Journal of Hydrology 530 (2015) 762–784

Contents lists available at ScienceDirect

Journal of Hydrology

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

The application of Dynamic Linear Bayesian Models in hydrological forecasting: Varying Coefficient Regression and Discount Weighted Regression

Maurycy Ciupak [a], Bogdan Ozga-Zielinski [b], Jan Adamowski [c,*], John Quilty [c], Bahaa Khalil [c]

[a] Hydrological Forecasting Office, Institute of Meteorology and Water Management – National Research Institute, ul. Piotra Borowego 14, Cracow 30-215, Poland
[b] Department of Environmental Protection and Development, Faculty of Environmental Engineering, Warsaw University of Technology, ul. Nowowiejska 20, Warsaw 00-653, Poland
[c] Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Sainte-Anne-de-Bellevue, QC H9X 3V9, Canada

ARTICLE INFO

Article history:
Received 30 June 2013
Received in revised form 14 January 2015
Accepted 7 October 2015
Available online 22 October 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Dynamic Linear Bayesian Model
Varying Coefficient Regression
Discount Weighted Regression
Hydrograph
Forecasting
Stream flow

SUMMARY

A novel implementation of Dynamic Linear Bayesian Models (DLBM), using either a Varying Coefficient Regression (VCR) or a Discount Weighted Regression (DWR) algorithm was used in the hydrological modeling of annual hydrographs as well as 1-, 2-, and 3-day lead time stream flow forecasting. Using hydrological data (daily discharge, rainfall, and mean, maximum and minimum air temperatures) from the Upper Narew River watershed in Poland, the forecasting performance of DLBM was compared to that of traditional multiple linear regression (MLR) and more recent artificial neural network (ANN) based models. Model performance was ranked DLBM-DWR > DLBM-VCR > MLR > ANN for both annual hydrograph modeling and 1-, 2-, and 3-day lead forecasting, indicating that the DWR and VCR algorithms, operating in a DLBM framework, represent promising new methods for both annual hydrograph modeling and short-term stream flow forecasting.

1. Introduction

Accurate stream flow forecasting is an important part of effective and sustainable water resources management, particularly in flood-prone areas (Adamowski et al., 2013; Rathinasamy et al., 2013; Adamowski and Prokoph, 2014; Karran et al., 2014). Accurate annual stream flow hydrograph simulations are very useful tools for reservoir operations and irrigation management, and contribute to improving our understanding of watershed science and management (Adamowski et al., 2012b; Nourani et al., 2013). Similarly, accurate short-term (*e.g.*, daily) stream flow forecasts are used in flood-prone areas for flood forecasting/warning systems and can be a valuable tool in providing advanced warning of an impending flood to reduce and mitigate the impacts of flooding on human health and infrastructure.

Data-driven hydrological methods are becoming increasingly popular in stream flow forecasting applications due to their rapid development times, minimum information requirements, and ease of real-time implementation (Adamowski and Prasher, 2012; Adamowski et al., 2012a; Tiwari and Adamowski, 2013). Although they may lack the ability to provide a physical interpretation or yield insights into catchment processes, they are nevertheless able to provide relatively accurate stream flow forecasts. Traditionally, multiple linear regression and autoregressive integrated moving average methods have been used in stream flow forecasting. More recently, newer methods have been explored, including artificial neural networks, support vector regression, and Bayesian-based methods, among others (Nourani et al., 2014).

Many applications of regression methods can be found in the hydrological literature (*e.g.*, Tangborn and Rasmussen, 1976; Curry and Bras, 1980; Phien et al., 1990; Tolland et al., 1998; Diop and Grimes, 2003; Chau et al., 2005; Archer and Fowler, 2008; Liping and Binghui, 2013; Haidary et al., 2013 amongst others). Examples of hydrological applications of autoregressive-moving-average (ARMA) and autoregressive integrated moving average (ARIMA) models include McKerchar and Delleur (1974), Noakes et al. (1985), Kember et al. (1993), Sun and Koch (2001), Yurekli and Kurunc (2005), Koutroumanidis et al. (2009), and Modarres

* Corresponding author.
E-mail address: jan.adamowski@mcgill.ca (J. Adamowski).

http://dx.doi.org/10.1016/j.jhydrol.2015.10.023
0022-1694/

and Ouarda (2013); amongst others. Artificial neural networks (ANN), or simply neural networks (NN), have recently gained wide-spread use in hydrological forecasting due to their ability to model any input–output relationship regardless of the degree of nonlinearity or lack of *a priori* knowledge of the physical system. In one of the first applications of NNs to river flow forecasting, Kang et al. (1993) used NNs and ARMA models to predict daily and hourly river flows and found that NNs could be effectively used for forecasting river flows. Since then, many studies have confirmed the usefulness of NN models in river flow forecasting (*e.g.*, Hsu et al., 2002; Tawfik, 2003; Piotrowski et al., 2006; Bhadra et al., 2010; Pramanik et al., 2011; Artigue et al., 2012; among others). For instance, Kisi (2004) used NNs to forecast mean monthly stream flow for reservoir operation, while Kisi (2007) used different NN algorithms for both short-term and long-term forecasts of stream flow in the North Platte River in the United States. These studies provide evidence in support of the ability of neural networks to serve as potentially useful forecasting tools in hydrology.

1.1. Bayesian methods in hydrological forecasting

Hydrological models include many uncertainties, which, according to the principles of Bayesian inference, should be quantitatively determined in the form of a probability distribution. Quantitatively assessing uncertainty allows researchers and practitioners alike to explore uncertainty from various sources (*e.g.*, collected data, data processing, model parameter selection, etc.). Moreover, quantitatively accounting for uncertainty can provide insight into the predictive power of a particular model and how suitable it might be in hydrological forecasting applications. In a study by Krzysztofowicz (1999), a methodological foundation and operational framework for probabilistic forecasting via any deterministic hydrologic model was introduced under the name of Bayesian Forecasting System (BFS). The author noted these main characteristics of a BFS: (i) total uncertainty is the result of both input uncertainty and hydrologic uncertainty, (ii) the probabilistic forecast occurs in the form of a Bayes density, (iii) the predictive density involves posterior updates, (iv) the BFS has a self-calibration property, and (v) the BFS guards the decision-maker against notoriously poor forecasts. In further studies, Krzysztofowicz (2002) and Krzysztofowicz and Maranzano (2004a, 2004b) used a short-term Probabilistic River Stage Forecast (PRSF) and Probabilistic Stage Transition Forecast (PSTF) within a Probabilistic Quantitative Precipitation Forecast (PQPF) to develop a deterministic hydrological model for rainfall–runoff transformation. The total predictive uncertainty was decomposed into two sources: (i) the precipitation uncertainty tied to the basin mean precipitation forecast, and (ii) the hydrological uncertainty, which treats all other sources of error as an aggregate.

Bayes's theorem has been further adapted in regression analysis, *e.g.*, the static regression model derived from Bayesian inference described by Carpenter (2003) and Howson and Urbach (2005). While this application of Bayesian theory was based on normally distributed random variables, it can, however, be used for any probability distribution. More recently, Bayesian approaches have been explored in the stream flow and flood forecasting literature with a focus on improving the quality of a forecast by providing the uncertainty surrounding the model parameters used to derive predictions. Fortin et al. (2004) used Gibbs sampling in a Bayesian framework for parameter estimation using a reformulated shifting level model for detection of abrupt regime changes and forecasting of annual stream flow series using data from the Senegal River in Africa.

In a study by Todini (2008), a Model Conditional Processor (MCP) approach was introduced for the assessment of predictive uncertainty and compared to a Hydrologic Uncertainty Processor (HUP) and Bayesian Model Averaging (BMA) for flood forecasting data from the Po River in Italy. The MCP approach combined several models' forecasts via a multivariate normal space by means of the Normal Quantile Transform which allowed for the assessment of the density of the predictand, conditional on all the model forecasts at the same time. Compared to both HUP and BMA, the MCP approach was deemed more useful in reducing predictive uncertainty variance. Wang et al. (2009) used a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of stream flows at multiple sites in the Murrumbidgee River catchment in Australia. The BJP approach used a Box–Cox transformed multivariate normal distribution to model the joint distribution of future stream flows and its predictors, while Bayesian inference of model parameters was carried out by Markov Chain Monte Carlo (MCMC) simulations. Finally, for several flood events in a small basin in the Calabria region of southern Italy, Biondi and De Luca (2013) evaluated the performance of a BFS, with the aim of evaluating total uncertainty in real-time flood forecasting. Their results highlighted the importance of using different diagnostic approaches to analyze the quality of the forecast.

1.2. Dynamic Linear Bayesian Models (DLBM)

Dynamic Linear Bayesian Models (DLBM) are useful for hydrological forecasting applications because they are able to reflect changing dynamics through linear updating of state variables and parameters, in a manner akin to the highly dynamic natural phenomena that is the hydrological cycle. Since hydrological cycle components occur simultaneously at various levels of complexity, forecasting them accurately using static models is difficult. The addition of a dynamic linear process matched with Bayesian inference in the form of DLBM helps bridge this gap by performing dynamic updates to state variables and parameters, allowing for the evolving dynamics of hydrological time series to be captured and effectively modeled.

Assuming that a variable is observed at regular intervals and that some error is associated with each observation, a dynamic model can be used. The DLBM in the form of Varying Coefficient Regression (VCR) was developed by Harvey (1986) for applications in modeling multivariate time series using algorithms for univariate cases, including for the analysis of international exchange rates. The prior distribution and the likelihood are combined via Bayesian dynamic regression into a new posterior distribution for the next step. The result is a regular updating of the conditional posterior density of the regression parameters and the predictive probability distribution. Dynamic Linear Bayesian Models have been applied for ecological modeling of the concentration of the marine toxic microalga *Dinophysis* cf. *acuminate* (Soudant et al., 1997). The basic assumption was the existence of an underlying and unobservable time series for the vector parameters whose distribution was sequentially estimated. The DLBM approach allows for the time-varying influence of the covariates. The evolution in time of the regression parameters indicates scales of influence in the environmental factors, and provides a segmentation of the time series into significant and non-significant phases.

The first DLBM approach introduced in this study for applications to hydrological modeling and forecasting is the VCR. To address network security issues involving breaches or unauthorized information manipulation in computer systems, Triantafyllopoulos and Pikoulas (2002) developed and used a VCR model based on a unique approach where the unknown observational variance matrix distribution was left unspecified. Thereby free of the Wishart limitation, they were able to provide faster and more reliable forecasts. Further exploring the VCR model, Salvador and Gargallo (2004) proposed an automatic monitoring