Vol. 17(10), pp. 252-274, October 2023
DOI: 10.5897/AJEST2023.3221
Article Number: ECF684971454
ISSN: 1996-0786

http://www.academicjournals.org/AJEST

ACADEMIC JOURNALS
expand your knowledge

African Journal of Environmental Science and Technology

Full Length Research Paper

An evaluation of the performance of imputation methods for missing meteorological data in Burkina Faso and Senegal

Semou Diouf[1], Abdoulaye Deme[1]*, El Hadji Deme[2], Papa Fall[1] and Ibrahima Diouf[3]

[1]LEITER, Applied Science and Technology Training and Research Unit, Gaston Berger University, BP 234, Saint-Louis 32000, Senegal.
[2]LERSTAD, Applied Science and Technology Training and Research Unit, Gaston Berger University, BP 234, Saint-Louis 32000, Senegal.
[3]LPAOSF, Polytechnic School, Cheikh Anta Diop University, BP 5085, Dakar 10700, Senegal.

Received 14 August, 2023; Accepted 24 October, 2023

Addressing data incompleteness issues is crucial for reliable climate studies, especially in regions like Africa that commonly experience data gaps. This study aims to evaluate the performance of five imputation methods (knn, ppca, mice, imputeTS, and missForest) on meteorological data from stations in Burkina Faso and Senegal. The imputed data is compared with ERA5 reanalysis data to validate its accuracy. Temperature, relative humidity, and precipitation observations from the GSOD dataset (1973-2020) were used, creating subsets with missing rates of 5, 10, 20, 30 and 40%. An evaluation was conducted using the Taylor diagram and Kling-Gupta Efficiency (KGE). The results show a good estimation of temperature and relative humidity time series, with missForest performing the best for handling missing values. Precipitation estimation was less accurate, but there was strong agreement between estimated and observed data. ImputeTS was recommended for precipitation. Spatial consistency between imputed data and ERA5 reanalysis products was found. This research improves the quality of meteorological data, provides essential information about climatic characteristics, and serves as a foundation for climate change and weather modeling studies.

Key words: Meteorological data, imputation methods, Senegal, Burkina Faso.

INTRODUCTION

Numerous studies have addressed the issue of missing data in time series, spanning across Africa, Europe, and various other regions worldwide (Moron et al., 2016; Kertali, 2019; Yozgatligil et al., 2013). Scientific disciplines such as meteorology, biology, social sciences, and medicine rely on complete and reliable data to conduct in-depth studies. However, the presence of missing data in a database can distort analysis results, introduce biases, and lead to essential information loss (Lotsi et al., 2017). This issue is common and inevitable in the data

*Corresponding author. E-mail: abdoulaye.deme@ugb.edu.sn. Tel: +221764149737.

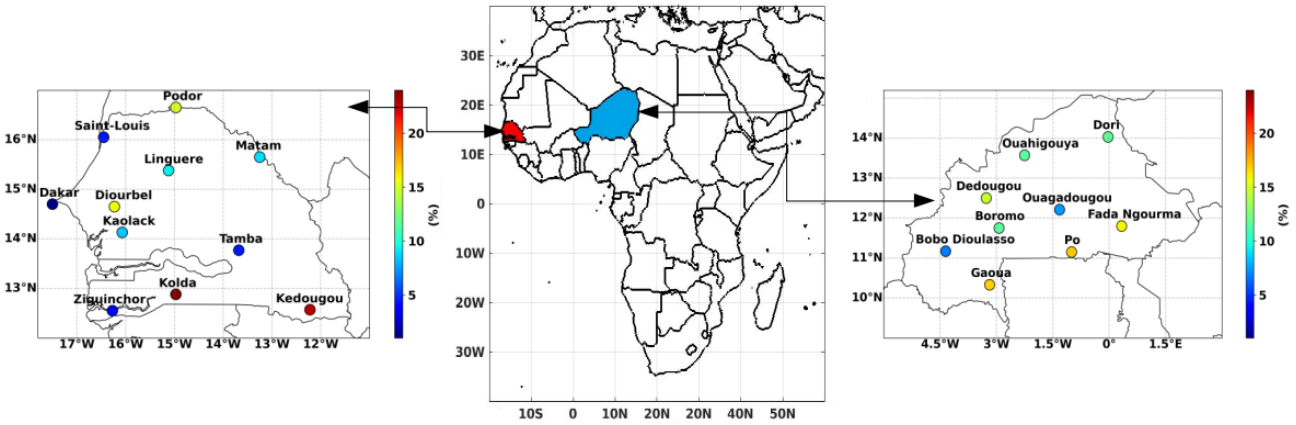

Figure 1. Study area and percentage of missing data for the 11 meteorological stations (Ziguinchor, Kolda, Kedougou, Kaolack, Tamba, Dakar, Diourbel, Linguere, Matam, Saint-Louis and Podor) in Senegal (left) and the 09 stations (Gaoua, Bobo Dioulasso, Pô, Boromo, Fada Ngourma, Ouagadougou, Dedougou, Ouahigouya and Dori) in Burkina Faso (right).

acquisition process, often resulting from faulty measurements or human errors during manual recordings (Schafer and Graham, 2002; Beaulieu et al., 2007).

The context of climate change underscores the importance of having comprehensive and reliable meteorological databases for conducting significant climate studies. For instance, climate studies that account for trends require time series of at least 30 years with minimal missing data. Unfortunately, the available meteorological data for several stations in Senegal and Burkina Faso, covering the period from 1973 to 2020, exhibit missing data rates exceeding 10% (Figure 1). This reality is common in meteorology and can limit the reliability of climate data. Faced with these challenges, it is essential to find appropriate methods to fill in the missing data in meteorological databases.

Numerous approaches have been developed to address missing values in time series data. Some statistical methods involve limiting the study to records with complete information, but this approach is often limited when missing data rates exceed 5% (Davey et al., 2001; Niass et al., 2015). To address these issues, various imputation techniques have been developed and successfully used. Moron et al. (2016) used probabilistic principal component analysis (ppca) to fill in missing values in daily temperature time series from the Global Surface Summary of the Day (GSOD) and Global Historical Climatology Network (GHCN) datasets. Soltani and Haouari (2017) reconstructed monthly series of maximum and minimum temperatures over western Algeria using the ppca method. Kertali (2019) applied the double mass method to fill in the missing annual precipitation series collected at the Dar El Beida meteorological station in Algeria over a period of 24 years, from 1983 to 2018. Dixneuf et al. (2021) conducted a comparative study between missForest, mice, and knn on 10 complete environmental databases of various natures (qualitative, quantitative, and mixed data). The relative performance of each imputation method was evaluated using two error indicators: the normalized Root Mean Square Error (NRMSE) quantified the errors associated with quantitative attributes, while the Proportion of False Classified (PFC) is the indicator used for qualitative attributes. More recently, Bousri et al. (2021) used Mean Absolute Bias estimation (MAB) to compare the performance of four multiple imputation methods of missing temperature data, namely: missForest, mice, knn, and principal component analysis (PCA). Comparing six imputation methods, Yozgatligil et al. (2013) showed the accuracy to fill the gaps in Turkish meteorological databases for temperature and precipitation. These authors applied a nonlinear dynamic approach that accounts for spatial and temporal dependencies in the time series to evaluate imputation performances. Costa et al. (2021) used the mice method to fill gaps in the meteorological databases of 96 stations in Brazil from 1961 to 2014.

The gap-filling techniques applied by the authors have demonstrated the feasibility of generating plausible values for each weather variable of interest. Typically, the performance of time series technical reconstruction is assessed using precision measures like Mean Absolute Error (MAE), Akaike Information Criterion (AIC), and Root Mean Square Error (RMSE) among others. In this article, we employed Taylor diagrams (Taylor, 2001) and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) to assess the performance of meteorological data imputation techniques.

The primary objective of this study is to assess the performance of five imputation methods, expanding our evaluation to encompass meteorological data from Burkina Faso in addition to the Senegalese stations. This extension enables us to compare the robustness of imputation methods in two countries with similar climatic

characteristics, but with a key difference: the presence of the ocean only in Senegal. Furthermore, the variable of relative humidity has been introduced to provide a more comprehensive analysis of the meteorological data. To enhance the reliability and accuracy of the results obtained, we have also employed advanced methodological approaches based on time series and spatial distributions to validate the imputed data.

Hence, this article represents an enriched application of our previous study, with a specific emphasis on meteorological aspects and a broader geographical scope. The incorporation of Burkina Faso and the inclusion of the relative humidity variable enhance the significance of our research for policymakers and researchers interested in the climate of West Africa. Moreover, the validation of the imputed data through comparison with reanalysis data adds an additional dimension to the credibility of the results obtained.

MATERIALS AND METHODS

In situ and reanalysis data sets

The daily observations of precipitation (Pr), relative humidity (Rh), maximum temperature (Tmax), minimum temperature (Tmin) and mean temperature (Tmean) spanning from 1973 to 2020 were obtained from 11 stations in Senegal (Figure 1; left) and 09 mainland stations in Burkina Faso (Figure 1; right). Notably, two Burkina Faso stations, Dedougou and Pô, are concerned for 1983-2020 period. These observed data were sourced from the Global Surface Summary of the Day (GSOD) data set. GSOD encompasses 18 distinct daily surface weather variables (Moron et al., 2016) each exhibiting varying percentages of missing values, ranging from 1 to 20%, across all the stations. To provide an overview of the data scarcity issue in Africa, Figure 1 illustrates the daily maximum temperature's percentage of missing values across all stations. This same comparison applies to all other variables.

The ERA5 reanalyses are employed to assess the spatial consistency of the imputed data. ERA5 relies on advanced modeling and data assimilation systems to integrate extensive historical observations into global estimates. This dataset covers the entire earth and provides atmospheric resolution across 137 levels, from the surface up to an altitude of 80 km. The ERA5 daily atmospheric fields were provided by the European Center for Medium-Range Weather Forecasts (ECMWF) at a 0.25° × 0.25° (25 × 25 km) resolution (Hersbach et al., 2020). In this study, we used precipitation (Pr), relative humidity (Rh), maximum temperature (Tmax), minimum temperature (Tmin), and 2 m temperature (T2m) over the period from 1973 to 2020. The study subsequently outlines various methods employed to address missing values within the dataset, each tailored to the specific dataset in consideration, along with their corresponding evaluation criteria.

Treatment of missing data

The missing data classification is subdivided into three categories depending on missing data probabilities: (i) when the probability of having missing data is not associated with the variable with missing values itself and is not related to any other measured variable: it is in the Missing Completely at Random (MCAR) category; (ii) when the probability of missing data is not related to the missing values of the variable itself and is related to some measured variable: it falls under the Missing at Random (MAR) category; (iii) when the probability of missing values depends on the values of variables in question: it belongs to the Missing Not at Random (MNAR) category. Two major imputation methods have been proposed here to deal with missing data issues: (i) simple imputation where the missing value is replaced with a single plausible value and (ii) multiple imputations where the missing value is replaced with some probable values. We tested five imputation methods: k-nearest neighbors (knn) (Kowarik and Templ, 2016), time series missing value imputation (imputeTS) (Moritz and Bartz-Beielstein, 2017), probabilistic principal components analysis (ppca) (Josse and Husson, 2016), multiple imputation by chained equations (mice) (Buuren and Groothuis-Oudshoorn, 2011), and random forest (missForest) (Stekhoven, 2011) on the daily climate database of temperature, relative humidity and rainfall. More details on these methods (knn, imputeTS, ppca, mice and missForest) can be found in Diouf et al. (2022). All the databases were incomplete, with missing data present. To compare different imputation methods, we first generated complete databases using the k-nearest neighbors (knn) approach. Then, we introduced random rates of missing values at five different percentages: 5, 10, 20, 30, and 40%. This procedure applied to all the 20 stations, resulting in 100 matrices of missing values (5 × 20).

Performance evaluation criteria

Taylor's diagram

Meteorologists often use Taylor's diagram (Taylor, 2001) to evaluate the performance of climate models by comparing the distance between observations and simulations. Taylor's diagram shows the quality of imputation methods predictions with respect to the actual values by presenting three complementary statistics: centered Root Mean Square Error (RMSEc), standard deviation of the simulation from the observation (σ), and correlation between estimated data and observations (r).

Kling-Gupta efficiency (KGE)

The Kling-Gupta Efficiency (KGE) metric is commonly used to evaluate model performance in hydrology. To provide a more comprehensive assessment, the KGE combines three components: correlation, bias, and the ratio of variances (Gupta et al., 2009; Osuch et al., 2015; Melsen et al., 2019). A KGE value of 1 indicates a perfect relationship between the simulations and observations. To assess the relationship between estimated values and observations, we used a modified version of the KGE statistic proposed by Kling et al. (2012). Based on the modifications according to Kling et al. (2012), the performance of the methods can be classified as follows: good (KGE ≥ 0.75), intermediate (0.75 > KGE ≥ 0.5), poor (0.5 > KGE > 0), and very poor (KGE ≤ 0). Taylor's diagram and the Kling-Gupta Efficiency criterion are detailed in Diouf et al. (2022).

RESULTS AND DISCUSSION

In this study, five widely used imputation methods were employed to complete databases with varying percentages of missing values (5, 10, 20, 30 and 40%). The databases consisted of daily temperature (Tmax, Tmin, and Tmean), rainfall (Pr), and relative humidity (Rh) data obtained from 11 stations in Senegal and 09

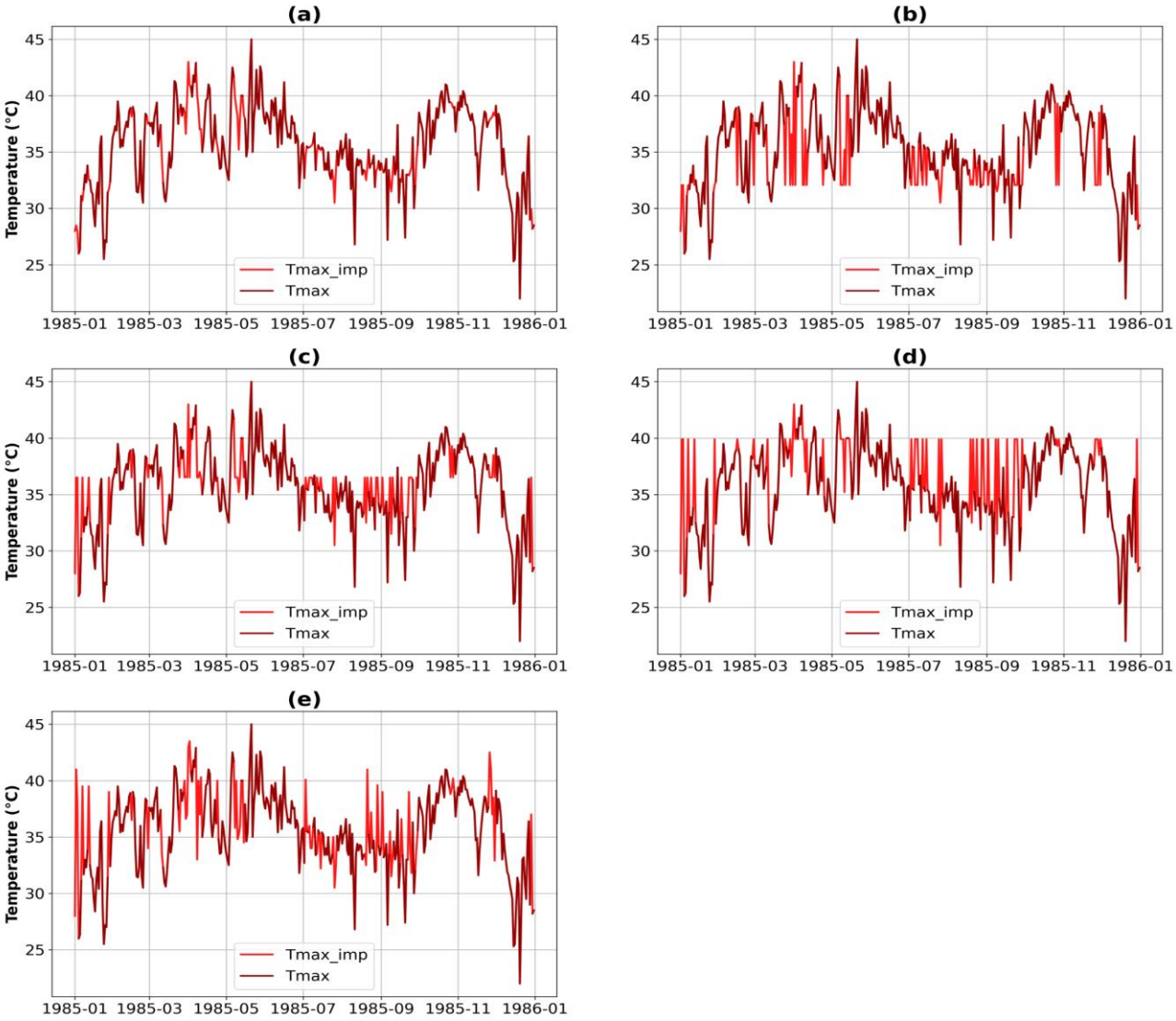

Figure 2. Time series of original (Tmax) and imputed (Tmax_imp) maximum temperature for Diourbel station (1985): (a) imputeTS; (b) missForest; (c) ppca; (d) *knn*; (e) mice.

stations in Burkina Faso.

Daily evolution of climate parameters after missing values imputation

Figure 2 presents an example of gap filling in the daily maximum temperature (Tmax) data for the Diourbel station located in central Senegal. A comparative study was conducted for the year 1985, during which missing values were recorded at the station for several days. The distribution of the incomplete original data is depicted in dark red, while the series imputed by different filling techniques are shown in light red. It was observed that the estimated Tmax (Tmax_imp) time series exhibits variable evolutions depending on the imputation method used. This variability is primarily attributed to the distribution of missing values in the original database over time. In Figure 2, where the gaps have been replaced by the imputeTS method, the evolution of Tmax appears reasonable in comparison to the series with missing values. However, at the beginning of 1985, the imputation methods provided different estimations for the missing values in the series. Some of them (ppca, mice, and knn) tend to strongly overestimate the values of Tmax (approximately 40°C) when compared with the first non-missing value of the original series (26°C). This discrepancy indicates that these methods might not accurately capture the true underlying pattern of Tmax for the missing days. Additionally, it becomes evident that the imputation methods diverge when multiple successive days have gaps. The ppca and knn methods, in

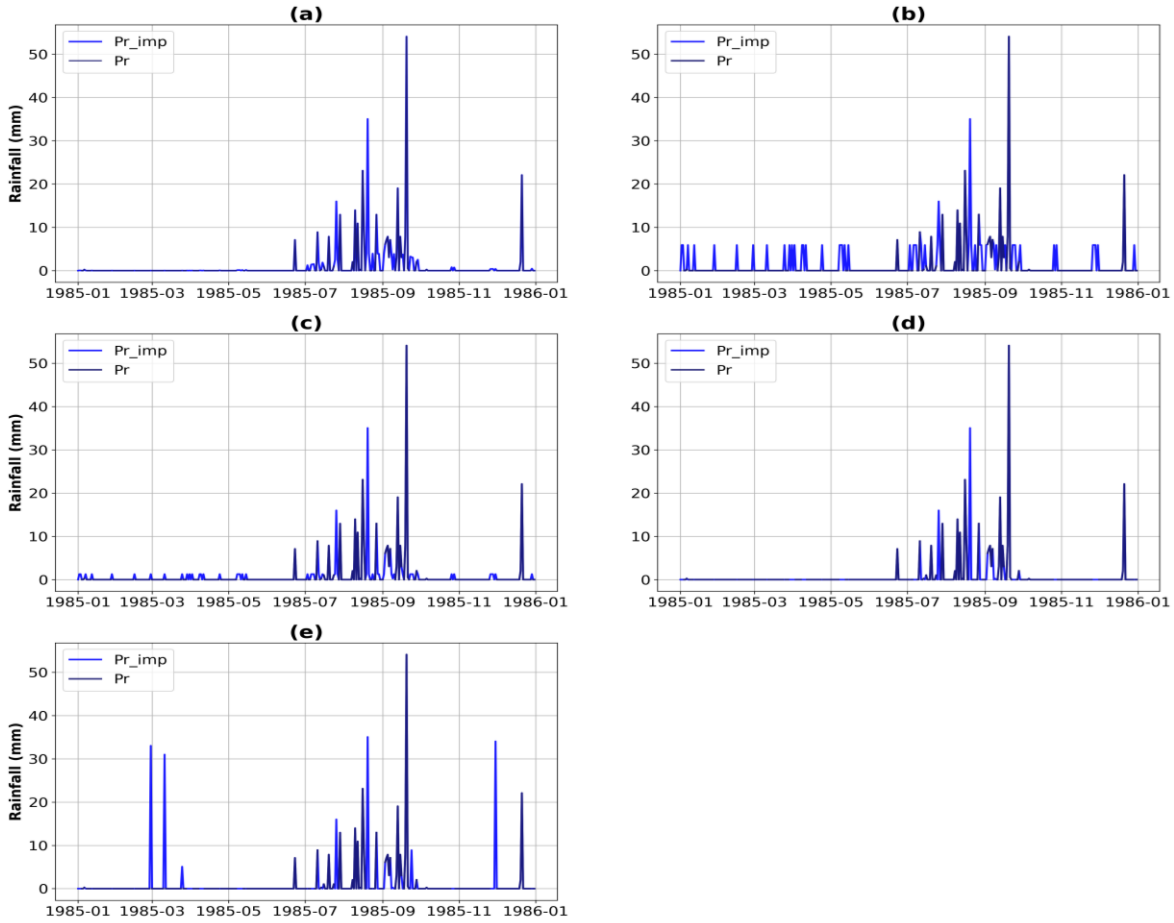

Figure 3. Time series of original (Pr) and imputed (Pr_imp) rainfall for Diourbel station (1985): (a) imputeTS; (b) missForest; (c) ppca; (d) knn; (e) mice.

particular, impute the same values to all consecutive missing values, which may not accurately reflect the actual temperature variations during those periods. Figure 3 displays the time series of original (dark blue) and imputed (light blue) rainfall data for the Diourbel station in 1985. The findings indicate that three out of the five studied methods (missForest, ppca, and knn) assign rainfall values throughout the year, which does not align with the realities of our study area. In the study area, rainfall generally begins in early May (in the southern part of the country) and concludes between October and November. However, these imputation methods seem to struggle in accurately reproducing the time series of rainfall, leading to significant overestimation of the values.

The results presented in Figures 2 and 3 suggest that the imputation methods perform better in filling the gaps in the Tmax data compared to the rainfall data. This difference in performance can be attributed to the small temporal variation in temperatures. Temperature variations are relatively minor from one day to another. However, this is in contrast to rainfall, which exhibits substantial variability. As a result, imputing rainfall values becomes a more challenging task due to the greater fluctuations in the data. This pattern holds for other climate parameters, such as Tmin, Tmean, and relative humidity, which are continuous variables, in comparison to rainfall. Continuous variables typically exhibit smoother and more predictable patterns, making the imputation process more accurate.

Furthermore, it is important to note that when missing values are distributed successively, the imputation methods deviate significantly from the original series. This indicates that the accuracy of the imputed values decreases when dealing with consecutive missing data points.

Performance evaluation of imputation methods

Subsequently, the Taylor diagram and the KGE metric will be used to evaluate performance of imputation methods. Figures 4 to 7 provide a summary of the relative performance of the five imputation methods using Taylor's diagram to estimate daily maximum temperatures

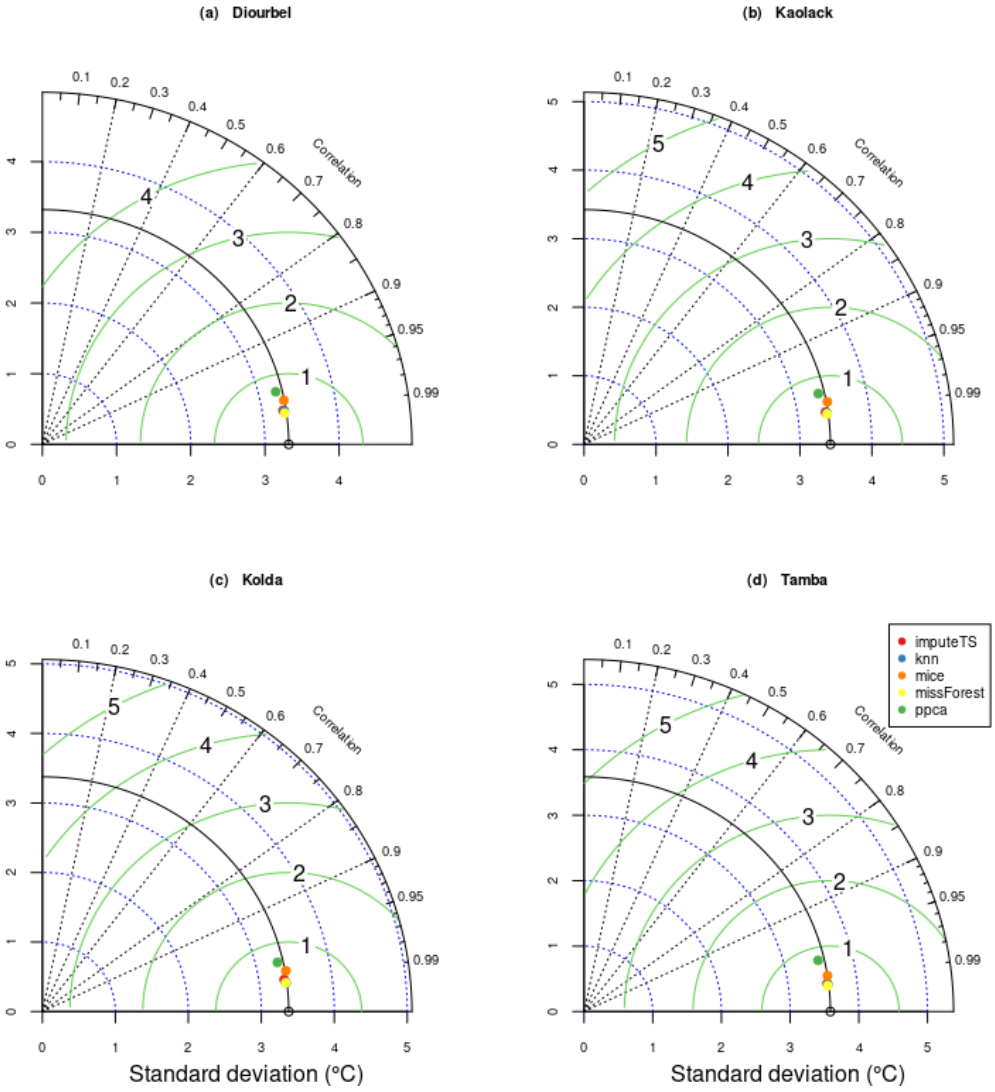

Figure 4. Taylor's diagram of maximum temperatures (Tmax) for Diourbel (a), Kaolack (b), Kolda (c) and Tamba (d) stations with 5% missing data**.**

at four stations in Senegal (Diourbel, Kaolack, Kolda, and Tamba) and Burkina Faso (Bobo Dioulasso, Dori, Fada Ngourma, and Ouagadougou). The figures compare the performance for scenarios with 5 and 20% missing data.

The results indicate a strong correlation between the estimated and observed temperature values, with r values very close to 1 for all stations. Additionally, the imputation methods exhibit less scattering (low standard deviation values) and have low root mean square error (RMSEc) values. The colored points (representing values estimated by the imputation methods) are almost merged and very close to the uncolored points located on the x-axis (representing the observations), suggesting that the imputation methods provide quite satisfactory estimates.

Among the imputation methods, the ppca method stands out slightly from the other data series reconstruction techniques with a slightly lower r value. However, the overall performance of all methods remains high, indicating the effectiveness of the imputation techniques in estimating missing temperature data. A noticeable dispersion of points is observed as the percentage of missing data increases (Figures 4 versus 5 and 6 versus 7). This corresponds to a decrease in r values and an increase in mean square error values and standard deviations, indicating reduced accuracy in the estimation results as the amount of missing data grows. The mice method shows the closest σ value to the observations. For a 20% shortage scenario, the imputeTS, knn, and missForest methods yield the best results. For percentages of missing values below 20%, all imputation methods perform well, with very strong correlations (> 0.9) and RMSEc values lower than 1°C. However, the performance of the imputation techniques deteriorates when dealing with 30% missing data or higher. In such

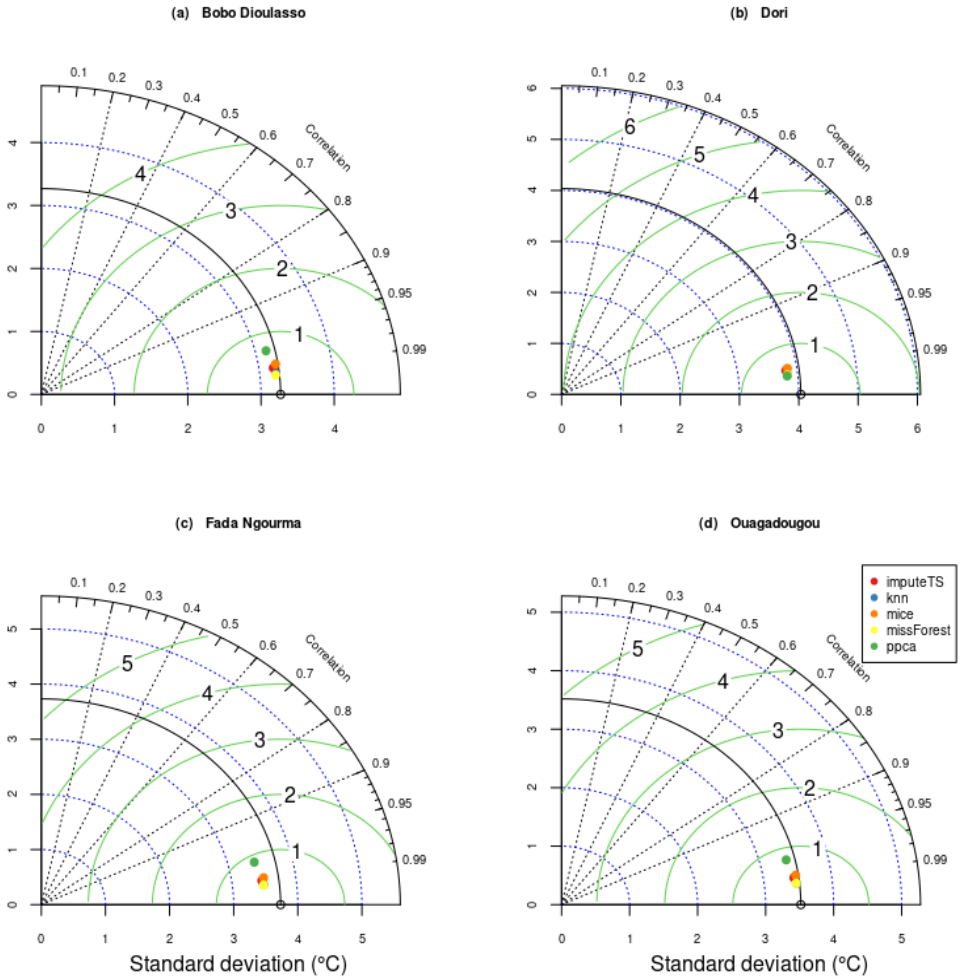

Figure 5. Taylor's diagram of maximum temperatures (Tmax) for Bobo Dioulasso (a), Dori (b) Fada Ngourma (c), Ouagadougou (d) stations with 5% missing data.

cases, the mean square error exceeds 1°C, indicating a decrease in accuracy and reliability of the estimated temperatures.

To evaluate the performance of the imputation methods for precipitation, a comparison is made between datasets with 5 and 20% missing data in Senegal and Burkina Faso, respectively. The results are presented in Figures 8 to 11. The correlations between observations and estimated precipitation values are strong, exceeding 0.9 in all scenarios. However, the root mean square error values for precipitation are higher compared to temperature and vary significantly between different imputation methods. For the 5% missing data scenario (Figures 8 and 9), the RMSEc values range from 1 to 3 mm, indicating a relatively good accuracy in the estimation of precipitation values. However, for the 20% missing data scenario (Figures 10 and 11), the RMSEc values increase to a range of 3 to 7.5 mm suggesting a larger discrepancy between the estimated values and the observations. Overall, the results show that precipitation is generally less accurate estimated compared to temperature. The high simulated variance and RMSEc values for the 20% missing data scenario indicate that imputing missing precipitation data becomes more challenging as the percentage of missing data increases. Among the imputation methods, the mice method stands out with the largest RMSEc values relative to the observed data in the estimation of precipitation time series. In contrast, the missforest and imputeTS methods appear to yield estimates that are closer to the observed values, indicating better performance for filling in missing precipitation data.

In summary, the study demonstrates positive correlations between estimated and observed precipitation values in all missing data scenarios. However, accurate estimation of precipitation remains more challenging compared to temperature. To decide between the methods that appear to be the most efficient in estimating the time series of temperature, relative humidity, and precipitation, we used the Kling-Gupta

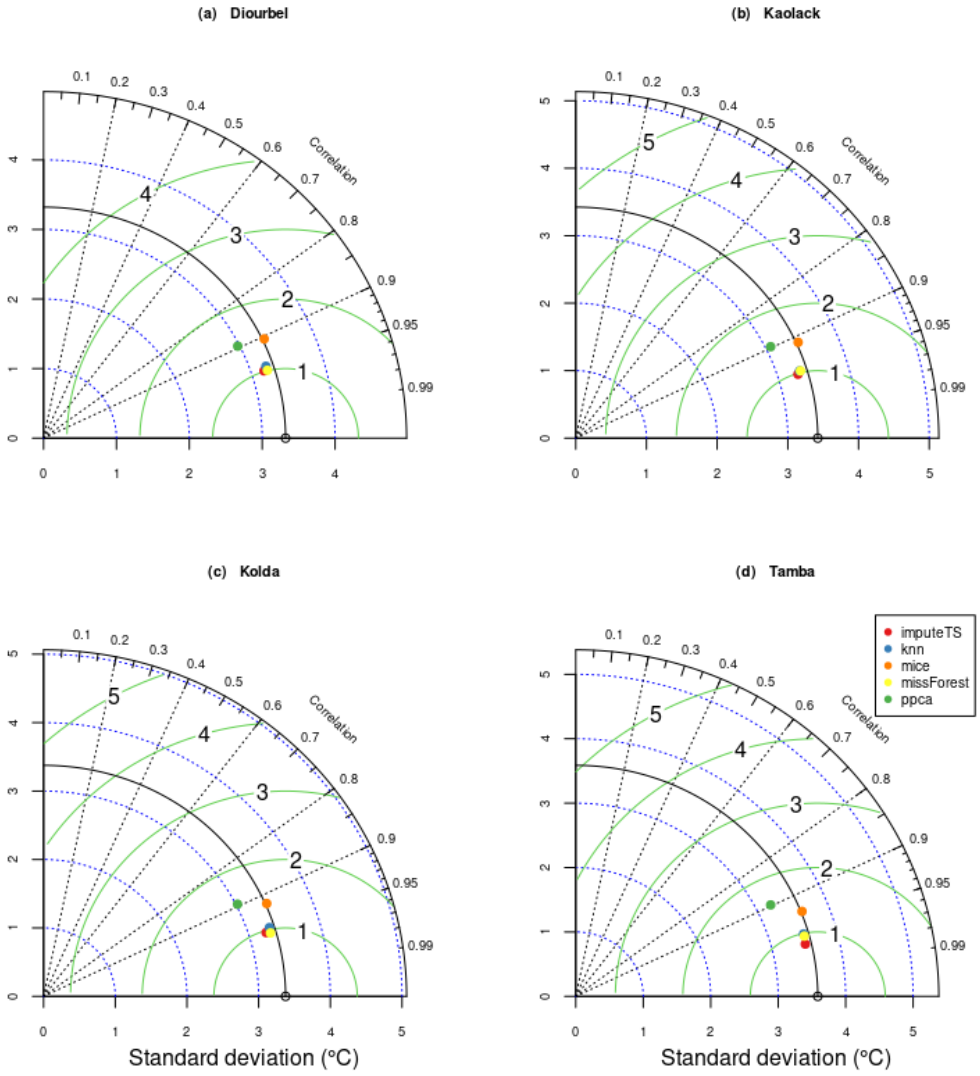

Figure 6. Taylor's diagram of maximum temperatures (Tmax) for Diourbel (a), Kaolack (b), Kolda (c) and Tamba (d) stations with 20% missing data.

Efficiency (KGE) criterion, which provides values ranging from 0 to 1. The effectiveness of a method is indicated by a KGE value close to 1. Tables 1 to 4 display the Kling-Gupta Efficiency values for all the variables analyzed under different percentages of missing data. Following the interpretation suggested by Kling et al. (2012), KGE values are considered good (KGE ≥ 0.75) for all temperature imputation methods when dealing with 5 to 20% missing data. Between 30 and 40% missing data, KGE values decrease but still remain above 0.75 for all methods except for the ppca method, which yields intermediate results (0.75 > KGE ≥ 0.5). The KGE scores obtained from the precipitation time series are relatively lower compared to those obtained from temperature, across all missing data scenarios. However, similar to the temperature analysis, the KGE criterion indicates good results for precipitation imputation methods when dealing with 5 to 20% missing data, with the exception of the knn and mice methods, which fall into the intermediate class for some stations. When the percentage of missing precipitation data increases to 30%, the scores remain in the intermediate range with a few exceptions. The Kling-Gupta Efficiency (KGE) scores obtained from the precipitation time series are not as good as those obtained from temperatures. This trend is consistent across all missing data scenarios. For temperature data, the KGE criterion results are considered good when the missing data ranges between 5 and 20%. However, for precipitation data, the KGE scores are lower, even within this 5 to 20% range, compared to the temperature data. For most imputation methods, the KGE scores for precipitation remain in the intermediate range when the percentage of missing data is 5 to 20%, with the exception of the knn and mice methods, which are in the

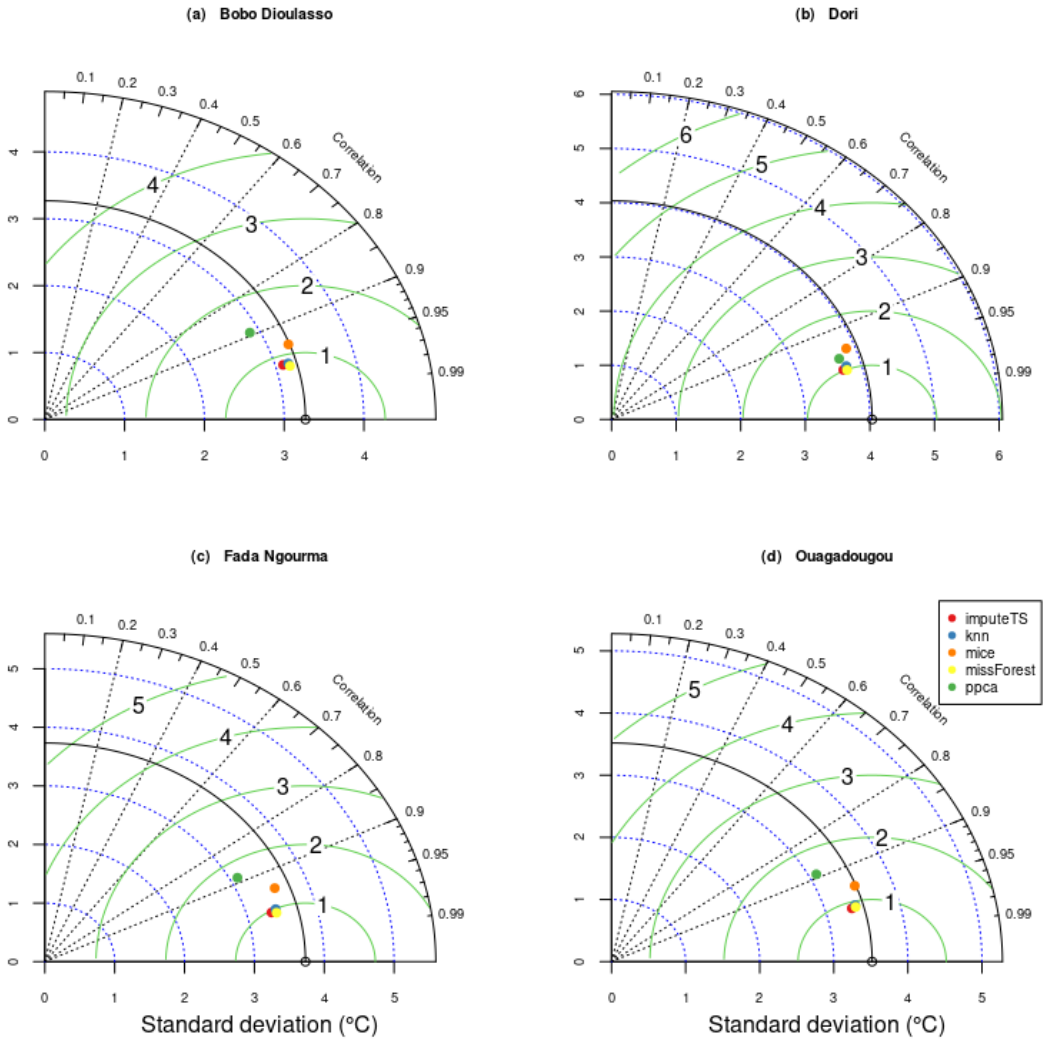

Figure 7. Taylor's diagram of maximum temperatures (Tmax) for Bobo Dioulasso (a), Dori (b) Fada Ngourma (c), Ouagadougou (d) stations with 20% missing data.

intermediate class for a few stations. As the percentage of missing precipitation data increases to 30%, the KGE scores become intermediate with a few exceptions. However, when the percentage of missing precipitation data reaches 40%, the KGE scores drop significantly (KGE < 0.5), especially for the missForest and knn imputation methods.

The KGE results (around 0.80 and 0.95 for 40 and 5% of missing temperature and humidity data, respectively) are in good agreement with those from the Taylor's diagrams. Accordingly, based on the KGE results, imputeTS (followed by ppca) is the best method for dealing with missing values in rainfall time series. In summary, irrespective of the percentage of missing values, the imputation methods have generally demonstrated success in estimating the observed temperature and relative humidity data. This success is attributed to the fact that temperature is a continuous variable and has very little spatiotemporal disparity.

According to the KGE and Taylor's diagram performance indicators, missForest is a suitable method for imputing relative humidity and temperature time series data. On the other hand, the imputeTS method is suitable for imputing precipitation data. Stekhoven and Bühlmann (2012) came to a similar conclusion in their study when comparing missForest with missPALasso, mice, and knn methods on 11 real databases. In environmental studies, Dixneuf et al. (2021) also showed that the missForest method is suitable for imputing missing data, while Bousri et al. (2021) found it to be more efficient than ppca, mice, and knn methods for missing temperature data imputation. Regarding the overall behavior for temperatures, the knn method produced estimates very close to the observations, while the ppca method was the least efficient with higher RMSEc values. However, all the methods, except for mice, underestimated the variability

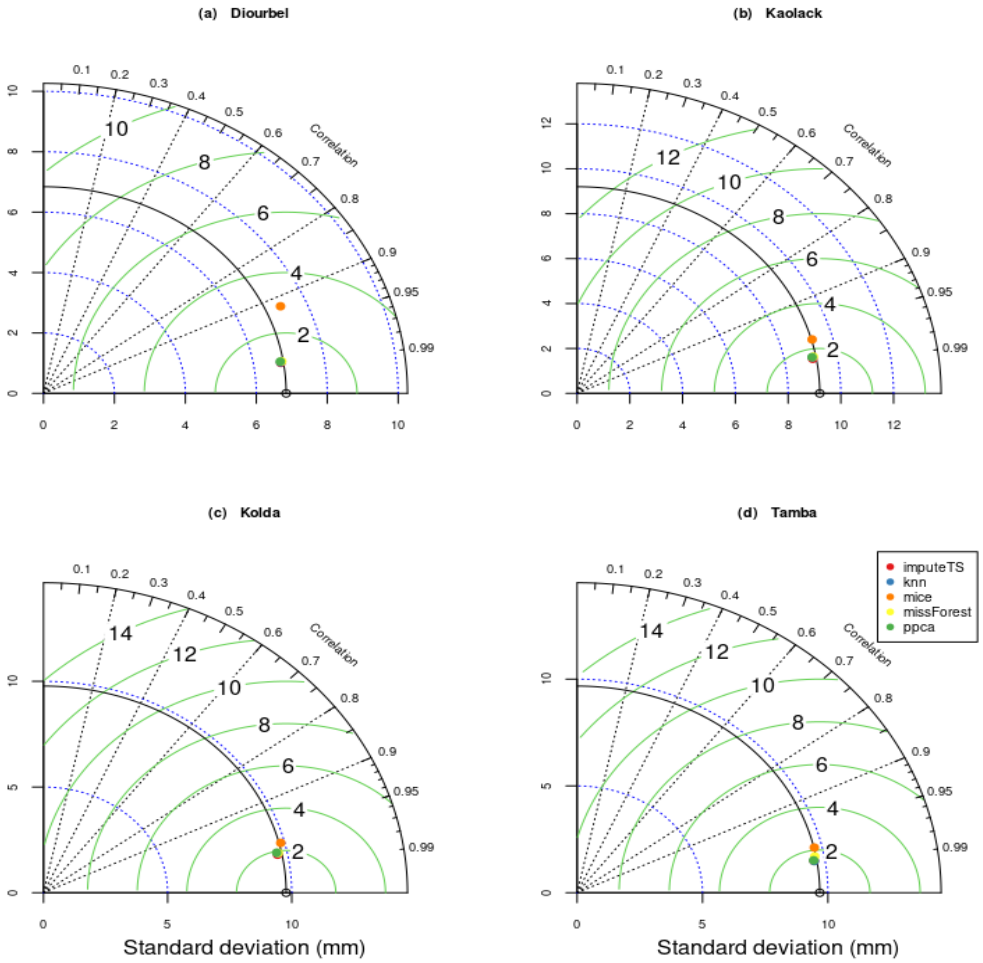

Figure 8. Taylor's diagram of rainfall (Pr) for Diourbel (a), Kaolack (b), Kolda (c) and Tamba (d) stations with 5% missing data.

of temperature since they fell within the standard deviation of the observed series (black contour). As the amount of missing data increased, the performance of missForest method decreased. When the percentage of missing values reached 40%, the estimated data moved further away from the observations. In this case, the imputeTS method performed better than missForest Similar results were found when applying the same analyses on Tmin, Tmean, and Rh.

Annual cycle of temperature, humidity, and rainfall

Figure 12 shows the average annual cycle of five meteorological parameters (Tmax, Tmin, Tmean, rainfall, and relative humidity) for the 11 stations in Senegal. Missing values in the different databases were previously filled in with the missForest method for temperature and relative humidity, and the imputeTS method for rainfall. An annual bi-modal distribution is observed in the continental stations in Senegal, with a main temperature peak in spring (March-April-May), and a secondary peak in autumn (October-November). The annual cycle of Tmax at coastal stations (Dakar and Saint-Louis) differs from that of continental stations, as they do not have a spring maximum. The maritime influence causes, compared to the rest of the country, a small difference in the seasonal regime marked by a warm period from March to October and a cold period from November to February. The Tmin seasonal cycle is characterized by two peaks in the continental stations: a spring peak delayed by about one month and a second, much weaker peak present in autumn.

Dakar, which is a coastal station, exhibits a single-mode cycle with its highest maximum, minimum, and mean temperatures recorded between July and October, coinciding with the rainy season. A mono-modal annual cycle, characteristic of Senegal's rainfall regime, is found in all stations. Additionally, the maximum amount of rainfall is observed in August.

The southern part of the country, particularly in Ziguinchor, receives the greatest amount of rainfall, as is

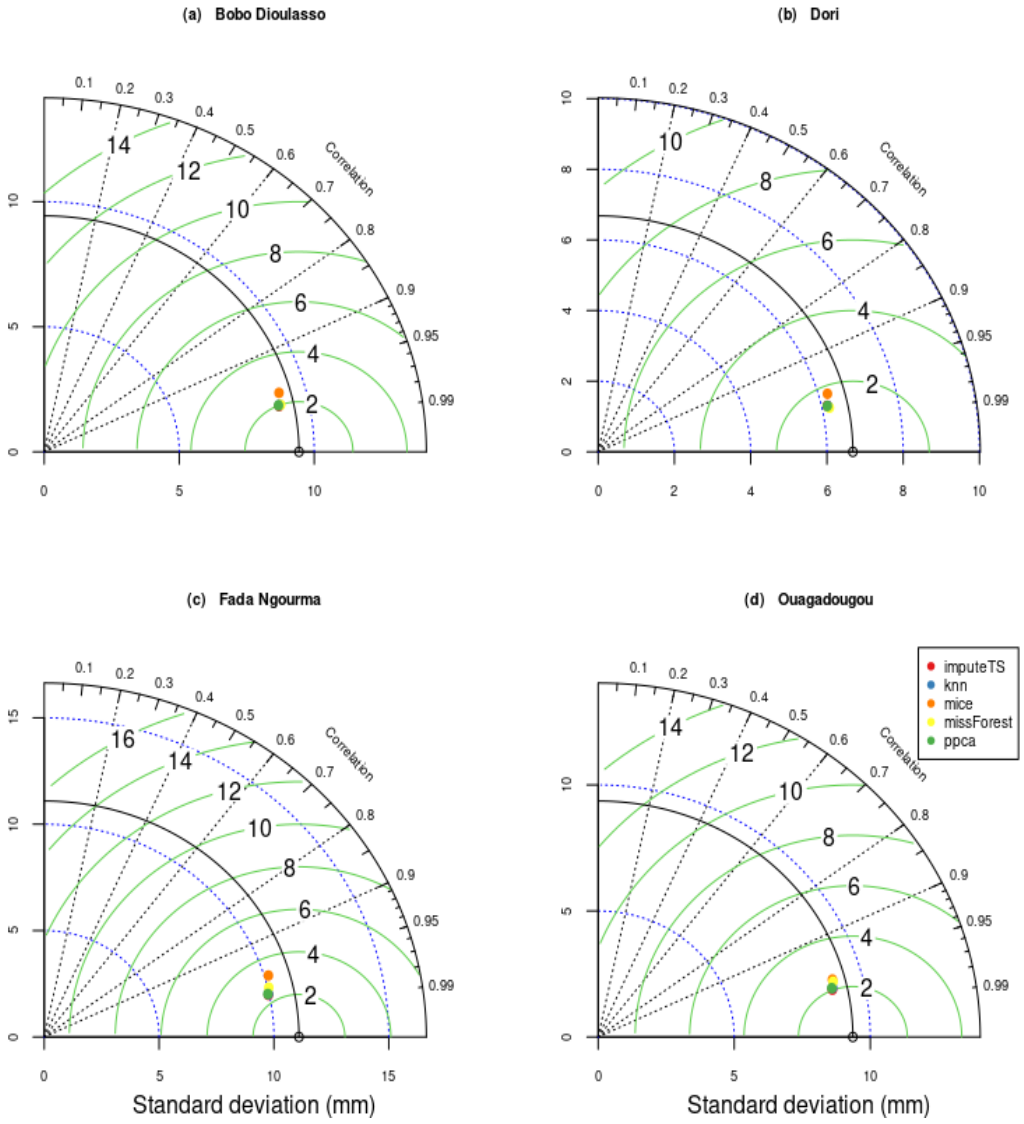

Figure 9. Taylor's diagram of maximum temperatures (Tmax) for Bobo Dioulasso (a), Dori (b) Fada Ngourma (c), Ouagadougou (d) stations with 5% missing data.

usually the case. The high relative humidity, depending on air temperature and hygrometric characteristics, coincides with the actual rainy season. Continental stations show a single peak in relative humidity in August. Dakar is characterized by high relative humidity values (above 70%) during most of the year. Saint-Louis and Ziguinchor show a fairly similar evolution of their seasonal cycle, with several months of the year experiencing high relative humidity. Overall, the major trends in temperature, relative humidity, and rainfall observed for all the stations in Senegal are represented in the annual cycle. According to Sagna (2007), the succession between a dry season (November to May) and a wet season (June to October) characterizes the Sudano-Sahelian zone. The west of the country has cooler temperatures than the east, which, along with the center, is the hottest part of the country. The cumulative annual rainfall follows a decreasing gradient from south to north of Senegal. It should also be noted that the peak of relative humidity is recorded in the middle of the rainy season.

Figure 13 displays the annual cycle of temperature (Tmax, Tmin, Tmean), relative humidity, and rainfall for 09 stations in Burkina Faso. These nine stations are well distributed throughout Burkina Faso, with at least two stations per climatic zone: the Sahelian zone in the north (Ouahigouya and Dori), the sub-Sahelian zone in the center (Dedougou, Ouagadougou, Fada Ngourma, and Boromo), and the northern Sudanian zone in the south (Gaoua, Pô, and Bobo Dioulasso). Burkina Faso experiences a tropical climate influence, with a distinct rainy season lasting from June to October and a dry

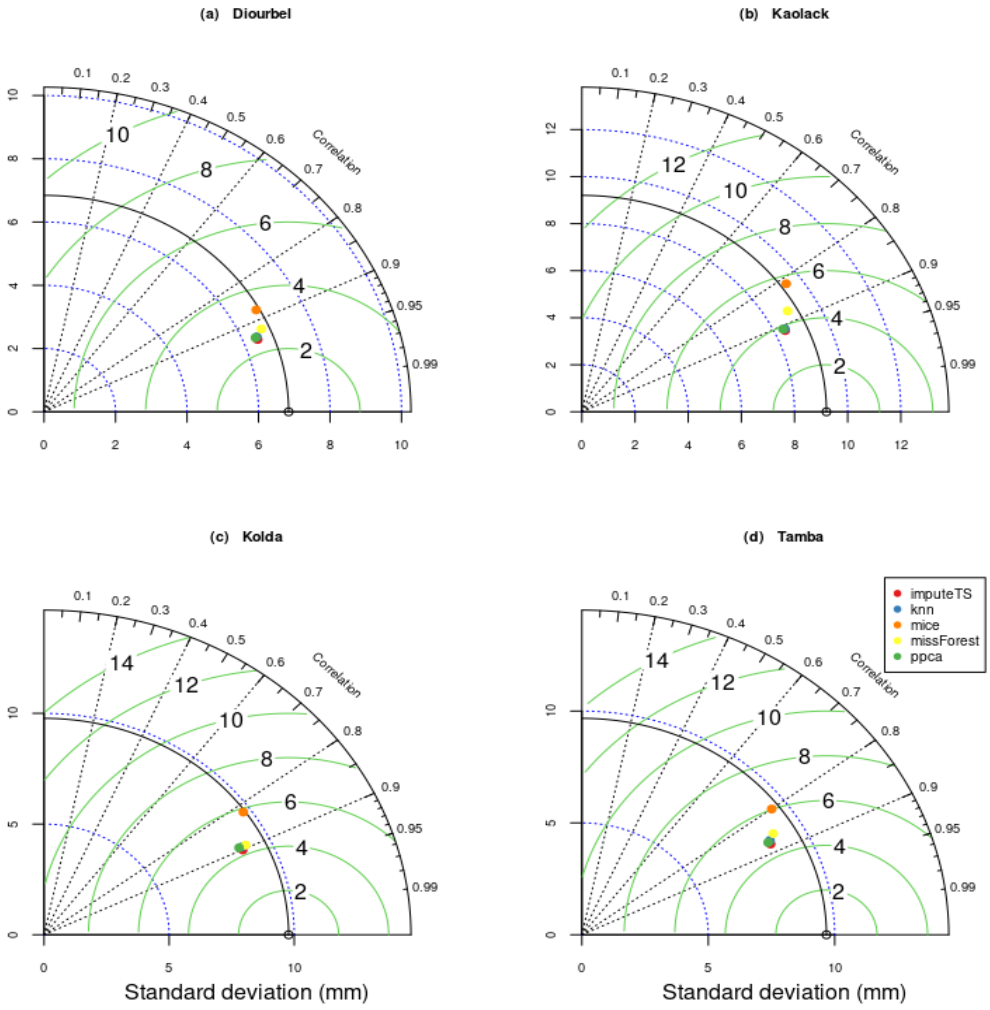

Figure 10. Taylor's diagram of rainfall (Pr) for Diourbel (a), Kaolack (b), Kolda (c) and Tamba (d) stations with 20% missing data.

season lasting from November to May. In the southern part of the country, the rainy season is longer and more intense, with more than 500 mm of rainfall per year. On the other hand, in the northern regions, the rains are limited to a shorter period of time. This climatic variation across Burkina Faso contributes to the diverse seasonal patterns of temperature, relative humidity, and rainfall observed in the different regions.

In Burkina Faso, January has the lowest Tmin and Tmean of the year, making it the coldest month. On the other hand, March and April are the warmest months, with the highest thermal averages in April. Depending on the season, the average maximum temperatures range between 31 and 40°C. During the coldest months, the temperature can drop to an average of 17°C per month, depending on the station. The bi-modal annual cycle is evident in the distribution of maximum, minimum, and average temperatures in Burkina Faso. However, the bi-modal evolution of minimum temperatures for the Burkina Faso stations is more noticeable compared to the Senegal stations. This means that in Burkina Faso, the variation in minimum temperatures between the coldest and warmest months is more pronounced compared to the pattern observed in Senegal.

In Burkina Faso, according to the main characteristics of its climate (Nakazawa and Matsueda, 2017), the highest relative humidity occurs in August, while the low relative humidity is obtained in February. This variation in relative humidity throughout the year is influenced by the seasonal changes in rainfall and temperature patterns in the region. The rainy season, which occurs from June to October, contributes to higher humidity levels, while the dry season, from November to May, leads to lower humidity levels, particularly in February. High temperatures are recorded almost all year round. From December to March, the presence of dry air is linked to the harmattan winds from the northeast. The maximum temperatures can reach 32°C during this time of year,

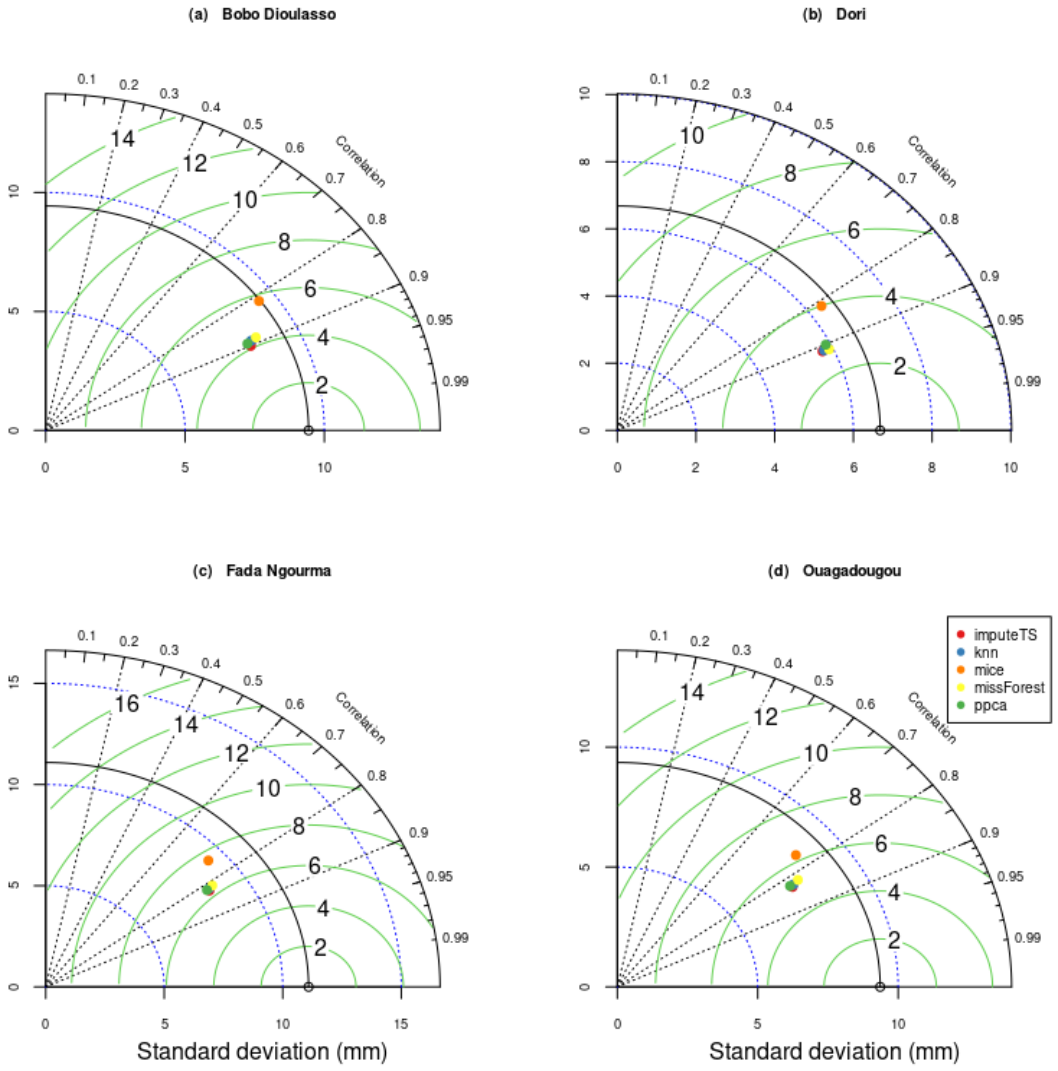

Figure 11. Taylor's diagram of rainfall (Pr) for Bobo Dioulasso (a), Dori (b) Fada Ngourma (c), Ouagadougou (d) stations with 20% missing data.

especially in the northern part of Burkina Faso. From February onwards, the temperature rises, and the heat becomes unbearable. The early arrival of the first precipitations in the south increases the relative humidity but also prevents the temperature from rising. Meanwhile, in the north and center, the heat is felt until April and May. In May and June, we witness the first intense showers from the south. Between July and September, when the monsoon is very intense, temperatures drop considerably throughout the country. Temperatures rise again as the monsoon recedes in October and November, more so in the north, where the maximum temperature returns to around 39°C, than in the south, where it remains around 36°C. Then, the main differences between Senegal and Burkina Faso are the presence of the Atlantic Ocean, which affects the temperature distribution of coastal stations in Senegal. As a continental country with stations sufficiently far from the ocean, all temperatures in Burkina Faso show a net bi-modal distribution (Figure 13).

Climatology of temperature, humidity, and rainfall

The most climate features for local populations in Sahel whose main activity is rainfed farming is seasonal cycle of monsoon. The seasonal cycle of the monsoon is characterized by the alternation between a dry season and a wet season which brings most of the precipitation across the Sahel. Figure 14 presents ERA5 maximum temperature (Tmax) climatology of the DJF (December-January-February), MAM (March-April-May), JJA (June-July-August), and SON (September-October-November) seasons in Senegal (top) and Burkina Faso (bottom).

Table 1. KGE scores for Senegal stations in a 5% missing data scenario.

Stations	Methods	Tmax	Tmin	Tmean	Pr	Rh
Diourbel	imputeTS	0.9835714	0.9831029	0.9824979	0.97872**	0.993394*
	knn	0.98728**	0.99220**	0.995849*	0.9533001	0.9887329
	mice	0.9820596	0.9857274	0.9929311	0.9075048	0.9779893
	missForest	0.988220*	0.992204*	0.99574**	0.9780226	0.98977**
	ppca	0.9614748	0.9631652	0.9651033	0.980264*	0.9724518
Kaolack	imputeTS	0.9857776	0.9831615	0.9815469	0.98168**	0.995729*
	knn	0.98983**	0.993495*	0.995836*	0.9423991	0.9918405
	mice	0.9830087	0.9899196	0.9934291	0.9588845	0.9817371
	missForest	0.990137*	0.99347**	0.99574**	0.982119*	0.99207**
	ppca	0.9643667	0.9635497	0.9630784	0.9801031	0.9724289
Kolda	imputeTS	0.9866557	0.9824853	0.9803995	0.976400*	0.982450*
	knn	0.991739*	0.99604**	0.997025*	0.9617717	0.9924523
	mice	0.9845515	0.993894	0.9953118	0.9703627	0.9820476
	missForest	0.99104**	0.996573*	0.99697**	0.97591**	0.99300**
	ppca	0.9667154	0.9632489	0.9632126	0.972823	0.9724476
Tamba	imputeTS	0.9905763	0.9859656	0.9905723	0.964945*	0.997212*
	knn	0.99348**	0.99261**	0.99626**	0.9444012	0.9909432
	mice	0.9903511	0.9901326	0.9946909	0.9226171	0.9798432
	missForest	0.993885*	0.993775*	0.996636*	0.9615129	0.99195**
	ppca	0.9643679	0.9662771	0.9645768	0.96340**	0.9719351

The method with the best scores is represented by one star (*) and the next best by two stars (**).

Circles superimposed represent the same climatology at 11 stations in Senegal and 09 stations in Burkina Faso, where the imputations have been applied. Spatially, the lowest temperature values in Senegal are recorded at stations located in coastal areas (Saint-Louis, Dakar). This drop in temperature is explained by the influence of maritime trade winds from the Azores anticyclone. Inland regions, on the other hand, do not benefit from the maritime influence. Instead, they are subject to the effects of continental conditions and thus record high temperatures. The regions located in the east and center of the country have the highest values. This is true for all seasons of the year. The spatial distribution shows a clear difference in temperature from one season to another. DJF is the season with the lowest temperatures. MAM is the warmest time of the year. The monthly mean temperature during JJA decreases with the onset of the rains. Shortly after the rainy season (SON season), temperatures vary a little. Average monthly temperatures in Burkina Faso during the MAM season are very high compared to other times of the year. Shortly before the hot period (that is, the DJF season), conditions are much milder. During the rainy season (JJA), there is a moderate decrease in temperatures, as well as in SON. Whatever the season, there is good agreement between ERA5 and Tmax observations, although Linguere (SN) and Ouahigouya (BF) seem to show slight differences. The amplitude of variation (around 16°C) observed for MAM and JJA seasons in Senegal do not alter the robustness of results. The heterogeneity of stations, comparing coastal and continental stations of Senegal or comparing Senegal vs. Burkina Faso, also does not seem to affect the robustness of imputation methods applied.

Climate warming has been extensively studied by many authors (Karl et al., 1993; Easterling et al., 1997), and they have demonstrated that it is largely influenced by the warming at night compared to the warming during the day. To capture the evolution of the diurnal cycle, we will study maximum and minimum temperatures separately. Figure 15 presents the ERA5 Tmin climatology for the four seasons, along with the associated imputed observations. Some slight discrepancies are noted for Burkina Faso at Bobo Dioulasso and Gaoua during the MAM season, as well as at Dori during the JJA season. Similar behavior is found in Senegal at Matam during the MAM season and Podor during the JJA season. Although these results might seem surprising, they are nonetheless in agreement with recent studies (Bao and Zhang, 2012; Barbier et al., 2018). In fact, numerous

Table 2. KGE scores for Senegal stations in a 20% missing data scenario.

Stations	Methods	Tmax	Tmin	Tmean	Pr	Rh
Diourbel	imputeTS	0.9340079	0.9354445	0.9242161	0.886251*	0.964294*
	knn	0.93894**	0.956278*	0.97297**	0.8003527	0.9301912
	mice	0.9042651	0.9263232	0.9524887	0.8779328	0.8750933
	missForest	0.943650*	0.95533**	0.974097*	0.8407321	0.93864**
	ppca	0.8432292	0.8435042	0.8431742	0.87652**	0.8389912
Kaolack	imputeTS	0.9395557	0.9255107	0.9241274	0.866969*	0.976151*
	knn	0.94444**	0.961467*	0.972078*	0.7978136	0.9471171
	mice	0.9110506	0.9322998	0.9526561	0.8150224	0.9038403
	missForest	0.94584*	0.95638**	0.96872**	0.8243893	0.95220**
	ppca	0.8459225	0.841268	0.8482069	0.85507**	0.8414453
Kolda	imputeTS	0.9405598	0.9284303	0.9246609	0.84158**	0.962767*
	knn	0.94848**	0.96952**	0.97500**	0.8248824	0.9534172
	mice	0.9164409	0.9391756	0.9611785	0.8201832	0.9001023
	missForest	0.953179*	0.976873*	0.979321*	0.845697*	0.95411**
	ppca	0.8421375	0.8412171	0.8424865	0.8251458	0.8371688
Tamba	imputeTS	0.962639*	0.9291145	0.9587488	0.821293*	0.986146*
	knn	0.9559973	0.93534**	0.97065**	0.7770106	0.9399982
	mice	0.9304526	0.9218699	0.9576508	0.7972693	0.8904773
	missForest	0.95833**	0.947257*	0.973431*	0.7812935	0.95838**
	ppca	0.8457186	0.838419	0.8429114	0.81101**	0.8407419

The method with the best scores is represented by one star (*) and the next best by two stars (**).

authors have previously highlighted differences in the Sahel region between observations and reanalyses, such as ERA-Interim (Dee et al., 2011), MERRA (Rienecker et al., 2011), and NCEP-CFSR (Saha et al., 2010). Despite these discrepancies, a good agreement between ERA5 and Tmin observations is confirmed in Figure 15. This suggests that despite some variations, the imputed data still align well with ERA5 in capturing the minimum temperature patterns.

Figure 16 illustrates the ERA5 mean surface temperature climatology for the four seasons in Senegal (top) and Burkina Faso (bottom), along with imputed observations represented as circles. The data reveals some discrepancies at three stations located in the center of Senegal, namely Kaolack, Diourbel, and Linguere. These discrepancies are more noticeable during the DJF season compared to the MAM and JJA seasons. Additionally, two stations in Burkina Faso, Gaoua and Dori, show slight discrepancies during the DJF and JJA seasons, which align with the findings from the Tmins climatology. In a study conducted by Barbier et al. (2018), they compared CRU temperature data with three widely used reanalyses, namely ERA-Interim, MERRA, and NCEP-CFSR, and observed significant differences in the Sahel region. Despite these differences, their results indicated that ERA-Interim and CRU demonstrate good global agreement, both in terms of annual average temperatures and trends. In contrast to ERA5, MERRA and NCEP-CSFR both tend to overestimate warming in the Sahel region across all seasons, and they exhibit a cold bias of a few degrees in winter, especially. Similar behavior to ERA5 is observed in the DJF season at the Kaolack, Diourbel, and Linguere stations. However, slight discrepancies found at these stations during the MAM and JJA seasons could be associated with the significant improvement of the warming representation by ERA5 in spring and rainy seasons. These findings on T2m (mean surface temperature) confirm the results shown in the Tmax and Tmin climatologies (Barbier et al., 2018).

The annual temperature cycle is the result of the combined influence of diurnal (Tmax) and nocturnal (Tmin) temperatures. The mechanisms governing the signatures of diurnal and nocturnal temperatures, observed in the Sahel and not entirely captured by ERA5, remain unclear. One potential explanation for the discrepancies between observations and ERA5 could be related to clouds or aerosols, which are not accurately, represented in the ERA5 data. These factors may impact temperature patterns and are particularly relevant for stations in the center of Senegal (Kaolack, Diourbel, and

Table 3. KGE scores for Burkina Faso stations in a 5% missing data scenario.

Stations	Methods	Tmax	Tmin	Tmean	Pr	Rh
Bobo Dioulasso	imputeTS	0.9872164	0.9828469	0.9898066	0.9738689	0.996151*
	knn	0.99514**	0.99196**	0.99806**	0.986721*	0.9906032
	mice	0.9929849	0.9881264	0.9963624	0.9742302	0.9831725
	missForest	0.995264*	0.992459*	0.998275*	0.97567**	0.99121**
	ppca	0.9677492	0.9672469	0.9734236	0.9733045	0.9702543
Dori	imputeTS	0.9878939	0.9911565	0.9917957	0.9725401	0.994430*
	knn	0.996445*	0.99711**	0.99900**	0.986164*	0.9929414
	mice	0.9943862	0.9964888	0.9982624	0.9031751	0.9808325
	missForest	0.99635**	0.997341*	0.999127*	0.97588**	0.99313**
	ppca	0.9705909	0.9713002	0.9730614	0.9713653	0.9711391
Fada Ngourma	imputeTS	0.9862052	0.9888173	0.9885409	0.969691*	0.994700*
	knn	0.99717**	0.99775**	0.99871**	0.96958**	0.9930543
	mice	0.9943225	0.9950144	0.9968945	0.9649539	0.9821287
	missForest	0.997243*	0.997794*	0.998776*	0.9690851	0.99356**
	ppca	0.9689393	0.9717281	0.9719625	0.9690284	0.9704246
Ouagadougou	imputeTS	0.9892712	0.9915593	0.9929666	0.9792713	0.997558*
	knn	0.99562**	0.99675**	0.99821**	0.98015**	0.99434**
	mice	0.9937532	0.9945344	0.9958028	0.9792421	0.9843125
	missForest	0.996231*	0.996933*	0.998369*	0.980642*	0.9943192
	ppca	0.9706874	0.9712022	0.9726659	0.9796095	0.9716285

The method with the best scores is represented by one star (*) and the next best by two stars (**).

Linguere), where they may be influenced by the monsoon flow from the south and aerosols advection from the north.

The West African Monsoon (WAM) represents the only rainfall event of the year in the Sahelian belt; all water resources depend on it, as do natural and cultivated plant resources. The WAM is characterized by the thermal contrast between the overheated Sahara during the summer and the relatively cooler ocean. Indeed, the ocean having a greater thermal inertia than the continent, the surface temperature of the continent is higher than that of the ocean. This difference creates a very marked meridian thermal gradient essential in the dynamics of the African monsoon, because it strengthens the trade winds of the southern hemisphere (southeast winds) which can then cross the equator (Diallo, 2018). The West African monsoon is therefore closely associated with the thermal depression formed on the West African continent due to the high insolation over the region. Senegal's climate is marked by a single rainy season or winter season, from June to October, which begins in the southeast (towards Kedougou) with the arrival of the monsoon and gradually invades the country. Precipitation increases slowly at first until mid-August when it reaches its maximum. The decrease is pronounced from mid-September onwards, and it becomes abrupt in October (Sagna, 2007).

Figure 17 (left column) shows the spatial distribution of rainfall over Senegal for the ERA5 reanalysis considering the seasons DJF, MAM, JJA and SON, for the period 1973-2020. A notable latitudinal gradient in JJA and SON separates the northern and southern parts of Senegal. The southwestern and southeastern parts of Senegal (Ziguinchor, Kolda, and Kedougou) receive more precipitation than the other regions of the country. The highest rainfall totals were recorded during the June-July-August period. The results show that the observations agree well with the ERA5 reanalysis, despite some deviations noted at the southern stations (Ziguinchor and Kedougou) in the JJA. Burkina Faso is under the influence of a tropical climate of the Sudano-Sahelian type. It is subject to seasonal alternation of humid monsoon air from high oceanic pressure and dry air from Saharan latitudes. The rainy season extends from May to the end of September and reaches its peak in August. There is a big difference between the northern and southern parts of Burkina Faso. The North is an arid zone, while the South has a long rainy season, which is sufficient to ensure a good harvest and more vegetation. Figure 17 (bottom) shows the spatial distribution of

Table 4. KGE scores for Burkina Faso stations in a 20% missing data scenario.

Stations	Methods	Tmax	Tmin	Tmean	Pr	Rh
Bobo Dioulasso	imputeTS	0.9457386	0.9218723	0.9411972	0.872509*	0.980788*
	knn	0.97344**	0.96413**	0.98314**	0.8613012	0.9447044
	mice	0.9533427	0.9407926	0.9671803	0.8345268	0.8917407
	missForest	0.975698*	0.967459*	0.984249*	0.8554356	0.94638**
	ppca	0.8462119	0.8384238	0.8452328	0.86475**	0.8422583
Dori	imputeTS	0.9414085	0.9494714	0.9470757	0.8478164	0.974299*
	knn	0.97509**	0.97788**	0.98720**	0.879913*	0.9542543
	mice	0.9593889	0.9643602	0.9767185	0.7477426	0.8923424
	missForest	0.977019*	0.980551*	0.987525*	0.87380**	0.95785**
	ppca	0.8397796	0.8411755	0.8463069	0.8436018	0.8447236
Fada Ngourma	imputeTS	0.9338015	0.9346789	0.9366333	0.7848675	0.974351*
	knn	0.981505*	0.97821**	0.988440*	0.78920**	0.9530212
	mice	0.9660728	0.9618837	0.9773704	0.793903*	0.8935671
	missForest	0.98045**	0.979247*	0.98827**	0.7861294	0.95932**
	ppca	0.8382011	0.8410722	0.8431608	0.7791571	0.8390588
Ouagadougou	imputeTS	0.9452872	0.9482752	0.9565578	0.7720296	0.984733*
	knn	0.97277**	0.97089**	0.98306**	0.807928*	0.9589387
	mice	0.9547706	0.9499016	0.9663506	0.7557566	0.9115751
	missForest	0.974311*	0.972741*	0.984247*	0.77302**	0.96077**
	ppca	0.8441102	0.8367376	0.8457214	0.7633141	0.8425379

The method with the best scores is represented by one star (*) and the next best by two stars (**).

precipitation in Burkina Faso according to the seasons DJF, MAM, JJA and SON for the ERA5 reanalysis. Significant rainfall variations from one season to the next can be seen. In the Sahelian zone (north), rainfall averaged 250 mm in JJA. In the south (Sudanian zone), the highest cumulative rainfall is recorded (about 600 mm). The central zone (between 11.6 and 13.5° north latitude), under the influence of a Sudano-Sahelian climate, recorded a cumulative rainfall of 350 mm. Burkina Faso, like Senegal, has a latitudinal rainfall gradient. JJA is the wettest season of the year. The climatology of the observational data is in line with the ERA5 reanalysis over several parts of Senegal and Burkina Faso. There are some rare exceptions, notably at the Bobo Dioulasso, Dedougou and Pô stations in JJA. For Dedougou and Pô, the difference noted in relation to the ERA5 reanalysis can be explained by the fact that the climatology of the rainfall data for these stations is made over the period 1983-2020, while for the ERA5 data, the study period is much longer (1973-2020).

In Senegal, the stations located in the south have a much stronger signal for the observations than for the ERA5 reanalysis during the wettest seasons (JJA and SON). According to Awange et al. (2016), the reliability of *in-situ* rainfall data is sometimes questioned. De Longueville et al. (2016), in their study of temperature and rainfall time series in Burkina Faso over the period 1950-2013, had to exclude stations from their study due to the presence of outliers. Relative humidity, expressed as a percentage, is the amount of water vapour present in the air. The possibility of rain occurs when Rh is at 100%: air is then saturated and can no longer hold water vapor. Figure 18 shows the spatial distribution of relative humidity in Senegal and Burkina Faso as a function of the seasons of the year (DJF, MAM, JJA and SON).

The results show that relative humidity varies relatively greatly from one season to another. There is an east-west gradient during the DJF and MAM seasons, with maximum relative humidity in the coastal regions (Saint-Louis, Dakar and Ziguinchor). In JJA, maximum values of relative humidity are recorded due to the arrival of the West African monsoon. Senegal and Burkina Faso have several similarities in terms of spatial distribution of seasonal relative humidity. Indeed, the highest relative humidity values are recorded during the rainy season. In addition, there is a clear distinction between the northern, southern and central zones. There is a very good spatial coherence between the imputed data from the Senegal and Burkina Faso stations and the ERA5 reanalysis data for most seasons. However, some discrepancies are

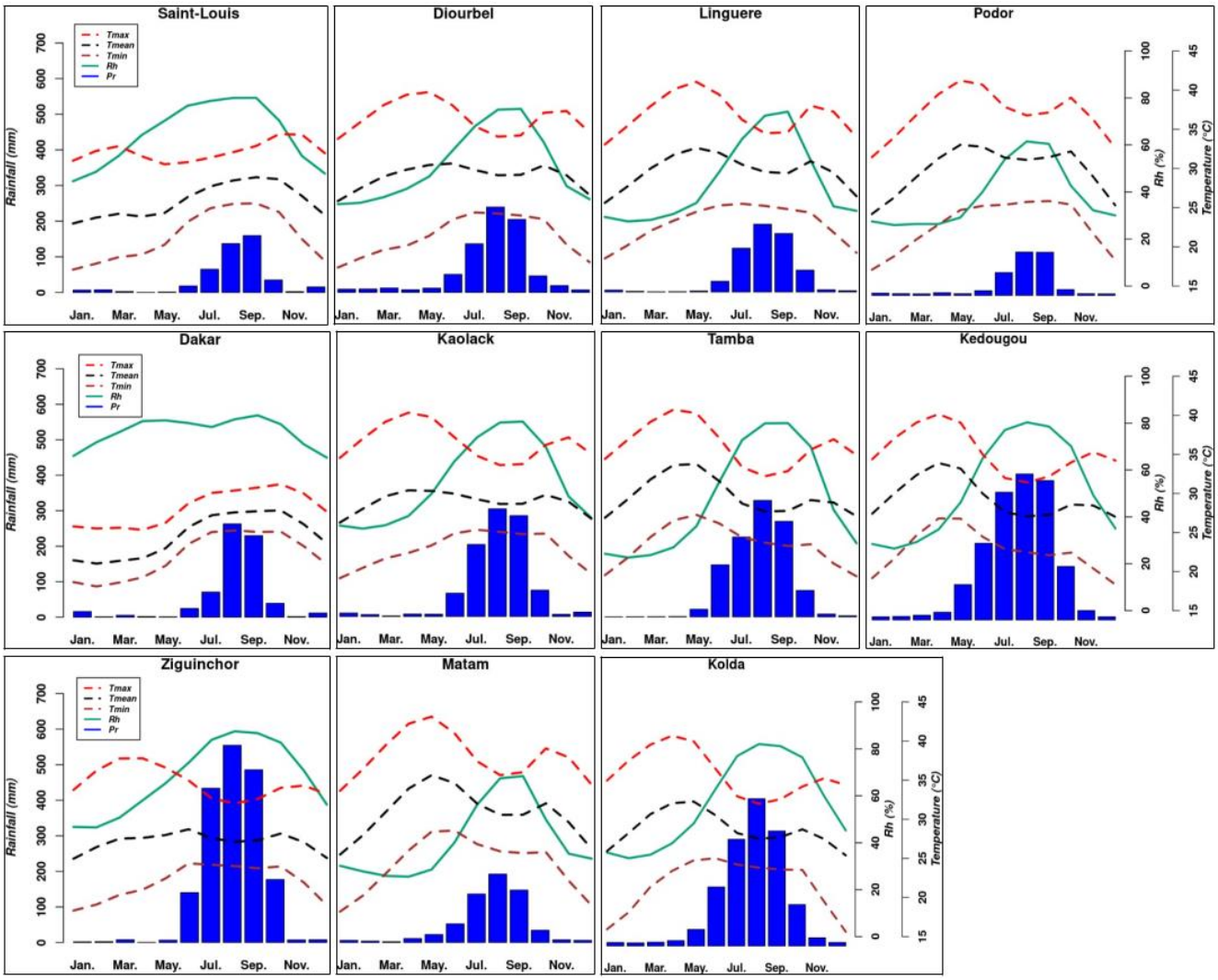

Figure 12. The mean seasonal cycle of monthly Tmax (red), Tmin (blue), Tmean (black), relative humidity (brown) and rainfall (green), for each of the 11 stations in Senegal. Missing data in the temperature and relative time series were first imputed by the missForest method, and the precipitation series by the imputeTS method.

noted at stations located in eastern Senegal, which show a strong signal over the MAM season. Relative humidity is also overestimated at the Dedougou station in MAM. In their study on extreme weather events, Chaney et al. (2014) indicate that despite the processing, errors are found in the GSOD data. These errors may be due to instrumentation errors and undetected station moves. As a result, further quality control tests of the climate data have been implemented to produce more reliable data [Hadley Center Integrated Surface Database (HadISD)] (Dunn et al., 2012).

Conclusion

In this study, observational data from the GSOD database was utilized, concentrating on synoptic stations classified by the World Meteorological Organization in Senegal and Burkina Faso. This data spans from 1973 to 2020. Notably, the dataset contains missing values, which have the potential to introduce biases into climate studies. High-quality meteorological data is indispensable for robustly monitoring and evaluating the impacts of climate change, including climate-related hazards and climate applications. Therefore, it is imperative to employ appropriate techniques for addressing missing data in these meteorological time series.

Thus, this study tested five widely used imputation methods, namely knn, ppca, imputeTS, mice, and missForest, under various missing data scenarios of 5, 10, 20, 30 and 40%. The performance of these models was compared using Taylor's diagram which simultaneously illustrates changes in the correlation coefficient, standard deviation and root mean square

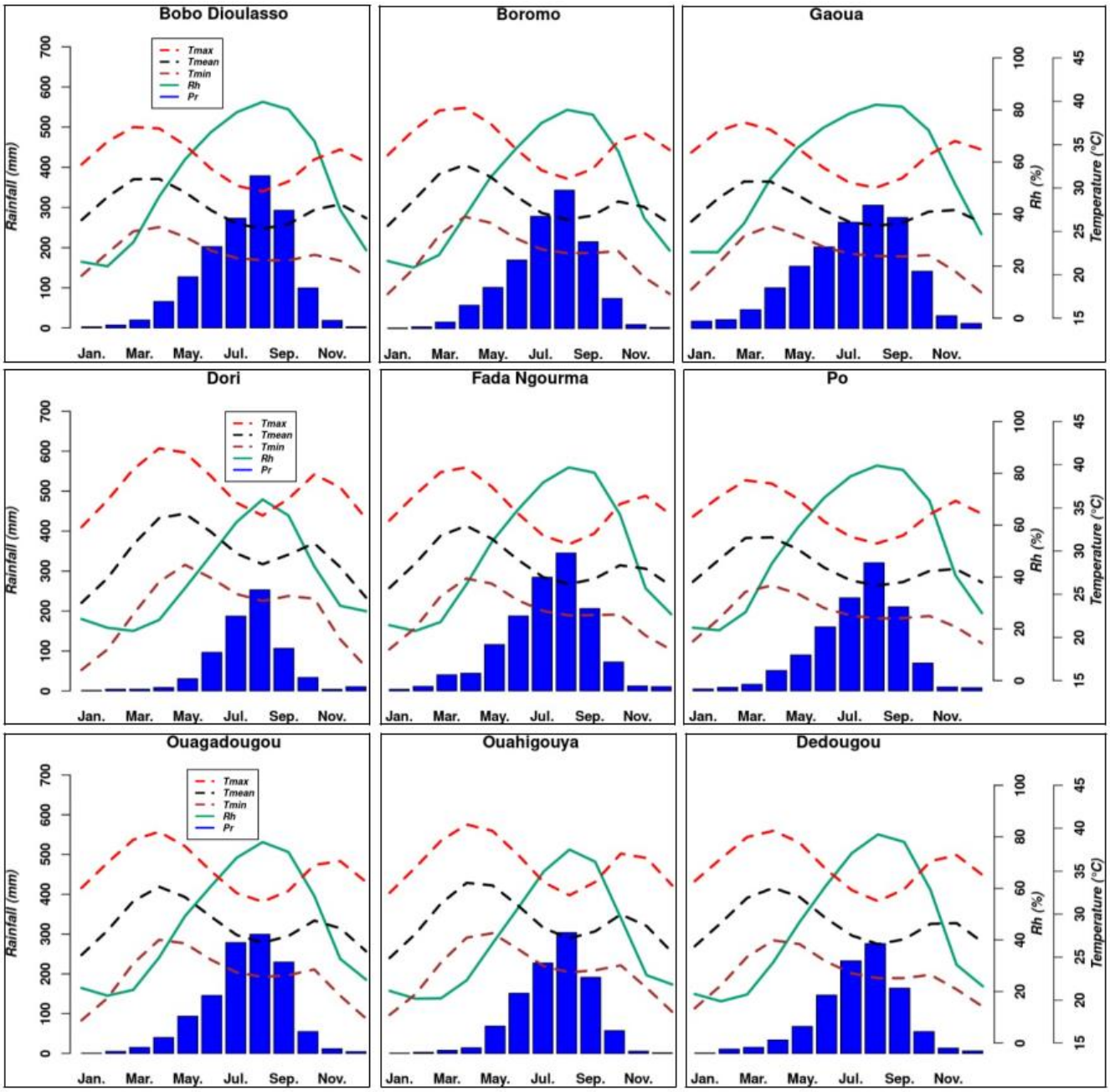

Figure 13. The mean seasonal cycle of monthly Tmax (red), Tmin (blue), Tmean (black), relative humidity (brown) and rainfall (green), for each of the 09 stations in Burkina Faso. Missing data in the temperature and relative humidity time series were first imputed by the missForest method, and the precipitation series by the imputeTS method.

error. These models were further compared using Kling-Gupta Efficiency which combines measures of correlation, bias, and variability.

The results revealed that the missForest method outperforms others in reconstructing temperature and relative humidity series while imputeTS is the preferred method for imputing rainfall time series. It is important to note that the performance of the imputation methods tends to decrease as the percentage of missing data increases. A closer examination of the daily evolution of the parameters after imputation reveals that missForest tends to overestimate rainfall during certain months of the year when climatic conditions are unfavorable for precipitation. However, when employing the best imputation method, the annual cycle representation closely aligns with the primary characteristics of the climate in our study area. A comparison was conducted between the ERA5 reanalysis and the observations from

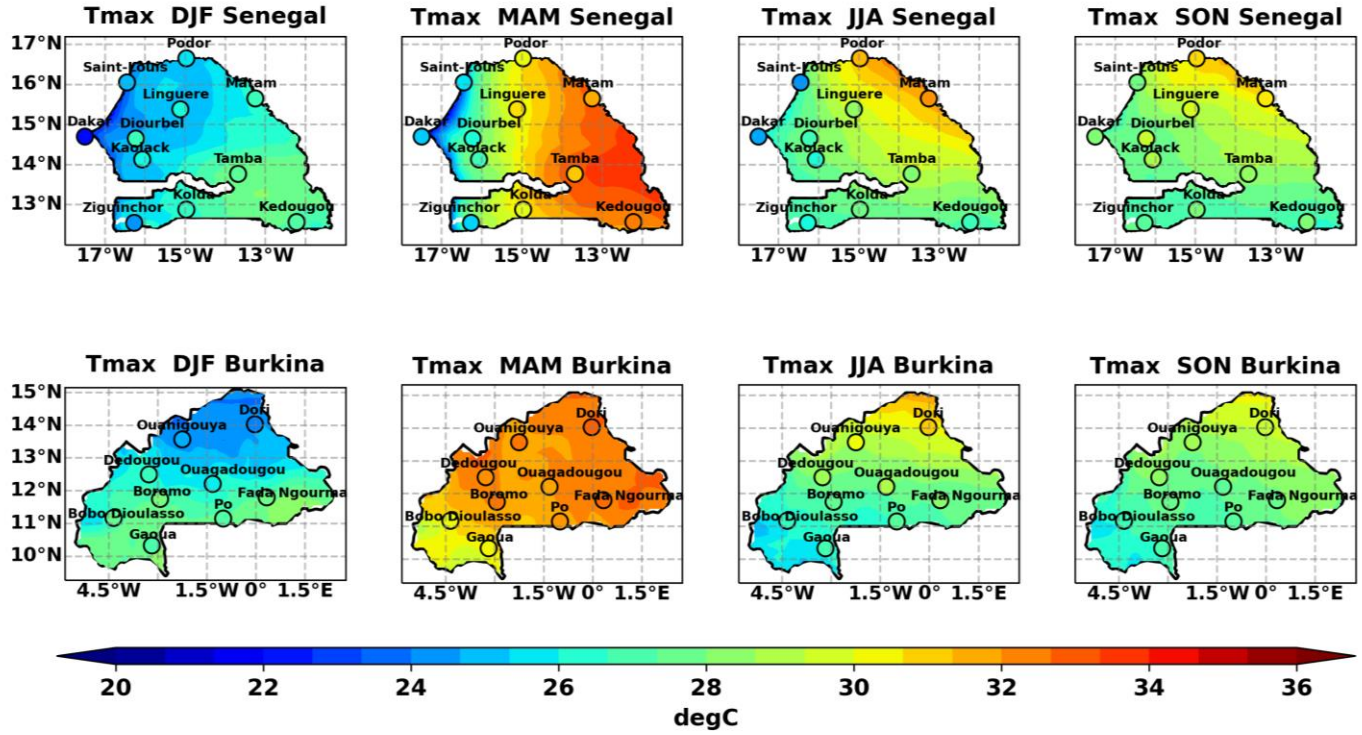

Figure 14. Spatial distribution of Tmax in ERA5 (imputed GSOD station data are represented by circles) over the season DJF, MAM, JJA and SON for Senegal (top) and Burkina Faso (bottom).

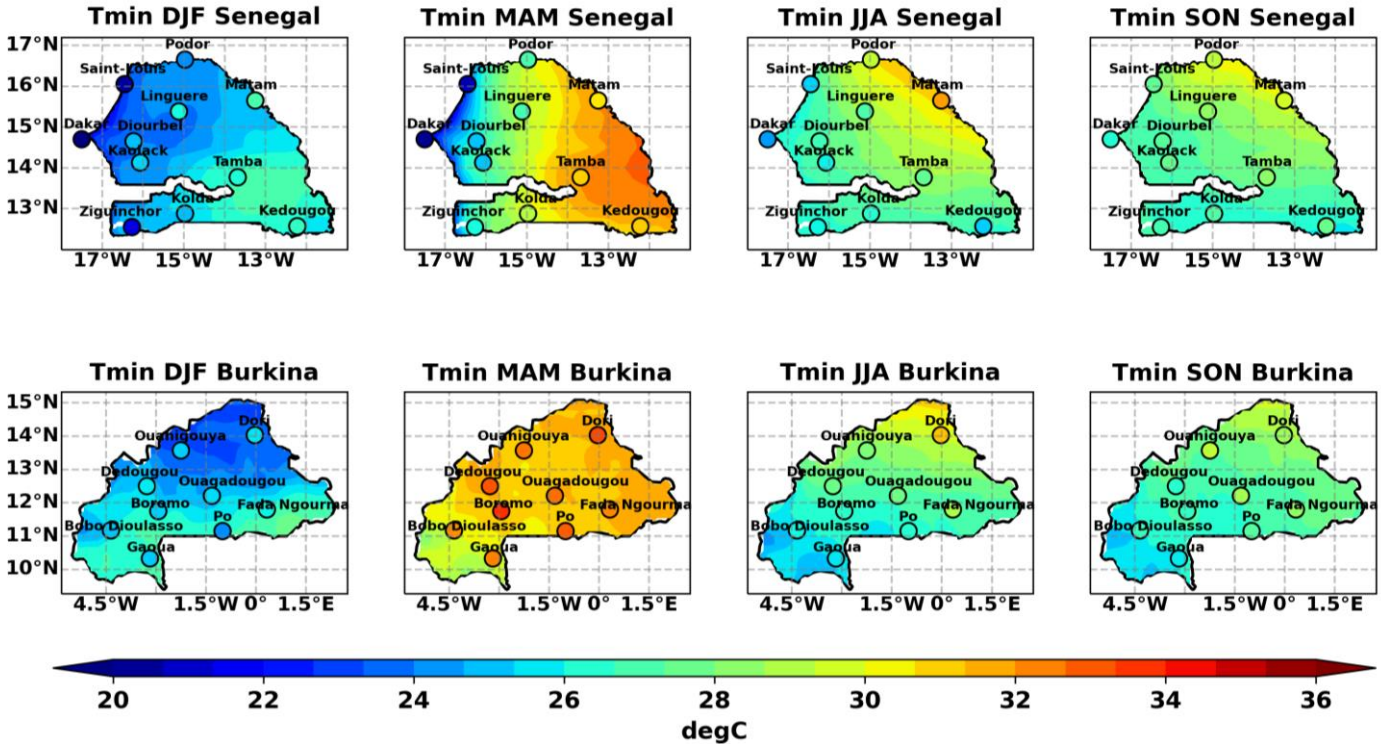

Figure 15. Spatial distribution of Tmin in ERA5 (imputed GSOD station data are represented by circles) over the season DJF, MAM, JJA and SON for Senegal (top) and Burkina Faso (bottom).

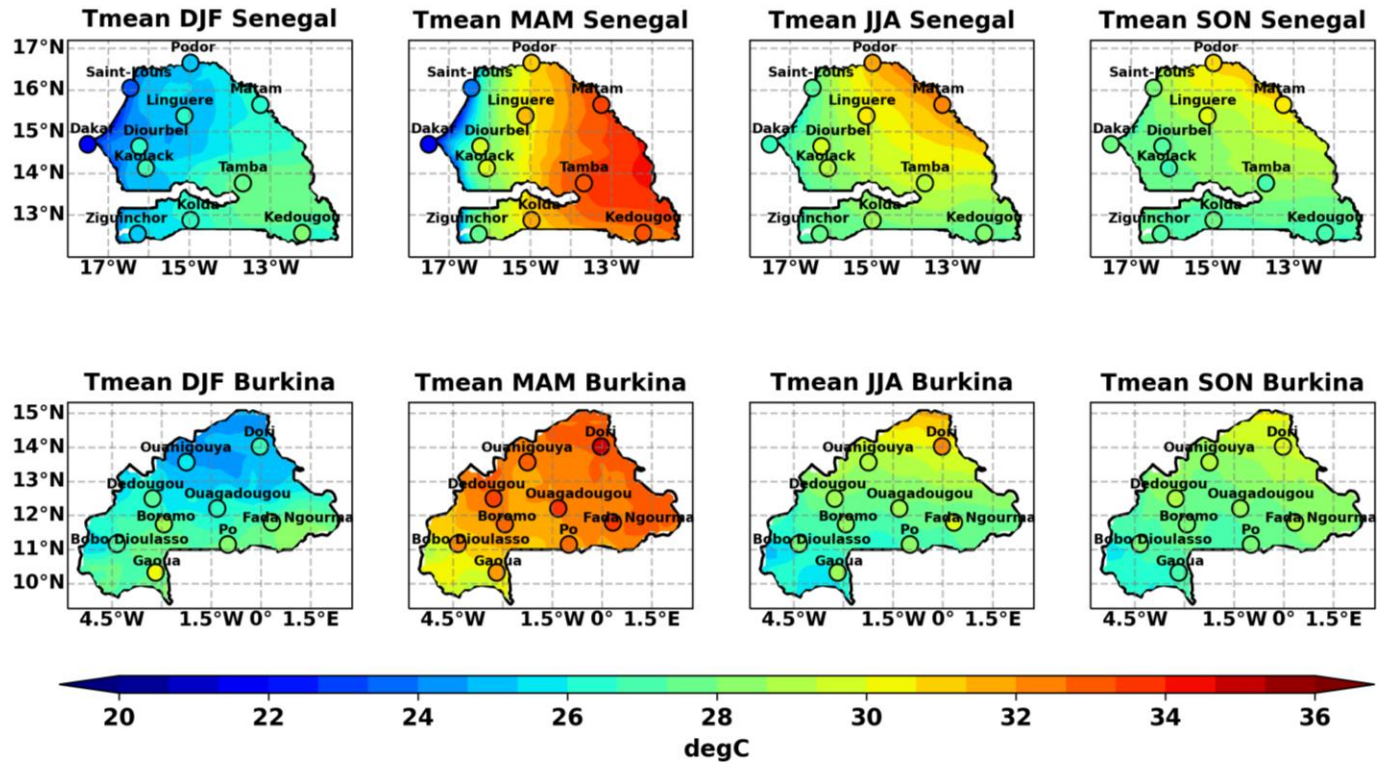

Figure 16. Spatial distribution of Tmean in ERA5 (imputed GSOD station data are represented by circles) over the season DJF, MAM, JJA and SON for Senegal (top) and Burkina Faso (bottom).

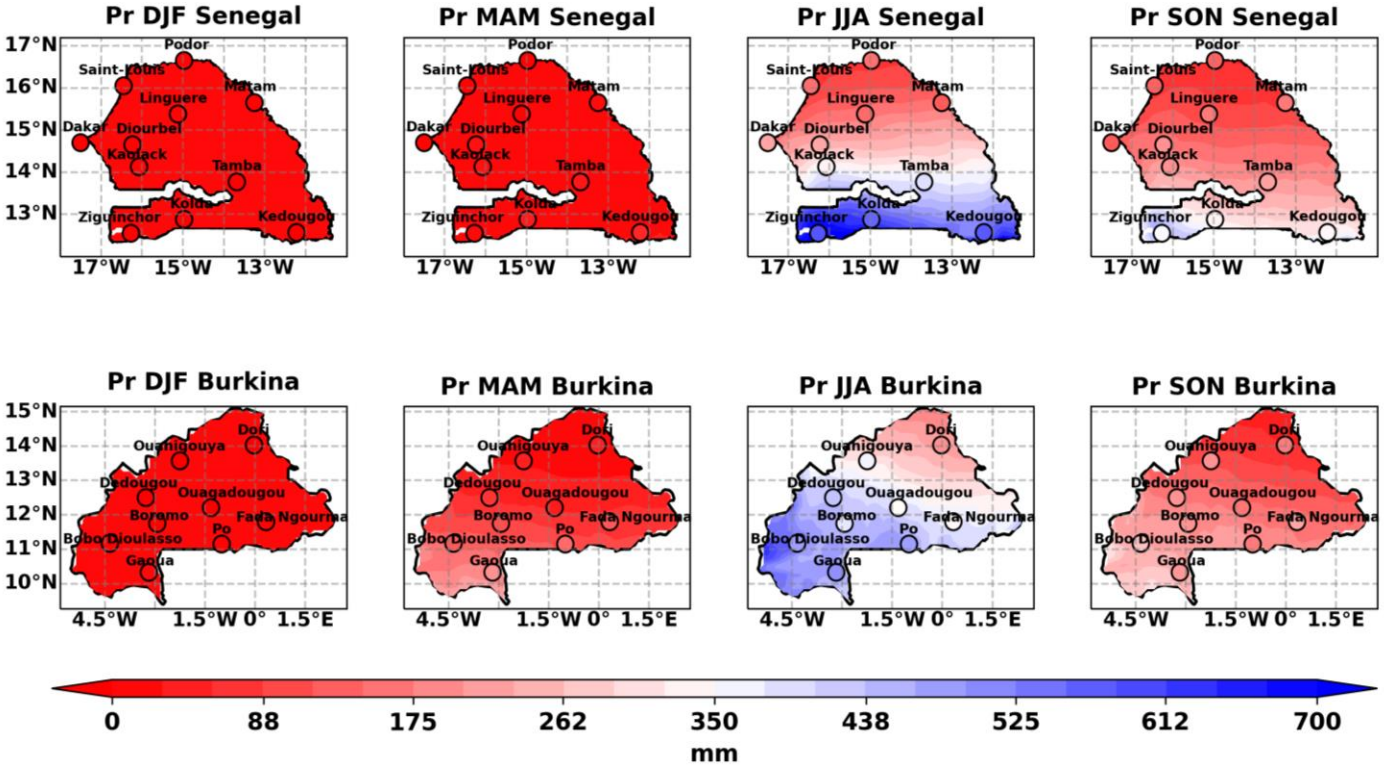

Figure 17. Spatial distribution of rainfall in ERA5 (imputed GSOD station data are represented by circles) over the season DJF, MAM, JJA and SON for Senegal (top) and Burkina Faso (bottom).

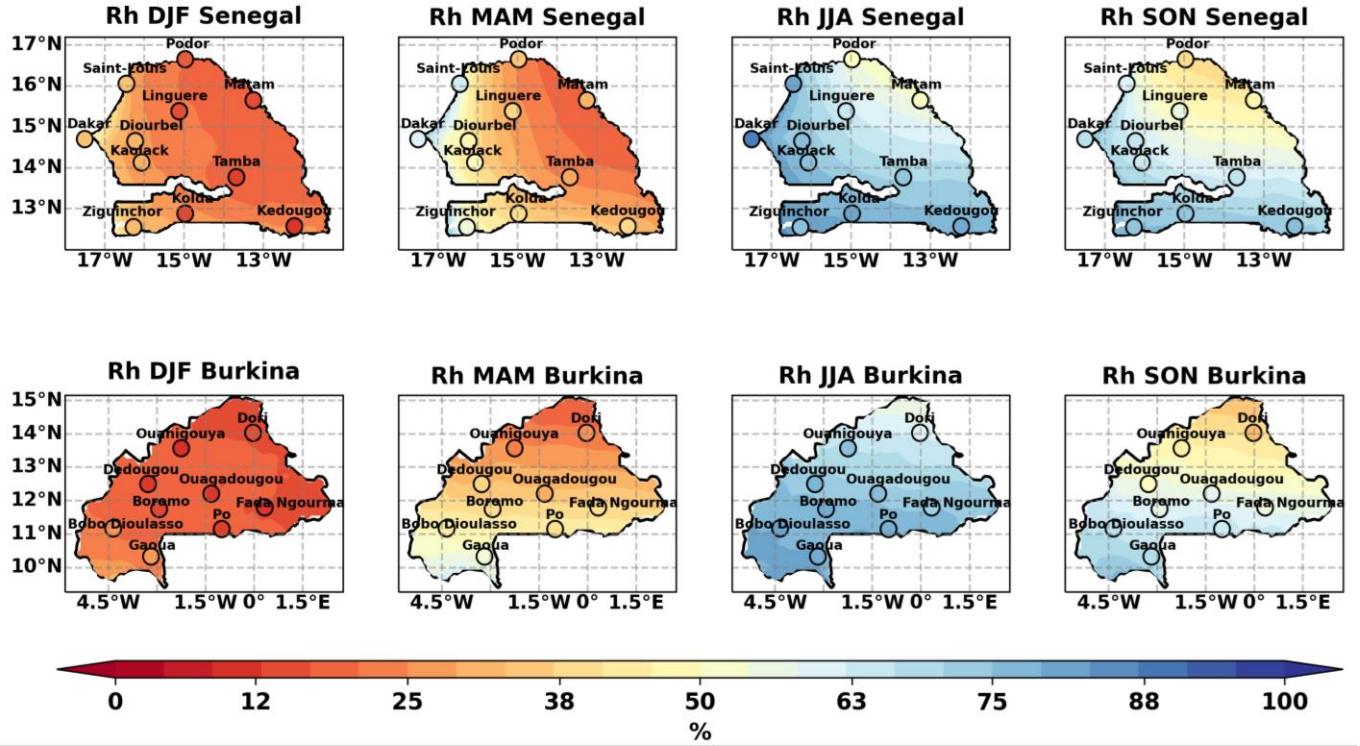

Figure 18. Spatial distribution of relative humidity in ERA5 (imputed GSOD station data are represented by circles) over the season DJF, MAM, JJA and SON for Senegal (top) and Burkina Faso (bottom).

a spatial representation of the climatic parameters of interest according to the seasons DJF, MAM, JJA and SON. This comparison revealed a very high degree of spatial consistency.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENT

The authors are grateful for of the ANR project ACASIS (2014–2018, grant: ANR-13-SENV-0007) and JEAI IRD program through JEAI-CLISAS (Jeune Equipe Associée à l'IRD Climat et santé au Sénégal).

REFERENCES

Awange JL, Ferreira VG, Forootan E, Andam-Akorful SA, Agutu NO, He XF (2016). Uncertainties in remotely sensed precipitation data over Africa. International Journal of Climatology 36(1):303-323.

Bao X, Zhang F (2012). Evaluation of NCEP/CFSR, NCEP/NCAR, ERA- Interim and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. Journal of Climate 26(1):206-214.

Barbier J, Guichard F, Bouniol D, Couvreux F, Roehrig R (2018). Detection of intraseasonal large-scale heat waves: Characteristics and historical trends during the Sahelian spring. Journal of Climate 31(1):61-80.

Beaulieu C, Ouarda TB, Seidou O (2007). Synthèse des techniques d'homogénéisation des series climatiques et analyse d'applicabilité aux séries de précipitations. Hydrological Sciences Journal 52(1):18-37.

Bousri I, Salah SA, Arab BM (2021). Validation d'une méthode d'imputation de données manquantes pour la reconstitution des séries de température. JAMA 5:28-32.

Buuren SV, Groothuis-Oudshoorn K (2011). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software 45:1-67.

Chaney NW, Sheffield J, Villarini G, Wood EF (2014). Development of a high-resolution gridded daily meteorological dataset over sub-Saharan Africa: Spatial analysis of trends in climate extremes. Journal of Climate 27(15):5815-5835.

Costa RL, Barros Gomes H, Cavalcante Pinto DD, da Rocha Júnior RL, dos Santos Silva FD, Barros Gomes H, Luís Herdies D (2021). Gap Filling and Quality Control Applied to Meteorological Variables Measured in the Northeast Region of Brazil. Atmosphere 12(10):1278.

Davey A, Shanahan MJ, Schafer JL (2001). Correcting for selective non response in the National Longitudinal Survey of Youth using multiple imputation. Journal of Human Resources, pp. 500-519.

Diallo FB (2018). Simulations multi-échelles couplées de la saisonnalité des vagues de chaleur et des pluies de mousson en Afrique de l'ouest. Thèse de Doctorat, Université Pierre et Marie Curie, Sorbonne, France.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Vitart F (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Journal of the Royal Meteorological Society 137(656):553-597.

De Longueville F, Hountondji YC, Kindo I, Gemenne F, Ozer P (2016). Long-term analysis of rainfall and temperature data in Burkina Faso (1950–2013). International Journal of Climatology 36(13):4393-4405.

Diouf S, Deme A, Deme EH (2022). Imputation methods for missing

values: the case of Senegalese meteorological data. African Journal of Applied Statistics 9(1):1245-1278.
Dixneuf P, Errico F, Glaus M (2021). A computational study on imputation methods for missing environmental data. arXiv preprint arXiv:2108.09500
Dunn RJ, Willett KM, Thorne PW, Woolley EV, Durre I, Dai A, Parker DE, Vose RS (2012). HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. Climate of the Past 8(5):1649-1679.
Easterling DR, Horton B, Jones PD, Peterson TC, Karl TR, Parker DE, Folland CK (1997). Maximum and minimum temperature trends for the globe. Science 277(5324):364-367.
Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009). Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology 377(1-2):80-91.
Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Thépaut JN (2020). The ERA5 Global Reanalysis, QJ Roy. Meteoritical Society 146 (730):1999-2049.
Josse J, Husson F (2016). missMDA: APackage for Handling Missing Values in Multivariate Data Analysis. Journal of statistical software 70:1-31.
Karl TR, Jones PD, Knight RW, Kukla G, Plummer N, Razuvayev V, Peterson TC (1993). A New Perspective on Recent Global Warming: Asymmetric Trends of Daily Maximum and Minimum Temperature. Bulletin of the American Meteorological Society 74(6):1007-1023.
Kertali F (2019). Étude de comblement de lacunes: Cas des séries pluviométriques observées du réseau de l'ONM. JAMA, pp. 49-58.
Kling H, Fuchs M, Paulin M (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology 424:264-277.
Kowarik A, Templ M (2016). Imputation with the R package VIM. Journal of Statistical Software 74(7):1-16.
Lotsi A, Asiedou L, Katsekpor J (2017). Comparison of Imputation Methods for Missing Values in Longitudinal Data Under Missing Completely at Random (MCAR) Mechanism. African Journal of Applied Statistics 4(1):241-258.
Melsen LA, Teuling AJ, Torfs PJ, Zappa M, Mizukami N, Mendoza PA, Clark MP, Uijlenhoet R (2019). Subjective modeling decisions can significantly impact the simulation of flood and drought events. Journal of Hydrology 568:1093-1104.
Moritz S, Bartz-Beielstein T (2017). imputeTS: Time Series Missing Value Imputation in R. The R Journal 9(1):207-218.
Moron V, Oueslati B, Pohl B, Rome S, Janicot S (2016). Trends of mean temperatures and warm extremes in Northern Tropical Africa (1961-2014) from observed and PPCA-reconstructed time series. Journal of Geophysical Research: Atmospheres 121(10):5298-5319.
Nakazawa T, Matsueda M (2017). Relationship between meteorological variables/ dust and the number of meningitis cases in Burkina Faso. Meteorological Applications 24(3):423-431.
Niass O, Diongue AK, Touré A (2015). Analysis of missing data in sereo-epidemiologic studies. African Journal of Applied Statistics 2(1):29-37.
Osuch M, Romanowicz RJ, Booij MJ (2015). The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. Hydrological Sciences Journal 60(7-8):1299-1316.
Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Woollen J (2011). MERRA-NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate 24(14):3624-3648.
Sagna P (2007). Caractéristique climatiques. Atlas du Sénégal, Paris, Les éditions J.A, pp. 66-69.
Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S, Goldberg M (2010). The NCEP Climate Forecast System Reanalysis. Bulletin of the American Meteorological Society 91(8):1015-1058.
Schafer JL, Graham JW (2002). Missing Data: Our View of the State of the Art. Psychological Methods 7(2):147-177.
Soltani K, Haouari M (2017). Reconstitution des séries mensuelles de températures maximales et minimales sur l'ouest Algérien. JAMA 1:83-87.
Stekhoven DJ (2011). Using the missForest package. R package pp. 1-11.
Stekhoven DJ, Bühlmann P (2012). Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1):112-118.
Taylor KE (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research Atmospheres 106:7183-7192.
Yozgatligil C, Aslan S, Iyigun C, Batmaz I (2013). Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. Theoretical and Applied Climatology 112(1-2):143-167.