Journal of Hydrology 386 (2010) 27–37

Contents lists available at ScienceDirect

Journal of Hydrology

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

Advances in ungauged streamflow prediction using artificial neural networks

Lance E. Besaw [a], Donna M. Rizzo [a,*], Paul R. Bierman [b], William R. Hackett [b]

[a] School of Engineering and Mathematical Sciences, University of Vermont, Votey Hall, 33 Colchester Ave., Burlington, VT, USA
[b] Department of Geology, University of Vermont, Delehanty Hall, 180 Colchester Ave., Burlington, VT, USA

ARTICLE INFO

Article history:
Received 18 June 2009
Received in revised form 24 December 2009
Accepted 20 February 2010

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Ungauged streamflow prediction
Artificial neural networks
Time-series analysis
Counterpropagation
Generalized regression neural network

SUMMARY

In this work, we develop and test two artificial neural networks (ANNs) to forecast streamflow in ungauged basins. The model inputs include time-lagged records of precipitation and temperature. In addition, recurrent feedback loops allow the ANN streamflow estimates to be used as model inputs. Publically available climate and US Geological Survey streamflow records from sub-basins in Northern Vermont are used to train and test the methods. Time-series analysis of the climate-flow data provides a transferable and systematic methodology to determine the appropriate number of time-lagged input data. To predict streamflow in an ungauged basin, the recurrent ANNs are trained on climate-flow data from one basin and used to forecast streamflow in a nearby basin with different (more representative) climate inputs. One of the key results of this work, and the reason why time-lagged predictions of steamflow improve forecasts, is these recurrent flow predictions are being driven by time-lagged locally-measured climate data. The successful demonstration of these flow prediction methods with publicly available USGS flow and NCDC climate datasets shows that the ANNs, trained on a climate-discharge record from one basin, prove capable of predicting streamflow in a nearby basin as accurately as in the basin on which they were trained. This suggests that the proposed methods are widely applicable, at least in the humid, temperate climate zones shown in this work. A scaling ratio, based on a relationship between bankfull discharge and basin drainage area, accounts for the change in drainage area from one basin to another. Hourly streamflow predictions were superior to those using daily data for the small streams tested due the loss of critical lag times through upscaling. The ANNs selected in this work always converge, avoid stochastic training algorithms, and are applicable in small ungauged basins.

Introduction

Accurate streamflow forecasts are an important component of watershed planning and sustainable water resource management (Brooks et al., 2003). Streams and rivers modify their channel and overflow their banks during flood events, sometimes inflicting catastrophic damage to human-built infrastructure; conversely riverine ecosystems are often most susceptible during protracted periods of low flow (Allen, 1995). The magnitude and locality of these extreme events can result in degraded surface water quality, loss of agricultural lands, damaged infrastructure, and the mobilization of phosphorus and sediment-related pollutants. Event frequency and severity are exacerbated by climate change and anthropogenic factors (Arnell et al., 2001). Accurate and timely predictions of high and low flow events at any watershed location (either gauged or ungauged) can provide stakeholders the information required to make strategic, informed decisions.

Current methods of forecasting gauged and/or ungauged streamflow fall into four categories: conceptual, metric, physics-based and data-driven. Conceptual models (*e.g.*, MODHYDROLOG) incorporate simplified conceptualizations of hydrological processes (Chiew and McMahon, 1994). Metric models (*e.g.*, IHACRES) do not rely on hydrological features or processes but rather are based on unit hydrograph theory (Jakeman et al., 1990). Physics-based rainfall-runoff models (*e.g.*, InHM) require considerable data and human effort to calibrate, validate, and test but are extremely useful in understanding the governing physics or processes (VanderKwaak and Loague, 2001). Because of the limited resources associated with developing and calibrating conceptual, metric, and physics models (Kokkonen and Jakeman, 2001), data-driven hydrological methods have been widely adopted for forecasting streamflow. Multiple linear regression (MLR), variations of autoregressive moving average (ARMA) models and artificial neural networks (ANNs) are commonly used methods (Wang et al., 2008). Although these data-driven techniques often require similar data as the aforementioned models, they require much less development time, are useful for real-time applications, and prove capable of accurately predicting stream flows (Govindaraju, 2000).

Despite the success of data-driven techniques, the number of ungauged streams greatly compounds the challenges associated with accurately forecasting streamflow. There are over 250,000

* Corresponding author. Tel.: +1 802 656 1495.
E-mail address: drizzo@cems.uvm.edu (D.M. Rizzo).

0022-1694/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2010.02.037

rivers in the US, of which less than 25,000 (<10%) are gauged daily by the USGS (Geological Survey, 2009). To accurately predict streamflow in an ungauged basin, streamflow observations must be available nearby. As an example, Mohamoud (2008) combined dominant landscape and climate descriptors from 29 catchments with multiple regression to develop flow duration curves capable of forecasting flow in nearby ungauged basins. We were only able to find one example of ANNs being adopted for forecasting ungauged streamflow (Yang et al., 2007).

In this work, we develop and test a method for predicting ungauged streamflow using two data-driven ANNs and publicly available climate and hydrologic data. A generalized regression neural network (GRNN) and a counterpropagation network (CPN) have been selected because the algorithms always converge, do not require stochastic training, and are applicable to small ungauged basins. Recurrent feedback loops are added to the CPN and GRNN algorithms, allowing future predictions to be based on time-lagged predictions (rather than time-lagged measurements). We use *predicted* flow along with locally-*measured*, time-lagged precipitation and temperature, data as model inputs. Time-series analyses are used to determine the appropriate number of model inputs (*i.e.*, precipitation lagged in time). We compare the GRNN and CPN networks with traditional data-driven methods (MLR and ARMA). We also show the importance of using correctly-scaled climate data by examining the ANN prediction accuracies with data collected on two time scales (daily and hourly). Climate and US Geological Survey streamflow records from sub-basins in Northern Vermont are used for training and testing the methods. Once trained, the predicted flows may be scaled by watershed area to allow for the prediction of streamflow in ungauged basins. To validate flow predictions in ungauged sub-basins, the ANNs are trained on climate-flow data from one sub-basin and used to forecast streamflow in a nearby sub-basin with alternate (nearest, and therefore more representative) climate inputs. Results reveal that predicting with climate data from nearby sub-basins produce accuracies that are not statistically different than those attained when training and predicting in the same sub-basin.

Background

Using an abundance of data from government sources (*e.g.*, US Geological Survey and National Climatic Data Center), data-driven methods are readily applicable and needed to model complex climate-flow relationships in all geographic regions (Walker et al., 2003). Over the past two decades, numerous data-driven ANN algorithms have been used for simulating and forecasting hydrological applications, including feed-forward backpropagation (FFBP) (Govindaraju and Ramachandra, 2000; Chang et al., 2002; Connor et al., 1994), radial basis function (Kisi, 2008; Moradkhani et al., 2004; Singh and Deo, 2007), self-organizing maps (Hsu et al., 1995, 2002), and the adaptive neuro-fuzzy inference system (Chang and Chen, 2001; Chang et al., 2001; Firat, 2008; Firat and Gungor, 2008).

Despite the diversity of ANN algorithms, the multilayer perceptrons (MLP) and the feed-forward backpropagation (FFBP) algorithms are, by far, the most common (accounting for more than 90 of the published applications). Both algorithms have been used to predict streamflow (*e.g.*, Khalil et al., 2005; Maier and Dandy, 2000; Rajurkar et al., 2002; Zealand et al., 1999) and more recently, the FFBP has been shown superior for predicting total sediment load concentration when compared to total sediment transport equations (Emrah et al., 2007), multi-linear regression (Alp and Cigizoglu, 2007; Rajaee et al., 2009), and conventional sediment rating curve models (Rajaee et al., 2009).

Unfortunately, the MLP and FFBP algorithms noted above: (1) require stochastic training, (2) do not always converge (*e.g.*, become trapped in local minima during training), and (3) are widely considered black-box approaches to hydrological modeling (Kingston et al., 2005). These challenges make their application (or transferability to other geographic locations) difficult for users not familiar with ANNs.

To circumvent the above challenges, we focus this research on two ANN algorithms that guarantee convergence (*i.e.*, find the correct weights) and are not stochastic in nature (*i.e.*, do not require iterative training procedures): the counterpropagation network (CPN) and the generalized regression neural network (GRNN). These algorithms have been used in a small number of studies to forecast streamflow (Aytek et al., 2008; Chang and Chen, 2001; Chang et al., 2001). The GRNN was found to outperform FFBP ANN methods when predicting daily (Cigizoglu, 2005a) and monthly streamflow (Cigizoglu, 2005b; Kisi, 2008). However, all previous studies (32 of 33 papers referenced in this manuscript), including those that use CPN and GRNN, use time-lagged flow *measurements* as model inputs. The one notable exception is Wang et al. (2006), who use modified FFBP ANNs to predict streamflow with a 1–10 day lead-time. Using measured streamflow is not suitable for our application, since the end goal is to model discharge in ungauged basins. The use of *predicted* discharge and locally available measured climate data, as model inputs, are key to successfully transferring this technology to ungauged streams.

Other data-driven methods multiple, such as linear regression and time series autoregressive moving average (MLR and ARMA), have been used prevalently throughout the literature for hydrological estimation applications (*e.g.*, Chaloulakou et al. (1999), McKerchar and Delleur (1974) and Yurekli et al. (2005)), streamflow forecasting (*e.g.*, Tangborn and Rasmussen (1976), Phien et al. (1990) and Schilling and Wolter (2005)) and to ANN model evaluation (Hsieh et al., 2003; Adamowski, 2008; Cigizoglu, 2003; Firat, 2008). The autoregressive moving average with exogenous input (ARMAX) models have been extended to incorporate precipitation data to forecast streamflow (Chang and Chen, 2001; Hsu et al., 1995). We use MLR and ARMAX to validate the predictive capabilities of our CPN and GRNN models.

Study site and available data

The Winooski River basin, located in northwestern Vermont, USA, was selected to demonstrate these forecasting algorithms because of the amount of available data (both in space and time) capturing the climate-hydrological relationship of the system. The Winooski basin ($\sim$2700 km^2) has a main branch length of 142 km, originates in the Green Mountains, and receives flow from five major tributaries before discharging into Lake Champlain (Fig. 1). The Mad River, Dog River, Little River, and North Branch and main stem of the Winooski River are all monitored by US Geological Survey (USGS) stream gauging stations, while the Huntington River remains ungauged.

The Winooski basin has a continental or Hemiboreal climate (Koppen classification Dfb), with warm, humid summers and cold winters. The average annual precipitation is about 100 cm (Hijmans et al., 2005). Basin land cover is largely forested in the higher elevations, while moderate development is primarily located in the stream valleys (Albers, 2000; Hackett, 2009). Bedrock is primarily schist and phyllite in the mountains with Cambro-Ordivician siliclastic rocks and carbonates to the west in the Champlain Valley (Doolan, 1996). There is an abundance of low permeability glacial till at elevation, with permeable and impermeable stratified glacial sediments in the valleys, and alluvium near river channels. Unconsolidated cover varies widely throughout the basin, with less material at the higher elevations and more in the valleys.