Supplementary Material

Periods of Missing Data

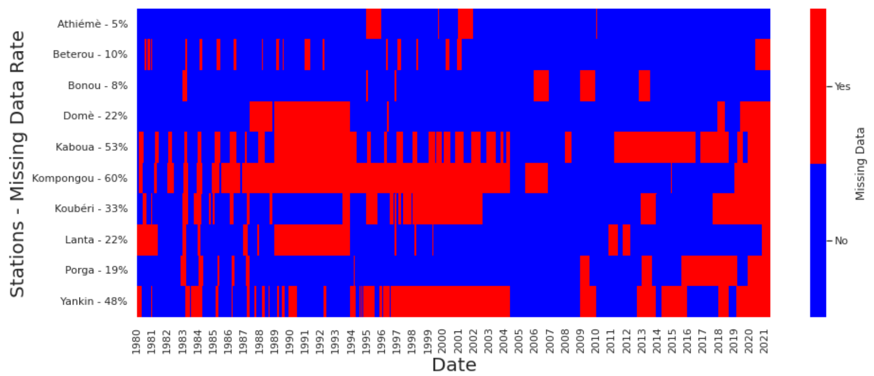


Figure 6. *Missing data periods at each of the ten hydrological gauging stations. Missing data periods are highlighted in red, and complete data periods are highlighted in blue.*

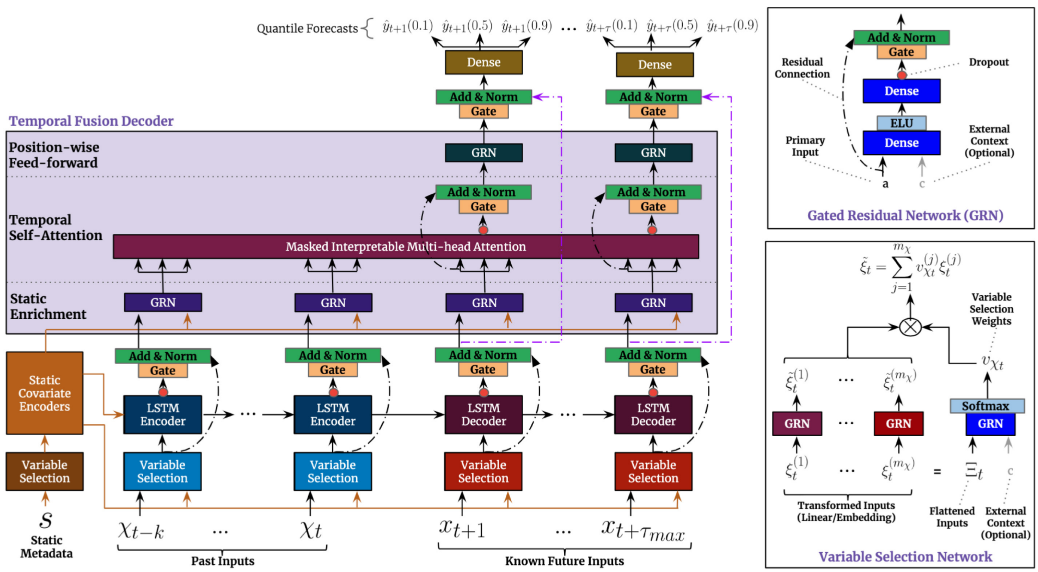


Figure 7. *A detailed depiction of the TFT architecture. The TFT architecture involves inputting static metadata, past inputs that vary with time, and future inputs that are known beforehand and also vary with time. Variable Selection is employed to choose the most important features based on the input data. The architecture utilizes Gated Residual Network blocks with skip connections and gating layers to ensure efficient information flow. LSTMs perform local processing, while multi-head attention is used to integrate information from any time step for time-dependent processing (Lim et al., 2021).*

Performance Evaluation

We measure the quality of the imputation using Kling-Gupta Efficiency (KGE) (Gupta et al., 2009)and Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970).

The KGE metric provides a diagnostically informative decomposition of the NSE and RMSE. It facilitates the analysis of the relative importance of its different components (correlation, bias and variability) in the context of hydrological modelling. The NSE measures the ability to predict variables different from the mean and gives the proportion of the initial variance accounted for by the model. The RMSE is frequently used to evaluate how closely the predicted values match the observed values based on the relative range of the data.

$$KGE = 1 - \sqrt{(r-1) + \left(\frac{\bar{\hat{x}}_t}{\bar{x}_t} - 1\right)^2 + \left(\frac{\sigma_{\hat{x}_t}}{\sigma_{x_t}} - 1\right)^2} \tag{10}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (x_t - \hat{x}_t)^2}{\sum_{i=1}^{n} (x_t - \bar{x}_t)^2} \tag{11}$$

$$R^2 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_t - \hat{x}_t)^2} \tag{12}$$

In the above equations, x_t represents the observed discharge at a time t, $\hat{x}_t$ represents the imputed discharge at a time t, $\bar{x}_t$ is the mean of the observed discharge, $\bar{\hat{x}}_t$ is the mean of the imputed discharge, $\sigma_{\hat{x}_t}$ is the variance of the simulated discharge, $\sigma_{\hat{x}_t}$ is the variance of the observed discharge, r is the Pearson correlation coefficient, and n is the number of observations. The KGE values range from $-\infty$ to 1, with values close to 1 indicating good agreement. Similarly to the KGE, the NSE values also range from $-\infty$ to 1, with the perfect model having the value 1. As this is a regression setting case the NSE is equivalent to the coefficient of determination R^2.

Details on variables

Except for the precipitation, maximum temperature, minimum temperature and discharge variables that are temporal datasets, all the other variables are static characteristics of catchments.

Variable name	Data source	Description
pr	METEO-BENIN	Daily precipitation in mm/day
tmax	METEO-BENIN	Daily maximum temperature in $^{\circ}C$
tmin	METEO-BENIN	Daily minimum temperature in $^{\circ}C$
hydro	DG-Eau	Daily river discharge in m^3/s
area	CGIAR-CSI	Area in km^2
perimeter	CGIAR-CSI	Perimeter in km
elev_mean	CGIAR-CSI	Mean elevation in m
slope_mean	CGIAR-CSI	Mean slope in m/km
brown	ORSTOM	Ratio of brown areas
ferralitic	ORSTOM	Ratio of ferralitic areas
ferruginous	ORSTOM	Ratio of ferruginous areas
hydromorphic	ORSTOM	Ratio of hydromorphic areas
river	ORSTOM	Ratio of river areas
slightly d.	ORSTOM	Ratio of slightly developed areas
vertisols	ORSTOM	Ratio of vertisols areas
agriculture	CENATEL	Ratio of agriculture areas
barren	CENATEL	Ratio of barren areas
built-up	CENATEL	Ratio of built-up areas
forest	CENATEL	Ratio of forest areas
grassland	CENATEL	Ratio of grassland areas
water	CENATEL	Ratio of water areas
wetland	CENATEL	Ratio of wetland areas
gneiss	OBEMINES	Ratio of gneiss areas
granite	OBEMINES	Ratio of granite areas
quartzite	OBEMINES	Ratio of quartzite areas
sandstone	OBEMINES	Ratio of sandstone areas
schist	OBEMINES	Ratio of schist areas
pr_mean	METEO-BENIN	Mean of daily precipitation
high_pr_freq	METEO-BENIN	Frequency of high precipitation days (>=5 times mean daily precipitation)
low_pr_freq	METEO-BENIN	Frequency of dry days (<1 mm/day) in $days/year$
tmax_mean	METEO-BENIN	Mean of maximum temperature
tmin_mean	METEO-BENIN	Mean of minimum temperature
q_mean	DG-Eau	Mean of river discharge in m^3/s
runoff_ratio	DG-Eau	Ratio of mean daily discharge to mean daily precipitation
Q5	DG-Eau	5% discharge quantile (low discharge)
Q95	DG-Eau	95% discharge quantile (high discharge)
high_q_freq	DG-Eau	Frequency of high discharge days (>9 times the mean daily discharge)
low_q_freq	DG-Eau	Frequency of low discharge days (<0.2 times the mean daily discharge)

Table 3. *A description of the variables and data sources used in this work.*

Additional Results

Method	NSE % Missing Data				
	5%	10%	20%	30%	50%
KNN	0.950	0.892	0.739	0.596	0.389
RF	0.947	0.888	0.736	0.593	0.372
Gaussian Process	**0.973**	**0.938**	0.840	0.751	0.623
GESS	0.799	0.655	0.447	0.121	-0.378
Elastic Net	**0.973**	**0.948**	**0.872**	**0.804**	**0.706**
Quantile Mapping	0.966	0.938	0.859	0.787	0.671

Table 4. *Mean NSE of each imputation method at varying levels of missingness across the ten stations.*

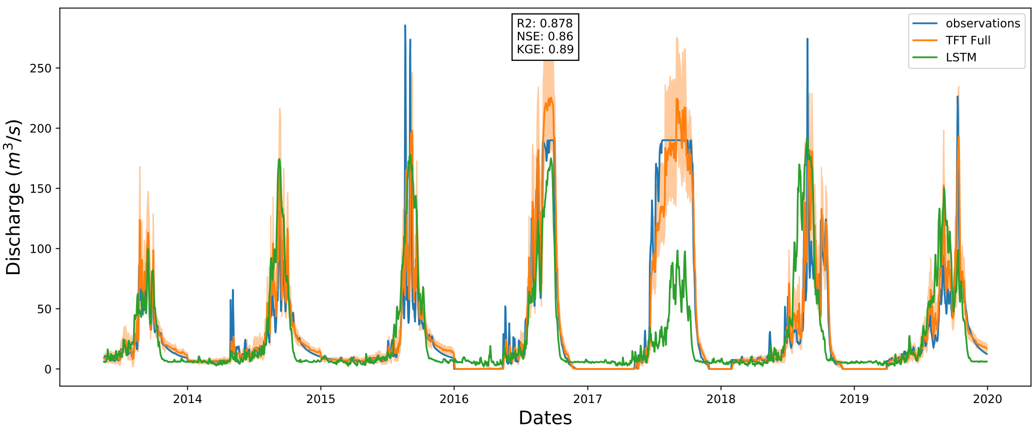


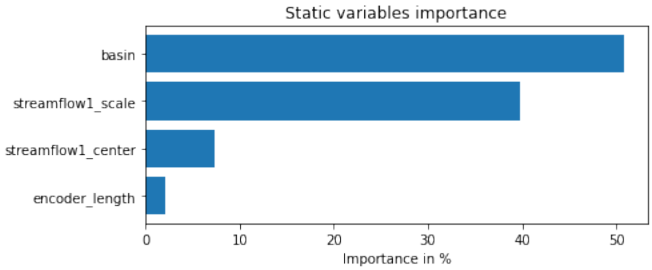
Figure 8. *An illustrative plot of the forecasts in the testing period for the Yankin station. The orange bands indicate the range between the 35th percentile and 65th percentile of the full TFT with the median full TFT prediction given by a solid orange line .*

Reference	Model	Station	Performance
Essou and Brissette (2013)	HMETS	Bonou	NSE between 0.823 and 0.937
Leroux et al. (2016)	DHSVM	Bétérou	R^2 between 0.82 and 0.87
Giertz et al. (2006)	UHP-HRU	Bétérou	R^2 of 0.75
Bossa et al. (2012)	SWAT	Domè	R^2 of 0.7 for weekly discharge
Hounkpè and Diekkrüger (2018)	WaSiM	Domè	KGE between 0.5 and 0.85
Ahouansou et al. (2015)	J2000	Porga	R^2 of 0.82 and NSE of 0.82
Sintondji, Dossou-Yovo, and Agbossou (2013)	SWAT	Kaboua	R^2 of 0.89
NOUNANGNONHOU et al. (2019)	GR2M	Kaboua	R^2 greater than 0.75 and NSE greater than 0.70
Biao et al. (2021)	ModHyPMA, GR4J, HBV and AWBM	Athiémé	NSE averaged over models between 0.5 and 0.76

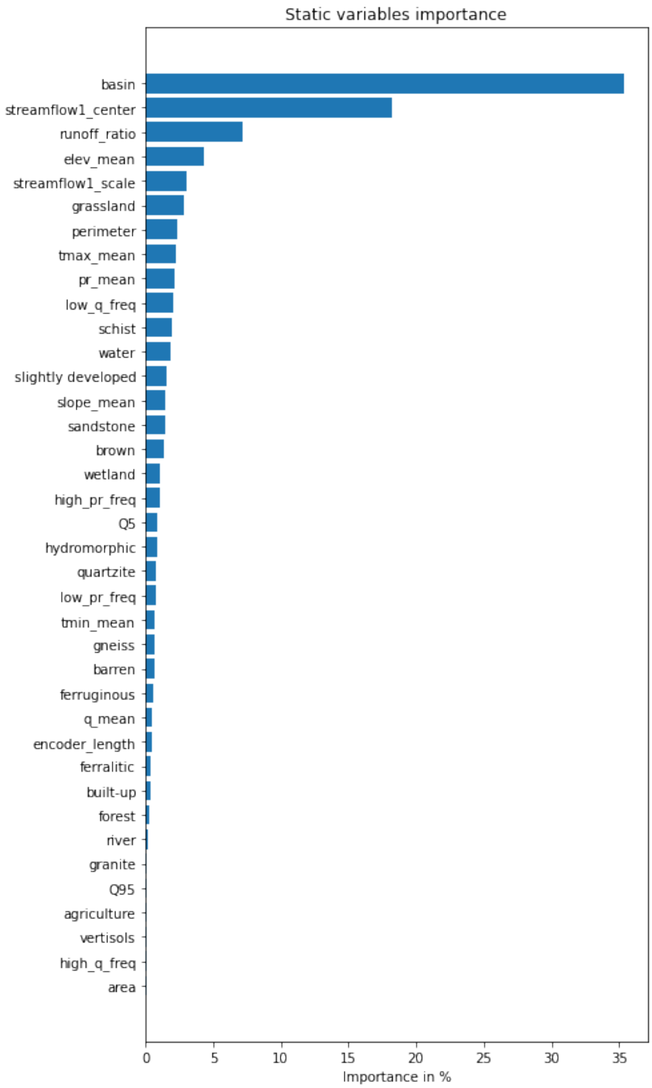
Table 5. *Additional performance of classical hydrological models on target stations*

Hydrological Model Name Lookup.

- AWBM: Australian Water Balance Model,
- DHSVM: Distributed Hydrology Soil Vegetation Model,
- GR2M: Groupe de Recherche en Microélectronique et Microsystèmes,
- GR4J: Génie Rural à 4 paramètres Journalier,
- HyMoLAP: Hydrological Model based on the Least Action Principle,
- HMETS: Hydrology Model-EST and
- WaSiM: Water balance Simulation Model



(a) Averaged Importance of Static Variables of the TFT-Lite model



(b) Averaged Importance of Static Variables of the TFT-Full model

Figure 9. *Average attention profiles of the TFT models in the testing period .*