Journal of Hydrology 508 (2014) 227–239

Contents lists available at ScienceDirect

Journal of Hydrology

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

Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments

D.J. Booker [a,*], R.A. Woods [b]

[a] National Institute of Water and Atmospheric Research, P O Box 8602, Riccarton, Christchurch, New Zealand
[b] Department of Civil Engineering, University of Bristol, Bristol, UK

ARTICLE INFO

Article history:
Received 22 May 2013
Received in revised form 27 September 2013
Accepted 6 November 2013
Available online 12 November 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Hydrological indices
Flow duration curves
Ungauged sites
Rainfall-runoff model
Random Forests

SUMMARY

Predictions of hydrological regimes at ungauged sites are required for various purposes such as setting environmental flows, assessing availability of water resources or predicting the probability of floods or droughts. Four contrasting methods for estimating mean flow, proportion of flow in February, 7-day mean annual low flow, mean annual high flow, the all-time flow duration curve and the February flow duration curve at ungauged sites across New Zealand were compared. The four methods comprised: (1) an uncalibrated national-coverage physically-based rainfall-runoff model (TopNet); (2) data-driven empirical approaches informed by hydrological theory (Hydrology of Ungauged Catchments); (3) a purely empirically-based machine learning regression model (Random Forests); and (4) correction of the TopNet estimates using flow duration curves estimated using Random Forests. Model performance was assessed through comparison with observed data from 485 gauging stations located across New Zealand. Three model performance metrics were calculated: Nash–Sutcliffe Efficiency, a normalised error index statistic (the ratio of the root mean square error to the standard deviation of observed data) and the percentage bias. Results showed that considerable gains in TopNet model performance could be made when TopNet time-series were corrected using flow duration curves estimated from Random Forests. This improvement in TopNet performance occurred regardless of two different parameterisations of the TopNet model. The Random Forests method provided the best estimates of the flow duration curves and all hydrological indices except mean flow. Mean flow was best estimated using the already published Hydrology of Ungauged Catchments method.

1. Introduction

River water provides a valuable resource for out-of-stream water use as well as for supporting in-stream environmental values. Alteration of natural river flow regimes is increasing globally as water is taken for human, agricultural and industrial use and power production, threatening both river biodiversity and security of human water use (Vörösmarty et al., 2010). Globally, this has led to a variety of legislative processes aimed at promoting prudent and rational use of natural water resources which seek to judge the trade-off between economic development and impact to the natural environment (e.g. EC, 2000; New Zealand Government, 2011). For example, default limits to water resource use for all rivers in New Zealand must comprise at least a minimum flow (the flow below which no water can be abstracted) and an allocation limit (a limit on the amount of abstraction taken from the resource) (New Zealand Government, 2011; Snelder et al., 2013).

Information summarising natural flow regimes is therefore required to assess both the in-stream environmental and out-of-stream economic effects of potential alterations to flow regimes. This information may take the form of various hydrological indices describing different aspects of the flow regime such as low flows, high flows or flow variability (Olden and Poff, 2003; Poff et al., 2010). Flow duration curves (FDCs) may also be utilised for various purposes including low flow analysis (Smakhtin, 2001), quantifying reliability of water supply (Snelder et al., 2011) and quantifying alterations to hydrological regimes (Vogel et al., 2007). This type of hydrological information is ideally derived from observed flow time-series at the site, or sites, of interest. However, flow time-series are only available at a small number of locations where flow gauges have been maintained and operated. Hydrological estimates are therefore often required at ungauged sites across a catchment or landscape (Sivapalan et al., 2003; Blöschl et al., 2013).

A variety of approaches can be used to provide estimates of hydrological indices at ungauged sites. In theory, these approaches range from purely physically-based to purely empirically-based. Physically-based approaches have also been referred to as

* Corresponding author. Tel.: +64 (0)3 348 8987; fax: +64 (0)3 348 7891.
E-mail addresses: d.booker@niwa.co.nz (D.J. Booker), ross.woods@bristol.ac.uk (R.A. Woods).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.11.007

deterministic (Chow et al., 1988), distributed (Beven and Binley, 1992), physics-based (Pechlivanidis et al., 2011), process-based or Newtonian (Yaeger et al., 2012). Empirically-based approaches have also been referred to as stochastic (Chow et al., 1988), metric (Pechlivanidis et al., 2011) data-based or Darwinian (Yaeger et al., 2012). Physically-based approaches are those that aim to estimate streamflow by utilising a conceptual understanding of the physics describing various parts of the hydrological cycle by approximating physical processes such as interception, evaporation, and storage (e.g. Beven and Kirkby, 1979; Clark et al., 2008). However, assumptions about physical processes are necessarily required to apply this understanding (Beven, 1997). For example, assumptions about continuity of volumes, discretisation of governing equations and some form of spatial averaging may be required for a physically-based approach to be spatially-distributed (Blöschl and Sivapalan, 1995; Singh and Frevert, 2006). Similarly, time dependence must be represented by updating state variables through a sequence of time steps (Singh, 1995). Physically-based approaches may also require spatially distributed input data such as information on soil characteristics such as water holding capacity, rainfall time-series or temperature time-series (e.g. Clark et al., 2008). This has led to much analysis and debate relating to data needs, parameter calibration and uncertainty in physically-based hydrological models (Beven, 1997; Gupta et al., 2006).

Empirically-based approaches are those that seek to estimate hydrological indices by quantifying patterns between observed hydrological indices and catchment characteristics. These patterns can be quantified using a variety of techniques including linear regression (e.g. Engeland and Hisdal, 2009), or machine learning techniques (e.g. Booker and Snelder, 2012). One advantage of empirically-based approaches is that their relative simplicity has allowed them to be transferred to ungauged catchments by way of regionalisation (e.g. Castellarin et al., 2004), generalisation or dissimilarity modelling (e.g. Booker and Snelder, 2012). An unexpected result from some regionalisation studies predicting hydrological statistics and hydrological model parameters is that spatial proximity can be a more effective predictor than catchment attributes (Merz and Blöschl, 2005; Parajka et al., 2005). This suggests that there is still much to learn from regionalisation studies, though it is not yet clear how to improve the performance of methods that use catchment attributes.

In practice, many physically-based models have empirical components and many empirical models incorporate some level of knowledge about physical processes. A balance between model complexity and data availability must be found for both physically-based (Fenicia et al., 2008) and empirically-based (Jakeman and Hornberger, 1993) approaches. All physically-based approaches require some parameterisation, and are known to perform best when calibrated against observed data (e.g. Clark et al., 2008; McMillan et al., 2013). Similarly, the independent variables used in empirically-based approaches are often chosen after consideration of physical principles and the form of fitted empirical relationships can also be interrogated to ensure consistency with physical principles (e.g. Booker and Snelder, 2012). Hybrid metric-conceptual models are those that seek to combine the strengths of empirically-based and physically-based conceptual models (Pechlivanidis et al., 2011).

Despite the variety of approaches available for estimating hydrological conditions at ungauged sites, few studies have compared estimates calculated using contrasting approaches. The aim of this work was to compare a variety of available methods for estimating several hydrological indices and flow duration curves at ungauged catchments across New Zealand. These methods employed a range of approaches from a physically-based rainfall-runoff model to empirically-based regressions. The primary aim was to objectively judge which method was best able to estimate several hydrological indices across New Zealand given current climatic and landcover conditions. The secondary aim was to assess the advantages of combining two approaches by correcting physically-based estimated time-series using empirically-based estimated FDCs.

2. Data description

2.1. Flow time-series

A flow time-series database was collated that comprised mean daily flows observed at 485 gauging stations with available records of 5 full years or longer. Available mean daily flow time-series from the National Institute of Water and Atmospheric Research's (NIWA) national database were collated alongside data supplied by particular regional councils (Northland Regional Council, Auckland Council, Waikato Regional Council, Greater Wellington Regional Council, and Environment Canterbury). The time-series database contained only sites that were not affected by large engineering projects such as dams, diversions or substantial abstractions, according to information given by each data provider. See Snelder et al. (2005) and Booker (2013) for further details on gauging station selection. These gauging stations were located throughout New Zealand (Fig. 1) and represented a wide range of hydrological conditions (Table 1). The observed time-series did not all cover the same time periods.

It is known that hydrological regimes may not be stationary (constant mean and constant variance through time; Hamilton, 1994) due to the presence of trends and temporal autocorrelations (Milly et al., 2008). This is because hydrological regimes may be influenced by a variety of factors including land cover change (e.g. Fahey and Jackson, 1997), inter-decadal climatic patterns (e.g. Kiem et al., 2003) and longer-term climate shifts (Parry

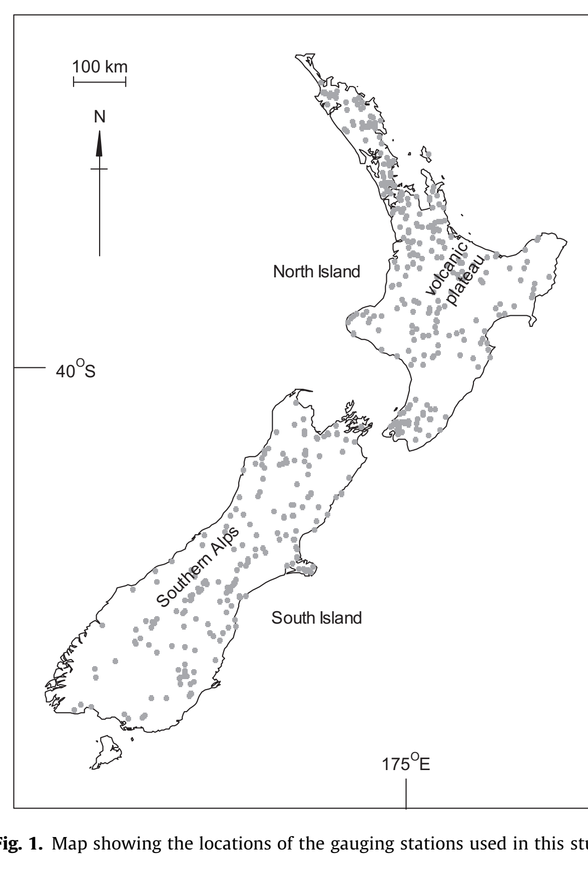

Fig. 1. Map showing the locations of the gauging stations used in this study.