Multi-variate infilling of missing daily discharge data on the Niger basin

Ganiyu Titilope Oyerindea,*, Agnide E. Lawina,b and Oluwafemi E. Adeyeric

a Graduate Research Program (GRP) Climate Change and Water Resources, West African Science Service Center on Climate Change and Adapted Land Use (WASCAL), University of Abomey-Calavi, Abomey-Calavi, BP 526 Cotonou, Benin
b Laboratory of Applied Hydrology, National Institute of Water, University of Abomey-Calavi, Abomey-Calavi, 01 BP 4521 Cotonou, Benin
c Department of Meteorology and Climate Science, Federal University of Technology, Akure, Nigeria

*Corresponding author. E-mail: ganiyuoyerinde@yahoo.com

ABSTRACT

The Niger basin have experienced historical drought episodes and floods in recent times. Reliable hydrological modelling has been hampered by missing values in daily river discharge data. We assessed the potential of using the Multivariate Imputation by Chained Equations (MICE) to estimate both continuous and discontinuous daily missing data across different spatial scales in the Niger basin. The study was conducted on 22 discharge stations that have missing data ranging from 2% to 70%. Four efficiency metrics were used to determine the effectiveness of MICE. The Flow Duration Curves (FDC) of observed and filled data were compared to determine how MICE captured the discharge patterns. Mann-Kendall, Modified Mann-Kendall, Pettit and Sen’s Slope were used to assess the complete discharge trends using the gap-filled data. Results shows that MICE near perfectly filled the missing discharge data with Nash-Sutcliffe Efficiency (NSE) range of 0.94–0.99 for the calibration (1992–1994) period. Good fits were obtained between FDC of observed and gap-filled data in all considered stations. All the catchments showed significantly increasing discharge trend since 1990s after gap filling. Consequently, the use of MICE in handling missing data challenges across spatial scales in the Niger basin was proposed.

Key words: hydrology, multiple imputation, river discharge, spatial scale, trend analysis, West Africa

HIGHLIGHTS

- Runoff data of the Niger basin has large amount of missing data.
- The quality of existing runoff data was improved by a multiple imputation technique.
- The effectiveness of the multiple imputation algorithm was influenced by the percentage of missing data.
- Trend analysis of gap-filled data shows significantly increasing trends from 1990s.
- Modified Mann-Kendall performs better than the original Man-Kendall test.

INTRODUCTION

West Africa have been ascribed with adverse climate change impacts. In past decades, the region experienced decline in food security due to warming, changing precipitation patterns, and greater frequency of some extreme events (IPCC 2019). Climate change has driven decreased discharge and increased drought in the Sahel since 1970, with observations showing that 1984 is the driest year on record (Biasutti 2019). High intensity rainfall and flood magnitudes was projected to increase in coming decades (Sylla et al. 2015; Aich et al. 2016). Increasing trend in the annual rainfall-runoff erosivity and soil loss are expected in the 21st century (Amanambu et al. 2019). Water resources are fundamental for several sectors such as hydropower, crop production and fisheries (Roudier et al. 2014). Sylla et al. (2018) disclosed that increase in temperature will lead to decrease in the potential to sustain large dams and irrigated agriculture in west Africa.

The Niger basin is ascribed with poorly documented historical hydrological data. There has been decrease in the amount of reliable rainfall stations since 1980 (Ali & Lebel 2009; Oyerinde et al. 2015) and discharge stations since 2000 (Schröder et al. 2019). These has been identified to be due to underfunding of data collection agencies, lack of technical capacity and commitment, inaccessibility to remote gauge stations due to logistical and security challenges and equipment malfunction (Ekeu-Wei et al. 2018). Radar altimetry measurements was found to aid in improving observed discharge data but generating reliable altimetry-discharge rating curves is still unresolved (Schröder et al. 2019). Temporal resolution of radar altimetry measurements due to satellite passing time also create challenges when there is consecutively gapped hydrological record (Ekeu-Wei et al. 2018). This has led
to insufficient research for development in water resources in west Africa (Vahid et al. 2015) and sometimes, contrasting future trends have been reported (Dosio et al. 2020); which makes policy making difficult.

Multiple imputation is now accepted as the best general method to deal with incomplete hydrological data (Little 1992; van der Heijden et al. 2006; van Buuren & Groothuis-Oudshoorn 2011; Ekeu-Wei et al. 2018; Sidibe et al. 2018). It was described as the method of choice for complex incomplete data problems (van Buuren & Groothuis-Oudshoorn 2011). Multiple Imputation performed better in filling missing data when compared to other methods such as: complete case analysis, available case methods, least squares on inputted data, maximum likelihood and Bayesian methods (Little 1992; van der Heijden et al. 2006). Ekeu-Wei et al. (2018) compared the performance of radar altimetry and a multiple imputation technique (MICE) on 5 hydrological stations located in the lower parts of the Niger basin (Nigeria). They concluded that MICE have potentials for ameliorating missing data challenges on the evaluated 5 catchments. These studies highlighted above have failed to assess the potential of multiple imputation across different scales on all parts of the Niger basin. The river basin has different flow regimes from upper, middle and lower parts (Oyerinde et al. 2017c) and it covers 10 countries. Therefore, this study assesses the efficacy of the Multiple imputation (MICE algorithm) in filling missing discharge data across different spatial scales on the whole Niger basin. The infilled data were subsequently used to assess complete discharge trends on 22 hydrological stations widely spread across the upper, middle and lower parts of the Niger basin. The objectives of the study are to:

1. Evaluate the efficacy of MICE in improving observed discharge data across different spatial scales on the Niger basin.
2. Estimate missing discharge data on 22 hydrological stations from 1980–2013.
3. Assess the effects of missing discharge data on hydrological trend analysis.

**Study area**

The Niger River Basin is the largest river basin in West Africa and it sustain livelihood of most countries in the Sahel. It covers 2.27 million km², with the active drainage area comprising less than 50% of the total basin (Ogilvie et al. 2010). The river has a length of 4,200 km and it is the third longest river in Africa. The basin has population of over 100 Million people based in ten countries (Algeria, Benin, Burkina Faso, Cameroon, Chad, Cote d’Ivoire, Guinea, Mali, Niger and Nigeria). The source of the Niger river is in the mountains of Guinea in an area with very high rainfall. Annual precipitation average over the whole Niger basin is 690 mm/year. This precipitation pattern varies across countries and regions. In the Sahelian parts, annual precipitation is about 280 mm while at Guinea parts precipitation is up to 1,635 mm. The average temperature ranges between 22 °C in the south to 27 °C in the northern parts. The basin topography shows high elevation up to 2,202 meters above sea level (MASL) in the Guinea mountains and lower elevation of 4 MASL at the basin exit into the ocean (Figure 1). Vegetation ranges from evergreen forests in the south to deserts in the Northern parts (Aich et al. 2016).

The Niger river flows Northeast from the source through the Upper Niger basin and enters the Inner Delta in Mali. During the rainy season, the delta forms a large flood plain of 20,000–30,000 km², facilitating the cultivation of rice, cotton and wheat as well as cattle herding and fishing (KIW 2010). The size of the flooded area is subject to strong annual variations, depending on the discharge of the Upper basin. A large part of the water is lost in the delta due to evaporation and seepage. Its main tributary, the Benue River, flows from highlands of Cameroon and joins the Niger at Lokoja, Nigeria, before reaching the Atlantic Ocean at the Gulf of Guinea (Oguntunde & Abiodun 2013). The Niger has an annual average flow of 1,005.83 m³/s (average 1980–2013) at Koulikoro (Mali) and up to 5,000 m³/s (average 1980–2013) close to the basin exit at Lokoja, Nigeria (Figure 1). The World Bank estimates that 30,000 gigawatt hours could be generated in the Niger River and its tributaries, but only 6,000 gigawatt hours have been developed so far. The Kainji Hydroelectric PLC (Figure 1) generate 22% of total hydroelectricity (KIW 2010). There are potential for increasing hydropower generation in the Niger basin.

**Data**

Daily river discharge data used in the study was obtained from the Niger River Basin Authority (NBA). We got data for 22 stations widely spread across the Niger River Basin from 1920 to 2013 (Figure 1). As shown in Table 1 and Figure 2, all the evaluated 22 hydrological stations were established at different times. Two stations...
Figure 1 | Spatial location of the 22 hydrological stations on the Niger Basin.

Table 1 | Length of available discharge data

| S/N | Stations          | Start | End  |
|-----|-------------------|-------|------|
| 1   | Koulikoro         | 1921  | 2013 |
| 2   | Douna             | 1923  | 2005 |
| 3   | Ansongo           | 1951  | 2013 |
| 4   | Couberi           | 1952  | 2013 |
| 5   | Kompongou         | 1952  | 2013 |
| 6   | Mopti             | 1952  | 2013 |
| 7   | Taoussa           | 1952  | 2013 |
| 8   | Yankin            | 1952  | 2013 |
| 9   | Malanville        | 1953  | 2013 |
| 10  | KeMacina          | 1954  | 2013 |
| 11  | Akka              | 1961  | 2007 |
| 12  | Dire              | 1961  | 2013 |
| 13  | Banakoro          | 1968  | 2013 |
| 14  | Alkongui          | 1980  | 2013 |
| 15  | Baro              | 1980  | 2013 |
| 16  | Garbey kourou     | 1980  | 2013 |
| 17  | Jidere            | 1980  | 2013 |
| 18  | Kakassi           | 1980  | 2013 |
| 19  | Kandadji          | 1980  | 2011 |
| 20  | Lokoja            | 1980  | 2013 |
| 21  | Makurdi           | 1980  | 2013 |
| 22  | Niamey            | 1980  | 2013 |
(Koulikoro and Dire) has data record since 1920s, 8 stations started in the 1950s, 3 stations were established in 1960s while 9 stations have records since 1980s. Clear visual assessments of intra and inter annual data quality shows most hydrological stations has good data from the 1980s (Figure 2). Thus, we chose the period of 1980–2013 for this study.

**METHODS**

**MICE Algorithm**

The MICE Algorithm was implemented on the R package MICE (van Buuren & Groothuis-Oudshoorn 2011). The multivariate imputation process involves three main steps: imputation, analysis and pooling that are well described by van Buuren & Groothuis-Oudshoorn (2011), Azur et al. (2011), and Sidibe et al. (2018). The method has been found to work well in varieties of hydrological applications and to perform better than other methods (Little 1992; Ekeu-Wei et al. 2018; Sidibe et al. 2018). In line with the study of Sidibe et al. (2018), missing values were estimated using the Predictive Mean Matching (PMM) approach. The missing data were simulated multiple times, in this case, five times, which is considered sufficient from the study of Ekeu-Wei et al. (2018).

The gap filled discharge data $Y$ is a partially observed random sample from the $p$-variate multivariate distribution $P(Y|\theta)$. We assume that the multivariate distribution of $Y$ is completely specified by $\theta$, a vector of unknown parameters. The MICE algorithm obtains the posterior distribution of $\theta$ by sampling iteratively from conditional distributions of the form

$$P(Y_1|Y_{-1}, \theta_1)$$

$$\vdots$$

$$P(Y_p|Y_{-p}, \theta_p)$$

The parameters $\theta_1, \ldots, \theta_p$ are specific to the respective conditional densities and are not necessarily the product of a factorization of the ‘true’ joint distribution $P(Y|\theta)$. Starting from a simple draw from observed marginal distributions, the $i$th iteration of chained equations is a Gibbs sampler that successively draws through its relation with other variables, and not directly. Convergence can therefore be quite fast, unlike many other MCMC methods. The name chained equations refers to the fact that the MICE algorithm can be easily implemented as a concatenation of univariate procedures to fill out the missing data (van Buuren & Groothuis-Oudshoorn 2011).
Calibration and validation of MICE gap filling method

MICE gap filling method was calibrated and validated at 3 discharge stations (Koulikoro, Dire and KeMacina). The 3 stations were selected because they have less than 5% missing data from 1980–2013 (Figure 3). The average amount of missing data on the 22 discharge stations were calculated as 27%. The same amount of missing data was artificially generated into the 3 discharge stations (Figure 4). Twenty percent discontinuous and random missing data were generated throughout the 3 discharge time series as shown in Figure 4. The remaining 7% missing gap was added through a continuous means from 1992–1994. Calibration was done by comparing the observed and inputted discharge data during the continuous missing years (1992–1994). Observed discharge data and Gap filled discharge data for the whole time series from 1980 to 2013 was used in data validation. Three efficiency metrics described below were used as an indicator of agreement between observed discharge and inputted discharge. The 3 metrics have been previously used in the Niger basin (Oyerinde et al. 2016, 2017b; Oyerinde & Diekkrüger 2017; Poméon et al. 2018).

Figure 3 | Percent missing data of 22 river discharge stations from 1980–2013.

a) Nash-Sutcliffe Efficiency \((\infty < NSE \leq 1)\) (Nash & Sutcliffe 1970) (NSE) is commonly used to assess the predictive power of river discharge. It is defined as:

\[
NSE = 1 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2}
\]  

(3)

Where \(O\) is the observed value and \(S\) is the predicted value at day \(i\). An efficiency of 1 corresponds to a perfect match between predicted and observations.

b) Kling-Gupta Efficiency (KGE) was developed by Gupta et al. (2009) to provide a diagnostically interesting decomposition of the NSE, which facilitates the analysis of the relative importance of its different components

\[
KGE = 1 - \sqrt{\frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2}}
\]
KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}

(4)

\beta = \frac{\mu_s}{\mu_o} \quad (5)

\gamma = \frac{CV_S}{CV_D} \quad (6)

r is the correlation coefficient between predicted and observed discharge (dimensionless), \beta is the bias ratio (dimensionless), \gamma is the variability ratio (dimensionless), \mu is the mean discharge in m$^3$/s, CV is the coefficient of variation (dimensionless). The KGE has its optimum at unity (Kling et al. 2012).
c) **Volumetric Efficiency (VE)** was proposed in order to circumvent some problems associated to the NSE which is not sensitive to differences in absolute discharge values. It represents the fraction of water delivered at the proper time; its compliment represents the fractional volumetric mismatch (Criss & Winston 2008).

\[
VE = 1 - \frac{\sum_{j=1}^{n} |S_j - O_i|}{\sum_{i=1}^{n} O_i}
\] (7)

**Flow duration curves (FDC)**

To evaluate the performance of the gap filling method, FDC was done at daily timestep between observed data and filled data. FDC have been widely applied in hydrological studies (Kling et al. 2012; Onyutha & Willems 2013; Abadzadesahraei & Sui 2016; Burgan & Aksoy 2020). It represents the relationship between the magnitude and frequency of daily, monthly (daily in this article) streamflow for a particular river basin. It provides an estimate of the percentage of time a given streamflow was equaled or exceeded over a historical period. FDC was used to provide a simple, yet comprehensive, graphical view of the overall historical variability associated with streamflow in a river basin. It is the complement of the cumulative distribution function (cdf) of daily streamflow. Each value of discharge Q has a corresponding exceedance probability p, and an FDC is simply a plot of the pth quantile or percentile of daily streamflow versus exceedance probability p, where p is defined by:

\[
p = 1 - P(Q \leq q)
\] (8)

\[
P = 1 - F_Q(q)
\] (9)

**Trend analysis**

**Mann-Kendall test (MK)**

To analyze for increasing/decreasing trends of river discharge, the Mann-Kendall test was employed. Mann-Kendall test evaluates the relative magnitudes of data trend and it is widely used in hydrology (Verstraeten et al. 2006; Meusburger et al. 2012; N’Tcha M’Po et al. 2017; Oyerinde et al. 2017a). The advantage of this test is that the data need not comply with any particular distribution. Mann-Kendall statistic (S) was computed as:

\[
S = \sum_{k-1}^{n} \sum_{j=k+1}^{n} \text{sign}(x_j - x_k)
\] (10)

\[
\text{sign}(x_j - x_k) = 1 \text{ if } x_j - x_k > 0
\]

\[
= 0 \text{ if } x_j - x_k = 0
\]

\[
= -1 \text{ if } x_j - x_k < 0
\] (11)

where \(x_j \) and \(x_k \) are the annual values in years \(j \) and \(k \), respectively and \(n \) is the number of years.

For an independent data sample without tied values, the mean and variance of \(S \) are given by:

\[
E[S] = 0
\] (12)

\[
\text{Var}(S) = \frac{n(n-1)(2n+5)}{18}
\] (13)

If tied values are present in the sample, \(\text{Var}(S) \) is computed by:

\[
\text{Var}(S) = \left[ n(n-1)(2n+5) - \sum_{i=1}^{n} t_i(i-1)(2i+5) \right] / 18
\] (14)
Then, the MK test statistic $Z$ for all those cases where $n$ is larger than 10 is given by:

$$Z = \begin{cases} \frac{S - 1}{\sigma} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S + 1}{\sigma} & \text{if } S < 0 \end{cases}$$

(15)

The trend was dignified as “no trend” when the change is not significant, ‘an increasing or a decreasing trend’ when $S$ is positive or negative respectively. Similarly, if $Z > 0$, it indicates an increasing trend, and vice versa.

**Modified Mann-Kendall test by variance correction (MMK)**

Study of Yue & Wang (2004) demonstrated that serial correlation in time series alters the variance of the MK statistic. Their study was able to use modification of variance to limit the effect of serial correlation. In addition, Chen et al. (2016) recommended the combine use of the MK and MMK tests when there are autocorrelations above lag 1. Due to these reasons, we assessed and compared the autocorrelation of observed and gap-filled discharge data. The results of the MK statistics before and after variance corrections of gap-filled data by MMK were also compared. Modified variance $V(S)^*$ for computing the MK statistic $Z$ is calculated as:

$$V(S)^* = V(S) \cdot \frac{n}{n^*}$$

(16)

where $n$ is the actual sample size of actual sample data, $n^*$ is the effective sample size, and $n/n^*$ is termed the correction factor.

**Sen’s slope**

Besides the magnitude of a time series, trend was evaluated by a simple non-parametric procedure developed by Sen (Sen 1968; Ali et al. 2019). The trend is calculated by:

$$\beta = \text{Median} \left( \frac{X_j - X_i}{j - i} \right), \ j > i$$

(17)

where $\beta$ is Sen’s slope estimate. $\beta > 0$ indicates upward trend in a time series. Otherwise the data series presents downward trend during the time period (Ali et al. 2019).

**Pettit test**

To evaluate the difference between cumulative distribution function before and after a time instant (K), the Pettit test was applied. Pettit test detects any significant change in the mean value in a time series (Kliment et al. 2011). The non-parametric Pettit rank test was reported to have good capabilities in handling outliers (Pettitt 1979; Verstraeten et al. 2006) and it was previously used in hydrological studies in the region (Nka et al. 2015; Oyerinde et al. 2017a). The significance of the analyzed trends in the dataset was tested at probability level $p \leq 0.05$ to show 99.5% experimental precision. It tests the $H_0$: The $T$ variables follow one or more distributions that have the same location parameter (no change), against the alternative: a change point exists. The non-parametric statistic is defined as:

$$K_T = \max |U_{t,T}|$$

(18)

where:

$$U_{t,T} = \sum_{i=1}^{T} \sum_{j=t+1}^{T} \text{sgn}(X_i - X_j)$$

(19)

The change-point of the series is located at $K_T$, provided that the statistic is significant.
### Table 2 | Efficiency metrics during calibration and validation of MICE gap filling method

|                | Calibration (Missing Years 1992–1994) | Validation (All the time Series from 1980–2013) |
|----------------|--------------------------------------|--------------------------------------------------|
|                | Koulikoro | Dire | KeMacina | Koulikoro | Dire | KeMacina |
| NSE            | 0.94      | 0.99 | 0.94     | 0.99      | 0.99 | 0.99     |
| KGE            | 0.96      | 0.98 | 0.94     | 0.99      | 1.00 | 0.99     |
| VE             | 0.82      | 0.93 | 0.80     | 0.96      | 0.98 | 0.96     |

### Figure 5 | Complete vs filled data at the 3 discharge stations.
RESULTS

Calibration and validation of MICE gap filling method

The MICE gap filling method was evaluated at three discharge stations (Koulikoro, KeMacina and Dire) which has less than 5% missing data (Figure 3). There was high efficiency metrics between the inputted data and observed data during both continuous missing years (1992–1994) and the randomly created missing data in all the time series from 1980 to 2013 (Table 2).

From Figure 5, visual assessments show that the MICE gap filling method well reproduced the observed seasonality of flow at the 3 discharge stations. The observed low and high flows were well reproduced in the gap filled discharge data.

Flow duration curves (FDC)

FDCs were plotted for all the 22 hydrological stations to show the fitness between the gap-filled and observed discharge data (Figure 6–8). The FDC of inputted and observed data nearly indicated a perfect fit at all stations. Figure 6 present FDCs that shows good fitness of observed and inputted data at selected stations on the main Niger river at the Upper Parts (Banakoro (11% missing data)), Inland Delta Parts at Mopti (7% missing data), Middle parts at Malanville (34% missing data) and the lower parts at Lokoja (19% missing data). From Figure 7,
there was an excellent fit between the FDC of inputted and observed discharge data at selected head water catchments at Couberi (30% missing data), Makurdi (35% missing data), Douna (30% missing data) and Kakassi (45% missing data). We assessed the FDC of four hydrological stations with highest percentage of missing data as shown in Figure 8 (Kompongou (72% missing data), Garbey Kourou (50% missing data), Baro (46% missing data) and Alkongui (63% missing data)). The pattern of flow of the observed was adequately reproduced by the inputted, although there was slight over estimation at Baro while slight underestimation was observed at Kompongou, Garbey Kourou and Alkongui.

Discharge trends

Figure 9 shows the autocorrelation plots for filled and observed discharge data of the 22 catchments using lags 1 to 10. Observed discharge data has very high autocorrelation values that ranges from 0.67 to 1.00 for all the 22 discharge stations. The gap-filled data however has decreased and more diverse autocorrelation values (∼0.2 to 0.77). We compared the MK tests statistics before and after filling the missing data (Table 3). Missing data decreased the magnitude of S and Z statistics due to decreased sample size while this problem was corrected in the gap filled MK test. Sen’s slope p value indicates that six discharge stations with high amounts of missing data (Alkongui (63%), Banakoro (11%), Couberi (30%), Douna (30%), Kompongou (70%) and Makurdi (35%)) shows no significant trend.
before gap filling. However, all the six catchments show significant increases after gap-filling. The significant levels and magnitude of the Sen’s slope were improved across all the 22 catchments after gap filling.

We compared results obtained from MK and MMK tests on gap-filled data of the 22 catchments and the results are shown in Table 4. MMK made substantial improvement of the Z and variance. The MK has problem with variance computation due to autocorrelation after gap-filling (Figure 9). Constant variance value of 4550.33 was computed by MK across the 22 stations while MMK was able to correct the problem.

Rapidly increasing trendline was observed in the graphical presentation shown in Figure 10. Pettit tests presented in Table 5 shows that all the 22 catchments witnessed significant increasing river discharge with break points of 14 catchments at 1993, 4 catchments had breaks at 1994, 3 catchments at 1997 and 1 at 1987 (Table 5).

Figure 11 presents the spatial map of the Sen’s slope on 22 stations on the Niger basin after gap-filling. Most (15) of the discharge stations showed significant increasing discharge at $p \leq 0.001$. Five discharge stations have been increasing since 1980s at $p \leq 0.01$ while 2 stations increase at $p \leq 0.05$.

**DISCUSSION**

The study shows that missing data is a major challenge in the Niger basin. Twenty-one discharge stations have missing values and incomplete records. This was due to the earlier discussed observed decrease in rainfall and
discharge observation stations in west Africa (Lebel & Ali 2009; Schröder et al. 2019). The problem of missing data often hinders precise forecast of extreme events and sustainable policy interventions. Ekeu-Wei et al. (2018) disclosed that incomplete hydrological data hinders flood management in many developing regions. Researchers

![Figure 9](https://iwaponline.com/wpt/article-pdf/doi/10.2166/wpt.2021.048/894272/wpt2021048.pdf)

**Figure 9** | Auto-correlation (ACF) Plots of the Observed and Gap-filled discharge data.

**Table 3** | Effects of missing data on Man-Kendall (MK) and Sen’s slope statistics.

| Trend Tests | Test Statistics | Observed | | | | Filled | | | |
|-------------|-----------------|---------|--|---|---------|---------|--|---|---|
| Test MK test | Sen’s Slope | p.value | Sig.Level | | | Test MK test | Sen’s Slope | p.value | Sig.Level |
| Trend Tests | Test Statistics | S | Z | p.value | Sig.Level | S | Z | p.value | Sig.Level |
| Akka | 134 | 2.63 | 14.28 | 8.6E-03 | 0.01 | 237 | 3.50 | 13.13 | 4.7E-04 | 0.001 |
| Alkongui | 46 | 0.84 | 0.32 | 4.0E-01 | NS | 177 | 2.61 | 0.51 | 9.1E-03 | 0.01 |
| Ansongo | 231 | 3.41 | 13.39 | 6.5E-04 | 0.001 | 277 | 4.09 | 11.25 | 4.3E-05 | 0.001 |
| Banakoro | 59 | 0.86 | 4.58 | 3.9E-01 | NS | 141 | 2.08 | 7.80 | 3.8E-02 | 0.05 |
| Baro | 133 | 2.91 | 62.99 | 3.6E-03 | 0.01 | 305 | 2.61 | 0.51 | 9.1E-03 | 0.01 |
| Coubéri | 103 | 1.82 | 0.52 | 6.9E-02 | NS | 189 | 2.79 | 0.51 | 5.3E-03 | 0.01 |
| Dire | 215 | 3.17 | 10.01 | 1.5E-03 | 0.01 | 253 | 3.74 | 10.58 | 1.9E-04 | 0.001 |
| Douna | 41 | 0.88 | 2.95 | 3.8E-01 | NS | 207 | 3.05 | 5.77 | 2.3E-03 | 0.01 |
| Garbé_y_kourou | 168 | 2.71 | 1.80 | 6.8E-01 | NS | 219 | 3.23 | 1.20 | 1.2E-03 | 0.01 |
| Jidere | 209 | 3.08 | 13.24 | 2.0E-03 | 0.01 | 271 | 4.00 | 16.23 | 6.3E-05 | 0.001 |
| Kakassi | 203 | 3.60 | 1.09 | 3.1E-04 | NS | 225 | 3.32 | 0.56 | 9.0E-04 | 0.01 |
| Kandadjì | 202 | 3.26 | 10.80 | 1.1E-03 | 0.01 | 285 | 4.21 | 11.67 | 2.6E-05 | 0.001 |
| KeMacina | 171 | 2.52 | 9.79 | 1.2E-02 | 0.05 | 185 | 2.75 | 10.89 | 6.4E-03 | 0.01 |
| Kompongou | 24 | 0.95 | 0.62 | 3.4E-01 | NS | 223 | 3.29 | 0.47 | 1.0E-03 | 0.01 |
| Koulikoro | 191 | 2.82 | 12.32 | 4.9E-03 | 0.01 | 191 | 2.82 | 12.32 | 4.9E-03 | 0.01 |
| Lokoja | 191 | 2.82 | 87.71 | 4.9E-03 | 0.01 | 261 | 3.85 | 85.51 | 1.2E-04 | 0.001 |
| Makurdi | 95 | 1.68 | 46.03 | 9.4E-02 | NS | 221 | 3.26 | 31.03 | 1.1E-03 | 0.01 |
| Malanville | 117 | 1.97 | 10.59 | 4.9E-02 | 0.05 | 243 | 3.62 | 11.92 | 3.0E-04 | 0.001 |
| Mopti | 128 | 1.97 | 8.70 | 4.9E-02 | 0.05 | 227 | 3.35 | 11.52 | 8.1E-04 | 0.001 |
| Niamey | 305 | 4.51 | 12.19 | 6.6E-06 | 0.001 | 317 | 4.68 | 13.38 | 2.8E-06 | 0.001 |
| Taoussa | 221 | 3.26 | 18.73 | 1.1E-03 | 0.01 | 253 | 3.74 | 9.62 | 1.9E-04 | 0.001 |
| Yankin | 100 | 2.31 | 1.13 | 2.1E-02 | 0.05 | 227 | 3.35 | 0.70 | 8.1E-04 | 0.001 |
Table 4 | Comparison of Man-Kendall (MK) and modified Man-Kendall (MMK) tests for the 22 discharge stations

| Station       | N/N* | MK Z | MMK Z | MK Variance | MMK Variance |
|---------------|------|------|-------|-------------|--------------|
| Akka          | 2.21 | 3.50 | 2.35  | 4,550.33    | 10,053.99    |
| Alkongui      | 0.64 | 2.61 | 3.27  | 4,550.33    | 2,899.36     |
| Ansongo       | 3.44 | 4.09 | 2.21  | 4,550.33    | 15,641.67    |
| Banakoro      | 2.18 | 2.08 | 1.41  | 4,550.33    | 9,908.12     |
| Baro          | 1.65 | 4.51 | 3.51  | 4,550.33    | 7,496.05     |
| Couberi       | 0.71 | 2.79 | 3.30  | 4,550.33    | 3,245.08     |
| Dire          | 2.85 | 3.74 | 2.21  | 4,550.33    | 12,947.53    |
| Douna         | 1.07 | 3.05 | 2.96  | 4,550.33    | 4,854.94     |
| Garbey kourou | 0.76 | 3.23 | 3.71  | 4,550.33    | 3,445.16     |
| Jidere        | 1.53 | 4.00 | 3.23  | 4,550.33    | 6,971.60     |
| Kakassi       | 2.25 | 3.32 | 2.21  | 4,550.33    | 10,260.67    |
| Kandadji      | 2.90 | 4.21 | 2.47  | 4,550.33    | 13,201.15    |
| KeMacina      | 1.69 | 2.73 | 2.10  | 4,550.33    | 7,708.06     |
| Kompongou     | 0.94 | 3.29 | 3.40  | 4,550.33    | 4,262.18     |
| Koulikoro     | 1.72 | 2.82 | 2.15  | 4,550.33    | 7,835.35     |
| Lokoja        | 1.26 | 3.85 | 3.43  | 4,550.33    | 5,732.18     |
| Makurdi       | 1.53 | 3.26 | 2.64  | 4,550.33    | 6,948.59     |
| Malanville    | 1.92 | 3.62 | 2.61  | 4,550.33    | 8,736.60     |
| Mopti         | 1.34 | 3.35 | 2.89  | 4,550.33    | 6,103.43     |
| Niamey        | 2.18 | 4.68 | 3.17  | 4,550.33    | 9,927.89     |
| Taoussa       | 3.44 | 3.74 | 2.02  | 4,550.33    | 15,635.95    |
| Yankin        | 0.66 | 3.35 | 4.12  | 4,550.33    | 3,015.10     |

Figure 10 | Annual mean discharge trends at 22 stations of the Niger basin.
Table 5 | Results of the Pettit test for the 22 discharge stations

| S/N | Stations       | Year | K       | P       | Significant Levels | n |
|-----|----------------|------|---------|---------|--------------------|---|
| 1   | Akka           | 1993 | 7.59E-05| 0.001   |                    | 34|
| 2   | Alkongui       | 1997 | 4.49E-02| 0.05    |                    | 34|
| 3   | Ansongo        | 1994 | 4.37E-05| 0.001   |                    | 34|
| 4   | Banakoro       | 1993 | 1.89E-04| 0.001   |                    | 34|
| 5   | Baro           | 1993 | 2.53E-04| 0.001   |                    | 34|
| 6   | Coubéri        | 1993 | 1.64E-02| 0.05    |                    | 34|
| 7   | Dire           | 1993 | 4.04E-05| 0.001   |                    | 34|
| 8   | Douna          | 1993 | 1.98E-03| 0.01    |                    | 34|
| 9   | Garbey kourou  | 1997 | 2.89E-03| 0.01    |                    | 34|
| 10  | Jidere         | 1993 | 5.55E-05| 0.001   |                    | 34|
| 11  | Kalkassi       | 1987 | 8.45E-03| 0.01    |                    | 34|
| 12  | Kandadji       | 1994 | 2.29E-05| 0.001   |                    | 34|
| 13  | KeMacina       | 1993 | 4.50E-04| 0.001   |                    | 34|
| 14  | Kompongou      | 1993 | 2.25E-03| 0.01    |                    | 34|
| 15  | Koulikoro      | 1993 | 1.40E-04| 0.001   |                    | 34|
| 16  | Lokoja         | 1993 | 3.38E-04| 0.001   |                    | 34|
| 17  | Makurdi        | 1993 | 8.98E-04| 0.001   |                    | 34|
| 18  | Malanville     | 1993 | 8.98E-04| 0.001   |                    | 34|
| 19  | Mopti          | 1993 | 1.20E-04| 0.001   |                    | 34|
| 20  | Niamey         | 1994 | 4.37E-05| 0.001   |                    | 34|
| 21  | Taouessa       | 1994 | 5.13E-05| 0.001   |                    | 34|
| 22  | Yankin         | 1993 | 4.18E-03| 0.01    |                    | 34|

Figure 11 | Spatial Sen’s Slope and Significant Levels; *** Significant at p ≤ 0.001, ** Significant at p ≤ 0.01 and * Significant at p ≤ 0.05.
such as Badou et al. (2016) have to exclude some important locations where inadequate data are recorded. These drives incomplete important information that will be useful for sustainable river basin management.

MICE gap filling method shows promising results on 22 river discharge stations widely spread on the Niger Basin. Gap filled discharge data have high efficiency metrics when compared with observed during both continuous and discontinuous missing discharge time series. These findings are all in agreements with previous studies where MICE was evaluated for filling missing river discharge data (Little 1992; Ekeu-Wei et al. 2018; Sidibe et al. 2018). Little (1992) reported that the use of regression based approaches in filling missing data doesn’t account for errors in the imputations thus leading to small standard errors. MICE overcome this challenge by drawing values from the predictive distribution and then repeating complete-data analyses. Ekeu-Wei et al. (2018) found out that MICE outperformed radar altimetry when filling continuous missing data and MICE filled data gives good similarity to natural floods. Sidibe et al. (2018) disclosed that MICE better estimate missing data more than the random forest at discharge stations with high amount of missing data at West and Central Africa. The use of hydrological models in simulating missing discharge data has been promising but poor observed meteorological data required for hydrological model calibration and validation have made setting up such models difficult (Poméon et al. 2017).

We evaluated the performance of MICE on all the 22 catchments by comparing FDC of inputted data with the FDC of observed. There were perfect fits on all the 22 stations evaluated on the river basin and MICE captures both low and high flows on the basin. These will enable MICE enhanced hydrological records to be applicable in extreme hydrological applications (Ekeu-Wei et al. 2018). FDC of stations with high amount of missing data captured the observed river discharge patterns with slight under and over-estimation. This shows that MICE should be applied with precautions in stations with daily missing data percentage that is more than 70%. Barnes et al. (2006) found out that small sample size constrains the generalization potential of the MICE, thus resulting in uncertain missing data estimates.

Gap-filling of missing discharge data with MICE decreased autocorrelation and significantly improved the MK statistics and Sen’s Slope. MK test has been reported to perform worse when there is autocorrelation (Chen et al. 2016). Yue & Wang (2004) have demonstrated that the existence of negative serial correlation will decrease the possibility of rejecting the null hypothesis of no trend. We compared the performance of the MK and MMK statistics and discovered that MK had challenges at variance computation on the Niger basin. This happens because negative serial correlation reduces the variance of the MK statistics, and hence a smaller number of samples falls in the critical regions (Yue & Wang 2004).

All the 22 discharge stations evaluated on the Niger basin shows significant increasing trend since 1980s with a break point at the 1990s for most of the stations. This result corroborates the findings of Amogu et al. (2010) who attributed the increases in discharge in Sahel catchments to land use changes. Descroix et al. (2018) also attributed the increases in discharge from 1990s in the Sahel to recovery from the great drought of West Africa (1968–1990). The authors also explained that increasing discharge is due to lower soil infiltration rate compared to the pre-drought era (Bichet & Diedhiou 2018; Descroix et al. 2018). Some authors found direct relationship among the trends in discharge, flood and some extreme rainfall indices (Nka et al. 2013; Adeyeri et al. 2019). They attributed the rainfall factors as major factor aggravating increase in runoff coefficients in the Sahelian region (Nka et al. 2015).

**CONCLUSIONS**

The challenge of missing discharge data has hindered reliable flow prediction and forecast in West Africa thus, hindering sustainable river basin management. This study assessed the multiple benefits of MICE gap filling method in ameliorating the challenges posed by high amounts of missing discharge data on the Niger basin. We evaluated the percentage of missing data on 22 discharge stations. MICE gap-filling method was assessed and used in filling missing data gaps on the discharge stations. It was observed that the basin has high percentage of missing data across different stations which has significant impacts on trend analysis. All the discharge stations show high negative autocorrelations before gap-filling. Comparison of autocorrelations of the observed and gap-filled data reveals gradual reduction in degree of autocorrelation after gap-filling. The performance of the MK and MMK statistics were compared on the gap-filled data and results showed very poor performance of the MK as compared to the MMK due to autocorrelation. Significantly increasing discharge trends was observed on all
the discharge stations after gap-filling from 1990s. Further studies should assess the possibility of using MICE to increase efficiency hydrological models and decrease modelling uncertainties.

**ACKNOWLEDGEMENTS**

This research was funded by the German Ministry of Education and Research (BMBF) through the DAAD ClimapAfrica post-doctoral program (https://www.daad.de/en/the-daad/what-we-do/sustainable-development/funding-programmes/climapfrica/). We thank the Niger Basin Authority for providing the hydrological data.

**AUTHOR CONTRIBUTIONS**

Data collection, running of MICE algorithm, graphics and tables were made by Ganiyu Titilope Oyerinde and Adeyeri Oluwafemi. The post-doctoral study was supervised by Prof Agnide E. Lawin. All authors contributed to the writing of the manuscript.

**DATA AVAILABILITY STATEMENT**

The Niger Basin Authority has strict policies in sharing daily data. However, we have tried our best to provide all relevant data in this paper or its supplementary information.

**REFERENCES**

Abadzadesahraei, S. & Sui, J. 2016 Potential impacts of climate change on water resources in British Columbia communities. (26), 1–24. https://doi.org/10.3390/w8010023.

Adeyeri, O. E., Laux, P., Lawin, A. E., Ige, S. O. & Kunstmann, H. 2019 Analysis of hydrometeorological variables over the transboundary komadugu-Yobe basin, West Africa. Journal of Water and Climate Change 1–16. https://doi.org/10.2166/wcc.2019.283.

Aich, V., Liersch, S., Vetter, T., Fournet, S., Andersson, J. C. M., Calmanti, S., van Weert, F. H. A., Hattermann, F. F. & Paton, E. N. 2016 Flood projections within the Niger river basin under future land use and climate change. Science of the Total Environment 562, 666–677. https://doi.org/10.1016/j.scitotenv.2016.04.021.

Ali, A. & Lebel, T. 2009 The sahelian standardized rainfall index revisited. International Journal of Climatology 29(12), 1705–1714. https://doi.org/10.1002/joc.

Ali, R., Kuriqi, A., Abubaker, S. & Kisi, O. 2019 Long-term trends and seasonality detection of the observed flow in Yangtze river using Mann-Kendall and Sen’s innovative trend method. Water (Switzerland) 11(9). https://doi.org/10.3390/w11091855.

Amanambu, A. C., Li, L., Egbionola, C. N., Obarin, O. A., Mupenzi, C. & Chen, D. 2019 Catena spatio-temporal variation in rainfall-runoff erosivity due to climate change in the lower Niger basin, West Africa. Catena 172, 324–334. https://doi.org/10.1016/j.catena.2018.09.003.

Amogu, O., Descroix, L., Yéro, K. S., Le Breton, E., Mamadou, I., Ali, A., Vischel, T., Bader, J.-C., Moussa, I. B., Gautier, E., Boubkraoui, S. & Belleudy, P. 2010 Increasing river flows in the Sahel? Water 2, 170–199. https://doi.org/10.3390/w2020170.

Badou, D. F., Kapangaziwiri, E., Diekkrüger, B., Hounkpè, J. & Afouda, A. 2016 Evaluation of recent hydro-climatic changes in four tributaries of the Niger river basin (West Africa). Hydrological Sciences Journal 62(5), 1–14. https://doi.org/10.1080/02626667.2016.1250898.

Biasutti, M. 2019 Rainfall trends in the african sahel: characteristics, processes, and causes. WIREs Clim Change 1–22. https://doi.org/10.1002/wcc.591.

Bichet, A. & Diedhiou, A. 2018 West African Sahel becomes wetter during the last 30 years but dry spells are shorter and more frequent. Climate Research (in press), 155–162. https://doi.org/10.3354/cr01515

Burgan, H. I. & Aksoy, H. 2020 Monthly flow duration curve model for ungauged river basins. Water (Switzerland) 12(2), 1–19. https://doi.org/10.3390/w12020338.

Chen, Y., Guan, Y., Shao, G. & Zhang, D. 2016 Investigating trends in streamflow and precipitation in Huangfuchuan basin with wavelet analysis and the Mann-Kendall test. Water (Switzerland) 8(3). https://doi.org/10.3390/w8030077.

Criss, R. E. & Winston, W. E. 2008 Do nash values have value? Discussion and alternate proposals. Hydrological Processes 22, 2723–2725. https://doi.org/10.1002/hyp.7072.
Oyerinde, G. T., Wisser, D., Hountondji, F. C. C., Odo, Descroix, L., Guichard, F., Grippa, M., Lambert, L., Panthou, G., Amogu, O., Nka, B. N., Oudin, L., Karambiri, H., Paturel, J. E. 2018 Evolution of surface hydrology in the sahelo-Sudanian strip: an updated review. Water (Switzerland) 10(6), 748. https://doi.org/10.3390/w10060748.

Dosio, A., Turner, A. G., Tamoffo, A. T., Sylla, M. B., Lennard, C., Jones, R. G., Terray, L., KfW 2010 Adaptation to Climate Change in the Upper and Middle Niger River Basin. Available from: http://www.ipcc.ch/site/assets/uploads/2019/08/4.-SPM_Approved_Microsite_FINAL.pdf (accessed 18 September 2019).

Kiw 2014 Climate change impacts on river discharge in West Africa: a review. Journal of Hydrology 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Intergovernmental Panel on Climate Change (IPCC) 2019 Climate change and land. Available from: https://www.ipcc.ch/site/assets/uploads/2019/08/6.-Summary_for_Policy-Makers_FINAL.pdf (accessed 18 September 2019).

Little, R. J. A. 1992 Regression with missing X’s: a review. Journal of the American Statistical Association 87(420), 1227–1237. https://doi.org/10.1080/01621459.1992.10476282.

Meusburger, K., Steel, A., Panagos, P., Montanarella, L. & Aalke, C. 2012 Spatial and temporal variability of rainfall erosivity factor for Switzerland. 167–177. https://doi.org/10.5194/hess-16-167-2012.

Nka, B. N., Lawin, E., Yao, B., Oyerinde, G. T., Attogouinon, A. & Afouda, A. 2017 Decreasing past and mid-century rainfall indices over the ouémé river basin, Benin (West Africa). Climate (3), 74. https://doi.org/10.3390/cli3040074.

Ogilvie, A., Mahéé, G., Ward, J., Serpantiéé, G., Lemoalle, J., Morand, P., Barbier, B., Kaczan, D., Lukasiewicz, A., Paturel, J., Ličenou, G. & Clanet, J. C. 2010 Water, agriculture and poverty in the Niger river basin. Water International 35(5), 594–622.

Oguntunde, P. G. & Abiodun, B. J. 2013 The impact of climate change on the Niger river basin hydroclimatology, West Africa. Climate Dynamics 40, 81–94. https://doi.org/10.1007/s00382-012-1498-6.

Onyutha, C. & Willems, P. 2013 Uncertainties in flow-Duration-Frequency relationships of high and low flow extremes in lake Victoria basin. Water (Switzerland) 5(4), 1561–1579. https://doi.org/10.3390/w5041561.

Oyerinde, G. T. & Diekkrüger, B. 2017c Farming responses to changing hydrological trends in the Niger river basin parts of Benin. Hydrology and Earth System Sciences Discussions 11(2), 2483–2514. https://doi.org/10.5194/hessd-11-2483-2014.
Schröder, S., Springer, A., Kusche, J., Uebbing, B., Fenoglio-Marc, L., Diekkrüger, B. & Pomeón, T. 2019 Niger discharge from radar altimetry: bridging gaps between gauge and altimetry time series. *Hydrology and Earth System Sciences* 23(10), 4113–4128. https://doi.org/10.5194/hess-23-4113-2019.

Sen, P. K. 1968 Оценки коэффициента регрессии по Тау Кендалла. Estimates of the regression coefficient based on Kendall’s Tau. *Journal of the American Statistical Association* 63(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934.

Sidibe, M., Dieppois, B., Mahe, G. & Paturel, J. 2018 Trend and variability in a new, reconstructed streamflow dataset for West and Central Africa, and climatic interactions, 1950–2005. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2018.04.024.

Sylla, M. B., Giorgi, F., Pal, J. S., Gibba, P., Kebe, I. & Nikiema, M. 2015 Projected changes in the annual cycle of high intensity precipitation events over West Africa for the late 21st century. *Journal of Climate* 28(16), 6475–6488. https://doi.org/10.1175/JCLI-D-14-00854.1.

Sylla, M. B., Pal, J. S., Faye, A., Dimobe, K. & Kunstmann, H. 2018 Climate change to severely impact West African basin scale irrigation in 2°C and 1.5°C global warming scenarios. *Scientific Reports* 8(1), 14395. https://doi.org/10.1038/s41598-018-32736-0.

Vahid, D. H., Noori, M., Heydari, M. & Sheikhi, M. 2015 A statistical review of top cited ISI papers regarding the different effects of climate change on rivers. *International Journal of Research and Innovations in Earth Science* 2(5), 2394–1375. Retrieved from http://ijries.org/administrator/components/com_jresearch/files/publications/IJRIES_96_Final.pdf

van Buuren, S. & Groothuis-Oudshoorn, K. 2011 Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* 45(3). https://doi.org/10.18637/jss.v045.i03

van der Heijden, G. J. M. G., Donders A, T., Stijnen T, R. & Moons, K. G. M. 2006 Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *Journal of Clinical Epidemiology* 59(10), 1102–1109. https://doi.org/10.1016/j.jclinepi.2006.01.015.

Verstraeten, G., Poesen, J., Demarée, G. & Salles, C. 2006 Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): implications for assessing soil erosion rates. *Journal of Geophysical Research Atmospheres* 111(22), 1–11. https://doi.org/10.1029/2006JD007169.

Yue, S. & Wang, C. Y. 2004 The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management* 18(5), 201–218. https://doi.org/10.1023/B:WARM.0000043140.61082.60.

Received 29 November 2020; accepted in revised form 14 May 2021. Available online 28 May 2021