EGU General Assembly 2023
23-28 April 2023 – Vienna, Austria

UNIVERSITY OF WEST LONDON

Real-time Flood Overflow Forecasting in Urban Drainage Systems by Using Time-series Multi-stacking of Data Mining Techniques

Farzad Piadeh[1], Kourosh Behzadian[1], Albert S. Chen[2], Luiza C. Campos[3], Joseph P. Rizzuto[1]

[1] School of Computing and Engineering, University of West London, UK

[2] Centre for Water Systems, Faculty of Environment, Science and Economy, University of Exeter, UK

[3] Department of Civil, Environmental and Geomatic Engineering, University College London, UK

Dr Kourosh Behzadian
Associate Professor, University of West London
BSc, MSc, PhD, CEng, FICE, FCIWEM, C.WEM, FHEA
kourosh.behzadian@uwl.ac.uk

UNIVERSITY OF EXETER

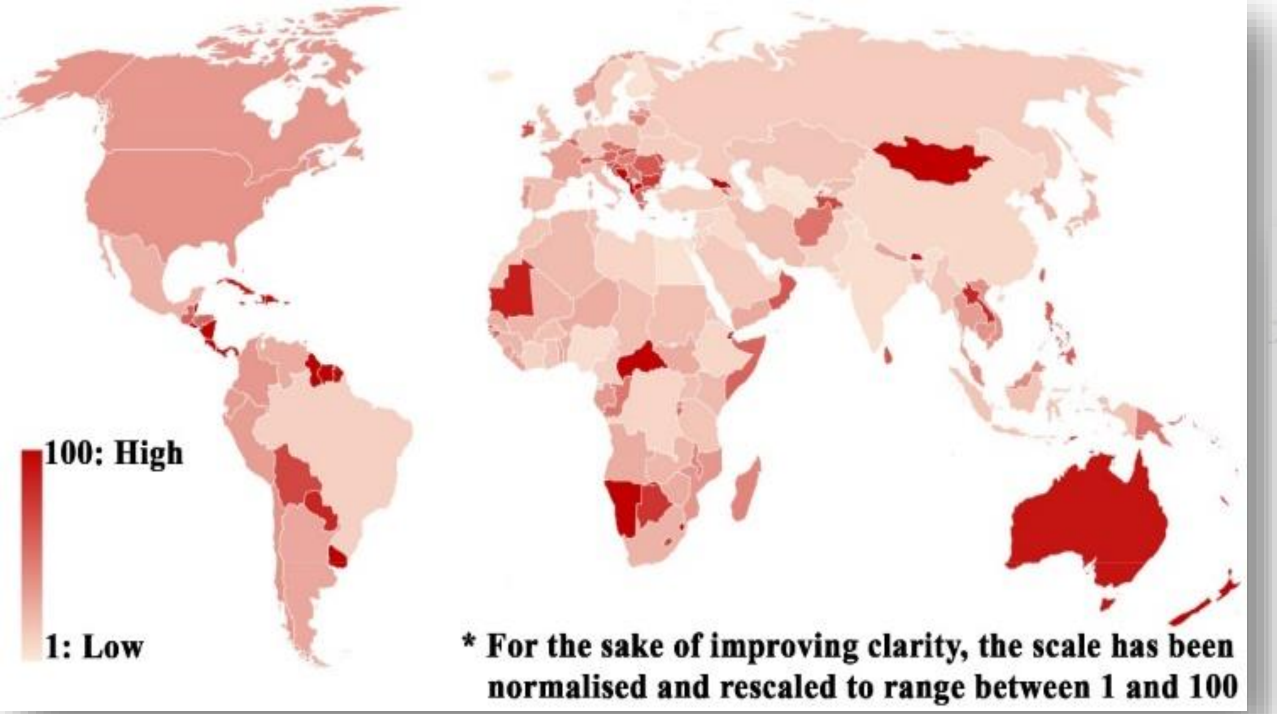
UCL

UNIVERSITY OF WEST LONDON

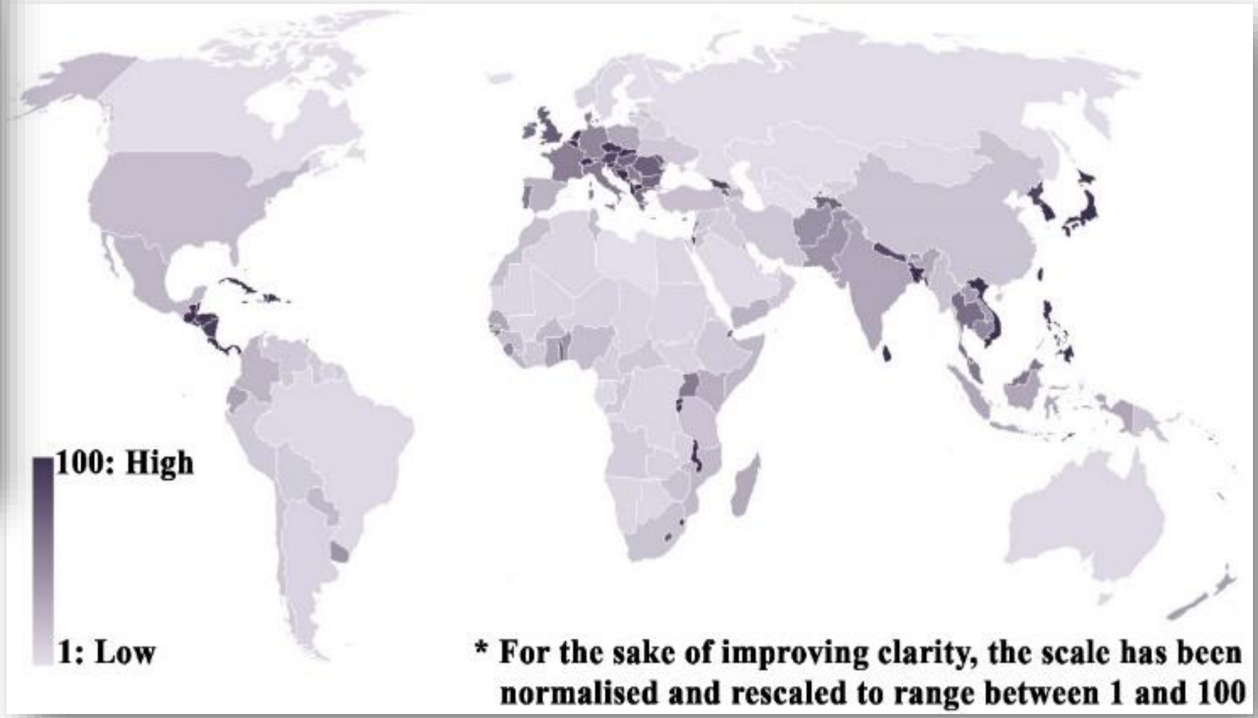
Outline

- Introduction
  - Concepts, necessity and gap finding
- Methodology
  - Defining proposed approaches
- case study/Results
  - Verifying proposed approach by real case study
- Conclusions
  - Key findings and future works

Urban flood Occurrence



Population-based



Area-based

Gaps & Aims

Research gaps for forecasting the water level in Urban Drainage Systems (UDS)

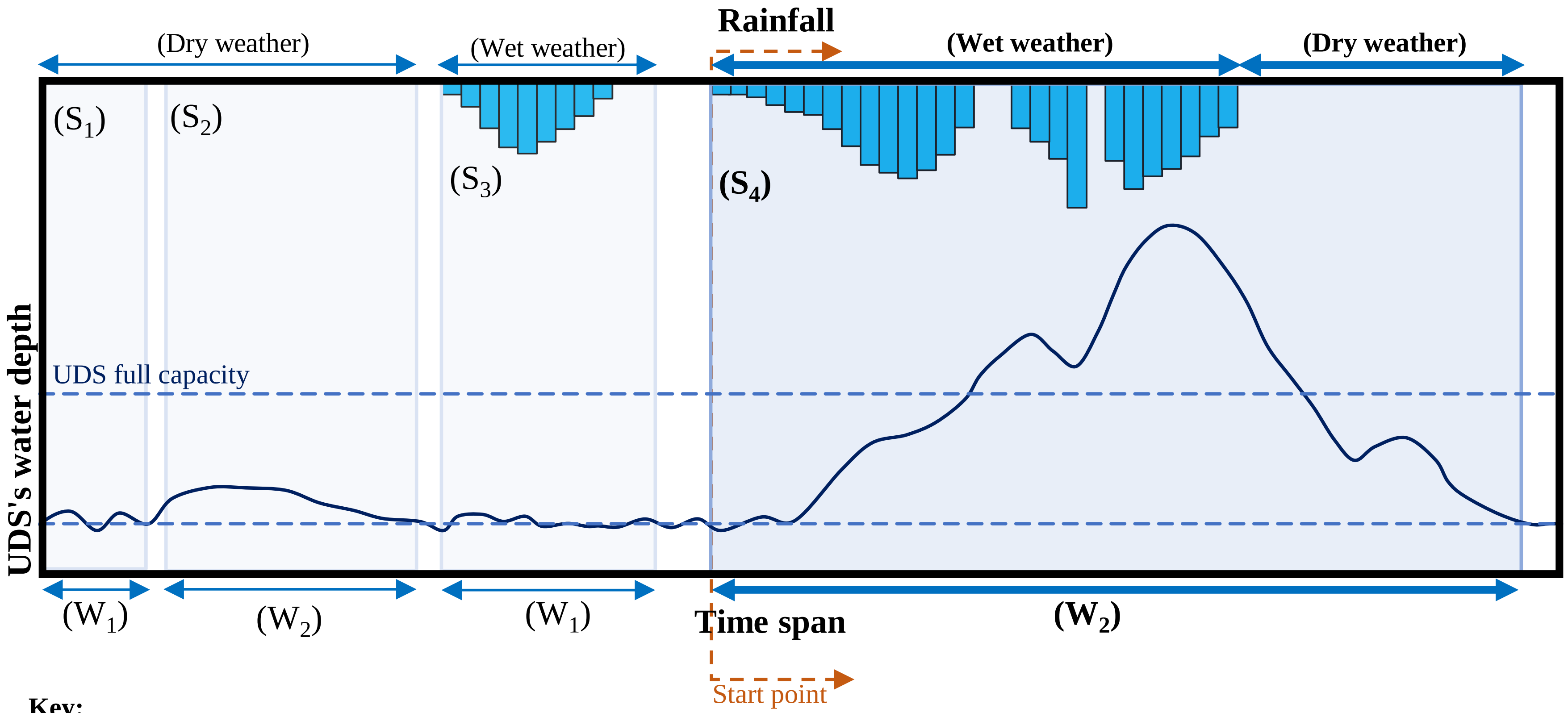
- Water Level forecasting (depth): Inaccuracy high for more than 90 min. ahead
- Classification forecasting (flood or non-flood): Inaccuracy high for more than 120 min.
- Lack of proper addressing time-series real-time operation

Research aim

Novel multi-stacking model integrating different decision tree frameworks by developing various weak learner data mining techniques and associated model performance indicators in the process of time-series blending of pre-trained stacked ensemble models.

UNIVERSITY OF WEST LONDON

Event Identification Method



Key:

State	Captured data	
	Rainfall intensity	**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.

Rainfall feature extraction

Group feature	Extracted rainfall feature	Description	Transformation key	Unit/class
Current rainfall characteristics	Duration	Time period of between the onset and end of the precipitation	Numerical	min
	Depth	Maximum water depth if all rainfall cumulated in saturated impervious surface	Numerical	mm
	Intensity	The ratio of total depth to the duration	Numerical	mm/hr
	Peak depth	Maximum rainfall intensity	Numerical	mm
Antecedent precipitation history	Occurrence	Previous rainfall occurred until maximum previous period equalled to time of concentration	Binary	0:No 1:Yes
	Average intensity	The average rainfall intensity of previous rainfall occurred until maximum previous period equalled to time of concentration	Numerical	mm/hr
Time occurrence	Season	A different class of humid temperate climate	Class	1:Dry 2:Mild 3:Rainy
	Long-term similarity	Average of past 10 years' rainfall intensity for a similar duration of current event	Numerical	mm/hr

Developing base models

- Discriminant analysis (DA)
- Decision tree (DT)
- Gaussian process regression (GPR)
- K-nearest neighbourhood (KNN)
- Naive bayes (NB)
- Supervised vector machine (SVM)

UNIVERSITY OF WEST LONDON

Developing base models

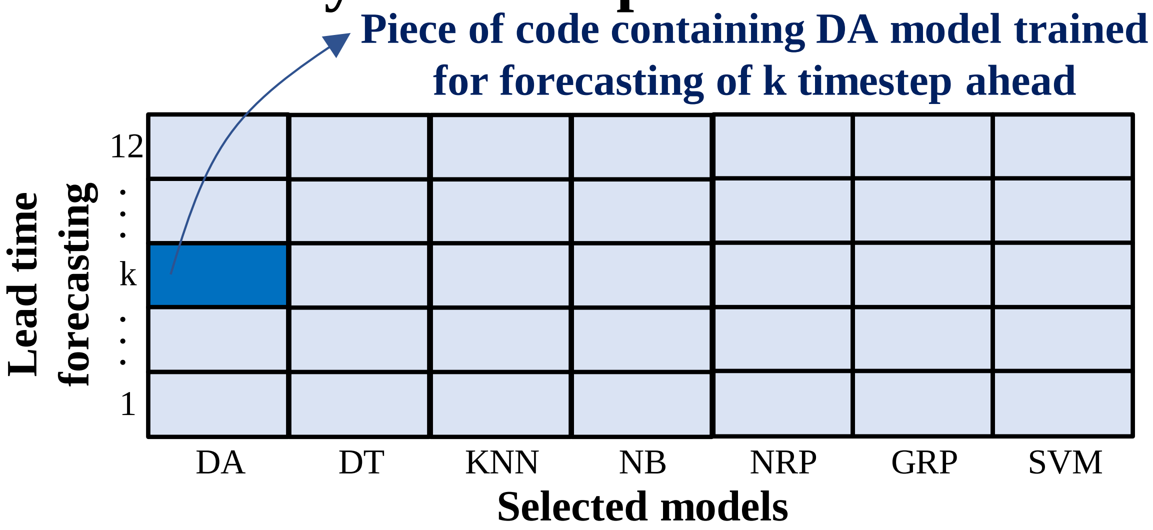
Table 2. Selected key performance indicators used for performance assessment of WLDMs

Code	Description	Formula
TPR	Model sensitivity in recalling actual flood condition, i.e., accuracy of flood class	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100$
TNR	Model specificity in selecting actual non-flood condition, i.e., accuracy of non-flood class	$\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \times 100$
ACC	Probability in that the model forecasting is correct, i.e., interested in forecasting the right classes without caring about the type of the class or class distribution	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{n}} \times 100$
MCC	Highlighting correlation and agreement between observed and predicted classes	$\frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})\times(\mathrm{TP}+\mathrm{FN})\times(\mathrm{TN}+\mathrm{FP})\times(\mathrm{TN}+\mathrm{FN})}}$
DP	Determining the likelihood of correct flood and non-flood conditions	$\frac{\sqrt{3}}{\pi} \times [\log(\frac{TPR}{1-TNR}) + \log(\frac{TNR}{1-TPR})]$
F_1 score	Revealing the best trade-off between overflow and not-flood forecasting by interpretation as a weighted average between PPV and TPR	$\frac{2\times\mathrm{TPR}\times\mathrm{PPV}}{\mathrm{TPR}+\mathrm{PPV}}$
CKR	Measuring the concordance between ACC, TPR, and TNR	$\frac{\mathrm{ACC}-[\mathrm{TPR}\times(1-\mathrm{TPR})+\mathrm{TNR}\times(1-\mathrm{TNR})]}{1-[\mathrm{TPR}\times(1-\mathrm{TPR})+\mathrm{TNR}\times(1-\mathrm{TPR})]}$

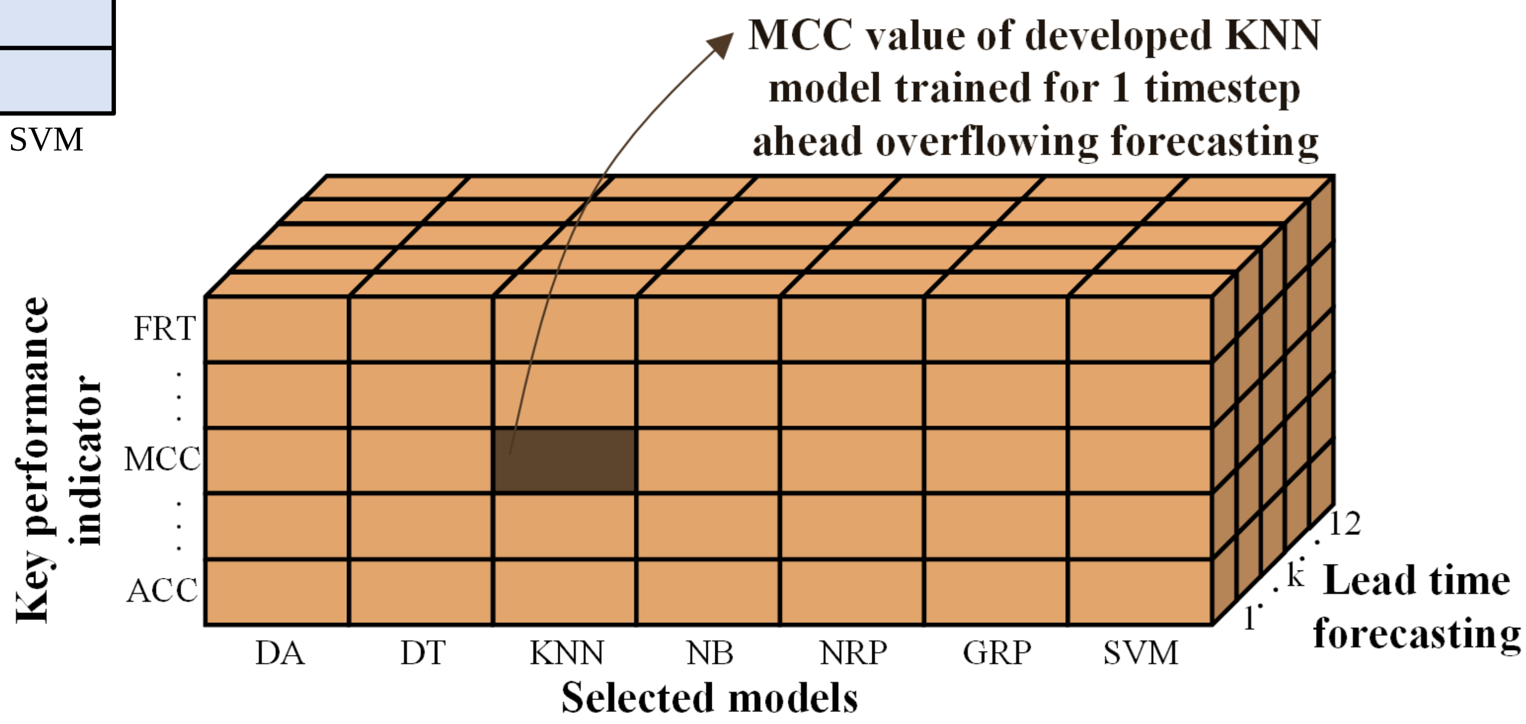
ACC: ACCuracy of true classification; CKR: Cohen's Kappa Rate; DP: Discriminant Power; F_1-score: Harmonic mean; MCC: Matthews Correlation Coefficient; PPV: Positive Predictive Value; TNR: True Negative Rate; TPR: Total Positive Rate

Data warehouse

Library of developed WLDM models



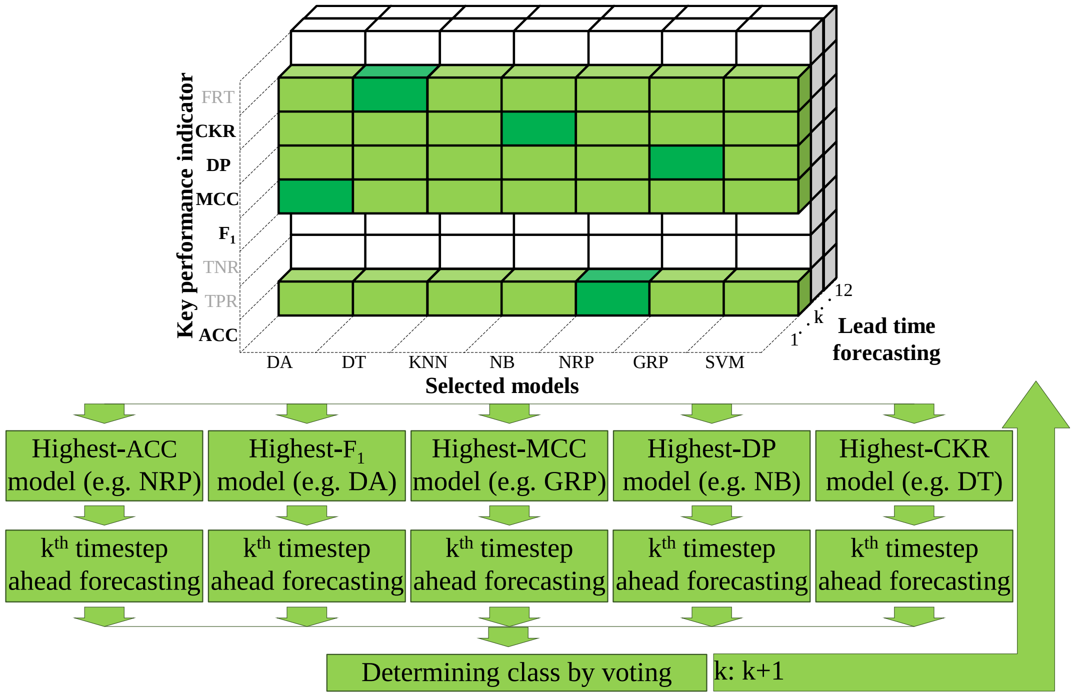
Structure of performance data cube



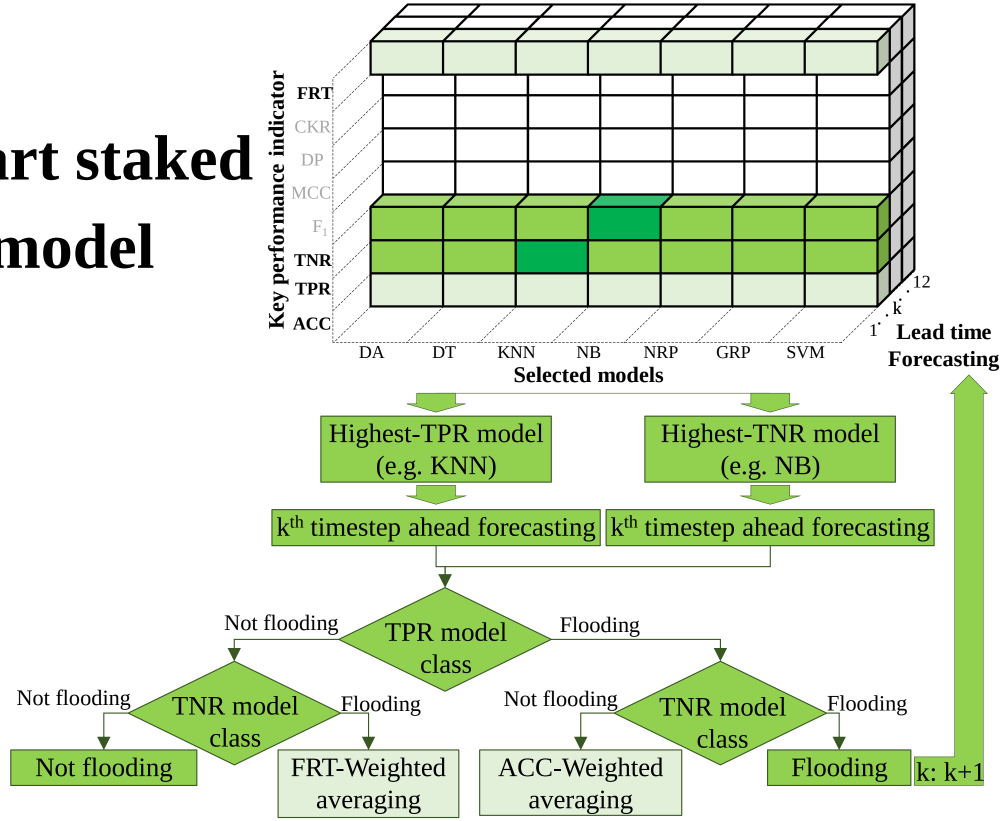
UNIVERSITY OF WEST LONDON

Developing multi-staked ensemble models

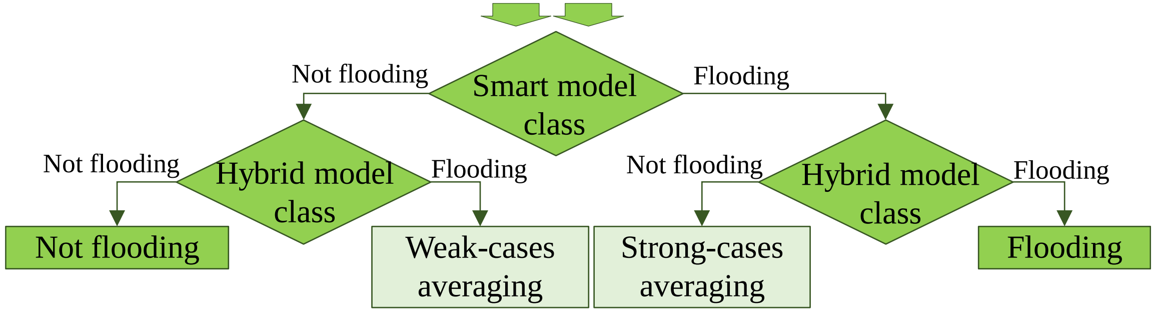
Hybrid staked model



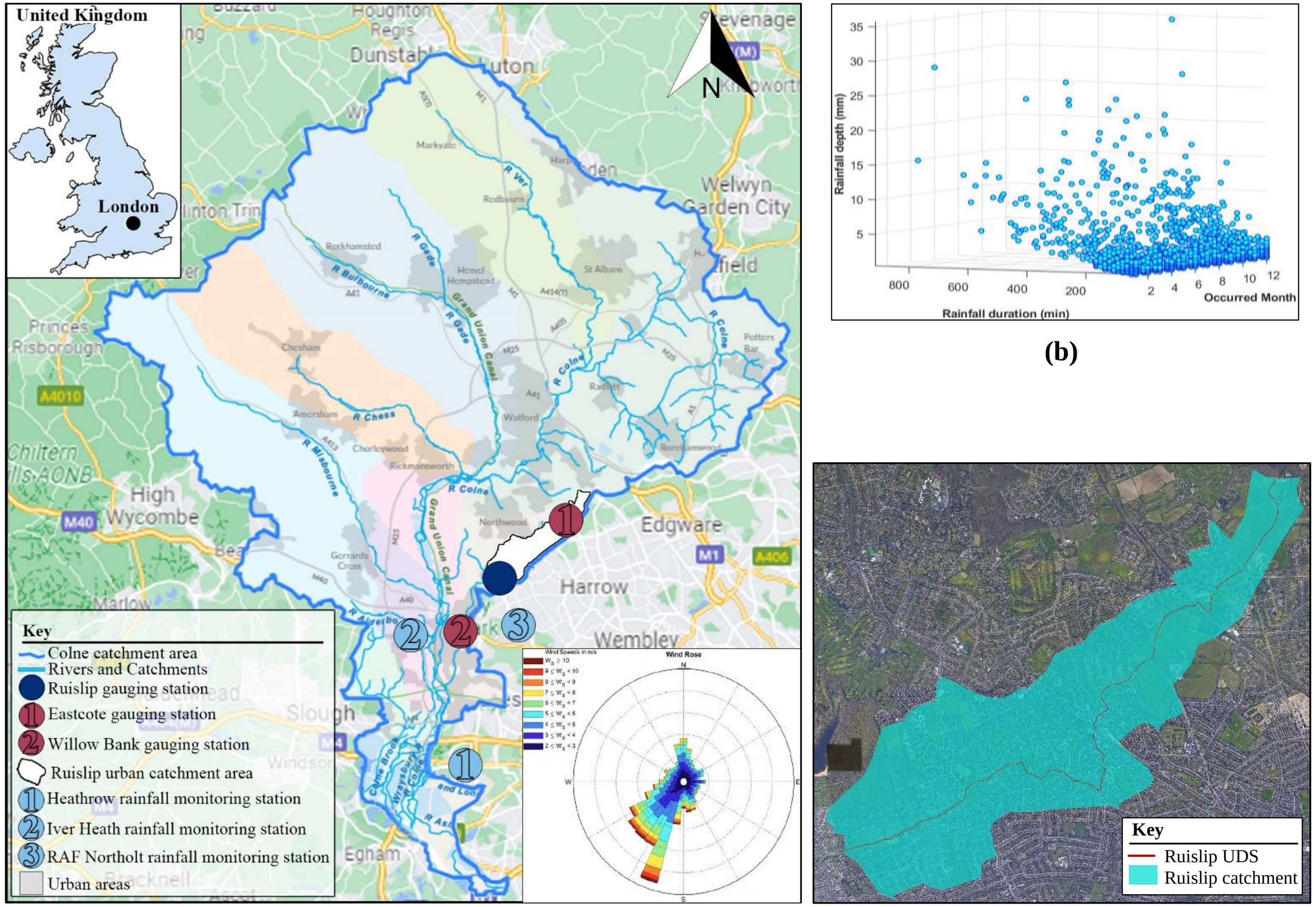
Smart staked model



Decision framework for multi-staking



Case study description



(a) (b) (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

Benchmark methods

TPR –based

Voting-based

Bayesian weighting-based

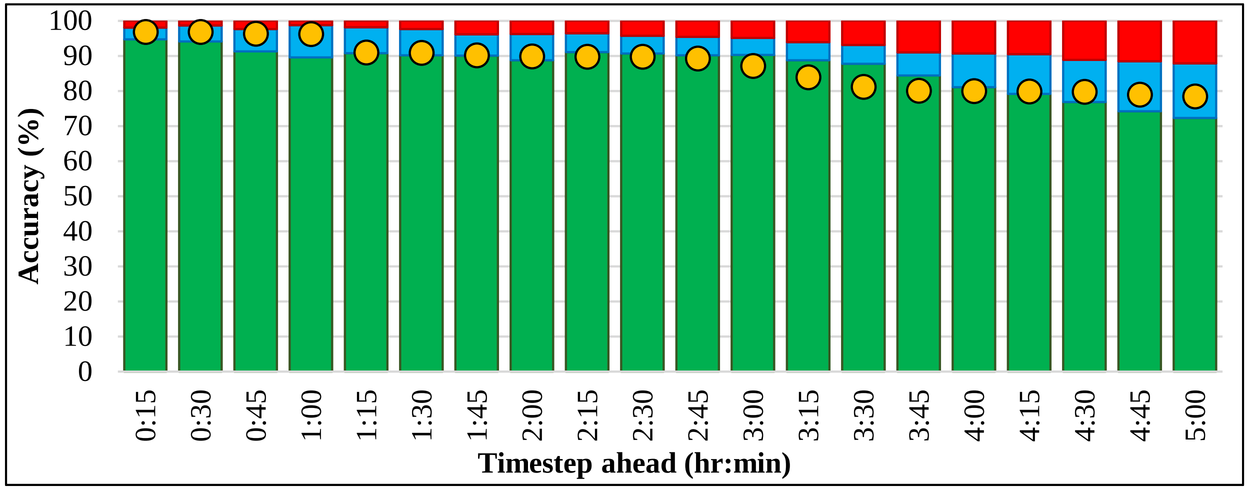
Hybrid model

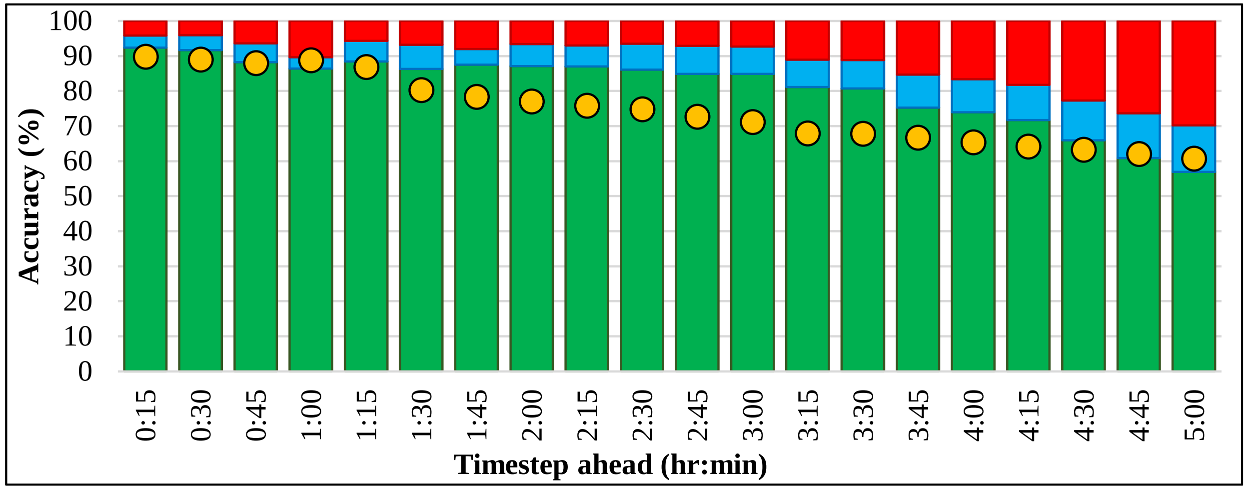
Smart model

UNIVERSITY OF WEST LONDON

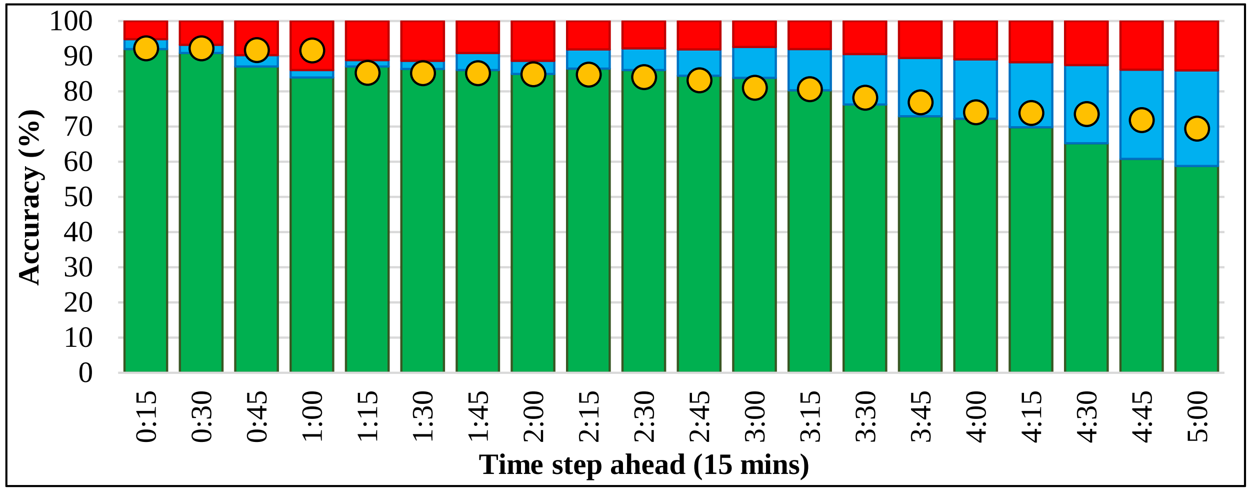
Performance of the time-series ensemble models

Proposed Model





Hybrid Model



Smart Model

Miss rate - Under estimation=FN/total events
Miss rate - Over estimation=FP/total events
Hit rate=(TP+TN)/total events
Flood forecasting=TP/(TP+FN)

Conclusions

*Comparison to best performed model in 5 hrs. ahead

01 Multi-step performance

2% improvement in miss rate
14% hit rate enhancement

02 Flood detection performance

13.5% improvement in flood detection accuracy

03 Future works

Integrating proposed categorised type model with numeral type water level forecasting in the concept of real-time early flood warning systems

Acknowledgement

Thanks for your attention!

Academic and Research team
& Funding Bodies

Q&A?