Deep learning for model emulation – an ML/AI FSP perspective

Dan Pagendam[1], Conrad Sanderson[1], Petra Kuhnert[1], Josh Bowden[2] and Eric Lehman[1]
[1] *CSIRO Data61*
[2] *CSIRO IM&T*
Email: Daniel.pagendam@data61.csiro.au
ORCID: orcid.org/0000-0002-8347-4767

Activities like uncertainty quantification, typically require physical process models to be run many times. This ensures that the outputs of a model have taken into account potential uncertainties in parameters (parametric uncertainty), errors in data collection (measurement uncertainty) and uncertainty in process dynamics (structural uncertainty). Bayesian statistical approaches are routinely used to jointly quantify predictive uncertainties from all of these sources (Cressie and Wikle, 2011), however, they can be impractical when a model for a physical process has a very long run-time. Model emulation is one way to try and circumvent this problem. The idea is to model the computationally burdensome physical model with a faster surrogate model, referred to as the "emulator". An emulator is trained using data obtained from a set of model runs that are typically executed in parallel and may use different parameters and forcing variables.

In recent years, a number of modelling approaches have been used for model emulation. Gaussian processes have been popular since they naturally quantify the uncertainty in the emulator's predictions. However, their computational efficiency diminishes rapidly when the emulation dataset is large and this is often the case for complex models with many inputs and outputs. For this reason, first-order emulators (Hooten et al, 2011) that focus on the mean of the emulator's prediction have been advocated as an alternative solution.

In this work, we focus on how Deep Neural Networks can be used as model emulators in the environmental sciences and how these offer a number of major advantages. Firstly, these methods have good computational performance, even for large emulation datasets generated from models with many inputs and many outputs. Secondly, these models are easily structured so that they can provide a quantification of the emulator's predictive error, making them a competitor with Gaussian Processes. Finally, Deep Neural Networks seem capable of mimicking functions with sharp transitions or thresholds which we often encounter in physical models.

Acknowledgement
This research was supported by the Digiscape and AI/ML Future Science Platforms.

References
Cressie, N. and Wikle, C.K. (2011). Statistics for spatio-temporal data. John Wiley & Sons.
Hooten, M., Leeds, W., Fietcher, J., and WIkle, C. (2011). Assessing first-order emulator inference for physical parameters in nonlinear mechanistic models. Journal of Agricultural, Biological and Environmental Statistics 16(4): 475-494.