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

Importance of scientific rigour in validating machine learning models: The Google global flood model example

Saman Razavi[a], Kailong Li[a,b], Holger R Maier[c], Markus Hrachowitz[d], Ehsan Nabavi[e], Natasha Harvey[f], Khaled Akhtar[g] and Fisaha Unduche[h]

[a] *School of Environment and Sustainability, and Global Institute for Water Security, University of Saskatchewan, Canada*
[b] *Desert Research Institute, Las Vegas, Nevada, USA*
[c] *School of Architecture and Civil Engineering, The University of Adelaide, Australia*
[d] *Department of Civil Engineering, Delft University of Technology, Delft, The Netherlands*
[e] *Responsible Innovation Lab, Australian National Center for Public Awareness of Sciences, Australian National University, Canberra, Australia*
[f] *Institute for Water Futures, Fenner School of Environment and Society, Australian National University, Canberra, Australia*
[g] *River Forecast Center, Alberta Environment and Parks, Edmonton, Canada*
[h] *Hydrologic Forecasting and Water Management, Manitoba Transportation and Infrastructure, Winnipeg, Canada*
Email: holger.maier@adelaide.edu.au

Abstract: Machine Learning (ML) and Artificial Intelligence (AI) are unquestionably the 'flavour of the month' in a variety of disciplines, including the earth and environmental sciences. Enthusiasm for these methods is fuelled by the widespread perception that they are a panacea for almost every problem known to humanity, resulting in broad support from government, industry and funding agencies. This has resulted in an unprecedented increase in research activity in ML and AI that invariably demonstrates the benefits of these methods, adding further fuel to the fire.

Despite the unquestioned potential of ML and AI, it is vital that we are not swept up in the accompanying excitement at the expense of following the well-trodden, but less exciting, path of applying the scientific method and following good modelling practice. By getting caught up in the hype and not viewing the results of ML and AI studies through a critical lens, we are at great risk of overstating the capabilities of these methods. This is because of the widespread tendency to ignore critical aspects of AI system functionality and to skip essential AI technology transition stages in favour of rapid, value-aligned deployment that can lead to "algorithmic harm" for affected communities. The resulting consequences can be significant, not only in terms of discrediting ML and AI methods unnecessarily, thereby setting the field back many years or even decades, but also in terms of having a range of negative real-life consequences.

One example of this is the recent paper of Nearing et al. (2024), who claim to have developed a global flood prediction model suitable for operational flood warnings, particularly for extreme events. However, our analysis clearly and unambiguously reveals multiple critical flaws that misleadingly and substantially inflate predictive skill, mask significant model errors, and thus fundamentally undermine the model's reliability for real-world deployment.

REFERENCES

Nearing G, Cohen D, Dube V, Gauch M, Gilon O, Harrigan S, Hassidim A, Klotz D, Kratzert F, Metzger A, Nevo S, Pappenberger F, Prudhomme C, Shalev G, Shenzis S, TekalignTY, Weitzneer D and Matias Y (2024) Global prediction of extreme floods in ungauged watersheds. Nature 627, 559-160, https://doi.org/10.1038/s41586-024-07145-1.

Keywords: *Artificial intelligence, Machine Learning, Model Validation, Model Credibility*

