Short-term predictions of transient shallow groundwater level at local scale using data-driven models

Jiangyue Ju, University of Waterloo; David Rudolph, University of Waterloo; Jimmy Lin, University of Waterloo; Morteza Behbooei, University of Waterloo.

Groundwater is a critical natural resource for human activities and a vital component of the hydrological cycle. The fluctuation of groundwater level (GWL) can be induced by the interplay of several natural processes and human activities. Although long-term variations of GWL have been extensively studied, short-term fluctuations of shallow GWL at a local scale remain understudied, despite their importance in groundwater resource management, contamination forecasting, and agricultural management. Physical-based numerical models have been used to simulate groundwater dynamics; however, they are often time-consuming, expensive, and require a comprehensive understanding of such non-linear natural processes, making data-driven models a more effective and efficient approach for GWL prediction. In this paper, seven data-driven models were investigated for short-term, local scale GWL prediction using time series of hydro-meteorological data. These models include Linear Regression (LR), ARIMAX, Feed-forward Neural Networks (FFNN), Recursive Neural Networks (RNN), Random Forest (RF), K-Nearest Neighbor (KNN), and Support-Vector Regressor (SVR). Inputs for the models involves daily historical GWL, precipitation, temperature and evapotranspiration recorded from 2015 to 2018 at Mannheim site within the Alder Creek watershed in Ontario, Canada. These datasets are pre-processed with wavelet transform and divided into training and testing groups. 1-day rolling forecast is applied to replicate a practical situation for decision-makers in both the agriculture and environmental sectors. Detailed modeling framework for each model is provided and the final calibrated models were validated using both training and testing data. The results suggested that all models produced acceptable predictions, but the RNN model comparatively performed better than other models. Simpler algorithms do not necessarily underperform, instead they provided adequate predictions. Neural networks are prone to overfitting, but they can outperform other algorithms if carefully tuned. The study demonstrates the ease and effectiveness of using data-driven models for short-term GWL prediction and provides valuable insights for practitioners who aim to use data-driven models in hydrological studies.