26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Harnessing the strengths of machine learning and geostatistics to improve streamflow prediction in ungauged basins; the best of both world

Vaughn Grey [a,b,c], **Tim D Fletcher** [a], **Kate Smith-Miles** [b], **Belinda E Hatt** [a,c], **Rhys A Coleman** [a,c]

[a] *School of Agriculture, Food and Ecosystem Science, University of Melbourne, Burnley VIC*
[b] *School of Mathematics and Statistics, University of Melbourne, Parkville VIC*
[c] *Melbourne Water Corporation, Docklands, VIC*
Email: Vaughn.Grey@unimelb.edu.au

Abstract: Streamflow directly influences stream conditions, such as water quality and physical form, and therefore the ability of the river to support waterway values. However, streamflow cannot be measured at all reaches across a catchment, hindering the understanding of flow dynamics (e.g. baseflows) and interpretation of flow-related variables (e.g. nutrients, total suspended solids). Modelling streamflow for reaches without measurements, generally termed prediction in ungauged basins, is therefore a key challenge of hydrology and waterway management more generally.

This study compared the performance of two contrasting statistical algorithms for predicting daily streamflow: a Long Short-Term Memory (LSTM) machine learning algorithm and a Spatial Stream Network (SSN) geostatistical algorithm. We applied these models to 114 hydrographic stations in the Greater Melbourne region of South-East Australia, a region characterised by large gradients in streamflow, land use and climate.

Both LSTM and SSN algorithms demonstrated high capability in predicting daily streamflow in a leave-one-out cross-validation. However, each algorithm had its own strengths and weaknesses, leading to variable performance over the region.

To harness the strengths of both approaches, we used a meta-modelling approach called Instance Space Analysis (ISA) to predict algorithm suitability across various catchment characteristics (Figure 1). ISA was highly effective in predicting the best-performing algorithm for ungauged reaches, correctly selecting the superior model at 86.8% of sites. This approach improved overall prediction performance compared to using either individual algorithm or as an ensemble.

The study demonstrates that harnessing the strengths of both machine learning and geostatistical approaches can significantly improve catchment wide predictions of streamflow for locations without observations. The methodology provides valuable insights into the drivers of algorithm performance.

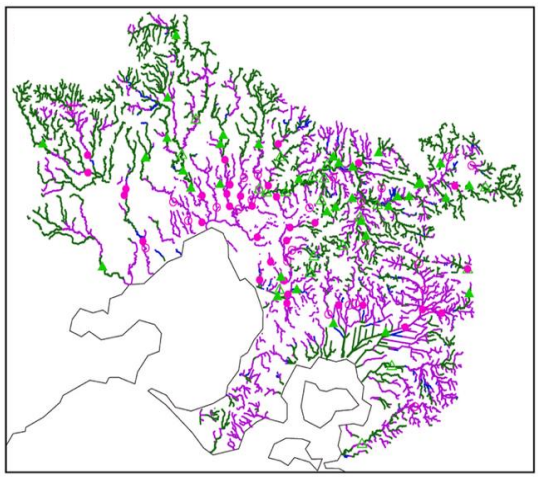

Figure 1: Prediction of model best suited for individual reaches over the stream network of the Greater Melbourne region. Pink indicates SSN model is the preferred algorithm. Green indicates LSTM is the preferred algorithm (figure from Grey et al. 2025)

The study has recently been published in the Journal of Hydrology. I will present the key aspects of the approach, findings for the Greater Melbourne region and implications for researchers and waterway managers.

REFERENCES

Grey, V., Fletcher, T., Smith-Miles, K., Hatt, B., Coleman, R., 2025b. Harnessing the strengths of machine learning and geostatistics to improve streamflow prediction in ungauged basins; the best of both worlds. Journal of Hydrology, 662. DOI:10.1016/j.jhydrol.2025.133936

Keywords: *Streamflow, predictions, machine learning, geostatistics*