophysical characteristics of the study area. The following sections provide a detailed exploration of these patterns and their underlying causes.

Precipitation

The analysis of precipitation (Figure 5) revealed a satisfactory correlation between TerraClimate data and weather station observations, with coefficients of determination (R^2) ranging from 0.26 to 0.94 (average of 0.81). The lowest R^2 values were observed in Azrou (0.26) and Ifrane (0.74), while the highest R^2 values – in Azzaba, Dar El Arsa, and Meknes (0.94, 0.94, and 0.86 respectively).

The poor performance in Azrou and Ifrane can be attributed to the complex topography of the Middle Atlas Mountains, where orographic effects and localized microclimates create significant spatial variability in rainfall patterns. TerraClimate's coarse resolution (~4 km) struggled to capture these fine-scale variations, leading to underestimation and low correlation. In contrast, the high R^2 values in Azzaba, Dar El Arsa, and Meknes reflect the relatively homogeneous climatic conditions in these lowland regions, where satellite data performed more reliably.

The Pearson correlation confirms a strong linear relationship between TerraClimate data and weather station observations for most stations, with values ranging from 0.51 to 0.97 (average of 0.89). The stations Azrou and Ifrane featured the lowest correlations (0.51 and 0.86), while Azzaba, Dar El Arsa, and Meknes – the highest correlations (0.97, 0.97, and 0.93 respectively).

The bias reveals a tendency to underestimate precipitation in most stations, except for Chelihet, Dar El Arsa, and Meknes, which showed a slight tendency to overestimate. The average bias is -3.74 mm/month, with the largest biases in Azrou and Ifrane (-19.90 mm/month and -17.2 8mm/month, respectively) and the smallest bias in Meknes (-1.9 6mm/month).

The RMSE and MAE revealed moderate prediction errors overall, with averages of 19.57 mm/month and 13.82 mm/month respectively. Azrou and Ifrane showed the largest prediction errors (RMSE of 28.75 mm/month and 27.11 mm/month, and MAE of 22.76 mm/month and 20.7 2mm/month, respectively), while Aguelmame Sidi Ali and Meknes showed the smallest prediction errors (RMSE of 11.37 mm/month and 18.4 3mm/month, and MAE of 9.17 mm/month and 12.93 mm/month, respectively).

The NSE denotes acceptable model performance for most stations, with an average of 0.56. Azrou and Ifrane showed the least favorable performances (NSE of -0.54 and 0.47, respectively), while the data from the Aguelmame Sidi Ali and Meknes stations demonstrated the best performances (NSE of 0.48 and 0.70, respectively).

Temperature

For temperature (Figure 6), the R^2 was significantly high, with an average of 0.99 for all stations. The Pearson correlation remained significantly high for all stations, with an average of 0.99.

The bias was generally low, with an average of -0.02 °C. The stations Azrou, Dar El Arsa, and Ifrane showed the smallest biases (0.02 °C, -0.02 °C, and -0.09 °C, respectively), while Chelihet featured the largest bias (0.01 °C).

The RMSE and MAE indicated very low prediction errors for all stations, with averages of 0.64 °C and 0.27 °C, respectively. Ifrane showed the smallest prediction errors (RMSE of 0.41 °C and MAE of 0.16 °C), while Chelihet showed the largest prediction errors (RMSE of 0.59 °C and MAE of 0.20 °C).

The NSE was remarkably high for all stations, with an average of 0.99, suggesting excellent model performance.

Evapotranspiration (aET and pET)

For actual evapotranspiration (Figure 7), the R^2 values ranged from 0.69 to 0.95, with the highest R^2 (0.95) in Ifrane, and the lowest (0.69) in Chelihet.

For potential evapotranspiration (Figure 8), the R^2 values indicated excellent data convergence, ranging from 0.88 to 0.98. Ifrane and Azrou showing the highest R^2 values (0.98), while Aguelmame Sidi Ali – the lowest R^2 (0.88).

The Pearson correlation remained high for all stations regarding actual evapotranspiration, with Ifrane and Meknes with the highest correlations (0.99), and Azrou – the lowest correlation (0.84). Similarly, the Pearson correlation remained high for potential evapotranspiration in almost all stations, with an average of 0.99.

The bias revealed a tendency to underestimate actual evapotranspiration in all stations, with Ifrane showing the largest bias (–0.43 mm/month) and Aguelmame Sidi Ali – the smallest (–0.03 mm/month). For potential evapotranspiration, a negative bias was observed for all stations, with Ifrane and Azzaba showing the largest biases (–0.13 mm/month), while Meknes showed the smallest bias (–0.16 mm/month).

The RMSE and MAE values indicate moderate to significant prediction errors for actual evapotranspiration, with the largest prediction errors (RMSE of 4.18 mm/month and MAE of 1.37 mm/month) for Ifrane, while Meknes showed the smallest prediction errors (RMSE of 2.72 mm/month and MAE of 0.76 mm/month). Similarly, for potential evapotranspiration, moderate to significant prediction errors were observed for all stations, the largest prediction errors (RMSE of 0.41 mm/month and MAE of 0.06 mm/month) in Ifrane, while Chelihet and Meknes showed the smallest prediction errors (RMSE of 0.32 mm/month and MAE of 0.25 mm/month).

Runoff

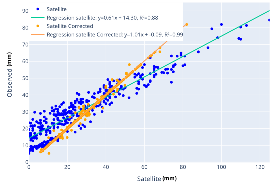
For runoff (Figure 9), the R^2 values ranged from 0.69 to 0.95, with Ifrane and El Hajeb showing the highest R^2 values (0.95) and Azrou and Chelihet – the lowest R^2 values (0.69).

The Pearson correlation remained high for all stations, confirming a strong linear relationship between measured data and raw TerraClimate satellite data, with an average of 0.99.

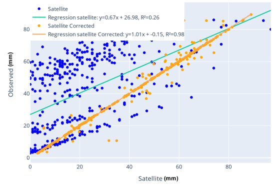
The bias revealed a tendency to underestimate runoff in most stations, except for Chelihet, which showed a positive bias of 0.10 mm/month, indicating a slight overestimation.

The RMSE and MAE values indicated moderate to significant prediction errors for all stations, with Ifrane with the largest prediction errors (RMSE of 0.11 mm/month and MAE of 0.03 mm/month), while El Hajeb and Meknes showed the smallest prediction errors (RMSE of 0.24 mm/month and MAE of 0.14 mm/month).

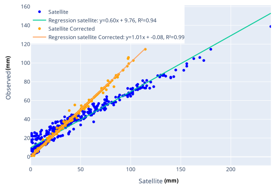
The spatial patterns in the performance metrics can be linked to the geophysical characteristics of the study area. Regions with complex topography, such as Azrou and Ifrane, tend to have lower R^2 values and higher biases due to the challenges in capturing orographic effects and microclimatic variations. In contrast, regions with more homogeneous climatic conditions, such as Azzaba and Meknes, show higher R^2 values and lower biases, indicating better performance of the TerraClimate data.



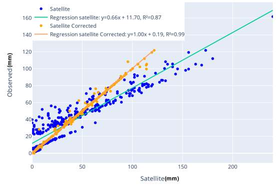
(a) Aguelmame Sidi Ali



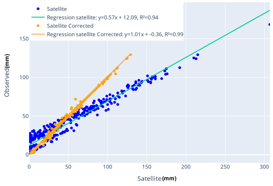
(b) Azrou



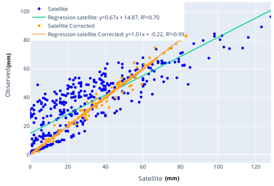
(c) Azzaba



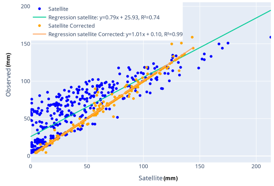
(d) Chelihet



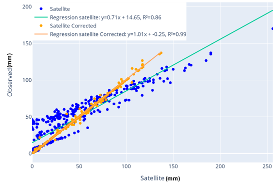
(e) Dar El Arsa



(f) El Hajeb

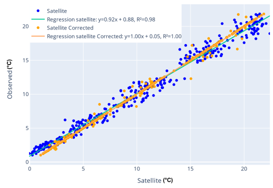


(g) Ifrane

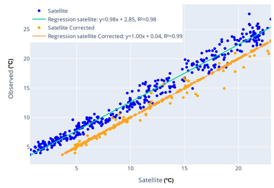


(h) Meknes

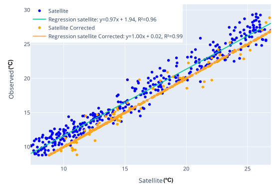
Figure 5. R^2 before and after boosting for precipitation



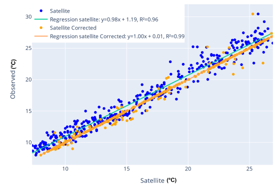
(a) Aguelmame Sidi Ali



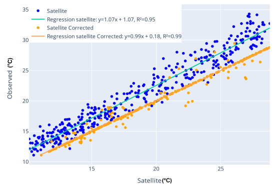
(b) Azrou



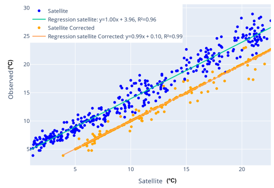
(c) Azzaba



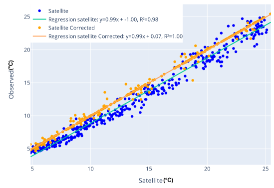
(d) Chelihet



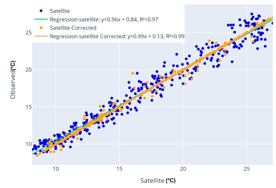
(e) Dar El Arsa



(f) El Hajeb



(g) Ifrane



(h) Meknes

Figure 6. R^2 before and after boosting for temperatures

3.2 Comparison after boosting with the XGBoost model between measured data and boosted TerraClimate data

The results obtained after applying the boosting method showed a remarkable performance of the XGBoost model in optimizing the values of the studied hydroclimatic variables. For all variables, the coefficient of determination (R^2) after boosting reached optimal values (between 0.99 and 1) for all stations, reflecting a near-optimal agreement between the measured data and the model estimates.

The Pearson correlation after boosting also reached the maximum value of 1, confirming a strong linear relationship between the observations and model estimates.

The evaluation of bias after boosting revealed the absence of systematic underestimation or overestimation for all the studied hydroclimatic variables. Aguelmame Sidi Ali, Azrou, and El Hajeb exhibited biases closest to the optimal (bias of 0), while Azzaba and Ifrane displayed values furthest from the optimal.

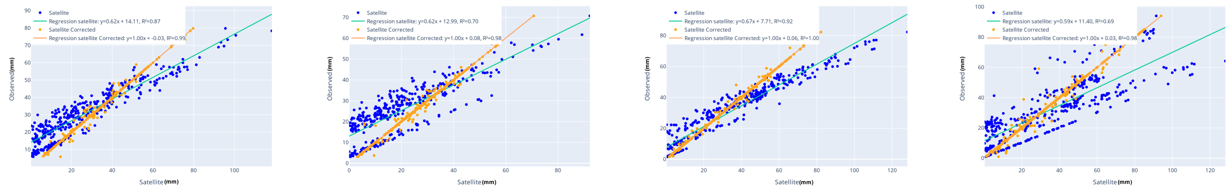
Regarding accuracy measures such as RMSE and MAE, these indicators remained very low for all stations, attesting to the high precision of the estimates. The stations Aguelmame Sidi Ali, Azzaba, and El Hajeb showed the lowest values, while the stations Chelihet and Meknes showed the highest values, although still at acceptable levels.

The NSE after boosting was near its optimal value of 1 for all hydroclimatic parameters, indicating perfect model performance in reproducing the temporal variability of these variables (Figure 10).

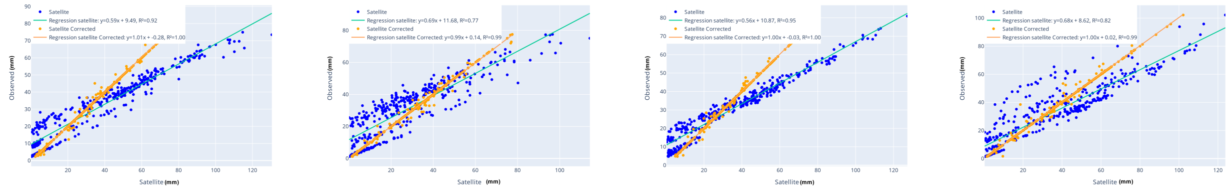
3.3 Monthly trends of hydroclimatic variables in the study area

The results of this study highlight the role of the XGBoost algorithm in optimizing TerraClimate data, preparing it for effective use in future predictions. This is supported by the overall performance metrics (averaged across the eight meteorological stations), which show significant improvements: the MAE decreased from 3.08 to 1.47, the RMSE decreased from 4.84 to 2.3, and the NSE increased from 0.82 to 0.99. These metrics confirm the effectiveness of XGBoost in enhancing hydroclimatic data quality.

The curves in Figures 11–15 illustrate the temporal evolution of the studied hydroclimatic parameters (precipitation, temperature, evapotranspiration, and runoff) between 1971–2001. These graphs compare the measured data, the raw TerraCli-

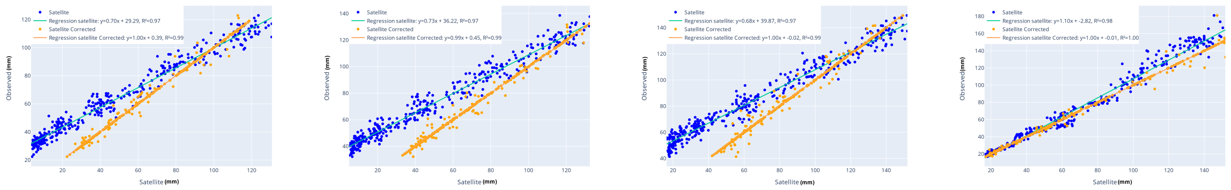


(a) Aguelmame Sidi Ali **(b)** Azrou **(c)** Azzaba **(d)** Chelihet

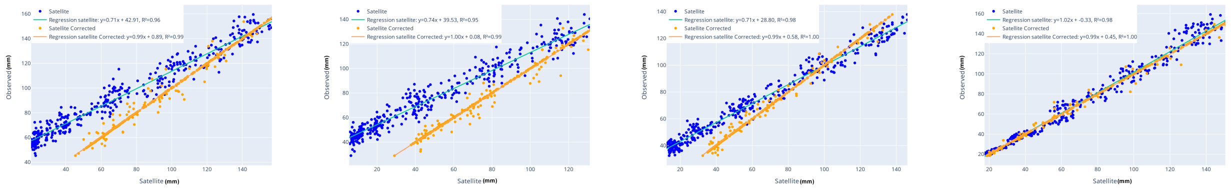


(e) Dar El Arsa **(f)** El Hajeb **(g)** Ifrane **(h)** Meknes

Figure 7. R^2 before and after boosting for aET

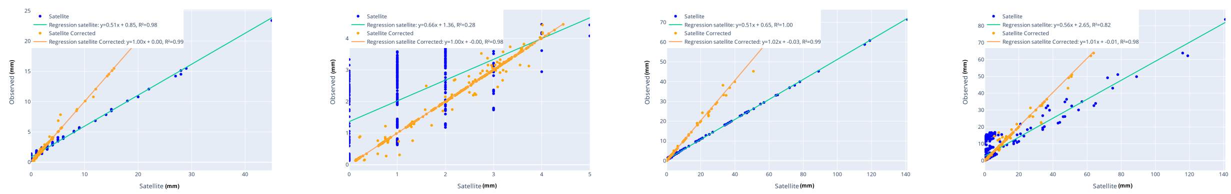


(a) Aguelmame Sidi Ali **(b)** Azrou **(c)** Azzaba **(d)** Chelihet

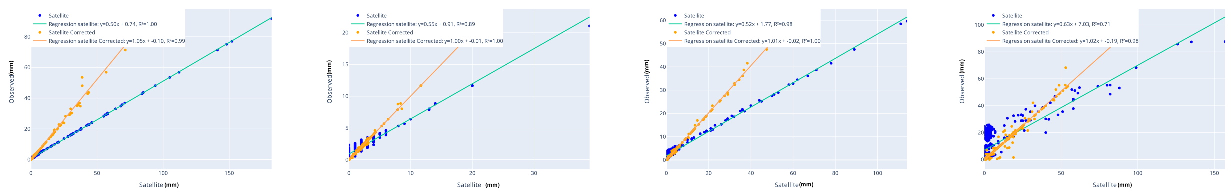


(e) Dar El Arsa **(f)** El Hajeb **(g)** Ifrane **(h)** Meknes

Figure 8. R^2 before and after boosting for *pET*

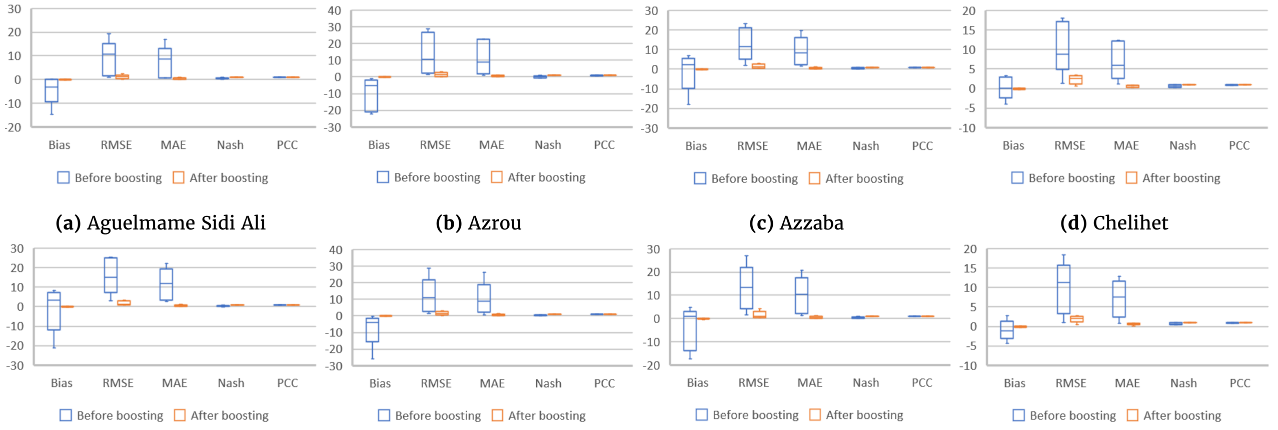


(a) Aguelmame Sidi Ali **(b)** Azrou **(c)** Azzaba **(d)** Chelihet



(e) Dar El Arsa **(f)** El Hajeb **(g)** Ifrane **(h)** Meknes

Figure 9. R^2 before and after boosting for runoff



(a) Aguelmame Sidi Ali **(b)** Azrou **(c)** Azzaba **(d)** Chelihet

(e) Dar El Arsa **(f)** El Hajeb **(g)** Ifrane **(h)** Meknes

Figure 10. Boxplots of the eight stations representing the improvement of TerraClimate data by the XGBoost algorithm

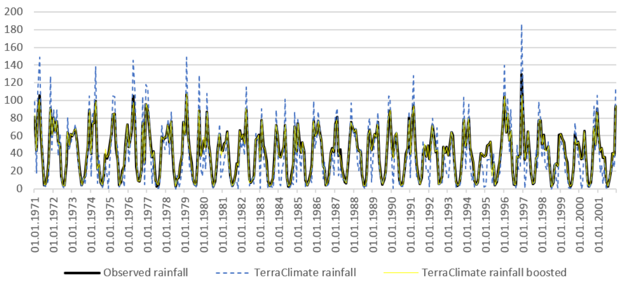


Figure 11. Trend comparison of measured precipitation data, TerraClimate data and TerraClimate boosted data

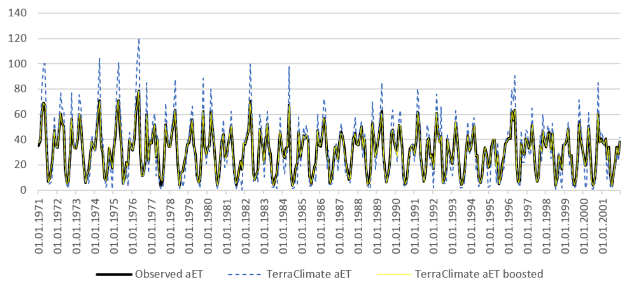


Figure 13. Trend comparison of measured aET data, TerraClimate data and TerraClimate boosted data

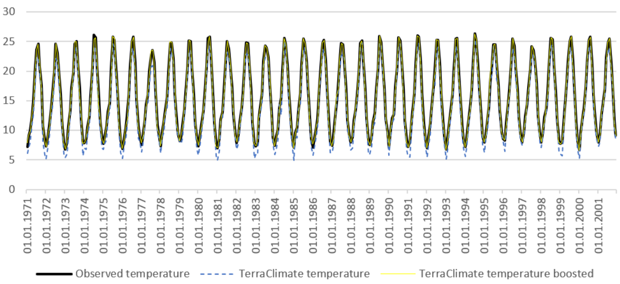


Figure 12. Trend comparison of measured temperature data, TerraClimate data and TerraClimate boosted data

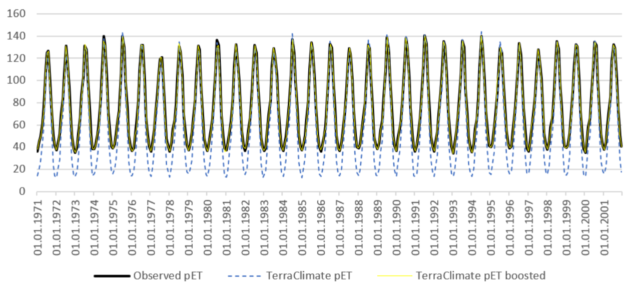


Figure 14. Trend comparison of measured pET data, TerraClimate data and TerraClimate boosted data

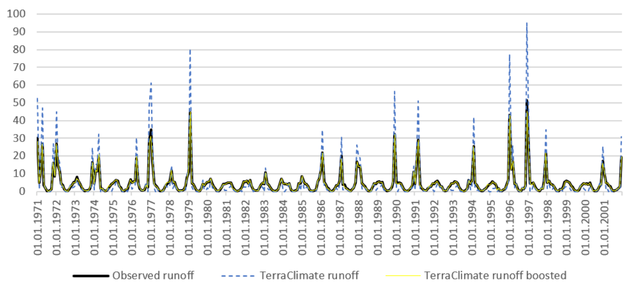


Figure 15. Trend comparison of measured runoff data, TerraClimate data and TerraClimate boosted data

mate data, and the TerraClimate data optimized by XGBoost. The time series are based on monthly averages calculated from data collected at the eight meteorological stations, providing a comprehensive overview of temporal trends across the region. The time annotations (e.g., '01.01.1971') indicate the start of each month, with values representing monthly averages for that period.

The graphs show an almost perfect overlap of the curves for the measured data and the optimized TerraClimate data. This consistency reinforces the reliability of the optimized data and paves the way for their use in various applications, including their integration into predictive models for the evolution of hydroclimatic variables, thus contributing to a more comrehensive understanding of water security.

While the results of this study are specific to the Meknes Plateau and the Middle Atlas Causse, the methodology and findings have potential transferability to other regions with different climatic and geological conditions. The XGBoost model's ability to correct spatial heterogeneity and improve the accuracy of TerraClimate data suggests that it could be applied to other mountainous or topographically complex regions where satellite data often underperforms.

Additionally, the model's success in lowland areas like Meknes indicates its applicability to flat, homogeneous regions. However, the specific performance of the model may vary depending on local climatic conditions, such as aridity, vegetation cover, and seasonal variability.

Future studies could explore the transferability of this methodology by applying it to other regions and comparing its performance across different climatic zones, which would help establish the generalizability of the approach and identify any necessary adjustments for specific environments.

3.4 Limitations

Although this study demonstrates promising results of optimizing hydroclimatic data, certain limitations must be considered. The accuracy and reliability of the XGBoost algorithm, while well-established for TerraClimate data are not universally guaranteed across all geographical and hydrological contexts. Variations in topography, vegetation cover, and land use can influence the model's performance, requiring localized adjustments.

Moreover, the methodology presented in this study has been developed and validated under specific conditions. Its application to other datasets or climatic regions may necessitate additional calibration steps. Future research should focus on refining the model's adaptability and evaluating its robustness across a wider range of environmental settings to enhance its generalization potential.

4 Conclusions

This study demonstrated the potential of the XGBoost algorithm to optimize TerraClimate hydroclimatic data for the Meknes Plateau and Middle Atlas Causse in Morocco. Comparative analysis revealed a significant correlation between raw TerraClimate data and ground observations, though imperfections in accuracy were noted. Applying XGBoost significantly improved data quality, as shown by the performance metrics: MAE decreased from 3.08 to 1.47, RMSE decreased from 4.84 to 2.3, and NSE increased from 0.82 to 0.99. These results highlight the effectiveness of machine learning in enhancing satellite-based hydroclimatic data, particularly in regions with sparse gauge networks.

The optimized TerraClimate data provide a reliable alternative to ground-based observations, addressing critical data gaps for water resource management in water-stressed regions like Morocco.

This study underscores the value of integrating satellite products with advanced machine learning techniques to improve the accuracy and usability of hydroclimatic data. However, the model's performance may vary in regions with different hydrological and climatic characteristics, emphasizing the need for region-specific adaptations and further validation.

In conclusion, this work contributes to the growing use of satellite data and machine learning in hydroclimatic studies. The findings offer a robust foundation for improving water resource management and resilience in Morocco and similar regions. Future research should explore the transferability of this approach to other areas and incorporate additional data sources to further enhance model performance and applicability.

Acknowledgements

We would like to extend our sincere gratitude to the TerraClimate team for providing their dataset, which was essential for the completion of this study. Additionally, we wish to thank the members of the Research Team and our Supervisor for their invaluable support and guidance throughout this project.

Data availability

The data used in this study were accessed from two primary sources: (1) observed hydroclimatic data (precipitation, temperature, evapotranspiration, and runoff) collected from eight meteorological stations in the Meknes Plateau and Middle Atlas Causse, provided by the Sebou Hydraulic Basin Agency (ABHS) and the Provincial Directorate of Agriculture (DPA) of Meknes, and (2) TerraClimate satellite data, which were made available by the TerraClimate dataset. The observed station data are available upon request from the respective agencies, while the TerraClimate dataset can be accessed online at https://www.climatologylab.org/terraclimate.html or through Google Earth Engine. Processed data and scripts used for analysis in this study are available from the corresponding author upon reasonable request.

Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data*, 5:1–12, doi:10.1038/sdata.2017.191.

Abdullah, M. and Al-Ansari, N. (2022). Missing rainfall data estimation—an approach to investigate different methods: case study of Baghdad. *Arabian Journal of Geosciences*, 15:1740, doi:10.1007/s12517-022-10995-6.

Adnan, R. M., Mirboluki, A., Mehraein, M., and et al. (2024). Improved prediction of monthly streamflow in a mountainous region by Metaheuristic-Enhanced deep learning and machine learning models using hydroclimatic data. *Theoretical and Applied Climatology*, 155:205–228, doi:10.1007/s00704-023-04624-9.

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J. (2003). Guidance on metadata and homogenization. *Wmo Td*, 1186:1–53.

Aguilar, E., Peterson, T. C., Obando, P. R., Frutos, R., Retana, J. A., Solera, M., Soley, J., Garcia, I. G., Araujo, R. M., Santos, A. R., et al. (2005). Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *Journal of Geophysical Research: Atmospheres*, 110, doi:10.1029/2005jd006119.

Aieb, A., Madani, K., Scarpa, M., Bonaccorso, B., and Lefsih, K. (2019). A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon*, 5:e01247,

doi:10.1016/j.heliyon.2019.e01247.
Alexandris, S. (2013). Evapotranspiration: An overview.
Ali, S., Liu, D., Fu, Q., Cheema, M. J. M., Pal, S. C., Arshad, A., Pham, Q. B., and Zhang, L. (2022). Constructing high-resolution groundwater drought at spatio-temporal scale using GRACE satellite data based on machine learning in the Indus Basin. *Journal of Hydrology*, 612:128295, doi:10.1016/j.jhydrol.2022.128295.
Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration – Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300:D05109.
Amiri, M. A. and Mesgari, M. S. (2017). Modeling the spatial and temporal variability of precipitation in northwest Iran. *Atmosphere*, 8:254, doi:10.3390/atmos8120254.
Amraoui, F. (2005). *Contribution à la connaissance des aquifères karstiques: Cas du Lias de la plaine du Sais et du Causse Moyen Atlasique tabulaire (Maroc)*. PhD thesis, Doctorat d'Atat, Université Hassan II, Casablanca.
Beaulieu, C., Ouarda, T. B. M. J., and Seidou, O. (2007). Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations. *Hydrological sciences journal*, 52:18–37, doi:10.1623/hysj.52.1.18.
Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Dijk, A. I. J. M. V., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21:6201–6217, doi:10.5194/hess-21-6201-2017.
Bentayeb, A. and Leclerc, C. (1977). Le Causse moyen atlasique. *Ressources en Eaux du Maroc, Tome3, Domaines Atlasiques et Sud-Atlasiques*, page 37–66.
Bonnet, O. (1970). *Calcul automatique des" bilans d'eau" mensuels et annuels par les méthodes de Thornthwaite et de Turc*. BRGM; Orleans-la Source.
Brandsma, T. and Buishand, T. A. (1998). Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrology and Earth System Sciences*, 2:195–209, doi:10.5194/hess-2-195-1998.
Castellani, C. (1986). Régionalisation des précipitations annuelles par la méthode de la régression linéaire simple: l'exemple des Alpes du Nord. *Revue de géographie alpine*, 74:393–403, doi:10.3406/rga.1986.2658.
Chen, H., Liu, H., Chen, X., and Qiao, Y. (2020). Analysis on impacts of hydro-climatic changes and human activities on available water changes in Central Asia. *Science of the Total Environment*, 737:139779, doi:10.1016/j.scitotenv.2020.139779.
Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, page 785–794. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, doi:10.1145/2939672.2939785.
Cherif, M., Diaz-Cassou, J., and Megevand, C. (2023). Climate Change and Development in Morocco. *Morocco's Quest for Stronger and Inclusive Growth*, page 121.
Chowdhury, R., Mohamed, M. M. A., and Murad, A. (2016). Variability of extreme hydro-climate parameters in the North-Eastern Region of United Arab Emirates. *Procedia Engineering*, 154:639–644, doi:10.1016/j.proeng.2016.07.563.
Daguin, F. (1927). Contribution à l'étude de la région préfiraine (Maroc occidental). Technical report.
Ebert, E. E. (2007). *Methods for verifying satellite precipitation estimates*. Springer, doi:10.1007/978-1-4020-5835-6_27.
Eddif, A., Ouazzani, H., Sadkaoui, I., Tajdi, A., Aghbal, A., and Hamid, A. (2018). La chaine volcanique du Moyen Atlas au Maroc: patrimoine géologique encore peu connu. *International Journal of Innovation and Applied Studies*, 24:117–129.
Er-Raki, S., Chehbouni, A., Guemouria, N., Duchemin, B., Ezzahar, J., and Hadria, R. (2007). Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region. *Agricultural water management*, 87:41–54, doi:10.1016/j.agwat.2006.02.004.
Ezzaouini, M. A., Mahé, G., Kacimi, I., Bilali, A. E., Zerouali, A., and Nafii, A. (2022). Predicting daily suspended sediment load using machine learning and NARX hydro-climatic inputs in semi-arid environment. *Water*, 14:862, doi:10.3390/w14060862.
Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., and Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy conversion and management*, 164:102–111, doi:10.1016/j.enconman.2018.02.087.
Faraj, H., Beaudet, G., Combes, M., Couvreur, G., Ionesco, T., Martin, J., Missante, G., Naciri, M., Smeyers, F., and Watteeuw, R. (1966). Le Sais.
Ghozat, A., Sharafati, A., and Hosseini, S. A. (2021). Long-term spatiotemporal evaluation of CHIRPS satellite precipitation product over different climatic regions of Iran. *Theoretical and applied climatology*, 143:211–225, doi:10.1007/s00704-020-03428-5.
Han, D., Wang, G., Liu, T., Xue, B.-L., Kuczera, G., and Xu, X. (2018). Hydroclimatic response of evapotranspiration partitioning to prolonged droughts in semiarid grassland. *Journal of Hydrology*, 563:766–777, doi:10.1016/j.jhydrol.2018.06.048.
Hanchane, M., Kessabi, R., Krakauer, N. Y., Sadiki, A., Kassioui, J. E., and Aboubi, I. (2023). Performance Evaluation of TerraClimate Monthly Rainfall Data after Bias Correction in the Fes-Meknes Region (Morocco). *Climate*, 11:120, doi:10.3390/cli11060120.
Hao, R. and Bai, Z. (2023). Comparative study for daily streamflow simulation with different machine learning methods. *Water*, 15:1179, doi:10.3390/w15061179.
Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, doi:10.1007/978-0-387-21606-5.
Herrera-Puebla, J., Meneses-Peralta, J., Duarte-Diaz, C., González-Robaina, F., and Hervis-Granda, G. (2020). Determination of Crop Coefficients for Estimating Evapotranspiration in a Paddy Field in Cuba. *Revista Ciencias Técnicas Agropecuarias*, 29:05–20.
Karmouda, N., Kacimi, I., ElKharrim, M., Brirhet, H., and Hamidi, M. (2022). Geo-statistical and hydrological assessment of three satellite precipitation products over Ouergha basin (Northern Morocco). *Arabian Journal of Geosciences*, 15:235, doi:10.1007/s12517-021-09124-6.
Kessabi, R., Hanchane, M., Caloiero, T., Pellicone, G., Addou, R., and Krakauer, N. Y. (2023). Analyzing Spatial Trends of Precipitation Using Gridded Data in the Fez-Meknes Region, Morocco. *Hydrology*, 10:37, doi:10.3390/hydrology10020037.
Kinouchi, T., Sayama, T., et al. (2021). A comprehensive assessment of water storage dynamics and hydroclimatic extremes in the Chao Phraya River Basin during 2002–2020. *Journal of Hydrology*, 603:126868, doi:10.1016/j.jhydrol.2021.126868.
Kofidou, M., Stathopoulos, S., and Gemitzi, A. (2023). Review on spatial downscaling of satellite derived precipitation estimates. *Environmental Earth Sciences*, 82:424, doi:10.1007/s12665-023-11115-7.
Kohavi, R. (1995). A study of cross-validation and bootstrap for

accuracy estimation and model selection. In *Ijcai*, volume 14, page 1137–1145. Ijcai.

Kohler, M. A. (1949). On the use of double-mass analysis for testing the consistency of meteorological records and for making required adjustments. *Bulletin of the American Meteorological Society*, 30:188–195, doi:10.1175/1520-0477-30.5.188.

Koppen, W. (1936). Das geographische System de Klimate. *Handbuch der klimatologie*.

Laurent, H., Jobard, I., and Toma, A. (1998). Validation of satellite and ground-based estimates of precipitation over the Sahel. *Atmospheric Research*, 47:651–670, doi:10.1016/s0169-8095(98)00051-9.

Lettenmaier, D. P. and Wood, E. F. (1993). Hydrologic forecasting.

Li, P., Stuart, E. A., and Allison, D. B. (2015). Multiple imputation: a flexible tool for handling missing data. *Jama*, 314:1966–1967, doi:10.1001/jama.2015.15281.

Mjejra, M., Dubreuil, V., and Hénia, L. (2014). Evaluation de l'évapotranspiration maximale par télédétection à moyenne résolution spatiale: cas du bassin versant de la Mejerda en Tunisie. In *Actes du 27ème colloque de l'Association Internationale de Climatologie*, page 54–60. Actes du 27ème colloque de l'Association Internationale de Climatologie.

Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P., and Righetti, M. (2024). Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environmental Modelling and Software*, page 105971, doi:10.1016/j.envsoft.2024.105971.

Ouatiki, H., Boudhar, A., and Chehbouni, A. (2021). The Suitability of Eight Spatial Rainfall Products to Simulate Daily Streamflow in Semi-Arid Watersheds using the HBV model. In *EGU General Assembly Conference Abstracts*, page EGU21-7428. EGU General Assembly Conference Abstracts, doi:10.5194/egusphere-egu21-7428.

Rachdane, M., Khalki, E. M. E., Saidi, M. E., Nehmadou, M., Ahbari, A., and Tramblay, Y. (2022). Comparison of high-resolution satellite precipitation products in sub-saharan Morocco. *Water*, 14:3336, doi:10.3390/w14203336.

Rachdane, M., Khalki, M. E. E., Saidi, M. E., Nehmadou, M., Ahbari, A., and Tramblay, Y. (2023). Assessment of six satellite precipitation products in a Moroccan arid area. In *EGU General Assembly Conference Abstracts*, page EGU-1808. EGU General Assembly Conference Abstracts, doi:10.5194/egusphere-egu23-1808.

Rachidi, S., Mazoudi, E., Houssine, E. L., Alami, J. E., Jadoud, M., and Er-Raki, S. (2023). Assessment and comparison of satellite-based rainfall products: validation by hydrological modeling using ANN in a semi-arid zone. *Water*, 15:1997, doi:10.3390/w15111997.

Runtunuwu, E. (2007). Penentuan koefisien tanaman menggunakan data pengamatan iklim dan index vegetasi dari setelit NOAA/AVHRR. *Jurnal Ilmu-Ilmu Pertanian Indonesia*, 9:165–171, doi:10.31186/jipi.9.2.165-171.

Saeedi, M., Sharafati, A., Brocca, L., and Tavakol, A. (2022). Estimating rainfall depth from satellite-based soil moisture data: A new algorithm by integrating SM2RAIN and the analytical net water flux models. *Journal of Hydrology*, 610:127868, doi:10.1016/j.jhydrol.2022.127868.

Sebou Hydraulic Basin Agency (2024). Nappe des Causses moyens Atlasiques. https://www.abhsebou.ma/presentation-du-bassin/eaux-souterraines/nappe-des-causses-moyens-atlasiques/.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, page 517–524. Proceedings of the 1968 23rd ACM national conference, doi:10.1145/800186.810616.

Shrestha, N. K. and Shukla, S. (2015). Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agricultural and forest meteorology*, 200:172–184, doi:10.1016/j.agrformet.2014.09.025.

Tabyaoui, F. (2005). Contribution à l'étude hydrogéochimique des eaux de la nappe phréatique du plateau de Meknès (MAROC). pages 53–62.

Tan, M. L., Gassman, P. W., and Cracknell, A. P. (2017). Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins. *Water*, 9:229, doi:10.3390/w9030229.

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

Vörösmarty, C. J., Federer, C. A., and Schloss, A. L. (1998). Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. *Journal of Hydrology*, 207:147–169, doi:10.1016/s0022-1694(98)00109-7.

Wang, H. and Xuan, Y. (2022). A spatial pattern extraction and recognition toolbox supporting machine learning applications on large hydroclimatic datasets. *Remote Sensing*, 14:3823, doi:10.3390/rs14153823.

World Bank Group (2022a). Note Technique Pénurie d'eau et sécheresses. https://documents1.worldbank.org/curated/en/099052223171010874/pdf/P177376089c18001f08b5c09e715216fca3.pdf.

World Bank Group (2022b). Rapport Climat et Développements. https://openknowledge.worldbank.org/server/api/core/bitstreams/0a549cdc-c5c8-53a0-a570-078800be7e02/content.

Xu, T., Deines, J. M., Kendall, A. D., Basso, B., and Hyndman, D. W. (2019). Addressing challenges for mapping irrigated fields in subhumid temperate regions by integrating remote sensing and hydroclimatic data. *Remote Sensing*, 11:370, doi:10.4211/hs.3766845be72d45969fca21530a67bb2d.

Yu, S., Chen, Z., Yu, B., Wang, L., Wu, B., Wu, J., and Zhao, F. (2020). Exploring the relationship between 2D/3D landscape pattern and land surface temperature based on explainable eXtreme Gradient Boosting tree: A case study of Shanghai, China. *Science of the total environment*, 725:138229, doi:10.1016/j.scitotenv.2020.138229.