Framework for interpretable Bayesian soil moisture modelling

J. Simmons [a], V. Pino [b], A. Graaf [b] and R.W. Vervoort [a,b]

[a] *Data Analytics for Resources and Environment (DARE), ARC Industrial Transformation Training Centre, University of Sydney, Australia*
[b] *Sydney Institute of Agriculture, Eveleigh, Australia*
Email: joshua.simmons@sydney.edu.au

Abstract: Soil moisture is a key driver of hydrological processes such as flooding during wet periods and vegetation growth in dry periods, but is highly variable in time and space. Soil moisture varies due to landcover, soil type, landscape position, and rainfall input. Disentangling the different drivers and spatial relationships in soil moisture are important to deliver forecasts. To facilitate this, the most widely-used approach to simulating soil moisture is numerical modelling. More recent approaches using neural networks display good predictive performance, but lose interpretability. In contrast, numerical models are often too rigid, and quantifying uncertainty can be difficult and computationally expensive.

The objective of this paper is to demonstrate a Bayesian modelling framework that is flexible, quantifies uncertainties, disentangles the relative importance of different drivers, and is able to make forecasts.

The data are derived from a dense soil moisture observation network at Llara farm in Narrabri (NSW), installed as part of a landscape rehydration project. This project consists of two 40 ha sites, each with control areas and treatment areas. The treatment involves the installation of contour banks with 1 m elevation to reduce overland flow velocities and increase infiltration. Given their spatial distribution, each of the gauges has varying topological features and soil characteristics that may also influence the relationship of soil moisture to forcing variables (e.g., rainfall and evaptranspiration). For this modelling, we focused on 16 months of 10 minute interval data across 32 gauges at a common depth below the surface (200 mm). These data were aggregated to a daily mean to align with rainfall data from a nearby rain gauge at Llara farm and evapotranspiration data downloaded from SILO.

The model is a Hierarchical Linear Model (HLM) fit using a Bayesian approach leveraging the No-U-Turn MCMC sampler implemented in NumPyro. HLMs are well suited to modelling nested data, such as in this case where gauges can be grouped by site and treatment, alongside gauge specific factors. Soil moisture at time t (SM_t) is modelled at each gauge as a simple linear regression of the rainfall (R), evapotranspiration (E), and an autoregressive term (the soil moisture from the previous timestep, SM_{t-1}). The corresponding coefficients of the linear regression are β_R, β_E and β_{AR}, respectively, along with an intercept term (β_0). By introducing group level parameters via hierarchical priors on our β terms, we can explore the effects of site and treatment on the dynamics of soil moisture over time.

$$SM_t = \beta_{AR} \cdot SM_{t-1} + \beta_R \cdot R + \beta_E \cdot E + \beta_0$$

Here we focus on a single layer, thus leaving extension to the depth dimension for later. The model performed adequately in reproducing the observations and we achieved a mean Brier Skill Score (against a baseline of persistence) of 0.53 across the gauges. 30 of the gauges achieved a Brier Skill Score greater than 0.3. In addition, the structure of the model allowed us to quantify group differences (by treatment and site) in infiltration rate, via shifts in the distribution of the regression parameters applied to the rainfall. The relevance of the framework is that it is interpretable, quantifies modelling uncertainty, can be easily extended to include more variables, and serves a model for other environmental monitoring networks. For example, the framework could be easily applied to a groundwater observation network, or an air quality monitoring network.

Keywords: *Data science, uncertainty, NumPyro, forecasting*