Machine Learning for Reconstructing Streamflow Time Series: An Application to the Nile River

UNIVERSITY OF CAMBRIDGE

Camilla Giulia Billari[1], Marc Girona-Mata[1], Kevin Wheeler[2], Andrea Marinoni[3,4], Edoardo Borgomeo[1]

[1]Department of Engineering, University of Cambridge, Cambridge, UK. [2]Water Balance Consulting, Boulder, USA. [3]Department of Physics and Technology, UiT the Arctic University of Norway, Tromsø, Norway. [4]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK.

AI4ER Application of AI to the Study of Environmental Risk

Motivation

1. Hydrological analysis and prediction with **sparse and discontinuous data** remain a key challenge for water resources planning and climate adaptation, especially in large river basins across the Global South.

2. Traditional stochastic hydrology methods and process-based models often fall short in their attempts to capture the complexity of these systems. Recent efforts to apply **machine learning** for river discharge **imputation** (assigning values to any data gaps in the target variable) and **reconstruction** (the inclusion of other proxy data to further inform imputation, such as climatic variables) show promise in creating complete historical datasets based on a limited set of discontinuous observations.

3. However, these methods have not been tested on datasets from large river basins with a **high proportion of missing values**.

We address this gap and investigate the suitability of machine learning methods for **streamflow imputation** and **reconstruction** in a case study of the **Nile River basin**.

Methodology

Two sets of **benchmarking experiments** were carried out to test the spatiotemporal gap-filling performance of different ML models: first for imputation, and then for reconstruction with climate forcings.

The models tested were a series of imputers (for imputation only; KNN, MICE), regressors (Random Forest, KNN, Bayesian ridge, AdaBoost, GradBoost and XGBoost) and conditional neural processes.

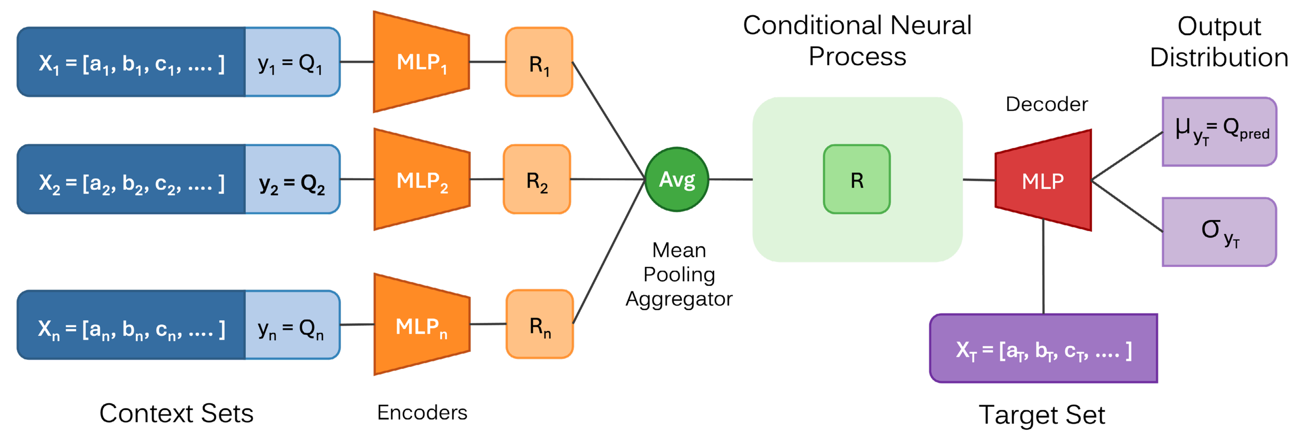

Fig 4: Conditional Neural Process (CNP) Architecture Diagram.

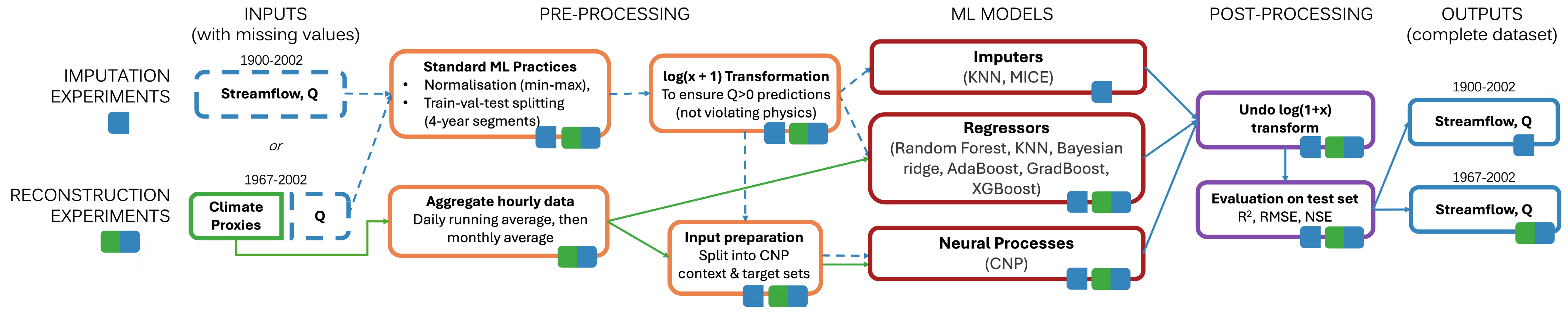

Fig 3: Experimental process for imputation and reconstruction experiments.

Data

Gauged streamflow dataset

- Time range: 1900-2002
- 13 stations (Uganda, South Sudan, Sudan, Ethiopia)
- 53% missing values

ECMWF ERA5 climate reanalysis

- Time range: 1967-2002
- Precipitation, temperature, relative humidity, wind speed, soil moisture data (monthly average)

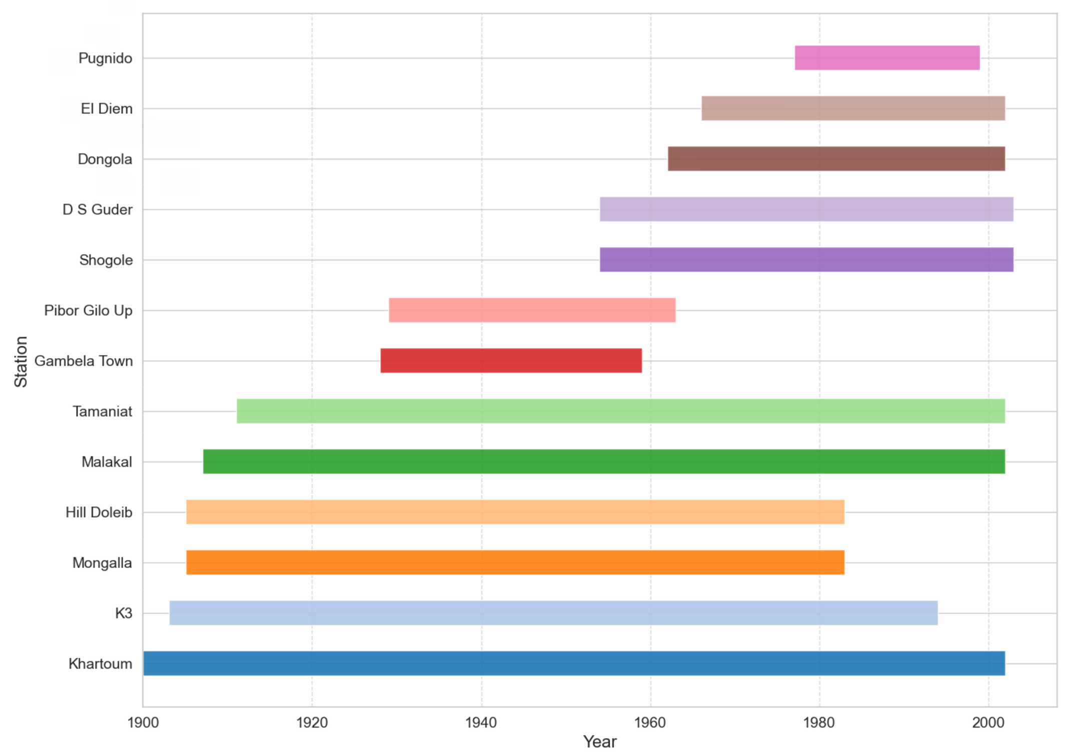

Fig 1: Active observational periods for each station, ordered by start date. Each bar represents the start and end date of the observed data, and white space represents its absence (missing values).

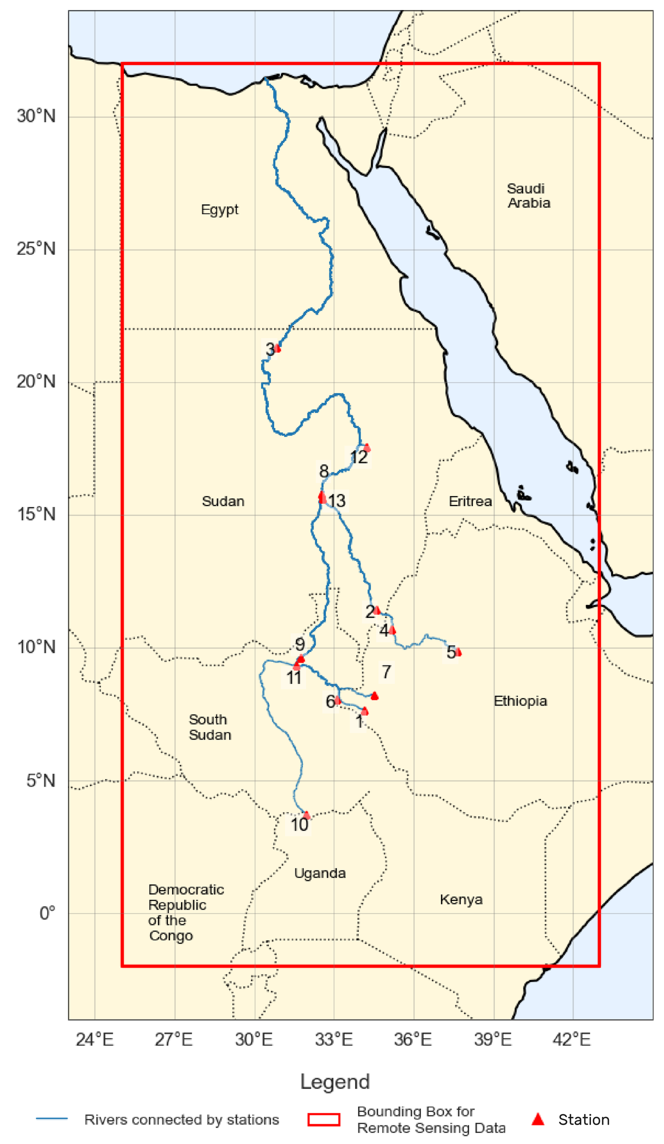

Fig 2: Map of North-East Africa; only tributaries downstream from stations shown.

Results and Conclusions

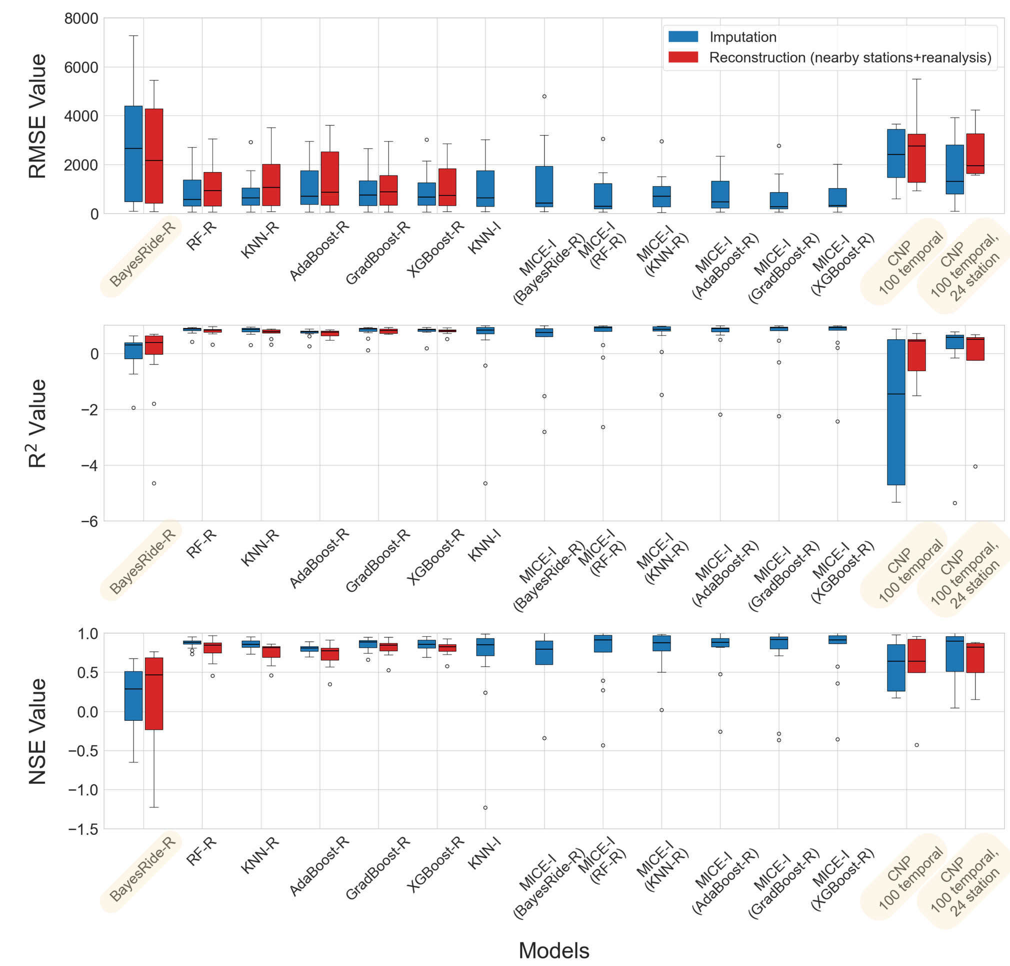

Fig 5: Boxplots of model performance distributions across all stations for each respective evaluation metric.

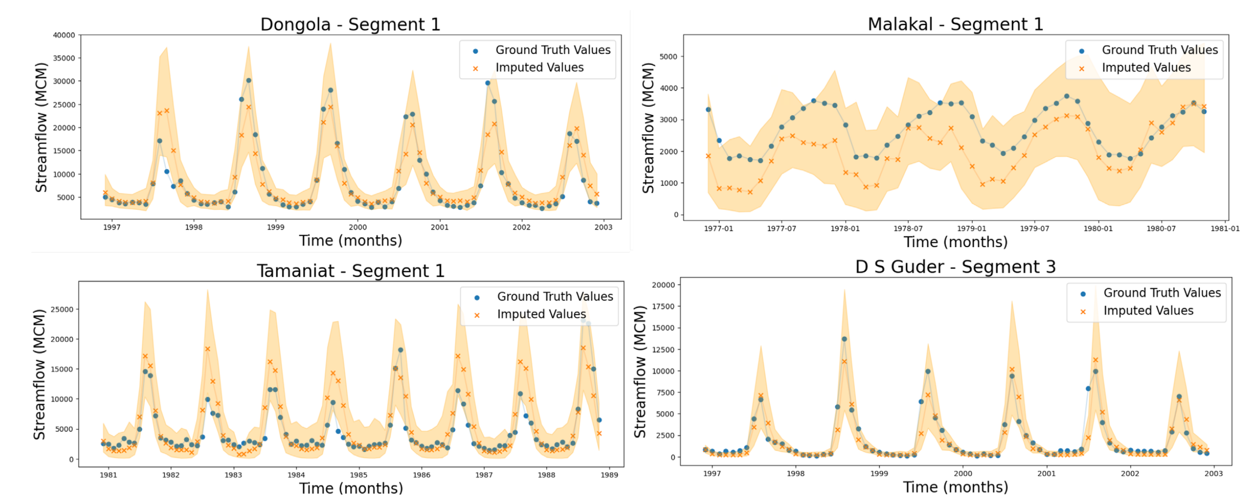

Fig 6: Examples of CNP reconstruction results with uncertainty quantification.

Tree-based regressors performed best across all experiments. Adding climate proxies decreased their accuracy in all metrics; their usefulness is limited to the quality of the initial dataset. They also do not provide uncertainty quantification. **CNPs show promise**, and benefitted from the addition of climate forcing data, but further work is needed for more extensidve model tuning and feature selection.

The approach developed in this study can be applied to other river basins with sparse observations to **build more complete hydrological datasets** for **water resources management** and **planning** applications.

EGU General Assembly, 01/05/25, Session: **HS3.6**

Abstract: https://doi.org/10.5194/egusphere-egu25-7429

References:

[1] Deltares. Annex A Eastern Nile Water Simulation Model: Hydrological boundary conditions. 2013 Jan.
[2] Copernicus Climate Change Service. ERA5 monthly averaged data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS); 2019. Available from: https://cds.climate.copernicus.eu/doi/10.24381/cds.6860a573.
[3] Gordon J, Bruinsma WP, Foong AYK, Requeima J, Dubois Y, Turner RE. Convolutional Conditional Neural Processes. arXiv; 2020. ArXiv:1910.13556 [cs, stat]. Available from: http://arxiv.org/abs/1910.13556.