Bull. Aust. Math. Soc. **100** (2019), 520–522
doi:10.1017/S0004972719000893

STREAMFLOW AND SOIL MOISTURE FORECASTING WITH HYBRID DATA INTELLIGENT MACHINE LEARNING APPROACHES: CASE STUDIES IN THE AUSTRALIAN MURRAY–DARLING BASIN

RAMENDRA PRASAD

(Received 27 June 2019; first published online 15 August 2019)

2010 *Mathematics subject classification*: primary 68T99.

Keywords and phrases: soil moisture, streamflow, forecasting, machine learning, artificial intelligence, ANN, random forest, EEMD, CEEMDAN, nonstationarity, stochastic data, feature selection.

For a drought-prone agricultural nation such as Australia, hydro-meteorological imbalances and increasing demand for water resources are immensely constraining terrestrial water reservoirs and regional-scale agricultural productivity. Two important components of the terrestrial water reservoir, that is, streamflow water level (SWL) and soil moisture (SM), are imperative both for agricultural and hydrological applications. Forecasted SWL and SM can enable prudent and sustainable decision making for agriculture and water resources management. To feasibly emulate SWL and SM, machine learning data-intelligent models are a promising tool in today's rapidly advancing data science era. Yet, the naturally chaotic characteristics of hydro-meteorological variables that can exhibit nonlinearity and nonstationarity behaviours within the model dataset are a key challenge for nontuned machine learning models. Another important issue that could confound model accuracy or applicability is the selection of relevant features to emulate SWL and SM, since the use of too few inputs can lead to insufficient information to construct an accurate model while the use of an excessive number and redundant model inputs could obscure the performance of the simulation algorithm.

This research thesis focusses on the development of hybridised data-intelligent models in forecasting SWL and SM in the upper layer (surface to 0.2 m) and the lower layer (0.2–1.5 m depth) within the agricultural region of the Murray–Darling basin, Australia. The SWL quantifies the availability of surface water resources, while the upper layer SM (or the surface SM) is important for surface run-off, evaporation and

---

Thesis submitted to the University of Southern Queensland in July 2018, degree awarded on 14 December 2018; principal supervisor Ravinesh Deo, associate supervisors Yan Li and Tek Moraseni.

energy exchange at the earth–atmosphere interface. The lower layer (or the root zone) SM is essential for groundwater recharge purposes, plant uptake and transpiration. This research study is constructed upon four primary objectives designed for the forecasting of SWL and SM with subsequent robust evaluations by means of statistical metrics, in tandem with the diagnostic plots of observed and modelled datasets.

The first objective establishes the importance of feature selection (or optimisation) in the forecasting of monthly SWL at three study sites within the Murray–Darling basin. An artificial neural network (ANN) model optimised with an iterative input selection (IIS) algorithm named IIS-ANN is developed whereby the IIS algorithm achieves feature optimisation. The IIS-ANN model outperforms the standalone models and a further hybridisation is performed by integrating a nondecimated and advanced maximum overlap discrete wavelet transformation (MODWT) technique. The IIS selected inputs are transformed into wavelet sub-series via MODWT to unveil the embedded features leading to an IIS-W-ANN model. The IIS-W-ANN model outperforms the comparative IIS-W-M5 model tree, IIS-based and standalone models.

In the second objective, improved self-adaptive multiresolution analysis (MRA) techniques, ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) are utilised to address the nonstationarity issues in forecasting monthly upper and lower layer soil moisture at seven sites. The SM time series are decomposed using EEMD/CEEMDAN into respective intrinsic mode functions (IMFs) and residual components. Then the partial auto correlation function based significant lags are utilised as inputs to the extreme learning machine (ELM) and random forest (RF) models. The hybrid EEMD-ELM model yielded better results in comparison to the CEEMDAN-ELM, EEMD-RF, CEEMDAN-RF and the classical ELM and RF models.

Since SM is contingent upon many influential meteorological, hydrological and atmospheric parameters, for the third objective 60 predictor inputs are collated in forecasting upper and lower layer soil moisture at four sites. An ANN-based ensemble committee of models (ANN-CoM) is developed integrating a two-phase feature optimisation via a neighbourhood component analysis based feature selection algorithm for regression (fsrnca) and a basic ELM. The ANN-CoM model shows better predictive performance in comparison to the standalone second-order Volterra, M5 model tree, RF and ELM models.

In the fourth objective, a new multivariate sequential EEMD-based modelling is developed. The establishment of multivariate sequential EEMD is an advancement of the classical single-input EEMD approach, achieving a further methodological improvement. This multivariate approach is developed to allow for the utilisation of multiple inputs in forecasting SM. The multivariate sequential EEMD optimised with a cross-correlation function and Boruta feature selection algorithm is integrated with the ELM model in emulating weekly SM at four sites. The resulting hybrid multivariate sequential EEMD-Boruta-ELM model attained a better performance in comparison with the multivariate adaptive regression splines (MARS) counterpart (EEMD-Boruta-MARS) and standalone ELM and MARS models.

The research study ascertains the applicability of feature selection algorithms integrated with appropriate MRA for improved hydrological forecasting. Forecasting at shorter and near-real-time horizons (that is, weekly) would help reinforce scientific tenets in designing knowledge-based systems for precision agriculture and climate change adaptation policy formulations.

Some of the research described in this thesis has been published in [1–4].

References

[1] R. Prasad, R. C. Deo, Y. Li and T. Maraseni, 'Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm', *Atmos. Res.* **197** (2017), 42–63.

[2] R. Prasad, R. C. Deo, Y. Li and T. Maraseni, 'Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors', *Soil Tillage Res.* **181** (2018), 63–81.

[3] R. Prasad, R. C. Deo, Y. Li and T. Maraseni, 'Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition', *Geoderma* **330** (2018), 136–161.

[4] R. Prasad, R. C. Deo, Y. Li and T. Maraseni, 'Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach', *Catena* **177** (2019), 149–166.

RAMENDRA PRASAD, Environmental Modelling and Simulation Group,
School of Agricultural, Computational and Environmental Sciences,
University of Southern Queensland, Queensland, Australia
e-mail: ramendrap23@gmail.com