Fine Flood Forecasts

Fine-tuning global flood forecasting models on local data

Emil Ryd[1], Grey Nearing[2]
[1]University of Oxford, [2]Google Research

ML models trained on large, global datasets are SOTA for flood forecasting. The requirement for global training results in a **loss of ownership for national forecasters** who can't incorporate their own data, sometimes **preventing models from being operationally deployed**. We quantify the expected value of **fine-tuning to address this problem**.

Globally pre-trained model

Global model

National forecaster

Local data

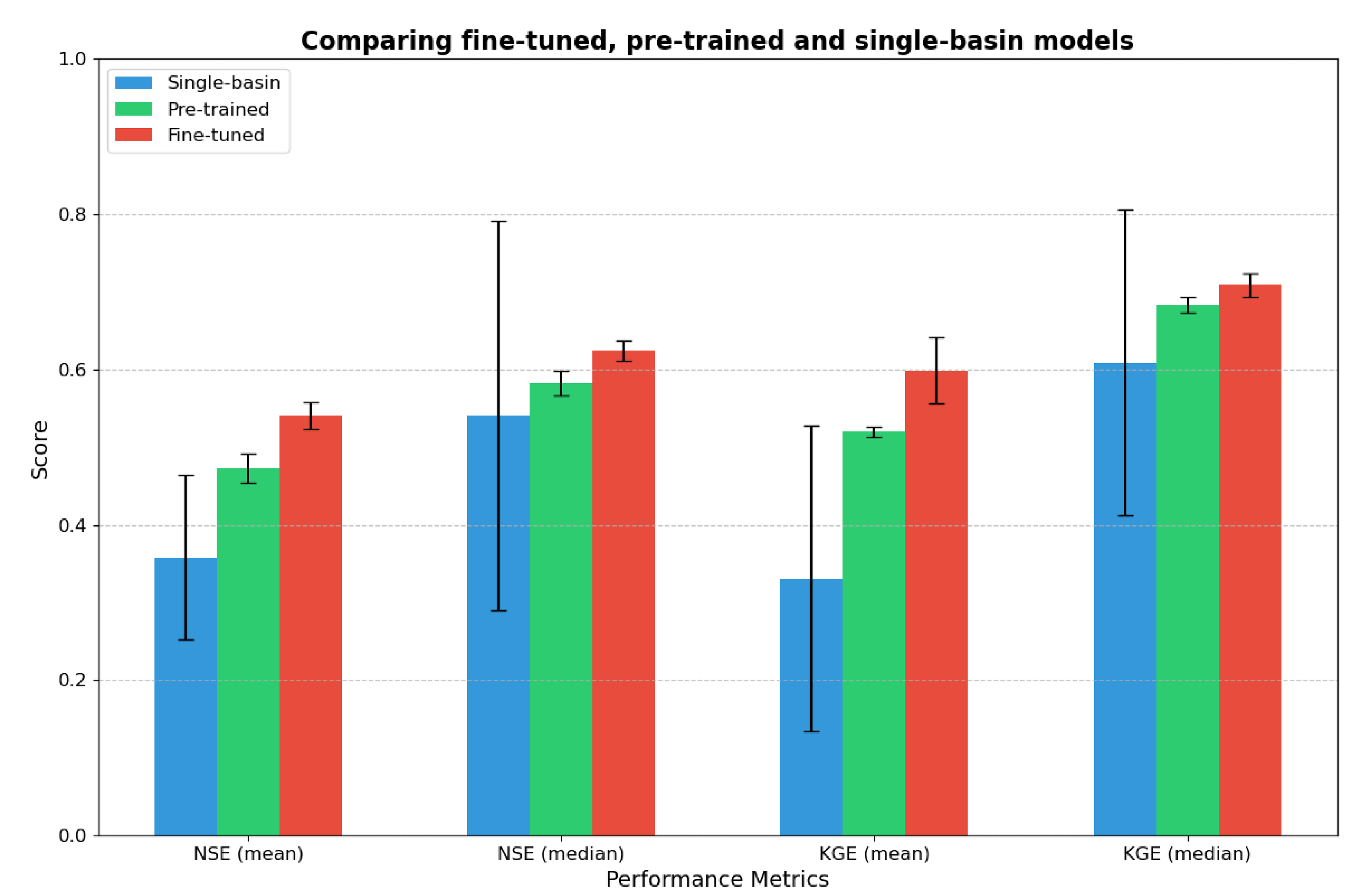
Local data of a single watershed

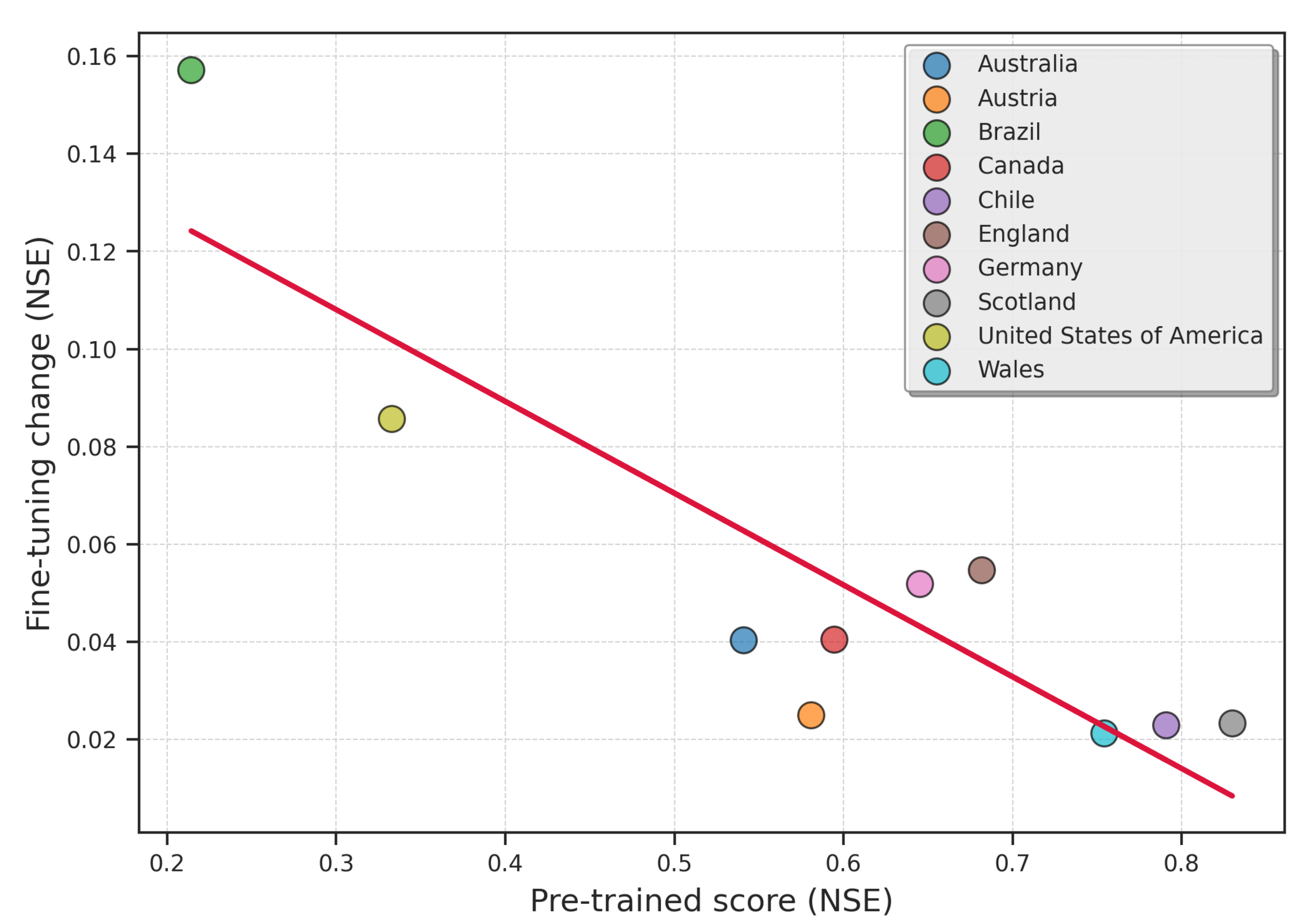
Fine-tuned local model

Our approach increases median performance† by 4-7%.

The largest previously published performance gain from incorporating local data was ~8%, using real time data assimilation.

Our method uses historical, not real time data - making it **less data demanding** for many agencies - and **can be combined with data assimilation** when real time data is available.





Fine-tuning leads to **larger improvements in underperforming locations.**

*Our model is an LSTM trained on tabular data of historical atmospheric variables (e.g. precipitation) to predict daily streamflow averages.

†We use the Nash Sutcliffe efficiency (NSE) and Kling-Gupta efficiency (KGE) which measure the accuracy of flood forecasts with a value between 0 and 1 (1 is better)

Schematic images above generated with OpenArt AI.

References

1. F. Kratzert, M. Gauch, D. Klotz, and G. Nearing. Hess opinions: Never train a long short-term memory (LSTM) network on a single basin. Hydrology and Earth System Sciences, 28(17):4187–4201, 2024.
2. G. S. Nearing, D. Klotz, J. M. Frame, M. Gauch, O. Gilon, F. Kratzert, A. K. Sampson, G. Shalev, and S. Nevo. Technical note: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks. Hydrology and Earth System Sciences, 26(21):5493–5513, 2022