EGU General Assembly 2023

23-28 April 2023 – Vienna, Austria

UNIVERSITY OF WEST LONDON
The Career University

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

Farzad Piadeh*, Kourosh Behzadian, Joseph P. Rizzuto

[1] School of Computing and Engineering, University of West London, UK
*: Researcher at University of west London - Farzad.piadeh@uwl.ac.uk

April 2023

Outline

01 **Introduction**

Concepts, necessity and gap finding

02 **Methodology**

Defining proposed approaches

03 **Applied method for real case study**

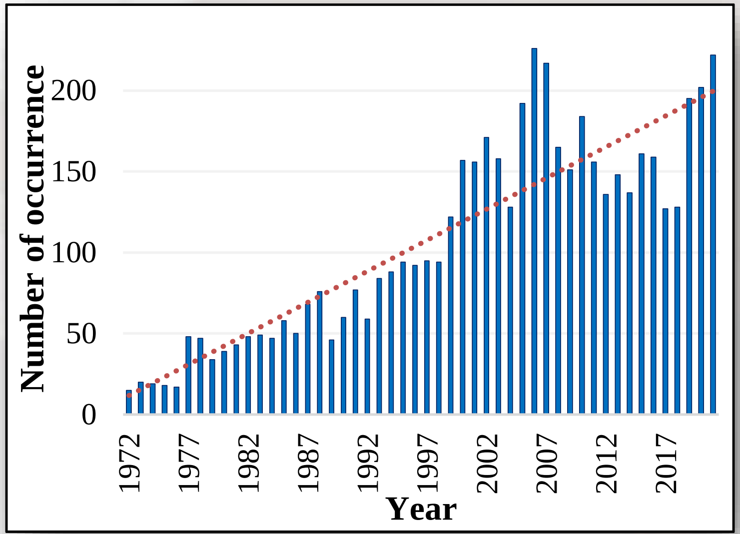
Verifying proposed approach by real case study

04 **Conclusion**

Points and further researches

Introduction

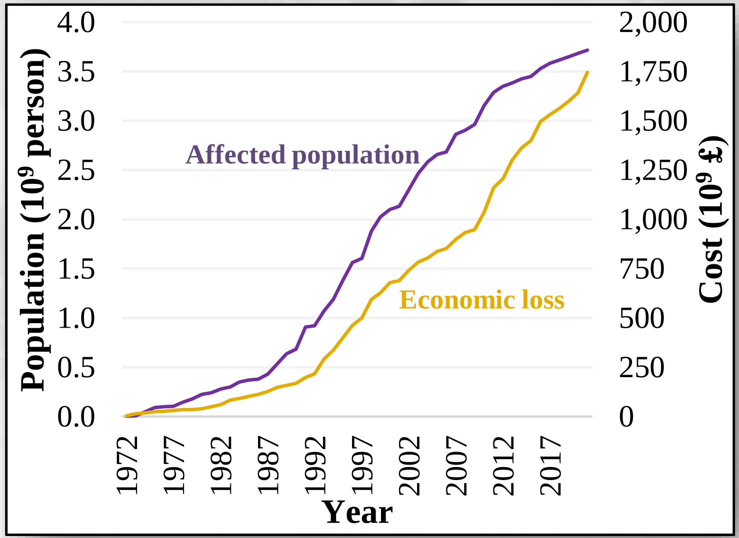
❖ *Urban flooding*



Number of flood occurrences

In recent 50 years, floods caused:

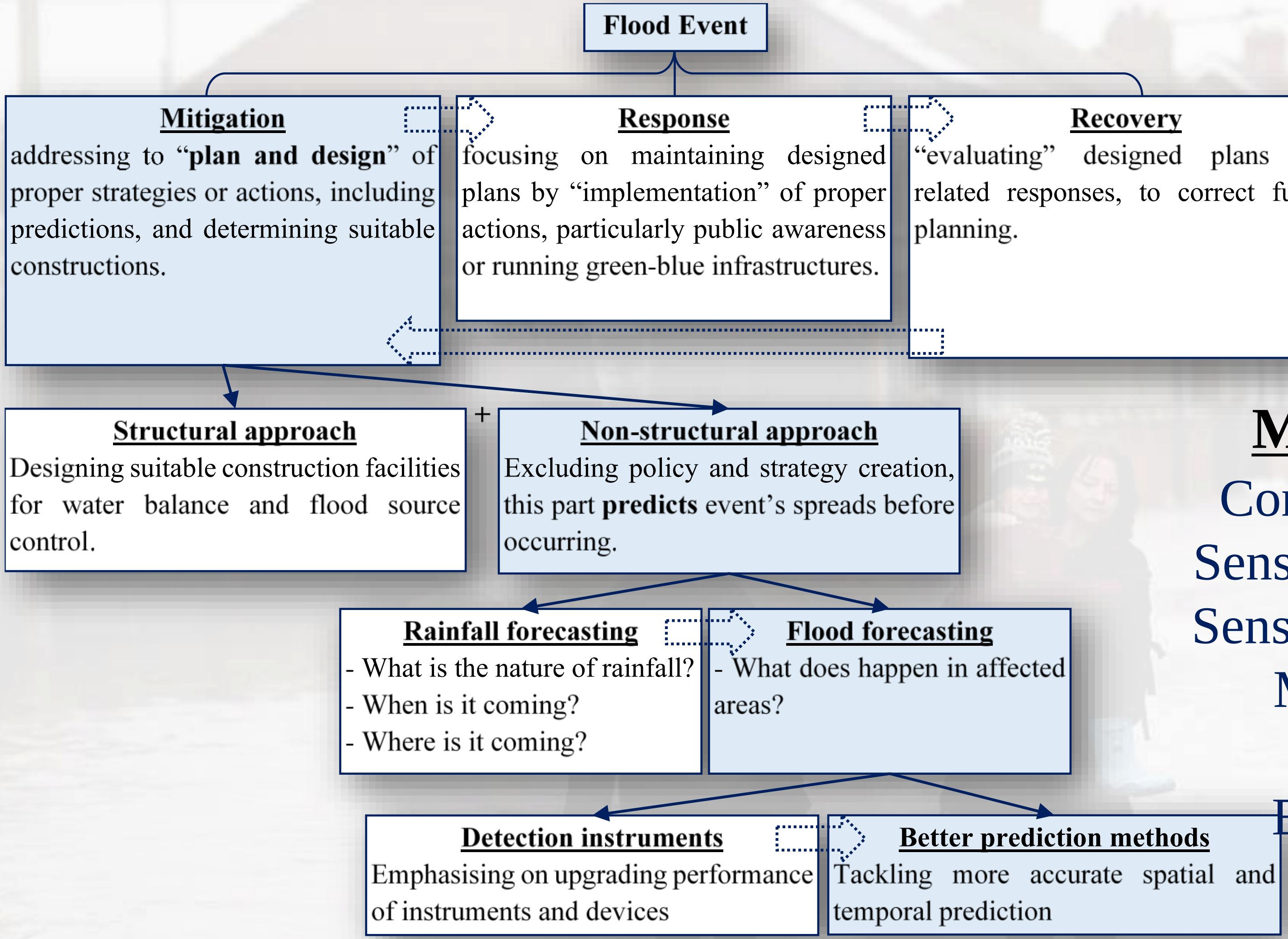
- ❖ **1,750 £ billion economy damages**
- ❖ **3.7 billion people are affected**
- ❖ **329,000 people are killed**



Cumulative social and economic loss

“Multi-Step Flood Prediction in Drainage Systems Using Time-series Data Mining Techniques”, Piadeh F., Behzadian K. Alani A.M., Water Efficiency Conference, West Indies, Trinidad and Tobago, 2022.

❖ *Solutions*



Realtime flood forecasting

Physically-based models
Conceptual models
Data-driven models

Missing data

Connection errors
Sensor malfunctions
Sensor displacement
Maintenance
Blunders
Environment

Data imputation

Linear regression
Kriging
K-Nearest neighbourhood
Copula-based
Inverse distance
Similar calendar

"A Critical Review of Real-Time Modelling of Flood Forecasting in Urban Drainage Systems", Piadeh F., Behzadian K. Alani A.M., *Journal of Hydrology*, 2022; 607: 127476

"Development of an Artificial Intelligence-Based Framework for Biogas Generation from a Micro Anaerobic Digestion Plant", Ikechukwu O., Piadeh F., Behzadian K., Campus L., Rokiah Y., Waste Management, 2023; 158, pp. 66-75.

Research gaps

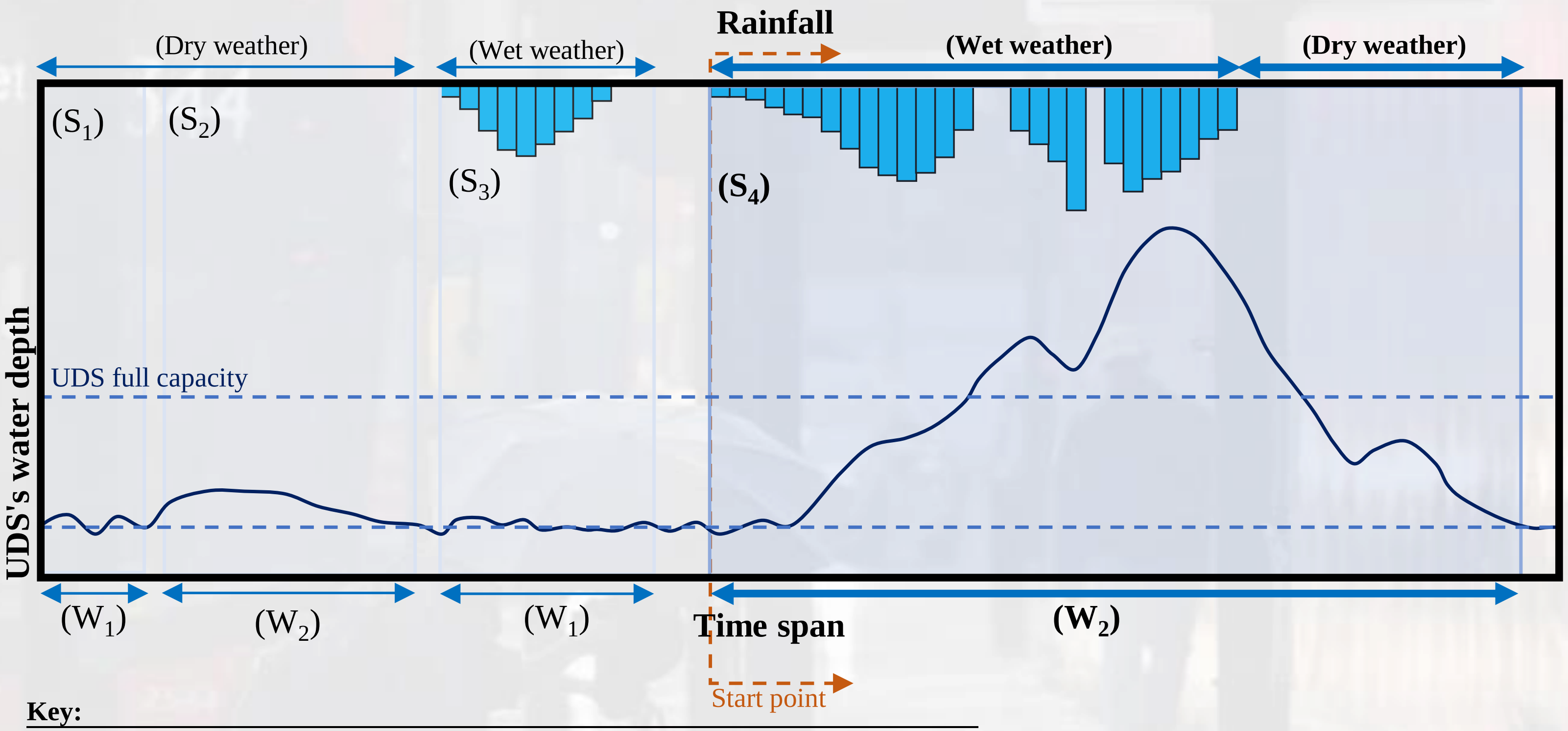
- ❖Inability to deal with sudden shift from negative to positive data gradient (flood events)
- ❖Inaccuracy for multistep and long period missing data
- ❖Lack of performance in database with huge missing data
- ❖Lack of understanding about earlier stages of rainfall or flood events

Aim

Event-based and external based data imputation method for infilling rainfall and water level missing data appearing in real-time operation of flood early warning systems

Method and material

❖ *Event Identification Method*



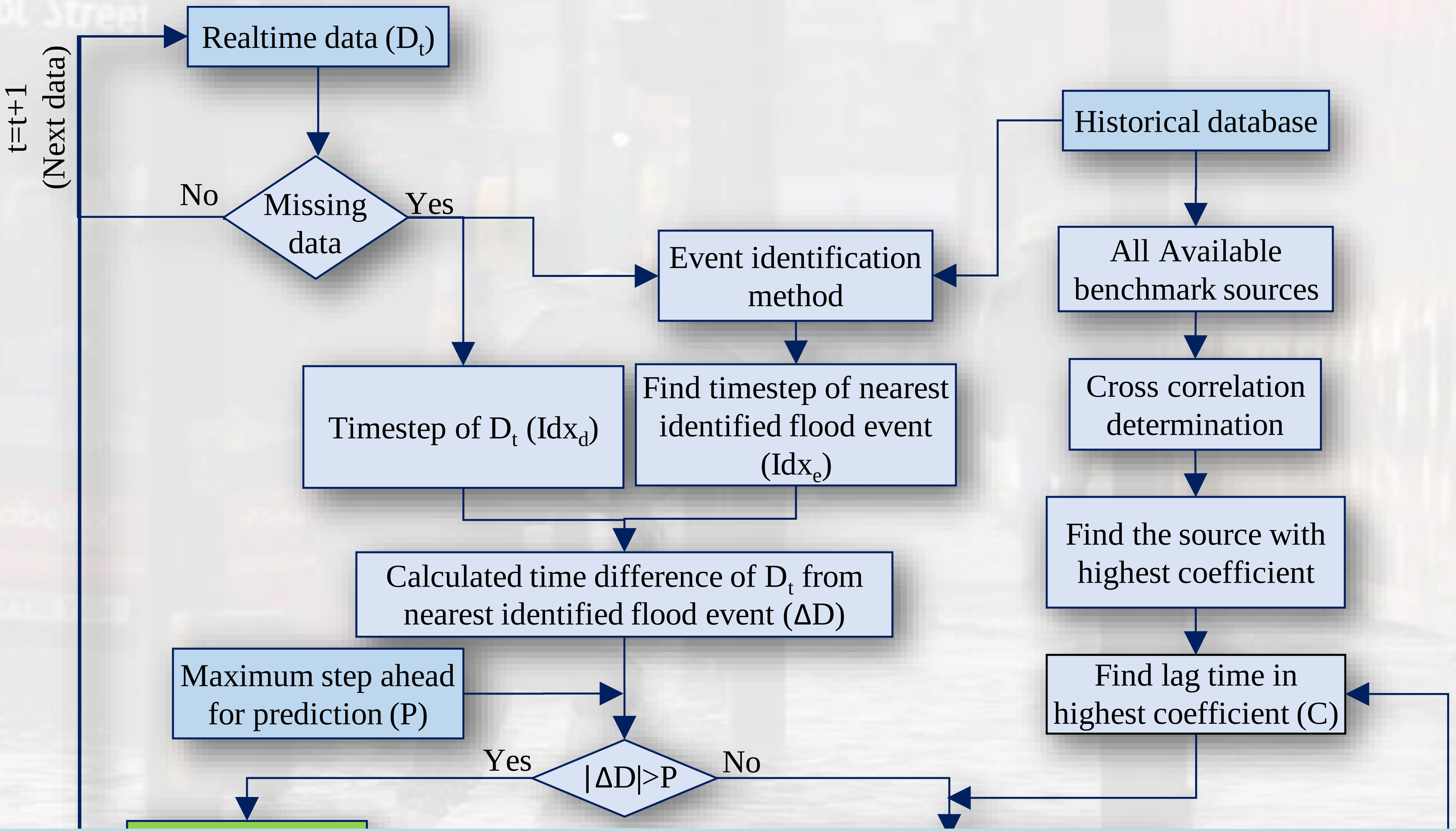
Key:

State	Captured data: Rainfall intensity	Captured data: Water depth
(S1): Dry weather, non-flood event	(R_1): -	(W_1): -
(S2): Sudden rising flow, non-flood event	(R_1): -	(W_2): +
(S3): Ineffective precipitation, non-flood event	(R_2): +	(W_1): -
(S4): Flood event	(R_2): +	(W_2): +

-: No rainfall, no change for water depth
+: Rainfall, net change (increase or decrease) for water depth

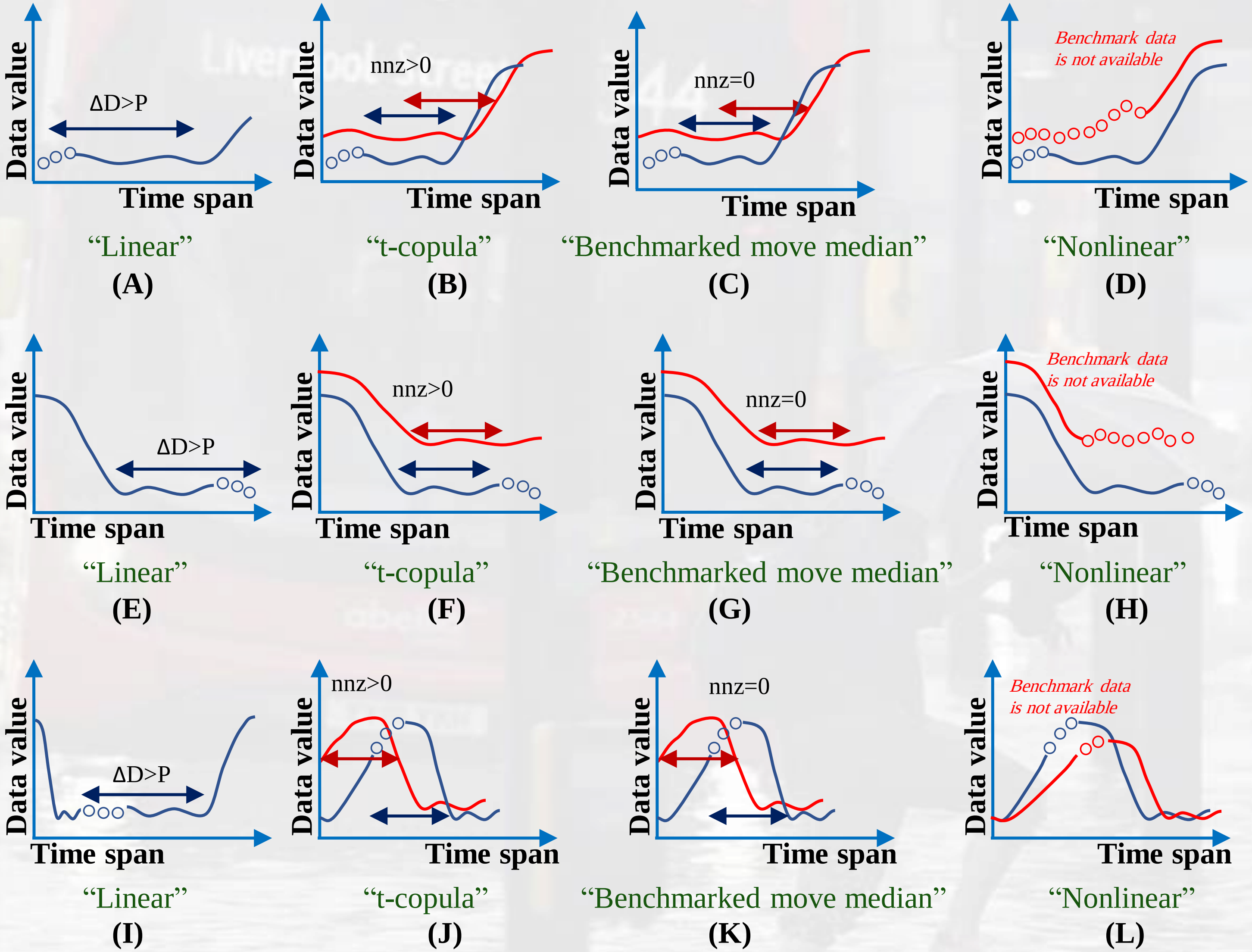
"The Role of Event Identification in Translating Performance Assessment of Time-Series Real-Time Urban Flood Forecasting", Piadeh F., Behzadian K. Alani A.M., 15th UWL Doctorial Conference, London, UK, 2021

❖ *Data Imputation Decision Framework*



Email: farzad.piadeh@student.uwl.ac.uk
https://doi.org/10.5194/egusphere-egu23-4524

❖ Different Applied Strategies in Proposed Methodology



(A-D): at the start of database

(E-H) at the end of database

(I-L): at the middle of database

Key

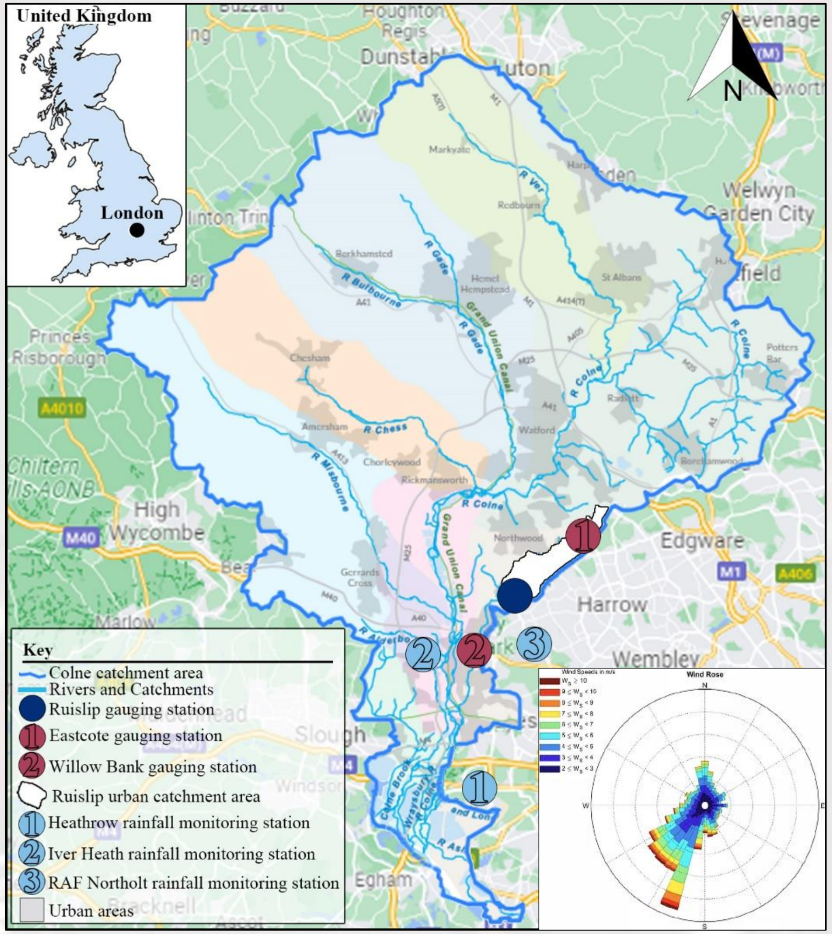
- ○ Missing data
- Available data
- Available Benchmark data

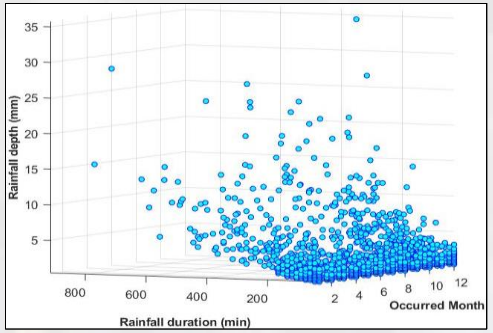
ΔD = Temporal distance of missing data from nearest identified flood event

P = Desired maximum time step-ahead of prediction

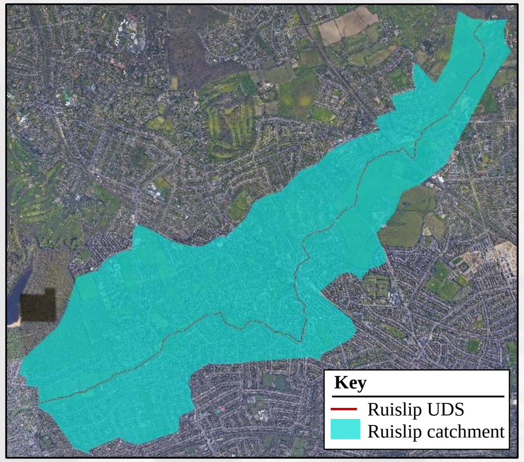
nnz= non-zero value of cross-covariance determination for P lags

❖ *Case study description*





(b)



(a) (c)

Geographical map and hydrological data of the pilot study: (a) location of stations and layout of catchment, (b) Characteristics of recorded rainfalls and (c) layout of Ruislip UDS and catchment

<u>Benchmark methods</u>

Linear regression

Kriging

K Nearest neighbourhood

Copula-based

Inverse distance

Similar calendar

Results and discussion

❖ *KPI Assessment*

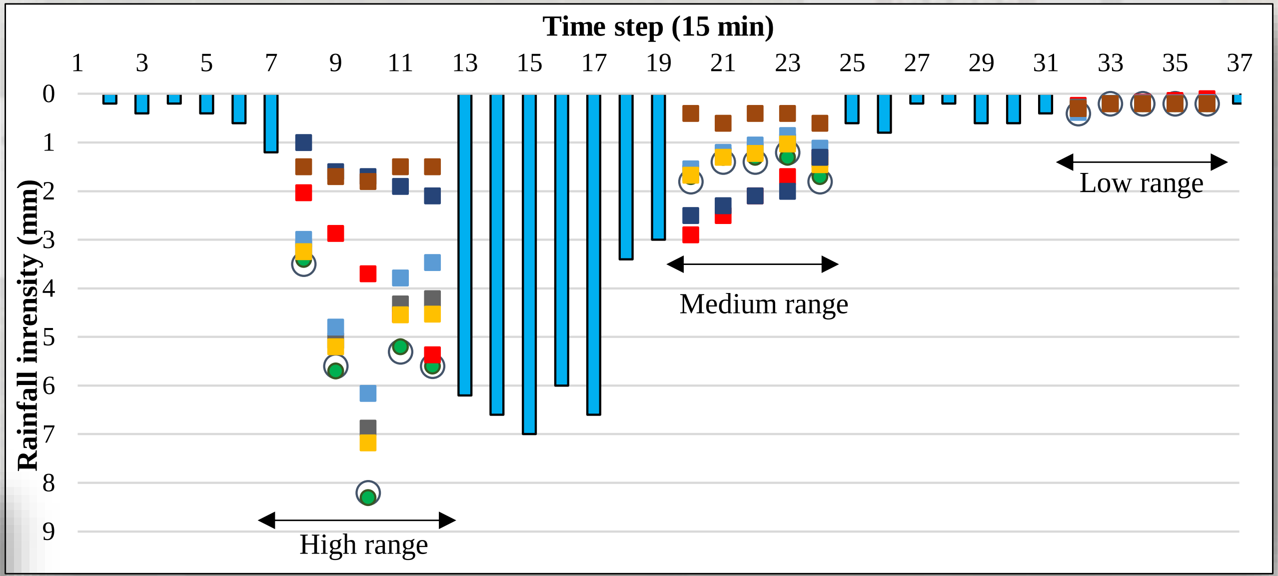
Performance indicator (RMSE in mm) of data imputation methods

Selected infilling method	Rainfall				Water level			
	NF[*]	**FE[**]**	**NF/FE**	**Rank**	**NF**	**FE**	**NF/FE**	**Rank**
The proposed method	**<0.01**	**0.08**	≅8	**1**	**1.16**	**3.95**	3	**1**
Linear regression	<0.01	8.77	≅877	5	1.16	63.67	55	6
Kriging	<0.01	1.39	≅139	3	1.19	13.54	11	2
Nearest neighbourhood	<0.01	3.29	≅329	4	1.33	35.50	30	4
Copula-based	<0.01	0.83	≅83	2	1.17	14.35	11	3
Inverse distance	<0.01	23.50	≅2350	6	1.34	47.59	36	5
Similar calendar	0.3	25.07	84	7	4.35	206.09	47	7

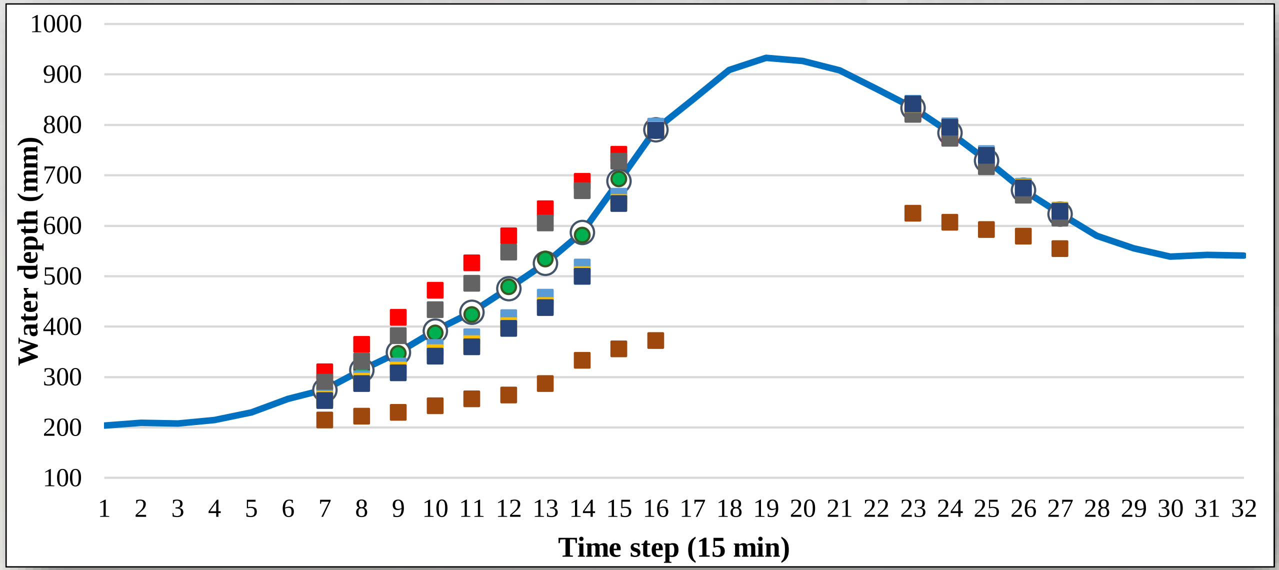
[*] NF: non-flood event [**] FE: flood event

❖ *Performance of Data Imputation Methods in Showcase*

- Proposed
- Missing data
- Observation
- Nearest neighborhood
- Kriging
- Linear regression
- Similar calender
- Inverse distance
- Copula-based

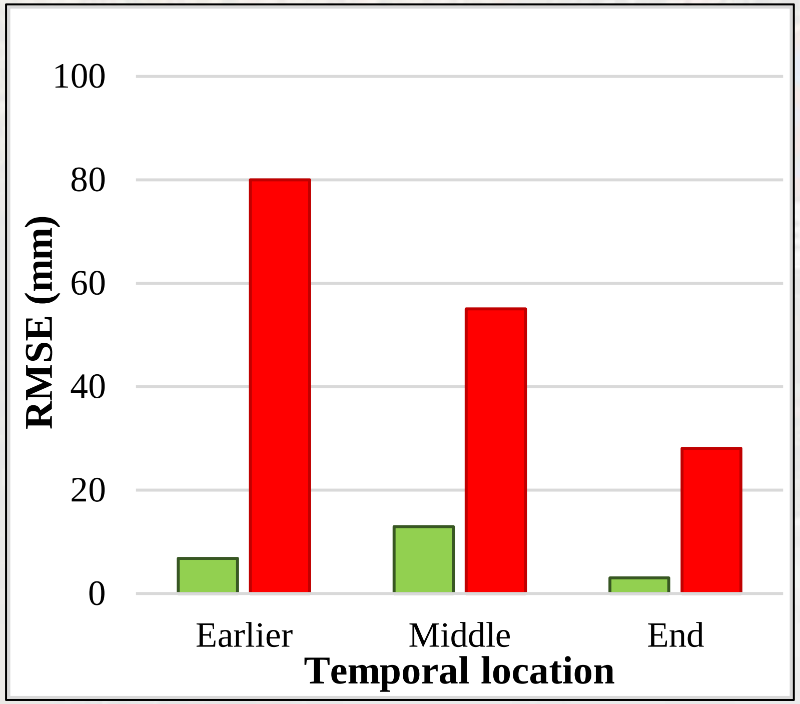


Rainfall data

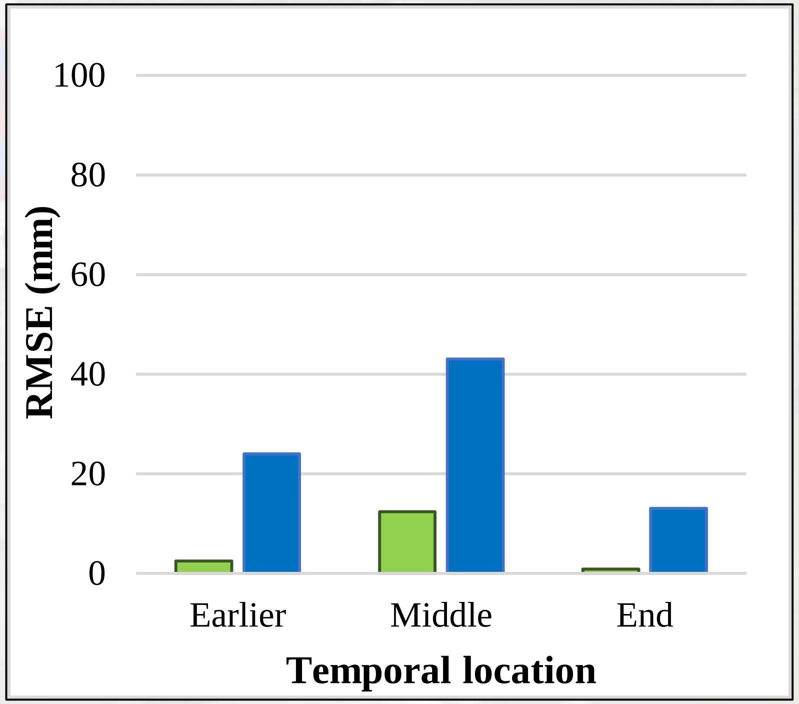


Water level data

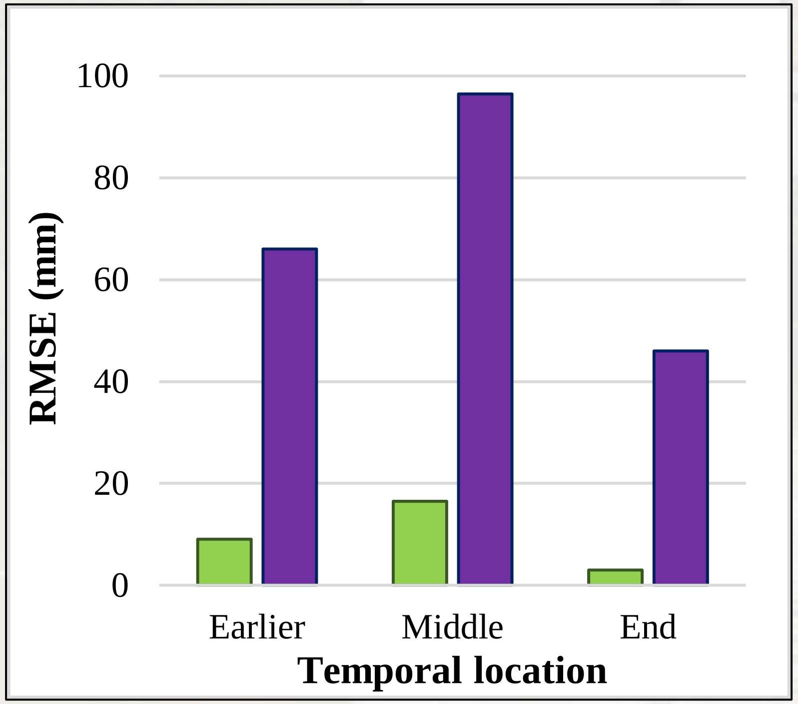
❖ *Performance of Data Imputation Methods in Flood Events*



Water level infilling



Rainfall infilling



Flood event infilling

Conclusion

01 Flexibility

Using range of data imputation methods based on temporal location of missing data in flood events

02 External benchmark

Huge advantages in real-time operation, especially at earlier stages of water level uprising and middle stages of flooding

03 Accuracy

Significant advantages when both rainfall and water level contain missing data


Questions?
Thank You For Your Attention