

APPLICATION PAPER

A novel workflow for streamflow prediction in the presence of missing gauge observations

Rendani Mbuva^{1,2} , Julien Y.P. Adoukpe³ , Mandela C.M. HOUNGNIPO⁴ , Wilson T. Mongwe⁵, Zahir Nikraftar¹, Tshilidzi Marwala⁵ and Nathaniel K. Newlands⁶

¹School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

²School of Statistics and Actuarial Science, University of Witwatersrand, Johannesburg, South Africa

³Institut National de l'Eau, Université d'Abomey-Calavi, Cotonou, Benin

⁴Agence Nationale de la Météorologie du Bénin, Cotonou, Benin

⁵Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC, Canada

⁶School of Electrical and Electronic Engineering, University of Johannesburg, Johannesburg, South Africa

Corresponding author: Rendani Mbuva; Email: r.mbuva@qmul.ac.uk

Received: 21 May 2023; **Revised:** 25 May 2023; **Accepted:** 25 May 2023

Keywords: Deep learning; imputation; missing data; streamflow; transformers; West Africa

Abstract

Streamflow predictions are vital for detecting flood and drought events. Such predictions are even more critical to Sub-Saharan African regions that are vulnerable to the increasing frequency and intensity of such events. These regions are sparsely gaged, with few available gaging stations that are often plagued with missing data due to various causes, such as harsh environmental conditions and constrained operational resources. This work presents a novel workflow for predicting streamflow in the presence of missing gauge observations. We leverage bias correction of the Group on Earth Observations Global Water and Sustainability Initiative ECMWF streamflow service (GESS) forecasts for missing data imputation and predict future streamflow using the state-of-the-art temporal fusion transformers (TFTs) at 10 river gaging stations in the Benin Republic. We show by simulating missingness in a testing period that GESS forecasts have a significant bias that results in poor imputation performance over the 10 Beninese stations. Our findings suggest that overall bias correction by Elastic Net and Gaussian Process regression achieves superior performance relative to traditional imputation by established methods. We also show that the TFT yields high predictive skill and further provides explanations for predictions through the weights of its attention mechanism. The findings of this work provide a basis for integrating Global streamflow prediction model data and the state-of-the-art machine learning models into operational early-warning decision-making systems in resource-constrained countries vulnerable to drought and flooding due to extreme weather events.

Impact Statement

This paper presents a state-of-the-art streamflow prediction workflow in poorly gaged areas that are plagued by missing data. We demonstrate the effectiveness of the workflow in forecasting daily streamflow at 10 gaging stations in the Benin Republic. The proposed workflow enhances decision-making based on streamflow model-based forecasts, reducing bias introduced by missing in-situ data and producing high-fidelity forecasts. This work is particularly relevant in areas where data collection is too costly, under-resourced, or not possible across Sub-Saharan Africa.

1. Introduction

In hydrology, streamflow refers to the volume of water that moves through a natural channel, such as a river or a stream, during a specific duration. This quantity is generally quantified in cubic meters per second or cubic feet per second and plays a crucial role in managing water resources, predicting floods, and monitoring the environment (Rantz, 1982).

Streamflow predictions are critical for decision-making in many areas of climate change adaptation, including flood and drought prevention, agricultural planning, and hydroelectric power system operations (Kratzert et al., 2019). Predictions can be generated by physical and statistical models of the hydrological process within the river catchment (Adoukpe et al., 2021). Regardless of the approach, observational data are required to calibrate models for the localized dynamics within catchments sufficiently. Model-based predictions are even more essential in vulnerable and underdeveloped regions that are less resilient to extreme weather events such as drought and flooding. However, localized observations for hydrological models are not always available due to physical sensor failure events (Hamzah et al., 2020). Physical sensor systems often endure harsh environmental conditions, destruction and power outages (Tencaliec et al., 2015). These phenomena result in prolonged periods of missing or invalid observation readings. In developing regions such as the Benin Republic, the effect of such missing data is further complemented by the already sparse distribution of stream gaging stations (Harrigan et al., 2020). Fitting statistical models based on incomplete observational data results in inaccurate models that do not capture the full dynamics of the underlying hydrological processes (Hamzah et al., 2020). Imputation of missing data, therefore, leads to much-needed improved flood and drought predictions in a climate change context where these events are expected to increase in intensity and frequency (Kalantari et al., 2018). Thus, missing data imputation has become a necessary first step in data preprocessing tasks (Pigott, 2001). Since the area of missing data imputation commands a relatively vast amount of statistical literature, a wide variety of techniques have been developed and applied with different data requirements and accuracy (Little and Rubin, 2019).

A simple approach to handling missing data is discarding records where any missingness arises. Gill et al. (2007) show that while deletion of missing data is common practice in hydrology, it is not necessarily optimal as significant amounts of predictive information can be lost. Single imputation methods attempt to replace missing values one feature at a time with a pre-specified summary statistic (e.g., mean or median) of the complete data (Baraldi and Enders, 2010). Regression-based methods, such as imputation by random forest (RF), *k*-nearest neighbors (KNNs), and multilayer perceptron, improve upon single imputation by exploiting correlations between variables in complete data (Pantanowitz and Marwala, 2009; Gao, 2017). Both regression and single imputation can lead to biases in the resultant analysis when the assumption that data are missing at random is violated (Gao (2017)). Multiple imputation methods, such as multivariate imputation by chained equations, reduce this bias by replacing missing data points with predictions from an ensemble of regressors obtained from multiple subsets of the complete data (Van Buuren, 2018). Each of the imputation techniques above is used extensively in hydrology (Adeloye and Rustum, 2012; Mwale et al., 2012; Hamzah et al., 2021).

Global streamflow prediction systems also present a potential solution to missing data and the sparse distribution of stream gage stations in developing countries. The Group on Earth Observations Global Water and Sustainability Initiative (GEOGloWS) produce one such system in the Global Flood Awareness System (GloFAS) with the GEOGloWS ECMWF streamflow service (GESS) Ashby et al. (2021). The GESS extends GloFAS to local sites using an updated digital elevation model that allows for increased resolutions in areas of the world with narrow watersheds (Ashby et al., 2021). Like GloFAS, runoff predictions from the HTESSSEL land surface model are further downscaled and routed through a high-resolution stream network using the Muskingum algorithm of the routing application for parallel-discharge computation (Gill, 1978). While GESS provides much-needed streamflow data, localized calibration remains a significant challenge. A global

calibration of these systems is biased primarily to developed regions with dense and complete observations data Ashby et al. (2021). Thus, the utility of streamflow forecasts from these systems can be significantly enhanced by bias and variance correction to minimize systemic biases between model output and gaging station observations. Statistical Learning methods are broadly popular when learning transfer functions between global model forecasts and local observations (Piani et al., 2010; Rischard et al., 2018; Wang et al., 2021). Some statistical learning methods that have been put forward for bias correction tasks include penalized regression, quantile mapping (QM), artificial neural networks, and Gaussian processes (GPs) (Pastén-Zapata et al., 2020; Xu and Liang, 2021; Hunt et al., 2022).

A natural next step from the imputation of missing data is streamflow forecasting. This area has recently been dominated by recurrent neural network (RNN) architectures given their successes in sequence-to-sequence modeling tasks such as natural language processing (Vaswani et al., 2017; Kratzert et al., 2019; Adoukpe et al., 2021; Castangia et al., 2023). This is mainly due to the ability of architectures like the long short-term memory (LSTM) networks to capture dependencies at longer time lags and the use of attention mechanisms that can focus on salient parts of the input sequence that are relevant to predictions for a particular time point (Alizadeh et al., 2021; Lim et al., 2021).

In this work, we address two gaps in streamflow prediction literature: (a) prior work on statistical learning approaches to streamflow prediction isolates the problems of missing data and forecasting and (b) most RNN approaches in this area are restricted to training and generating forecasts for a single catchment at a time due to the inability to incorporate encodings of categorical variables.

This paper closes these gaps by presenting and evaluating a novel workflow for streamflow forecasting that (a) performs missing data imputation through the bias correction of GESS river discharge estimates and (b) employs a state-of-the-art forecasting model, the temporal fusion transformer (TFT) for day head streamflow prediction for 10 catchments in the Benin Republic using a single interpretable model. The benefits of the proposed workflow are that it provides a principled method for dealing with missing observations while also minimizing training time for the forecasting module; competing architectures would require the training of 10 different models (one for each catchment).

2. Study Area

For the purposes of this study, 10 hydrological stations located on the outlets of Benin's major river catchments were selected and depicted in Figure 1: Athiémè, Beterou, Bonou, Domè, Kaboua, Kompongou, Koubéri, Lanta, Porga, and Yakin. Benin's catchments are not limited to only Benin's administrative boundary but extend to its neighboring countries: Togo, Burkina Faso, and Nigeria. The Kouffo, Mono, Okpara, Ouémé, and Zou rivers originate from central Benin and Togo and generally flow into the Atlantic Ocean. The Alibori, Mékrou, and Sota rivers also originate from the center of Benin but have their outlet up north in the Niger river. The Pendjari river (sometimes referred to as the Oti river) originates from Northern Benin and flows into the Volta river in Togo and Ghana.

3. Data

In-situ hydrological data (daily river discharge record at 10 locations from 1980 to 2021) was acquired from Benin's hydrological service, Direction Générale de l'Eau (DG-Eau). The missing data percentage of the selected hydrological stations is 27.9%, with the station of Athiémè having the least missing data and the Kompongou station having the most missing data. Supplementary Figure 6 details the periods of missing data for each station. On the other hand, the GESS forecasts present no missing data and provide daily river discharge data ranging from 1979 to the present at targeted river sections (Ashby et al., 2021).



Figure 1. Benin’s catchments, rivers, and hydrological stations and missing data rate indicated in the map legend.

Daily precipitation, maximum temperature, and minimum temperature were obtained from Benin’s meteorological agency, Agence Nationale de la Météorologie du Bénin. The static datasets of elevation, soil, land use, and geology were, respectively, obtained from the CGIAR Consortium for Spatial Information (Reuter et al., 2007), Office de la Recherche Scientifique et Technique Outre-Mer (Fauck, 1977), Centre National de Télédétection, and Office Béninois des Mines. Further details on the data used can be found in the Supplementary Material.

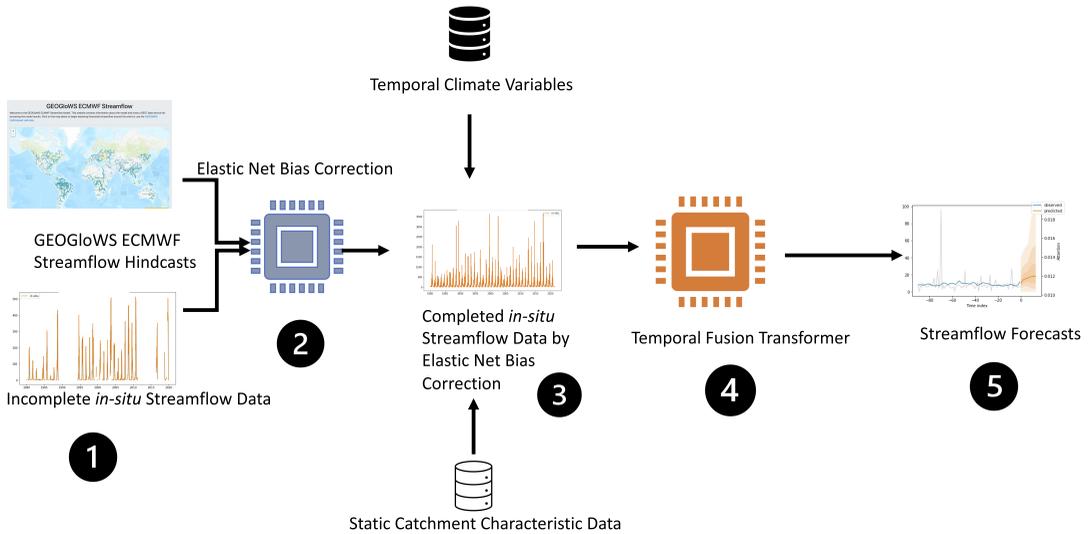


Figure 2. Our proposed workflow for missing data imputation and streamflow forecasting in poorly gaged areas. Step (1) includes collecting historical in-situ observations and the Group on Earth Observations Global Water and Sustainability Initiative ECMWF streamflow service (GESS) hindcasts for overlapping periods. (2) An elastic net is trained to correct biases of the GESS hindcasts onto in-situ observations to result in a completed dataset in (3). The completed in-situ observational dataset is then augmented with temporal climate variables and static catchment characteristic data. A temporal fusion transformer is trained in (4) to produce streamflow forecasts in (5).

4. Proposed Workflow

We propose the workflow depicted in Figure 2 that allows for both imputations of missing in-situ observations and multi-horizon streamflow forecasting. We now provide a detailed exposition of our imputation and forecasting methods.

4.1. Imputation of missing observations

We investigate streamflow imputation by bias-correcting GESS forecasts. We propose an imputation scheme where we replace a missing in-situ observation at time t with

$$\tilde{x}_t = h(\mathbf{x}'_t), \tag{1}$$

where \mathbf{x}'_t is the GESS forecast of streamflow at time t . h is a regressor trained on periods where \mathbf{x}'_t and \mathbf{x}_t are both available. We investigate settings where h takes various forms, including a simple lookup from GESS, a QM, elastic net, and GP.

4.1.1. Elastic net

The elastic net is a penalized linear regression model that employs weighted l_1 and l_2 norm regularization terms. The regularization terms serve as conservative priors (bias toward zero) on the coefficients that prevent overfitting to training data. Similar to the GP, we utilized a multi-output formulation of the Elastic Net that allows information sharing between stations. In our imputation formulation,

$$\tilde{\mathbf{x}}_t = h(\mathbf{x}'_t) = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

where β is a matrix of coefficients on the GESS predictions x'_t and ϵ is an error term of independent identically distributed Gaussian noise. The matrix of coefficients is obtained by optimizing the posterior:

$$\|X - X'\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1 \tag{3}$$

with the hyperparameters λ_1 and λ_2 obtained by cross-validation with the constraint $\lambda_1 + \lambda_2 = 1$. Given limited in-situ data, the elastic net has the advantage of limiting overfitting the bias correction relative to a regularized regression.

4.1.2. Gaussian processes

GPs are nonparametric probabilistic models that are well known for regression problems (Cohen et al., 2020). A GP is a collection of random variables, where each of its finite subsets is jointly Gaussian (Rasmussen and Williams, 2006). GPs are sufficiently defined by a mean function $\mu(x')$ and a kernel function $k(x'_i, x'_j)$.

Given a training dataset $\{x'_t, x_t\}_{t=1}^t$ with t noisy observations, $x_t = h(x'_t) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_X^2)$. The GP prior on h is such that $h(x') \sim \mathcal{N}(\mu_{x'}, K_{x'} + \sigma_X^2 I)$. The mean and variance of the Gaussian posterior predictive distribution of the function $h(x'_*)$ at a test point x'_* , are given by

$$\mathbb{E}(h_*) = \mu_{x'_*} + k_*^T (K_{x'} + \sigma_X^2 I)^{-1} (X - \mu_{x'}), \tag{4}$$

$$\text{var}(h_*) = k_{**} - k_*^T (K_{x'} + \sigma_X^2 I)^{-1} k_*, \tag{5}$$

where $k_{**} = k(x'_*, x'_*)$ and $k_* = k(x', x'_*)$. The kernel hyperparameters θ and the noise parameter σ_X are obtained by maximizing the log-marginal likelihood

$$\log p(X|X') = \log \mathcal{N}(\mu_{x'}, K_{x'} + \sigma_X^2 I). \tag{6}$$

In this work, we use a multi-output formulation of the GP where all 10 stations share a squared exponential kernel, thus allowing for implicit inference of connectivity relationships between stations.

4.1.3. Quantile mapping

QM is a well-known method for the bias correction of physical model output (Maraun, 2013; Ringard et al., 2017). In QM, we seek to learn a mapping between the empirical cumulative distribution functions (CDFs) of the in-situ and GESS prediction. In this case, the bias correction transfer function is

$$\tilde{x}_t = h(x'_t) = F_X^{-1} F_{X'}(x'_t), \tag{7}$$

where F_X^{-1} is the inverse empirical CDF of the in-situ data and $F_{X'}$ is the empirical CDF of the GESS data. Both empirical CDFs are obtained during a specified training period.

4.2. Temporal fusion transformer

We compare imputation by bias correction using the abovementioned methods to traditional regression imputation methods that rely only on complete in-situ data. We consider regression-based imputation by RF (Pantanowitz and Marwala, 2009), QM (Maraun, 2013; Ringard et al., 2017), GP (Rasmussen and Williams, 2006) regression, and KNNs (Hamzah et al., 2021) as our baselines. A description of these methods can be found in the Supplementary Material.

4.3. Forecasting methods

We investigate the LSTM network and TFT RNN models as candidates for our forecasting module.

The LSTM (Hochreiter and Schmidhuber, 1997) is an RNN architecture that utilizes a memory cell state to represent long-term dependencies in sequential data. The LSTM memory cell state at time t can be formulated as

$$\{c_t, h_t\} = f_{\text{LSTM}}(x_t, c_{t-1}, h_{t-1}, W_k), \quad (8)$$

where x are the inputs, c are the cell states, and h are the hidden states indexed by time t . W_k are the model weights. At the end of the input sequence, the cell and hidden states are connected to an output (y_{out}) via a dense neural network layer to make predictions:

$$y_{out} = \Phi(c_t, h_t, W_D). \quad (9)$$

Φ is the dense layer, and W_D is its weights.

The TFT (Lim et al., 2021) is a state-of-the-art multi-horizon forecasting model that takes a different approach to learn long-term dependencies in sequential data. The TFT uses a transformer encoder to create an abstract representation of the input sequence; this representation (encoding) is then transformed into predictions by a decoder. In general, a vital element of the TFT and transformer architectures is the self-attention mechanism which adds sparsity to the encoding. The sparsity added by the self-attention allows the decoder to focus on parts of the input sequence that are influential for prediction. Attention can be considered a differentiable dictionary lookup that matches a query to a key linked to a value in the encoding (a representation of past inputs). The TFT augments the transformer architecture with variable selection networks and static covariate encoders, increasing its utility for multivariate time series forecasting with heterogeneous inputs. The static covariate encoders are particularly important in streamflow forecasting. This allows us to use the catchment name as a feature and train one model for all catchments rather than multiple models (one for each catchment) in the case of traditional RNN models like the LSTM. A detailed depiction of the TFT is provided in Supplementary Figure 7. We consider two TFT models a TFT-Lite, which considers only the temporal inputs, that is, historical streamflow and climate variables and a larger TFT-full model that adds static catchment characteristics to the inputs. This allows us to test for the additional value of the static input encoder in the TFT.

We measure the quality of the imputation using the Kling–Gupta efficiency (KGE) (Gupta et al., 2009) and the Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970). These are discussed in more detail in the Supplementary Material.

5. Results and Discussion

We start by presenting the results of the missing data imputation and then proceed to discuss the results of forecasting elements.

5.1. Missing data imputation

We partition the data such that 60% of the period (January 1980 to November 2004) is for training and 40% (November 2004 to June 2021) is for testing. We evaluate the performance of the imputation methods by randomly simulating missingness on the complete testing data at the rates of 5, 10, 20, 30, and 50%.

Figure 3 shows the mean KGE and NSE of the imputation methods across gaging stations and levels of missingness. It can be seen that a simple lookup of the GESS predictions yields the worst imputation performance.

The hypothesis around significant bias in the GESS predictions is reinforced by the outperformance of the GP and elastic net bias correction imputation. Importantly the imputation by bias correction yields better performance than simply employing RF, KNN, and MLP regressors on the complete in-situ data. This can be reasonably expected as there are limited periods where the data are jointly complete for training the multiple imputation regressors. Thus, the GESS data provide the necessary signals regarding discharge seasonality and hydrologic connectivity to enhance imputation quality significantly.

Table 1 shows the KGE values at each of the stations at a level of 20% missingness. It can be seen that GESS yields the best performance at Yankin and Kompongou stations. Yankin and Kompongou had the

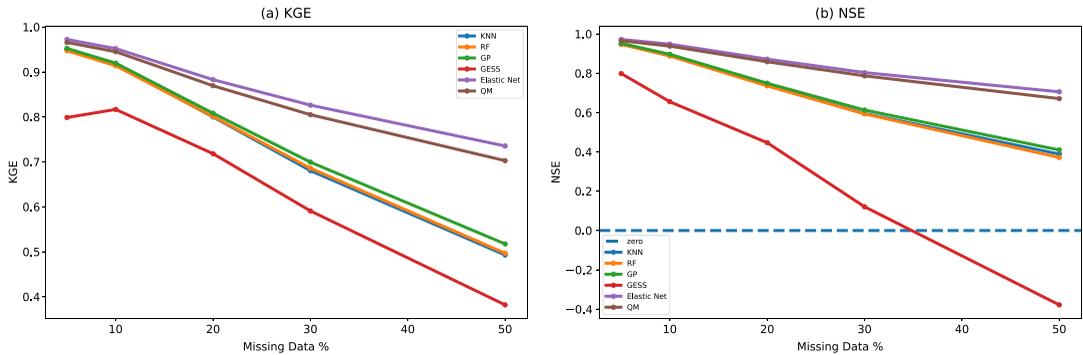


Figure 3. Plot of mean Kling–Gupta efficiency (KGE), and Nash–Sutcliffe efficiency (NSE) measures from each imputation method at varying levels of missingness across the 10 stations. The dashed blue line represents the zero NSE line. It can clearly be seen that the Group on Earth Observations Global Water and Sustainability Initiative ECMWF streamflow service imputation (in red) produces the lowest KGE and NSE values compared with all competing methods.

highest prevalence of missing data at 47.67 and 59.48%, respectively. A possible reason for this result is that, given the elevated levels of missingness at these stations, the complete data could not sufficiently calibrate the bias correction. GP and Elastic Net imputation methods provide the best performance in 8 of the 10 stations with the highest KGE values obtained at stations with highly complete data in the Athiémè and Bonou stations. QM as bias correction yields KGEs superior to multiple imputations by KNN and RF, suggesting that bias correction would remain the most efficient approach even when constrained by computational resources.

Figure 4a depicts the embedded bias in the GESS forecasts at the Koubéri station. The positive bias in the GESS manifests in that GESS lookups lead to new discharge extremes that are six times the extremes recorded in the in-situ observations. In operational settings, such significant positive biases can lead to false flood alerts and suboptimal use of already limited resources. The remedy to this extreme bias is seen in Figure 4b, where bias correction removes the significant positive bias while retaining the seasonal periodicity in discharge.

5.2. Day ahead streamflow forecasting

In testing the performance of the forecasting models, we trained all models with 30 years of data (January 1983 to April 2013) and tested the models on data from May 2013 to December 2019. The date range for training and testing of the forecasting models was adjusted due to the lack of in-situ observations of climate variables in periods prior to 1983. Table 2 summarizes the testing performance of forecasting models across each of the 10 gaging stations. It can be seen that the TFT models outperform the LSTM significantly across all gaging stations. The TFT-full model outperforms the much smaller TFT-lite model in 6 of the 10 stations. This suggests that the additional static inputs on specific catchment characteristics add predictive skill in forecasting streamflow. An essential consideration in explaining the relative performance of TFT lite is that an encoding of the catchment name is added as an input to the model. Thus, some catchment characteristics are implicitly represented in the input. It can also be seen that the models underperform on the Lanta catchment—this result is most likely explained by the size of the catchment being tiny relative to the others. Our results significantly improve on previous studies at overlapping catchments in the study area, mainly based on physical hydrological models. These include Biao (2017), which employs the hydrological model based on the least action principle at Beterou and Bonou, and Badou (2016) uses Soil and Water Assessment Tool, Hydrologiska Byråns Vattenavdelning (HBV light), and Universal Hydrological Program—

Table 1. KGE of each imputation method at 20% missingness at each respective gaging station.

Method	Stations									
	Yankin	Lanta	Kompongou	Athiémè	Kaboua	Beterou	Bonou	Porga	Koubéri	Domè
KNN	0.786	0.728	0.725	0.864	0.772	0.847	0.846	0.851	0.784	0.797
RF	0.821	0.784	0.780	0.859	0.841	0.807	0.832	0.839	0.626	0.822
Gaussian process	0.863	0.857	0.829	0.919	0.833	0.831	0.877	0.913	0.914	0.894
GESS	0.896	0.651	0.905	0.643	0.870	0.865	0.608	0.853	0.443	0.448
Elastic net	0.848	0.872	0.809	0.930	0.881	0.881	0.908	0.894	0.901	0.905
Quantile mapping	0.831	0.871	0.823	0.936	0.841	0.838	0.906	0.872	0.865	0.912

Note. The highest KGE for each station is in bold.

Abbreviations: GEOGloWS, Group on Earth Observations Global Water and Sustainability Initiative; GESS, GEOGloWS ECMWF streamflow service; KGE, Kling–Gupta efficiency.

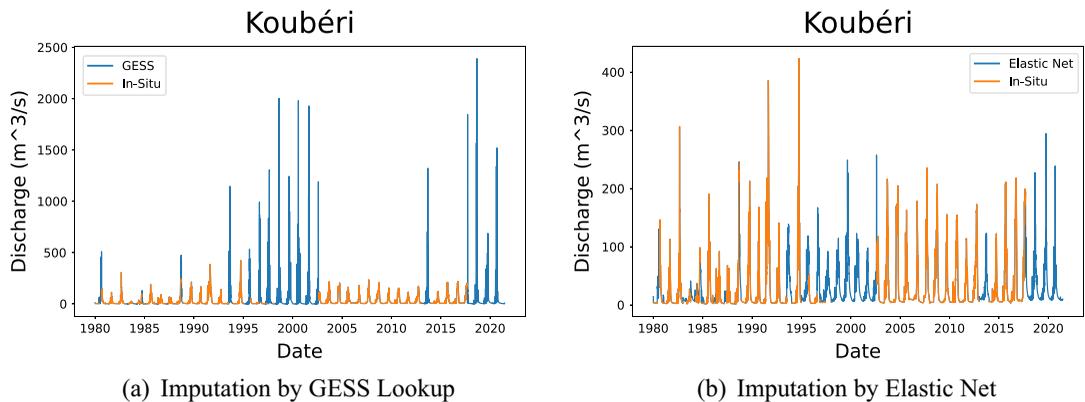


Figure 4. The infilled streamflow time series at the Koubéri station using the Group on Earth Observations Global Water and Sustainability Initiative ECMWF streamflow service (GESS) Lookup Bias Correction (a) and Elastic Net Bias Correction (b). The difference in scale of the y-axis between the two figures illustrates a bias in the GESS forecasts of a factor of about six times that of the in-situ data.

Hydrological Response Unit at Kouber, Yakin, and Kompongou. Details on other overlapping studies are contained in the Supplementary Material.

The attention mechanism of the TFT further allows us to endow our forecasts with some explainability of predictions. Figure 5 depicts the average temporal attention over the testing period of each TFT model. From Figure 5a, it can be seen that the TFT-lite model focuses more on inputs at lags of around 80–90 days with very sparse attention at earlier lags. On the other hand, in Figure 5b, the TFT full spreads attention of a similar scale to the TFT-lite model with a minor kink around 40 days with a similar pick around 80–90 days. This suggests that the TFT model has the potential to capture seasonal influences that typically have the delays of up to 90 days. The variable importance plots in the static variables in Supplementary Figure 9 show an agreement between the TFT models that the catchment encoding is an essential feature. This further supports the hypothesis around the advantages of using TFT with a static encoder for this task. Other variables that have high importance necessarily include the previous day's streamflow, runoff ratios, and mean elevation.

6. Conclusion

We evaluated the state-of-the-art imputation methods for streamflow prediction, demonstrating the need and importance of bias-correcting GESS streamflow forecast information. Our findings show that the bias introduced by GESS forecasts can be significant and result in possible false flooding alerts. We minimize such bias using Elastic Net regressions trained in periods where in-situ observations are available. The resultant imputation under simulated missingness shows that GESS bias correction outperforms train KNN and RF imputation trained on available data alone. We also implement a TFT architecture in streamflow prediction for the first time in literature and show that it significantly outperforms the LSTM and previous work on catchments in the Benin Republic.

Our findings enhance decision-making based on streamflow model-based forecasts. This enhancement is critical, particularly in areas where data collection is too costly, under-resourced, or not possible across Sub-Saharan Africa. Continued work aims to enhance and expand on the operational implementation of our bias-correction-based methodology. Specifically, we aim to interpolate and extrapolate the best-performing bias-correction model to unengaged areas. This is anticipated to vastly increase the utility of GESS to large areas across Sub-Saharan Africa and Central America that are currently unengaged.

Table 2. A summary of the performance of each forecasting model during the testing period at each respective gaging station.

Model	Metric	Stations									
		Yankin	Lanta	Kompongou	Athiémè	Kaboua	Beterou	Bonou	Porga	Koubéri	Domè
Area (km ²)		8.163	1.702	5.711	21.063	9.464	10.08	39.121	22.987	13.22	8.263
LSTM	R^2	0.409	0.306	0.188	0.315	0.183	0.794	0.678	0.620	0.746	0.619
	NSE	0.369	0.302	0.125	0.270	−0.087	0.792	0.629	0.049	0.744	0.525
	KGE	0.5792	0.409	0.0102	0.401	0.391	0.823	0.574	0.392	0.809	0.423
TFT lite	R^2	0.896	0.328	0.937	0.930	0.925	0.914	0.986	0.873	0.897	0.844
	NSE	0.892	0.302	0.892	0.930	0.923	0.906	0.981	0.851	0.856	0.808
	KGE	0.867	0.252	0.696	0.956	0.950	0.920	0.922	0.891	0.842	0.681
TFT full	R^2	0.877	0.371	0.935	0.925	0.915	0.923	0.989	0.894	0.926	0.885
	NSE	0.864	0.365	0.827	0.924	0.912	0.923	0.987	0.893	0.924	0.880
	KGE	0.890	0.495	0.572	0.943	0.889	0.938	0.950	0.932	0.907	0.855
Badou (2016)	R^2	0.59		0.41						0.64	
	NSE	0.57		0.37						0.53	
Biao (2017)	R^2						0.87	0.69			

Note. Best-performing models for each Station (and metric) are highlighted in bold.

Abbreviations: KGE, Kling–Gupta efficiency; LSTM, long short-term memory; NSE, Nash–Sutcliffe efficiency; TFT, temporal fusion transformer.

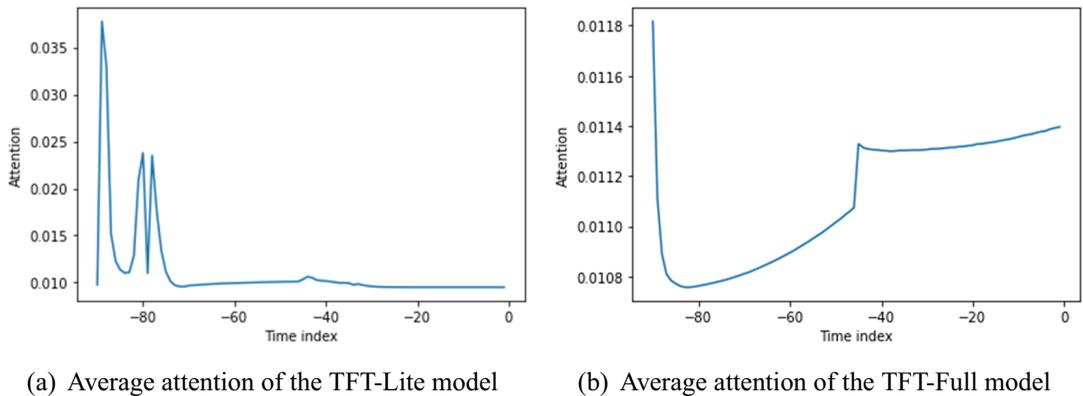


Figure 5. Average attention profiles of the temporal fusion transformer models in the testing period. This indicates that the observation time points are most influential in the making predictions.

Acknowledgments. We are grateful to Benin’s hydrological service, Direction Générale de l’Eau (DG-Eau) and Agence Nationale de la Météorologie du Bénin (METEO-BENIN) for the provision of hydrological and meteorological data for this work.

Author contribution. Conceptualization: R.M., J.Y.P.A., N.K.N.; Data curation: J.Y.P.A., M.C.M.H.; Investigation: R.M., J.Y.P.A.; Methodology: R.M., J.Y.P.A.; Writing—original draft: R.M., J.Y.P.A.; Writing—review and editing: T.M., N.K.N., W.T.M., Z.N., W.T.M. All authors approved the final submitted draft.

Competing interest. The authors of this manuscript have no competing interests.

Data availability statement. The data used in this study are available from Benin’s Hydrological service, Direction Générale de l’Eau (DG-Eau) for research purposes only.

Funding statement. R.M. is supported by a DeepMind Academic fellowship in Machine Learning.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/eds.2023.11>.

Provenance. This article is part of the Climate Informatics 2023 proceedings and was accepted in *Environmental Data Science* on the basis of the Climate Informatics peer-review process.

References

- Adeloye AJ and Rustum R (2012) Self-organising map rainfall-runoff multivariate modelling for runoff reconstruction in inadequately gauged basins. *Hydrology Research* 43(5), 603–617.
- Adoukpe PJY, Alamou E, Diallo B and Ali A (2021) Predicting discharge in catchment outlet using deep learning: Case study of the Ansongo–Niamey basin. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*.
- Alizadeh B, Bafti AG, Kamangir H, Zhang Y, Wright DB and Franz KJ (2021) A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *Journal of Hydrology* 601, 126526.
- Ashby KR, Hales RC, Nelson J, Ames DP and Williams GP (2021) Hydroviewer: A web application to localize global hydrologic forecasts. *Open Water Journal* 7(1), 9.
- Badou DF (2016) *Multi-Model Evaluation of Blue and Green Water Availability under Climate Change in Four Non-Saharan Basins of the Niger River Basin*. PhD Thesis, National Water Institute, University of Abomey–Calavi.
- Baraldi AN and Enders CK (2010) An introduction to modern missing data analyses. *Journal of School Psychology* 48(1), 5–37.
- Biao EI (2017) Assessing the impacts of climate change on river discharge dynamics in Oueme River basin (Benin, West Africa). *Hydrology* 4(4), 47.
- Castangia M, Grajales LMM, Aliberti A, Rossi C, Macii A, Macii E and Patti E (2023) Transformer neural networks for interpretable flood forecasting. *Environmental Modelling & Software* 160, 105581.
- Cohen S, Mbuva R, Marwala T and Deisenroth M (2020) Healing products of Gaussian process experts. In *International Conference on Machine Learning, Proceedings of Machine Learning Research*, pp. 2068–2077.
- Fauk R (1977) Evaluation of soil resources by ORSTOM. In *Soil Resource Inventories and Development Planning, Proceedings of Workshops at Cornell University*, vol. 1978, pp. 263–270.

- Gao Y** (2017) *Dealing with Missing Data in Hydrology: Data Analysis of Discharge and Groundwater Time-Series in Northeast Germany*. PhD Thesis, Freie Universität Berlin.
- Gill MA** (1978) Flood routing by the Muskingum method. *Journal of Hydrology* 36(3–4), 353–363.
- Gill MK, Asefa T, Kaheil Y and McKee M** (2007) Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. *Water Resources Research* 43(7). <https://doi.org/10.1029/2006WR005298>
- Gupta HV, Kling H, Yilmaz KK and Martinez GF** (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377(1), 80–91.
- Hamzah FB, Hamzah FM, Razali SM and Samad H** (2021) A comparison of multiple imputation methods for recovering missing data in hydrological studies. *Civil Engineering Journal* 7(9), 1608–1619.
- Hamzah FB, Mohd Hamzah F, Mohd Razali SF, Jaafar O and Abdul Jamil N** (2020) Imputation methods for recovering streamflow observation: A methodological review. *Cogent Environmental Science* 6(1), 1745133.
- Harrigan S, Zsoter E, Alfieri L, Prudhomme C, Salamon P, Wetterhall F, Barnard C, Cloke H and Pappenberger F** (2020) GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth System Science Data* 12(3), 2043–2060.
- Hochreiter S and Schmidhuber J** (1997) Long short-term memory. *Neural Computation* 9(8), 1735–1780.
- Hunt KMR, Matthews GR, Pappenberger F and Prudhomme C** (2022) Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. *Hydrology and Earth System Sciences Discussions* 2022, 1–30.
- Kalantari Z, Ferreira CSS, Keesstra S and Destouni G** (2018) Nature-based solutions for flood-drought risk mitigation in vulnerable urbanizing parts of East-Africa. *Current Opinion in Environmental Science & Health* 5, 73–78.
- Kratzert F, Herrnegger M, Klotz D, Hochreiter S and Klambauer G** (2019) NeuralHydrology—Interpreting LSTMs in hydrology. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham: Springer, pp. 347–362.
- Lim B, Arik SÖ, Loeff N and Pfister T** (2021) Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* 37(4), 1748–1764.
- Little RJ and Rubin DB** (2019) *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, Vol. 793. New York: John Wiley & Sons.
- Maraun D** (2013) Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate* 26(6), 2137–2143.
- Mwale FD, Adeloye AJ and Rustum R** (2012) Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi—A self organizing map approach. *Physics and Chemistry of the Earth, Parts A/B/C* 50, 34–43.
- Nash J and Sutcliffe J** (1970) River flow forecasting through conceptual models, part I—A discussion of principles. *Journal of Hydrology* 10(3), 282–290.
- Pantanowitz A and Marwala T** (2009) Missing data imputation through the use of the random forest algorithm. In Yu W and Sanchez EN (eds), *Advances in Computational Intelligence*. Berlin–Heidelberg: Springer, pp. 53–62.
- Pastén-Zapata E, Jones JM, Moggridge H and Widmann M** (2020) Evaluation of the performance of euro-CORDEX regional climate models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods. *Journal of Hydrology* 584, 124653.
- Piani C, Weedon G, Best M, Gomes S, Viterbo P, Hagemann S and Haerter J** (2010) Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology* 395(3), 199–215.
- Pigott TD** (2001) A review of methods for missing data. *Educational Research and Evaluation* 7(4), 353–383.
- Rantz SE** (1982) *Measurement and Computation of Streamflow*. Water Supply Paper 2175. U.S. Department of the Interior, Geological Survey.
- Rasmussen CE and Williams CKI** (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.
- Reuter HI, Nelson A and Jarvis A** (2007) An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science* 21(9), 983–1008.
- Ringard J, Seyler F and Linguet L** (2017) A quantile mapping bias correction method based on hydroclimatic classification of the Guiana shield. *Sensors* 17(6), 1413.
- Rischard M, Pillai N and McKinnon KA** (2018) Bias correction in daily maximum and minimum temperature measurements through Gaussian process modeling. Preprint, [arXiv:1805.10214](https://arxiv.org/abs/1805.10214).
- Tencaliec P, Favre A-C, Prieur C and Mathevet T** (2015) Reconstruction of missing daily streamflow data using dynamic regression models. *Water Resources Research* 51(12), 9447–9463.
- Van Buuren S** (2018) *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Press.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł and Polosukhin I** (2017) Attention is all you need. In *Advances in Neural Information Processing Systems*, Vol. 30.
- Wang C, Zhang W and Villarini G** (2021) On the use of convolutional Gaussian processes to improve the seasonal forecasting of precipitation and temperature. *Journal of Hydrology* 593, 125862.
- Xu T and Liang F** (2021) Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews: Water* 8(5), e1533.

Cite this article: Mbuva R, Adoukpe JYP, Hounnibo MCM, Mongwe WT, Nikraftar Z, Marwala T and Newlands NK (2023). A novel workflow for streamflow prediction in the presence of missing gauge observations. *Environmental Data Science*, 2: e23. doi:10.1017/eds.2023.11