Geoderma 330 (2018) 136–161

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition

Ramendra Prasad*, Ravinesh C. Deo*, Yan Li, Tek Maraseni

School of Agricultural, Computational, and Environmental Sciences, Institute of Agriculture and Environment, University of Southern Queensland, Springfield, Australia

A R T I C L E I N F O

Handling Editor: Morgan Cristine L.S.

Keywords:
Hybrid models
EEMD
CEEMDAN
ELM
Extreme learning machine
Random forest
Soil moisture forecasting
Drought-prone Murray-Darling Basin

A B S T R A C T

Soil moisture (*SM*) is an essential component of the environmental and the agricultural system. Continuous monitoring and forecasting of soil moisture is a desirable strategy to understand the soil dynamics for proactive planning and decision-making measures for agriculture and related fields. In this study hybrid data-intelligent, extreme learning machine (ELM) models are designed and explored for monthly *SM* forecasting. The chaotic, complex and dynamical behavior of *SM* can compound the accuracy of data-driven models. Consequently, two versatile, computationally efficient and self-adaptive multi-resolution utilities namely, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and the ensemble empirical mode decomposition (EEMD) algorithms are utilized to address these data non-stationarity issues, which if not resolved can lead to model prediction inaccuracies. The difference in these approaches is that, during the EEMD process, a Gaussian white noise is added to the intact (*i.e.*, unresolved) time series only, while, the CEEMDAN requires sequential additions at each decomposition phase. Integration of these multi-resolution tools with the ELM model led to the hybrid CEEMDAN-ELM and the EEMD-ELM models, that were benchmarked with random forest (RF) equivalent models. Using *WaterDyn* model's hind-simulated *SM* data, these models were applied (without any climate inputs) to forecast the upper (0.2 m) and the lower layer (0.2–1.5 m depth) soil moisture in Australia's agricultural-hub, the Murray-Darling Basin. The standalone ELM and RF model has similar computation efficiency and model performances. However, despite the implementation of computationally expensive ensemble techniques (*i.e.*, EEMD and CEEMDAN, the hybrid ensembles EEMD-ELM and CEEMDAN-ELM were highly efficient with improved performances. The research outcomes showed that the CEEMDAN-ELM model outperformed the alternative models at three (out of the seven) sites applied for upper layer *SM* forecasts, while the EEMD-ELM hybrid model was superior at all seven sites for the lower layer soil moisture forecasts. The study signifies the important role of the self-adaptive multi-resolution utility (CEEMDAN) hybridized with the ELM algorithm to potentially develop automated prediction systems for forecasting soil moisture, with potential applications in agriculture.

1. Introduction

The structure and functioning of the natural hydrological system is contingent upon soil moisture (*SM*) which is the principal regulating element of groundwater hydrology, biogeochemical balance, partitioning of the mass and energy fluxes in between land-atmosphere system (Brocca et al., 2017; Brocca et al., 2010; Petropoulos, 2014), and nutrient and greenhouse gas fluxes. On the other hand, the agricultural yield is also explicitly dependent on *SM* content and any unprecedented fluctuations could be deleterious for this volatile industry. To devise sustainable planning, and scheduling of specialized agricultural tasks, efficient and effective temporal predictive systems are essential tools. Advanced or forecasted knowledge of this important variable, *SM*, is pivotal for proactive sustainable decisions in efficient irrigation scheduling, grazing scheduling, water quality monitoring, yield predictions (Gill et al., 2006), water resource management (Zhang et al., 2017a) and soil carbon loss prediction (Rey et al., 2017). Intelligent agricultural decision support systems based on artificial intelligence used in monitoring and forecasting *SM* can provide useful and tangible solutions in enhancing sustainability and productivity of farming systems.

Envisioning this, *SM* forecast models have been established that includes empirical formulations, the water balance approach, the dynamic soil-water models, time series models, remote sensing models and neural network models (Huang et al., 2011). However, these models have limitations in practical applications. For instance, water balance, soil-water dynamic, and time series model require an intensive volume of spatial and temporal (measured) data as initialization

* Corresponding authors.
E-mail addresses: ramendra.prasad@usq.edu.au (R. Prasad), ravinesh.deo@usq.edu.au (R.C. Deo).

https://doi.org/10.1016/j.geoderma.2018.05.035
Received 7 December 2017; Received in revised form 2 March 2018; Accepted 30 May 2018
0016-7061/

conditions. In addition, the remote sensing model has a poor stability (*i.e.*, plagued by dew) while the empirical model parameters lack practical scope (Huang et al., 2011; Mahmood and Hubbard, 2004; Weimann et al., 1998). The problem is further exacerbated by the perplexing association between *SM* and its derivative factors, such as climate dynamics and geomorphologic properties (*e.g.*, topography, soil properties, vegetation type and density, depth to water table and land use) (Famiglietti et al., 1998; Zhang et al., 2017a). Moreover, sophisticated programs and rigorous optimization techniques are required for model calibration (Jain and Srinivasulu, 2004).

To surmount the difficulties, the preciseness of extreme learning machine (ELM) pioneered by Huang et al. (2004) is evaluated in forecasting *SM*-derived from the physical *WaterDyn* model (AWAP, 2016; Raupach et al., 2009). ELM is a recent state-of-the-art data intelligent model. It is convenient to use single layer feed-forward neural network (SLFN) with better generalization capability (Shamshirband et al., 2015; Sun et al., 2008). ELM has demonstrated high accuracy at a lower computational expense for forecasting water demand (Mouatadid and Adamowski, 2016; Tiwari et al., 2016), stream-flow (Deo and Sahin, 2016; Yaseen et al., 2016), wind speed (Shamshirband et al., 2015), dew-point temperature (Mohammadi et al., 2015) and evapotranspiration (Patil and Deka, 2016). The SLFN modeling framework of ELM is similar to that of the feed-forward neural network with random weights (Schmidt et al., 1992) and random vector functional-links (RVFL) (Pao et al., 1994) where the input weights and biases are also randomly assigned. ELM is occasionally referred to as a variant of RVFL (Cecotti, 2016; Scardapane et al., 2015). Yet, ELM has subtle but important variations. In comparison to feed-forward neural network with random weights, the ELM has added output biases which were lacking in the former model (Huang, 2014; Schmidt et al., 1992). In addition, there is no direct connection in between inputs and outputs in ELM, which is the case with RVFL (Huang, 2014; Pao et al., 1994; Wang and Wan, 2008). ELM also provides added versatility for implementations of various nonlinear activation and kernel functions (Huang, 2014; Shamshirband et al., 2015; Sun et al., 2008). However, the literature shows that ELM has not been fully explored in *SM* forecasting, and hybrid models of ELM integrating multi-resolution analysis are relatively scarce. One study by Liu et al. (2014) forecasted *SM* in Dookie apple orchard, Victoria, Australia using ELM and support vector machines (SVM), revealing the superiority of ELM in *SM* forecasting at a soil depth of 20, 40 and 60 cm. Yet, that study period spanned across a very short period (14 months) and apparently lacked the inclusion of significant seasonal and long-term climate dynamics derived from realistic, physically-based inputs.

To benchmark ELM, a bootstrapped-aggregated tree approach, random forest (RF) has been designed. RF has proven to yield good performance with reasonable prediction accuracy in forecasting hydrometeorological variables, such as temperature variation (Naing and Htike, 2015), wind power (Lahouar and Ben Hadj Slama, 2017) and standardized precipitation index (Chen et al., 2012a). Similar to ELM, RF is uncommon in *SM* forecasting. The study by Matei et al. (2017) used RF to forecast *SM* at soil depths of 10 cm, 30 cm and 50 cm in Transylvania plain, Romania, while no such study has been carried out in Australia so far.

Despite the ability to handle dynamicity and nonlinearity, so far no single data-intelligent approach has been able to provide aptest forecasts under erratic hydrological conditions (Yaseen et al., 2016). The chaotic, complex and dynamical behavior of pedologic and hydrological processes leads to non-stationarities (varying mean) and seasonality (changes in variance) within the model input series (Hu and Si, 2013; Kim and Valdes, 2003; Nourani et al., 2014). This behavior can compound the ability of conventional data-intelligent models in accurately simulating the soil moisture. With the insight to address this issue, two relatively new and advanced versions of empirical mode decomposition (EMD) has been utilized to resolve the embedded frequency information (*i.e.*, related to the physical structure of data) in the model inputs. Multi-resolution analyses (MRA) tool, ensemble-EMD (EEMD), was proposed by Wu and Huang (2009) and the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was proposed by Torres et al. (2011). Both aim to segregate higher frequency input series into lower frequency resolved parts to extract and isolate salient features representing the physical structure of the data. Both of these techniques have merits over conventional approaches (*e.g.*, wavelet transform (WT) (Mallat, 1989; Mallat, 1998; Nourani et al., 2014a; Nourani et al., 2009)), singular value decomposition (SVD), singular spectrum analysis (SSA) (Chau and Wu, 2010; Chitsaz et al., 2016) and principal component analysis (PCA) (Hu et al., 2007)). Among these, WT has been widely used (*e.g.*, (Anctil and Tape, 2004; Deo et al., 2017a; Deo et al., 2016; Labat et al., 2000; Nourani et al., 2009; Wen et al., 2016)). In particular, the non-decimated wavelet function (*i.e.*, maximum overlap discrete wavelet transform, MODWT) is able to retain the downsampled values but the choice of the mother wavelet with MODWT is a major concern. There is no explicit rule to select an optimal wavelet other than by an iterative trial and error process (Prasad et al., 2017). The EEMD and CEEMDAN decompositions does not require prescribed frequency bands or imposed basis functions, thus making the decomposition completely self-adaptive. This offers a significant advantage over wavelets. Both EEMD and CEEMDAN solve the 'mode mixing' issue of EMD, achieved by the addition of a Gaussian white noise to the intact (*i.e.*, undecomposed) series. EEMD has been found to reduce the difficulties in the forecasting process, by reducing the complexity of a time series (Di et al., 2014). During CEEMDAN-based decomposition, a Gaussian white noise with unit variance and noise coefficient is added sequentially at each decomposition stage. Although this does have limitations on parallel computing, the reconstruction of CEEMDAN decomposed data is complete and noise-free (Ren et al., 2015; Zhang et al., 2017b). In spite of the advantages and self-adaptability making it suited for practical applications, neither EEMD nor CEEMDAN has been broadly applied in soil moisture forecasting applications.

EEMD-based data-driven models have been explored in forecasting precipitation (Beltran-Castro et al., 2013; Jiao et al., 2016; Ouyang et al., 2016), reservoir inflows (Bai et al., 2015) and daily river data (Seo and Kim, 2016). Although these studies found that the models generated improved forecasts, very limited application of EEMD in *SM* forecasting has been carried out. Basha et al. (2015) carried out forecasting of temperature, precipitation and *SM* patterns for the United Arab Emirates using EEMD coupled Non-Stationary Oscillation Resampling (NSOR) model and compared it with Coupled Model Intercomparison Project phase 5 (CMIP5) projections. They found that the EEMD-NSOR model had a better forecasting capability. Likewise, CEEMDAN has been found to be more effective than EMD to forecast wind speed (Ren et al., 2015; Zhang et al., 2017b), power load (Palaninathan et al., 2016) and electricity markets (Afanasyev and Fedorova, 2016). In terms of estimation error, CEEMDAN was comparable with wavelet-decomposition (Afanasyev and Fedorova, 2016) but to the best of the authors' knowledge, the application of the technique is yet to be explored in forecasting *SM* at large.

The purpose of this research study is to develop a new and precisely tuned hybrid data-intelligent model overcoming non-stationarity issues in forecasting upper and lower layer soil moisture with potential for practical applications. Temporal hind-casted *SM* data generated from the physically-driven hydrological model (*i.e.*, Commonwealth Scientific and Industrial Research Organisation's (CSIRO's) *WaterDyn* model) incorporating climatic forcing (*e.g.*, solar radiation, temperature, rainfall, *etc.*) (AWAP, 2016; Raupach et al., 2009) are utilized. Two self-adaptive multi-resolution analysis methods (*i.e.*, EEMD and CEEMDAN) are embedded into an extreme learning machine (ELM) algorithm to resolve the frequencies and to unveil the physical structure of the input variable before the model is applied in the actual forecasting of soil moisture. The resulting hybrid EEMD-ELM and the CEEMDAN-ELM based ensemble models are designed and then