Science of the Total Environment 599–600 (2017) 20–31

Contents lists available at ScienceDirect

Science of the Total Environment

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

Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models

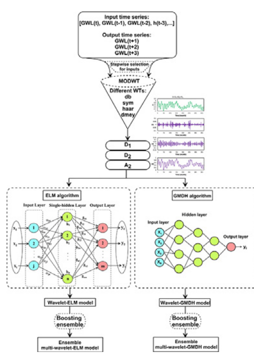
Rahim Barzegar [a,*], Elham Fijani [b], Asghar Asghari Moghaddam [a], Evangelos Tziritis [c]

[a] Department of Earth Sciences, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran
[b] School of Geology, College of Science, University of Tehran, Tehran, Iran
[c] Hellenic Agricultural Organization, Soil and Water Resources Institute, 57400 Sindos, Greece

HIGHLIGHTS

- The study is the first to couple wavelet transforms and GMDH model for the purposes of groundwater science.
- Boosting ensemble multi-wavelet GMDH and ELM models were developed for multi-step ahead forecasting of groundwater level.
- The boosting multi–wavelet models increased the performance of single wavelet based models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 February 2017
Received in revised form 24 April 2017
Accepted 25 April 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
Groundwater level
Forecast
GMDH
ELM
MODWT
Iran

ABSTRACT

Accurate prediction of groundwater level (GWL) fluctuations can play an important role in water resources management. The aims of the research are to evaluate the performance of different hybrid wavelet-group method of data handling (WA-GMDH) and wavelet-extreme learning machine (WA-ELM) models and to combine different wavelet based models for forecasting the GWL for one, two and three months step-ahead in the Maragheh–Bonab plain, NW Iran, as a case study. The research used totally 367 monthly GWLs (m) datasets (Sep 1985–Mar 2016) which were split into two subsets; the first 312 datasets (85% of total) were used for model development (training) and the remaining 55 ones (15% of total) for model evaluation (testing). The stepwise selection was used to select appropriate lag times as the inputs of the proposed models. The performance criteria such as coefficient of determination (R^2), root mean square error (RMSE) and Nash-Sutcliffe efficiency coefficient (NSC) were used for assessing the efficiency of the models. The results indicated that the ELM models outperformed GMDH models. To construct the hybrid wavelet based models, the inputs and outputs were decomposed into sub-time series employing different maximal overlap discrete wavelet transform (MODWT) functions, namely Daubechies, Symlet, Haar and Dmeyer of different orders at level two. Subsequently, these sub-time series were served in the GMDH and ELM models as an input dataset to forecast the multi-step-ahead GWL. The wavelet based models improved the performances of GMDH and ELM models for multi-step-ahead GWL forecasting. To combine the advantages of different wavelets, a least squares boosting (LSBoost) algorithm was applied. The use of the boosting multi-WA-neural network models provided the best performances for GWL forecasts in comparison with single WA-neural network-based models.

* Corresponding author.
E-mail addresses: rm.barzegar@yahoo.com, rm.barzegar@tabrizu.ac.ir (R. Barzegar).

http://dx.doi.org/10.1016/j.scitotenv.2017.04.189
0048-9697/

1. Introduction

Groundwater is an important fresh water resource for drinking, agricultural, and industrial purposes in many countries, especially in arid and semi-arid regions (Barzegar and Asghari Moghaddam, 2016; Barzegar et al., 2017b). Reliable forecasting of groundwater level (GWL) can prevent overexploiting groundwater and help more effective water resource management (Yoon et al., 2011). However, the GWL fluctuations are complicated and highly nonlinear because of complexity and uncertainties in the hydrogeological factors such as recharge, precipitation, temperature, etc. Therefore, developing effective models which can forecast the GWL accurately is essential for sustainable water resources management. The GWL forecasting can be used for evaluation of changes in GW storage, estimation of recharge rates, management of groundwater for different demands, to ensure ecological sustainability of a watershed and sustainable use of groundwater resources (Raghavendra and Deka, 2016).

Physical-based numerical models have been widely used for simulation of GWLs in several studies (Kim et al., 2008; Borsi et al., 2013). Developing the numerical models for groundwater requires collecting data such as thickness and hydraulic properties of each hydrogeological units, groundwater recharge/discharge, topography information, geological coverage, soil properties, land use map, vegetation distribution, evapotranspiration information, hydrologic and climatic data, etc. Therefore, numerical models are required a large quantity of precise data to assign physical properties of the domain and model parameters to calibrate the model simulations (Yoon et al., 2011). Also, numerical models makes many assumptions on the governing equation in porous media and are less competent in forecasting as most of the system forcings (e.g. evapotranspiration, rainfall) are less predictable. As a result, numerical models tend to produce imperfect results in spite of the perfect knowledge of the governing laws (Sun et al., 2016). Unlike the traditional physical-based models, the artificial intelligence (AI) models do not require explicit characterization of the physical properties, or accurate representation of the physical parameters, but rather simply determine the system patterns based on the relationships between inputs and outputs mapped in the training process. AI models typically use input variables that are more accessible to make predictions, and therefore circumvent the data reliance inherent to the numerical models. The advantages of ANN models over physically based models are discussed in French et al. (1992).

Over the last decades, AI models have gained considerable attention due to their advantages over numerical models and have been proved efficient in predicting complex hydrologic systems (Daliakopoulose et al., 2005; Shiri et al., 2013; Gong et al., 2016). Neural networks (NNs), as a major AI tool, have been used successfully in forecasting the GWLs due to its capability of solving complex nonlinear problems which are difficult to solve by classic parametric methods (Daliakopoulose et al., 2005; Chang and Chang, 2006; Mohanty et al., 2010; Suryanarayana et al., 2014). In addition to the use of AI techniques, researchers have started to couple wavelet transforms (WT) with the aforementioned model types. In coupled models, WT are applied as a pre-processing tool for the data before serving the data into the data driven models (Belayneh et al., 2016a). Coupled wavelet-artificial neural network (WA-ANN) models have been recently used in hydrological studies for different applications such as trend analysis (Kallache et al., 2005; Adamowski et al., 2009; Nalley et al., 2012), prediction and forecasting (Adamowski and Chan, 2011; Adamowski and Sun, 2010; Tiwari and Chatterjee, 2010; Barzegar et al., 2016) and multiscale correlation (Labat, 2005; Sang et al., 2010). The WT provides considerable information about the structure of the physical process to be modeled and has been very successful for hydrologic forecasting purposes.

An important issue in the use of the NN models is the required time for training; it can be trained by several algorithms such as the gradient descent (GD); gradient descent with momentum (GDM); gradient descent with Adaptive learning rate back propagation (GDA); one-step secant (OSS); scaled conjugate gradient (SCG); Fletcher-Reeves conjugate gradient (CGF); Powell-Beale conjugate gradient (CGB) and Polak-Ribiere conjugate gradient methods (CGP) (Sudheer and Mathur, 2012). Huang et al. (2004) introduced a single layer feedforward NN (SLFN) which is known as extreme learning machine (ELM). The algorithm uses a random process on weights to link between the input layer and the hidden layer, and the threshold of the neurons in the hidden layer. ELM model retains the advantages of fast learning, good ability to generalize and convenience in parameter selection in comparison with the back-propagation (BP) neural network, the support vector machine and other traditional models (Huang et al., 2006; Lian et al., 2014). Accordingly, a number of investigations have been carried out successfully to employ ELM algorithm for solving the problems in various scientific fields (Ghouti et al., 2013; Nian et al., 2014; Wang and Han, 2014; Wang et al., 2014; Yu et al., 2014; Wong et al., 2015).

This study applied a hybrid WA-NN approach with the use of the ELM training algorithm in forecasting GWLs. Inspired by the combined strength of NN and WT, in addition to the use of ELM, the study proposes a fast, accurate and reliable forecasting model. Furthermore, the study proposes another hybrid WA-NN model using a novel training method known as the group method of data handling (GMDH). The GMDH is an inductive learning algorithm that finds a relation between input and output variables, selecting an optimal structure of the model or network, through quadratic regression polynomials with two input variables (DeGiorgi et al., 2016). In order to solve the limitations of prestructured models, the GMDH which was originally developed by Ivakhnenko (1971), provides an objective model of a high-order polynomial in the input variables to solve prediction, identification, control, and other problems. During model self-organization, GMDH generates, validates, and selects many alternative networks of growing complexity (i.e. with increasing number of parameters, interactions between these parameters and/or nonlinearity) until an 'optimally' complex model has been found (i.e. when it begins to over-fit the design data) (Lambert et al., 2016). The use of GMDH algorithm in AI models is a growing research area (e.g. DeGiorgi et al., 2016; Lambert et al., 2016), especially in hydrological science, although its application for groundwater studies is still lacking. To the best knowledge of the authors, this study is the first to couple wavelet transforms and ANN with GMDH for the purposes of groundwater science. The novelty of this research is to develop a WA-ELM and WA-GMDH models and combine the advantages of different wavelets for groundwater levels forecasting. The objectives of this research are to study the performance of different hybrid wavelet-neural network based models and to develop ensemble multi-wavelet neural network based models for forecasting of groundwater level fluctuations.

2. Materials and methods

2.1. Group method of data handling (GMDH)

The GMDH network is a fast and adaptive learning machine which is founded on the principle of heuristic self-organizing and explicit

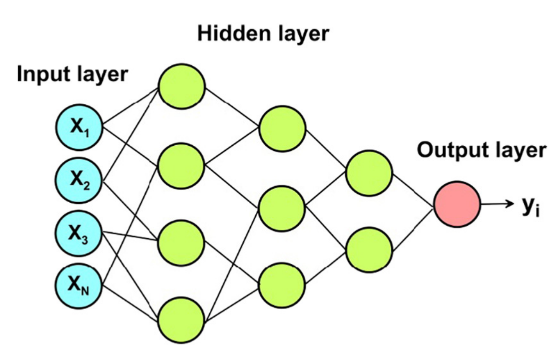


Fig. 1. An illustration of GMDH model.