Improving high-flow forecasting using dynamic multimodal feature fusion

Konstantina Theodosiadou[1], Andrew Paul Barnes[1], Thomas Rodding Kjeldsen[2]

{[1]Department of Computer Science, Faculty of Science, [2] Department of Architecture & Civil Engineering, Faculty of Engineering}, University of Bath, UK

What

We leverage **dynamic multimodal fusion methods** applying them in LSTM-based backbone models.

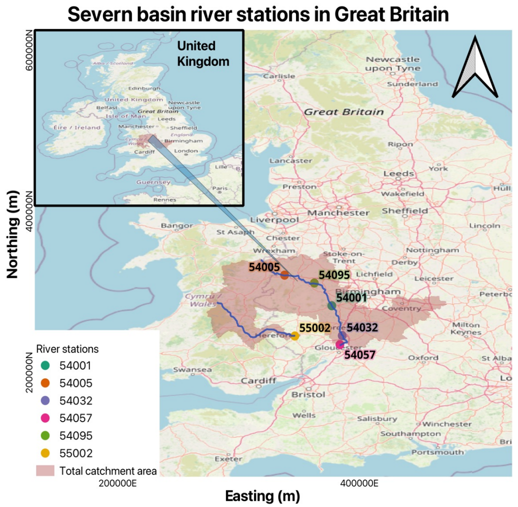
Why? For high-flow, flood-prone rivers, **improving short-term forecasts** is vital for effective warning alerts.

Wider problem: Climate change leads to more intense rainfall, imposing greater flood-risk in existing high-flow areas.

How

- We compare **operation-level** and **attention-based** dynamic multimodal fusion.
- This is tested in the neural network backbones of : **LSTM Encoder-Decoder** and **1DCNN-BiLSTM.**
- 2 model types were created: a **univariate** (station-specific) & a **multivariate** model (trained to all stations).

1. Study area



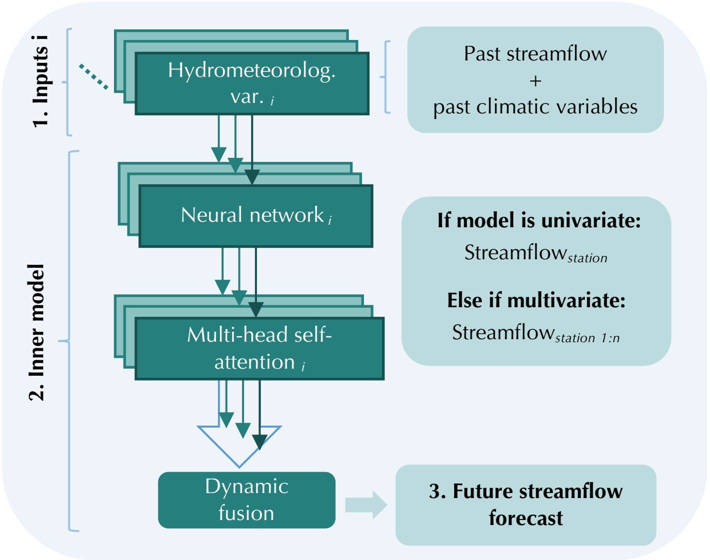
Rivers Severn and Wye are the most high-flow in Great Britain, responding intensely to rainfall in steep catchments. The average annual rainfall in the area lies between 700 and 1000 mm.

4. Conclusions

- Operation-level fusion for both univariate and multivariate models is better compared to attention-based by 3.96% for MAE, 7.40% for MAE_{HIGH}, 1.74% for NSE, 3.59% for MASE.
- Multivariate models are better by 2.86% in terms of MAE and faster by 74% than univariate, but twice more unstable.
- **Next steps**: we will add the spatial dimensions and we will focus on reducing uncertainty.

2. Model & data

Model structure



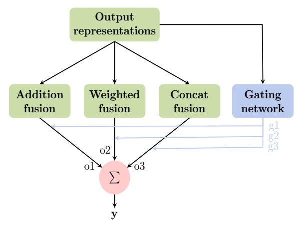
Multisource hydrometeorological data: Streamflow observations from 6 river stations (obtained from DEFRA Hydrology API) and reanalysis climatic data (obtained from ERA5 single-levels) are used.

***Climatic variables (ERA5 single-levels):**

- *Rainfall*
- *Air temperature*
- *Temperature at dew point*
- *Wind speed at u-component*
- *Wind speed at v-component*
- *Soil temperature at layer 1*
- *Soil temperature at layer 2*

Model components

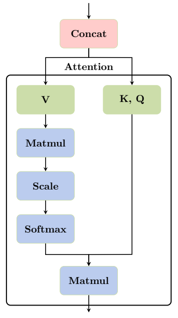
Dynamic fusion (a)
Operation-level



A gating network adaptively combines different fusion operations (e.g., addition, concatenation & weighted fusion)

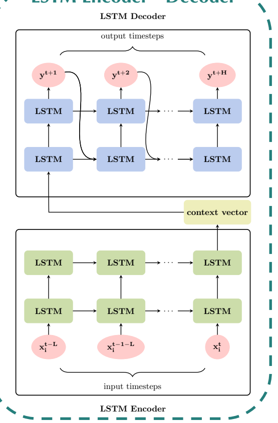
Dynamic fusion (b)
Attention-based

Inter-attention:

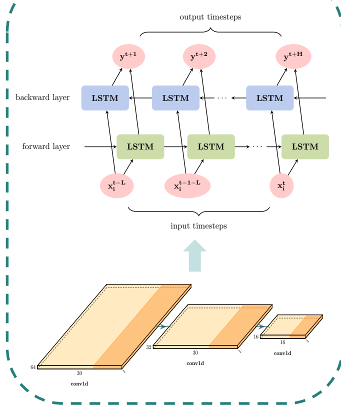


*Less memory-intensive than cross-attention

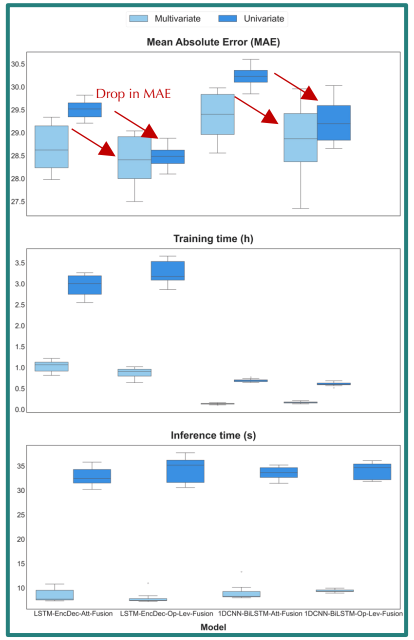
Neural network (a)
LSTM Encoder - Decoder



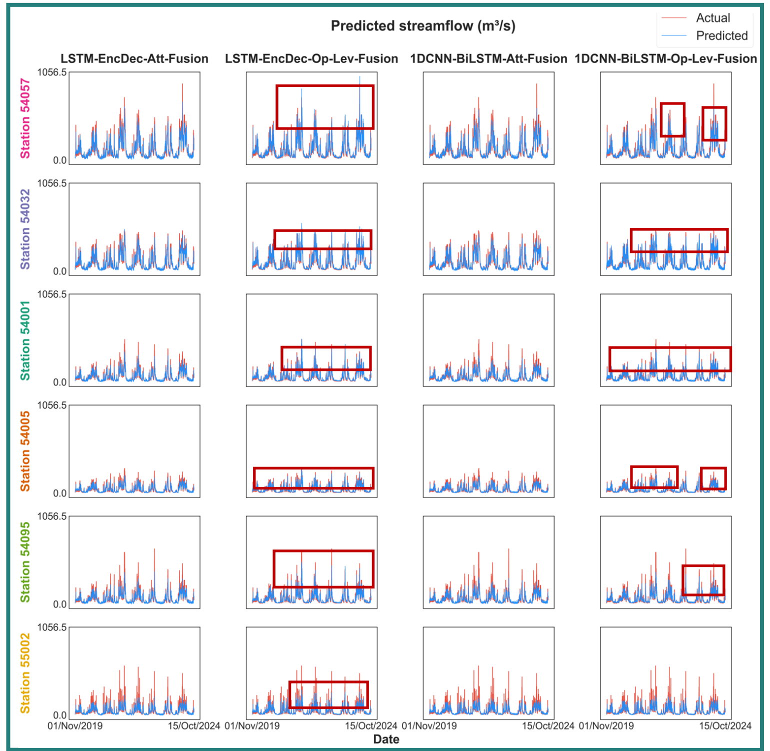
Neural network (b)
1DCNN - BiLSTM



3. Results



- Operation-level fusion is lowering MAE by 3.96% compared to attention-based fusion.
- Operation-level fusion captures better near-peak regions by 7.40% (MAE_{HIGH}) compared to attention-based fusion (red box areas below).



- According to the boxplot above, univariate models have higher MAE by 2.86% than multivariate.
- Univariate models are substantially slower than multivariate by 74%.
- But they yield twice tighter IQRs than multivariate models.

Univariate Models					Multivariate Models			
Metric	MAE	NSE	MASE	MAE_{High}	MAE	NSE	MASE	MAE_{High}
LSTM-Enc-Dec								
Att-Fusion	29.20	0.72	0.56	71.01	28.20	0.73	0.54	71.49
Op-Lev-Fusion	28.30	0.73	0.54	64.29	27.63	0.73	0.53	67.07
1DCNN-BiLSTM								
Att-Fusion	29.72	0.71	0.57	69.18	28.85	0.71	0.55	74.14
Op-Lev-Fusion	28.33	0.72	0.54	69.85	27.57	0.73	0.53	63.07

- For small catchment stations 54095 & 55002 all models underperform on the above hydrographs (latter 2 rows).
- LSTM Encoder-Decoder compared to 1DCNN-BiLSTM increase MAE and NSE a little by 1.10% and 1.37% .

Acknowledgements: This research is funded through a University Research Studentship Award from the University of Bath.