Event-based flood data imputation for infilling missing data in real-time flood warning systems

EGU23-4524
EGU General Assembly 2023

Event-based Flood Data Imputation for Infilling Missing Data in Real-time Flood Warning Systems

Farzad Piadeh, Kourosh Behzadian, and Joseph P. Rizzuto
School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, UK

Real-time flood warning systems as part of digital and innovative non-structural solutions have been widely used to prepare decision makers, operators, and affected population to alleviate socio-economic flooding consequences [1]. Many models have been introduced recently to provide more accurate flood forecasts with longer lead times. However, they rely highly on availability of input data which may contain missing values in measurement for one or more timesteps mainly due to wide range of reasons such as random/systematic errors and blunders. Hence, real-time early warning systems cannot be operated properly unless these missing data are properly infilled [2]. Despite data imputation techniques have been mainly employed in pre-processing step of historical data i.e., models training and validation, they have not been properly elaborated in real-time operation practically [3].

This paper aims to propose a new event-based data imputation method for infilling rainfall and water level missing data appearing in real-time operation of flood early warning systems. Event identification is first used to divide the real-time data into the wet or dry weather conditions which then are used for selecting the best strategy of infilling missing data. Imputation decision framework takes advantage of various imputation techniques including t-copula, move-median, and kriging based on external available benchmarks and temporal location of missing data. Proposed methodology is tested in real-world case study of urban drainage system in London, UK. Conventional techniques such as linear regression, kriging, nearest neighbourhood, t-copula, inverse distance, and similar calendar are first compared together and best techniques are then tested with proposed methodology in three real-time scenarios as (1) missing rainfall intensity, (2) missing water level, (3) missing both rainfall and water level. Recurrent neural network model used for flood forecasting and results are demonstrated for the next 3hr-ahead predictions.

Results show the proposed method can reduce root mean square error (RMSE) from 55% to 13%, 43% to 12%, and 97% to 17% for the above scenarios, respectively. Furthermore, using external benchmark data resources, i.e. other near rainfall/water level stations, shows very efficient when missing data appears at early steps of rainfall events where selected conventional techniques suffer from predicting rainfall pattern. Finally, when both water level and rainfall intensity were missing, the proposed imputation method can reduce RMSE from 197mm to 117mm (RMSE was originally 100 for no missing data) for 3hr-ahead predictions. Generally, this study shows the proposed imputation method can better infill the missing data, especially those in the flood event by using correlated data in other weather/gauging stations and flexibility in applying different

methods.

References

[1] Piadeh, F., Behzadian, K., Alani, A.M. (2022). Multi-Step Flood Forecasting in Urban Drainage Systems Using Time-series Data Mining Techniques. *Water Efficiency Conference*, West Indies, Trinidad and Tobago, repository.uwl.ac.uk/id/eprint/9690 [Accessed 31/12/2022].

[2] Piadeh, F., Behzadian, K., Alani, A. (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. *Journal of Hydrology*, 607, 127476.

[3] Ben Aissia, M., Chebana, F., Ouarda, T. (2017). Multivariate missing data in hydrology–Review and applications. *Advances in Water Resources*, 110, pp.299-309.