Journal of Hydrology 540 (2016) 623–640

Contents lists available at ScienceDirect

Journal of Hydrology

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

A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network

Greer B. Humphrey [a,*], Matthew S. Gibbs [a,b], Graeme C. Dandy [a], Holger R. Maier [a]

[a] School of Civil, Environmental and Mining Engineering, The University of Adelaide, Adelaide, SA 5005, Australia
[b] Department of Environment, Water and Natural Resources, GPO Box 2384, Adelaide, SA 5001, Australia

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 7 June 2016
Accepted 14 June 2016
Available online 16 June 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Monthly streamflow forecasting
Bayesian artificial neural networks
Conceptual hydrological models
Uncertainty
Hybrid modelling
South Australia

ABSTRACT

Monthly streamflow forecasts are needed to support water resources decision making in the South East of South Australia, where baseflow represents a significant proportion of the total streamflow and soil moisture and groundwater are important predictors of runoff. To address this requirement, the utility of a hybrid monthly streamflow forecasting approach is explored, whereby simulated soil moisture from the GR4J conceptual rainfall-runoff model is used to represent initial catchment conditions in a Bayesian artificial neural network (ANN) statistical forecasting model. To assess the performance of this hybrid forecasting method, a comparison is undertaken of the relative performances of the Bayesian ANN, the GR4J conceptual model and the hybrid streamflow forecasting approach for producing 1-month ahead streamflow forecasts at three key locations in the South East of South Australia. Particular attention is paid to the quantification of uncertainty in each of the forecast models and the potential for reducing forecast uncertainty by using the hybrid approach is considered. Case study results suggest that the hybrid models developed in this study are able to take advantage of the complementary strengths of both the ANN models and the GR4J conceptual models. This was particularly the case when forecasting high flows, where the hybrid models were shown to outperform the two individual modelling approaches in terms of the accuracy of the median forecasts, as well as reliability and resolution of the forecast distributions. In addition, the forecast distributions generated by the hybrid models were up to 8 times more precise than those based on climatology; thus, providing a significant improvement on the information currently available to decision makers.

1. Introduction

Accurate and reliable monthly streamflow forecasts can be extremely valuable for the proper management and allocation of water resources, particularly in a highly variable climate where the historical data alone have limited value in supporting decision making. This is the case in the South East of South Australia, where water resources are under pressure from changing land uses, yet highly variable flow regimes make these resources difficult to predict and manage. However, the competing environmental and agricultural demands on water resources in this region mean that the optimal management of flows is needed in order to ensure maximum benefit is derived from the water that is available (Gibbs et al., 2014).

In monthly streamflow forecasting, two sources of predictability are typically exploited: catchment conditions (wetness) at the time of the forecast and the effect of climate over the forecast period (Pokhrel et al., 2013). As discussed in Wang et al. (2009), there are essentially two distinct approaches for doing this. The first involves the use of hydrological models that are driven by dynamic climate forecasts (i.e. forecasts of rainfall and other weather variables) and represent hydrological processes related to soil water balance and the evolution of the flow to the outlet of the basin. The second approach relies on predictors for representing initial catchment conditions (e.g. antecedent streamflow or soil moisture data) and climate during the forecast period (e.g. large scale climate indices or climate model predictions), together with statistical relationships derived from data that relate these predictors to upcoming streamflows (Plummer et al., 2009; Robertson and Wang, 2012). Hydrologic models employed in the former 'dynamical' forecasting approach typically operate on a daily or sub-daily time scale and can range from simple lumped conceptual rainfall–runoff (R-R) models to more physically-based fully distributed models. Simple conceptual R-R models have been widely employed for modelling streamflow in Australia, as they generally provide

* Corresponding author.
E-mail address: greer.humphrey@adelaide.edu.au (G.B. Humphrey).

http://dx.doi.org/10.1016/j.jhydrol.2016.06.026
0022-1694/

good prediction accuracy, provided good climate data are available, and have relatively few parameters to calibrate (see Boughton (2005) for a review of hydrological models used in Australia and Wang et al. (2011b) for a discussion on the use of such models for monthly streamflow forecasting). These models attempt to explicitly simulate the dominant processes occurring within a hydrological system through the simplified representation of the system, typically as a series of conceptual water stores with simple empirical relationships used to describe the recharge and depletion processes that occur within and between them (Jain and Srinivasulu, 2006; Kokkonen and Jakeman, 2001). In the 'statistical' flow forecasting approach, on the other hand, system response is characterised primarily through the extraction of information implicitly contained in a set of observed data (e.g. monthly totals or averages), without directly taking into account the physical processes occurring within the hydrological system (Kokkonen and Jakeman, 2001; Toth and Brath, 2002).

A perceived advantage of the dynamical forecasting approach, when compared with statistical streamflow forecasting models, is that given the physical basis of the hydrological models, they are able to capture catchment dynamics that predictors used in statistical models cannot (e.g. those related to catchment wetness). Therefore, they should be more faithful in simulating the rainfall-runoff process (Chen and Adams, 2006; Robertson et al., 2013). This is particularly considered to be the case under nonstationary conditions where 'past' predictor–response relationships derived by statistical models may no longer represent those at the time of the forecast (Wang et al., 2011a). However, the transformation of rainfall into runoff is an extremely complex, dynamic, and nonlinear process (Hsu et al., 1995), which can be difficult to fully understand and represent, particularly by means of a simple, conceptual model. Furthermore, similar to statistical forecasting models, hydrological models generally require calibration using historical rainfall and streamflow data. The choice of the calibration period and its length can have a significant impact on the estimated conceptual model parameters and, hence, the relationships modelled. In addition, dynamical forecasting models generally require statistical post-processing to remove systematic biases and to quantify uncertainty not represented directly by the calibrated model (Robertson et al., 2013).

While knowledge-based hydrologic models are important for understanding hydrological processes, the main concern in many practical applications of monthly streamflow forecasting models is the accuracy and reliability of the forecasts; therefore, in such situations, statistical forecasting may be more suitable. These models do not require explicit consideration of the processes occurring within a hydrological system and, therefore, are not limited by an incomplete or unsuitable description of the complex R-R transformation processes as simple hydrological models may be. Furthermore, in contrast to conceptual R-R models, which typically require daily rainfall and potential evapotranspiration (PET) data as inputs, statistical models are generally not based on a prescribed (and possibly limited or prohibitive) set of input information, but rather they are able to take advantage of whatever relevant data are available. The potential to include auxiliary data, for example, those related to possible land use and climate impacts, may allow statistical streamflow forecasting models to characterise changes in the hydrological behaviour of a catchment that cannot be easily represented by simple conceptual R-R models (provided data are available that describe the change in rainfall-runoff relationship; e.g. data related to changes in land use, extractions or groundwater levels). Examples of statistical streamflow forecasting approaches include linear regression and time series models (Garen, 1992; Pagano et al., 2009; Valipour et al., 2012, 2013), non-parametric fitting (Sharma, 2000), independent component analysis (Westra et al., 2008), joint probability modelling (Wang et al., 2009) and artificial intelligence based approaches such as support vector machines (SVM), fuzzy logic and evolutionary computation based methods, Wavelet-Artificial Intelligence (W-AI) models and artificial neural networks (ANNs) (see Yaseen et al. (2015) for a review of such artificial intelligence based methods).

ANNs are an extremely versatile type of data-driven model that have become widely adopted for hydrological modelling applications over the past two decades (see Abrahart et al., 2012; Maier et al., 2010). An advantage of these models over more traditional statistical modelling approaches is their flexible model structure, which enables them to capture arbitrarily complex and nonlinear input-output relationships from data without any restrictive assumptions about the functional form of the underlying process. However, despite their appeal, the performance of an ANN, like all statistical streamflow forecasting models, is highly dependent on the availability and quality of observed data. Ideally, to develop a reliable and robust ANN model, concurrent observations of all relevant predictors (i.e. those representing catchment wetness and climate effects) and the streamflow response would be required, with records sufficiently long to include a wide range of conditions; while in reality, ANN models usually need to make do with whatever data are available. For example, antecedent rainfall and streamflows are typically used as rather crude proxies for representing initial catchment wetness in ANNs and other statistical streamflow forecasting models, due to the limited availability of soil moisture observations (Robertson et al., 2013). Furthermore, in comparison with conceptual R-R models, ANNs tend to have many more parameters requiring calibration and, consequently, they are more likely to be overparameterised with respect to the available data. As such, there is a greater risk that ANNs will not be capable of producing reliable forecasts beyond the range of the calibration data, unless models are updated as new data become available (e.g. Bowden et al., 2012).

In order to improve the accuracy and reliability of monthly streamflow forecasts, it would seem opportune to integrate or hybridise dynamical and statistical streamflow forecasting models in some way so as to exploit the strengths and eliminate the weaknesses of the respective modelling methodologies, rather than continuing to choose between the individual techniques and using them in isolation (Maier et al., 2010; Srinivasulu and Jain, 2009; Mount et al., 2016). There are a number of ways in which conceptual R-R models and ANNs can and have been combined in order to take advantage of their complementary strengths. These include the use of ANNs for the statistical post-processing of conceptual R-R model outputs and their associated uncertainty (Shamseldin and O'Connor, 2001; Brath et al., 2002; Abebe and Price, 2003; Anctil et al., 2003; Shrestha et al., 2009); the replacement of runoff generation and routing algorithms within both lumped and semi-distributed conceptual R-R models with ANNs (Chen and Adams, 2006; Corzo et al., 2009; Song et al., 2012; Liu et al., 2013; Loukas and Vasiliades, 2014); and the use of non-standard outputs from conceptual R-R models to expand the predictor set for ANNs (see Abrahart et al. (2012) for a more thorough discussion). The latter approach was taken by Anctil et al. (2004), Srinivasulu and Jain (2009), Isik et al. (2013) and Noori and Kalin (2016) who incorporated simulated data including soil moisture, effective rainfall, surface runoff and infiltration depths, baseflow and stormflow information derived from conceptual models into ANNs used for forecasting daily river flows. Similarly, Nilsson et al. (2006) used information about soil moisture and snow accumulation derived from a conceptual R-R model as auxiliary inputs to an ANN used for simulating monthly runoff. Recently, Rosenberg et al. (2011) and Robertson et al. (2013) investigated the benefits of hybrid seasonal forecasting systems where outputs from hydrological models were used as predictors in a statistical forecasting system (although not ANN-based). In both cases, it was found that