23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia, 1 to 6 December 2019
mssanz.org.au/modsim2019

KEYNOTE

EXTENDED ABSTRACT ONLY

Bridging between mathematical and physically based models

Lance M. Leslie and Milton S. Speer

School of Mathematical and Physical Sciences, University of Technology Sydney
Email: Lance.Leslie@uts.edu.au

Abstract: In this presentation, the mathematical models will largely, but not exclusively, be represented by statistical and artificial intelligence techniques, whereas the physical model will be represented by process-based models. Examples of the different approaches are predictions derived from multiple linear regression or support vector machine regression, and the physical models by methods that involve existing knowledge of the underlying processes of the system that is being modelled. The mathematical models often are characterized by the lack of a physical interpretation of the reasons for the skill of the models. For example, high skill and the predictive variables are chosen by error minimization techniques, but no physical understanding of the predictor selection is explicitly provided. Hence domain knowledge is critical for understanding why the predictors are effective. On the other hand, physical models are built largely on domain knowledge, which then typically is converted to predictive formulae based on the understanding of the processes involved. In hydrology, and in many other geophysical applications the mathematical models include those derived from either archived observed data or by the mathematical representation of various conservation laws. That is, they can either be data-driven or by solving idealized governing equations. Alternatively, the physical models are derived from an understanding of the physical processes involved. Clearly, both approaches have had good success in modelling geophysical systems, suggesting the third path, which is a hybrid system based on combining mathematical and physically based models possibly could outperform either technique applied independently. Examples are presented, in both hydrological and meteorological contexts, and in a warming climate, of possible gains from using hybrid models.

Examples are from two very different locations. One is the impact of climate change on the heights of the Murrumbidgee River during 1961-2017. It is part of the Murray-Darling Basin (MDB). Its headwaters are in the mountains of southeast New South Wales (NSW), from which it flows westward until it joins the Murray River. Over the past 30 years, the Murrumbidgee River heights downstream at Hay, decreased noticeably. This steep decline is linked with decreased rainfall and increased evaporation, resulting from climate warming, in the Murrumbidgee source region wet season rainfall. A permutation test showed a statistically significant difference between the mean river heights at Hay between the periods 1961-1988 and 1999-2017 (p-value =0.032). A number of attributes were candidates were assessed as predictors, for the test data set (2000-2017). These attributes included ENSO-4, Global Temperature (GlobalT), Niño4 and Indian Ocean Dipole Index (DMI). Neural Nets (NN) uses the same attributes as Linear Regression (LR), whereas Support Vector Regression (SVR) uses GlobalT, Pacific Decadal Oscillation (PDO) and DMI. Random Forest (ERF) uses GlobalT and DMI. The correlations, after cross-validation, for predicting the river heights were, in descending order, 0.69 (SVR), 0.67 (NN), 0.53 (LR) and 0.47 (RF). Possible explanations for the relative performance of the four approaches will be discussed. The second example is the Ethiopian drought, which began in 2011 and continues to the present, as the Horn of Africa's recurring drought has been influence first by an El Nino event and, more recently, by a positive phase of the Indian Ocean Dipole. The scale of the Ethiopian drought is immense, affecting at least 10 million people. The entire Horn of Africa is facing a humanitarian crisis requiring emergency international support. For Northern and Central Ethiopia, the wet season is the months of June-September (JJAS). The best Machine Learning (ML) predictors for the upcoming JJAS rainfall are the August rainfall of the predicted season, which cannot be provided by the ML methods, for which the highest weights were assigned, in order, to the August precipitation anomaly, the Indian Ocean Dipole Index (DMI), and the ENSO region Nino34 SST anomaly. However, high resolution climate model projections, from CMIP-5, provide August rainfall predictions. Hence a hybrid model combination of ML and a CMIP-5 ensemble can provide JJAS seasonal rainfall predictions, using a combination of March-May data and the August CMIP-5 prediction. The predictive skill, was 80% correctly predicted above or below normal precipitation (Probability of Detection; POD) and 20% incorrectly predicted (False Alarm Rate; FAR). Further work on predicting both precipitation and mean temperatures is planned for both regions and also globally, for other drought affected areas.

Keywords: *Drought, river heights, machine learning, climate models, hybrid models*