Journal of Hydrology 529 (2015) 1007–1020

Contents lists available at ScienceDirect

Journal of Hydrology

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

Entropy-based neural networks model for flow duration curves at ungauged sites

Maya Atieh, Bahram Gharabaghi *, Ramesh Rudra

School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada

ARTICLE INFO

Article history:
Received 26 May 2015
Received in revised form 27 August 2015
Accepted 31 August 2015
Available online 16 September 2015
This manuscript was handled by Geoff Syme, Editor-in-Chief, with the assistance of Ashok Mishra, Associate Editor

Keywords:
Flow duration curve
Artificial neural networks
Apportionment entropy disorder index
Ungauged sites

SUMMARY

An apportionment entropy disorder index (AEDI), capturing both temporal and spatial variability of precipitation, was introduced as a new input parameter to an artificial neural networks (ANN) model to more accurately predict flow duration curves (FDCs) at ungauged sites. The ANN model was trained on the randomly selected 2/3 of the dataset of 147 gauged streams in Ontario, and validated on the remaining 1/3. Both location and scale parameters that define the lognormal distribution for the FDCs were highly sensitive to the driving climatic factors, such as, mean annual precipitation, mean annual snowfall, and AEDI. Of the long list of watershed characteristics, the location parameter was most sensitive to drainage area, shape factor and percent area covered by natural vegetation that enhanced evapotranspiration. However, scale parameter was sensitive to drainage area, watershed slope and the baseflow index. Incorporation of the AEDI in the ANN model improved prediction performance of the location and scale parameters by 7% and 21%, respectively. A case study application of the new model for the design of micro-hydropower generators in ungauged rural headwater streams was presented.

1. Introduction

Flow duration curves (FDCs) are widely used for flow prediction as they contain information regarding the characteristics of hydrological regimes and flow variability (Castellarin et al., 2004; Booker and Snelder, 2012; Lyons and Lubitz, 2013; Casadei et al., 2014). For micro-hydropower generation, reliable FDCs are key for optimal design of generators, and are often produced using continuous historic flow data (Warnick, 1984; Metcalfe et al., 2005). Construction of FDCs is a simple task at gauged stations using the method described by Vogel and Fennessey (1994). The generation of FDCs is more challenging, however, for ungauged sites with limited to no flow data available.

The number of ungauged and poorly gauged sites around the world has been increasing (Sivaplan et al., 2003; Mishra and Coulibaly, 2010; Hrachowitz et al., 2013). To overcome the lack of flow data, methods for FDC estimation at ungauged sites have been widely researched (Shu and Ouarda, 2012; Razavi and Coulibaly, 2013); however, such methods do not address the complex nonlinear effects of physio-climatic parameters, which include climatic, land use and topographic parameters. As such, the following study presents a novel methodology for prediction of FDC at ungauged sites through the use of new spatial data generation tools, the entropy theory, and artificial neural networks.

1.1. Flow prediction methods

Current flow prediction methods at ungauged sites include the region of influence approach (Holmes et al., 2002), drainage area ratio method (Archfield and Vogel, 2010; Farmer and Vogel, 2013), kriging (Skoien and Bloschl, 2007; Castellarin, 2014), and an index flow model (Castellarin et al., 2007; Li et al., 2010). These methods are based on regionalization, which is the most common technique for predicting flow at ungauged sites. The process involves transferring data from gauged to ungauged catchments (Castellarin et al., 2004; Booker and Snelder, 2012; Razavi and Coulibaly, 2013).

Samuel et al. (2011) identified three broad categories for regionalization approaches: process-based, regression-based, and spatial proximity. Process-based methods, such as rainfall runoff models, emulate physical processes. Parameters of a developed and calibrated hydrological model are transferred from gauged to

* Corresponding author at: Room 2417, Thornbrough Building, School of Engineering, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada. Tel.: +1 (519) 824 4120x58451, cell: +1 226 971 3813; fax: +1 (519) 836 0227.

E-mail address: bgharaba@uoguelph.ca (B. Gharabaghi).

URL: http://www.uoguelph.ca/~bgharaba/ (B. Gharabaghi).

http://dx.doi.org/10.1016/j.jhydrol.2015.08.068
0022-1694/

ungauged catchments (Merz and Bloschl, 2004; Cheng et al., 2006; Oudin et al., 2008; Masih et al., 2010). These methods are time and data intensive, and as such were not explored in this study (Ahmed et al., 2013; Chapi et al., 2014; Liu et al., 2015; Asnaashari et al., 2015). Regression-based methods link flow data to catchment characteristics at gauged locations. The flow data is then estimated at ungauged locations with known catchment characteristics. These characteristics are related to flow quantiles using regional regression equations (Mohamoud, 2008; Archfield et al., 2010, 2013; Shu and Ouarda, 2012). Hashmi and Shamseldin (2014) developed equations to relate FDC and catchment characteristics using Gene Expression; however, their results did not present the sensitivity of parameters to the different flow quantiles. Spatial proximity is often used in which the nearest gauged catchment to an ungauged site is determined. Then nonlinear spatial interpolation techniques are used to relate flow data from the gauged stream to the ungauged site (Mohamoud, 2008; Archfield et al., 2010). Archfield and Vogel (2010) showed that selecting a reference gauge based on correlation between daily flow time series rather than proximity improved the estimate of daily flow at ungauged sites.

1.2. Data generation tools

There are currently many new online hydrological data tools that allow for more precise and detailed analyses of watersheds. Ontario Flow Assessment tools (OFAT III, 2013) is a new online spatial application that automates a series of labor-intensive hydrology tasks and provides a collection of data for hydrological applications in the province of Ontario (MNR, 2013). This includes the ability to delineate watersheds using a pour point method, extract the watershed's physiographic characteristics, and generate flow statistics for the selected watershed (MNR, 2013). Though, OFAT III provides data for Ontario alone, Streamstat, is a similar tool for the United States of America, developed by the U.S. Geological Survey (USGS) (Ries et al., 2008). It uses a map-based interface and point tool to delineate watersheds and generate catchment characteristics. Streamstat also estimates flow statistics for ungauged streams using regression and spatial proximity based methods and provides national coverage (Archfield et al., 2013).

The Government of Canada also provides online hydrologic data. The Water Survey of Canada provides real-time and historic hydrometric data for gauged streams across the country. This includes flow as well as water level data for active and discontinued stations (ECDE, 2012). Further, Environment Canada provides historic climate normals for weather stations across Canada (EC, 2012).

1.3. Apportionment entropy disorder index (AEDI)

The use of the entropy theory in water resource engineering is an evolving concept that has attracted much research in the past decade. Entropy is a measure of uncertainty associated with a random hydrologic process. Its application has made progress in different fields and for versatile purposes including, the assessment of model performance, parameter estimation, and derivation of functional relationships, development of FDCs, streamflow forecasting, uncertainty estimation, and assessing the efficiency of monitoring networks (Singh, 1997, 2013; Moramarco and Singh, 2010; Hao and Singh, 2011). The distribution of precipitation spatially and temporally has an integral role in influencing the hydrologic cycle. When flow regimes are altered accurate prediction of flow is affected (Coulibaly, 2006; Boyer et al., 2010; Khan and Coulibaly, 2010). Adoption of entropy for estimation of precipitation variability remains a novel and appealing approach that has not been fully unraveled. For the purpose of this study, the methodology developed by Mishra et al. (2009) was adopted to generate an index of precipitation variability, the apportionment entropy disorder index (AEDI).

1.4. Artificial neural networks

Artificial neural networks (ANNs) mimic the human brain function: they consist of highly interconnected neurons that receive information, elaborate on it through linear or nonlinear mathematical functions and then pass it to other neurons. Each link between neurons has an associated weight that is identified through training and validation processes.

There have been various studies regarding the applications of ANN in hydrology (Saliha et al., 2011; Kalteh, 2013; Sabouri et al., 2013; Trenouth and Gharabaghi, 2015). The major advantages of ANN include their capability of modeling non-linear relationships, providing flexibility and robustness in structure, and the ease of implementation. ANN models take into consideration both temporal changes in climatic conditions and the spatial variation of the watershed; therefore, they have gained popularity in flow estimation (Besaw et al., 2010; Maier et al., 2010). In many cases, ANN have outperformed other methods for prediction of flood flow, peak flow (Demirel et al., 2009) and volume of surface runoff (Mondal et al., 2012). Besaw et al. (2010) concluded that an ANN trained at one basin is capable of accurately estimating the streamflow at a nearby basin.

Three major sources of uncertainty in ANN include, model structure (parameters and architecture), input data, and output data. Previous studies addressed uncertainty associated with model parameters through the alteration, training, and testing of datasets. This involves training the network a number of times while varying initial weights and bias values. Similarly, the model architecture (type of ANN model, training time) is selected based on a trial and error approach, which optimizes the model. A key challenge for reliable adoption of ANN is embedded in the need to determine which input parameters significantly influence predictions. The sensitivity of these inputs for flow prediction is equally as important. Output uncertainty is typically addressed through prediction intervals (Kingston et al., 2005; Khan and Coulibaly, 2006; Srivastav et al., 2007; Solomatine and Shrestha, 2009; Talebizadeh et al., 2010; Khosravi et al., 2011).

Gene expression programming (GEP) has also gained much attention in improving the prediction accuracy of models (Liong et al., 2000; Makkearson et al., 2008; Sattar and Gharabaghi, 2015); however, the current study utilized ANN for FDC prediction at ungauged sites, particularly for the purpose of micro-hydropower generation. Future work will address the use of GEP for estimation of FDC at ungauged sites, and can be used to compare the accuracy of GEP and ANN.

2. Research objectives

This study aimed to overcome some of the major limitations of previous methods. First, OFAT III was employed as a simple, reliable, and accurate method of input data generation. The reliability and accuracy of OFAT III was assessed for data generation. Next, the effect of key topographic, land cover and climatic parameters on FDC was investigated. Additionally, an apportionment entropy disorder index, which addresses both temporal and spatial variability in precipitation, was introduced as a new input that has not yet been explored in current flow prediction models. Finally, this study employed complex nonlinear trained ANN model for FDC prediction at ungauged sites.