Booker, D. J., & Woods, R. A. (2014). Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. Journal of Hydrology, 508, 227-239.
DOI: 10.1016/j.jhydrol.2013.11.007

Elsevier Editorial System(tm) for Journal of Hydrology
Manuscript Draft

Manuscript Number: HYDROL14928

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

Article Type: Research Paper

Keywords: hydrological indices; flow duration curves; ungauged sites; rainfall-runoff model; random forests.

Corresponding Author: Dr. Douglas James Booker, Ph.D.

Corresponding Author's Institution: National Institute of Water and Atmospheric Research

First Author: Douglas James Booker, Ph.D.

Order of Authors: Douglas James Booker, Ph.D.; Ross A Woods, Ph.D.

Suggested Reviewers: Neil Viney PhD
Principal Research Scientist, Modelling catchment hydrology, CSIRO
Neil.Viney@csiro.au
Expertise in hydrological modelling, prediction in ungauged basins and ensemble modelling.

David Post PhD
Senior Research Scientist, Land and Water, CSIRO
David.Post@csiro.au
Expertise in catchment-scale hydrology, regionalisation to ungauged catchments, land use and climate change impacts on hydrology.

Parajka Juraj PhD
Research and Lecturer, Centre for Water Resource Systems, Institute of Hydraulic Engineering and Water, Vienna University of Technology
parajka@hydro.tuwien.ac.at
Expertise in Experimental Hydrology, Hydrology, Snow modelling and water balance modelling.

Stacey Archfield PhD
Research Hydrologist, Massachusetts-Rhode Island Water Science Center, USGS
sarch@usgs.gov
Expertise in flood prediction in ungauged catchments, environmental flow setting, hydrological regionalisation

Mike Acreman PhD
Senior scientist, Ecohydrological processes, entre for Ecology and Hydrolgy, UK
man@ceh.ac.uk

Expertise in water resource assessment, environmental flow setting and flow duration curve analysis

Highlights for **"**Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments" by Booker and Woods

- Methods for estimating various hydrological indices at ungauged sites were compared.
- Methods included a TopNet rainfall-runoff model and a Random Forest empirical model.
- TopNet estimates were improved through correction using Random Forest estimates.
- Random Forests provided the best estimates of all indices except mean flow.
- Mean flow was best estimated using an already published empirical method.

Research article submitted to Journal of Hydrology

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

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

[a] National Institute of Water and Atmospheric Research,
P O Box 8602
Riccarton,
Christchurch,
New Zealand
Email address: d.booker@niwa.co.nz
Phone: +64 (0)3 348 8987
Fax: +64 (0)3 348 7891

[b] Department of Civil Engineering
University of Bristol,
Bristol,
UK
Email address: ross.woods@bristol.ac.uk

[*]Corresponding author

Short title: hydrological estimates for ungauged catchments

Abstract

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.

Key words: hydrological indices; flow duration curves; ungauged sites; rainfall-runoff model; random forests.

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).

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 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 (Beven, 1989; Bloschl 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; Beven, 1989; Singh and Woolhiser, 2002; 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).

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

a. 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 (in press) for further details on gauging station selection. These gauging stations were located throughout New Zealand (Figure 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 & Jackson, 1997), inter-decadal climatic patterns (e.g. Kiem et al., 2003) and longer-term climate shifts (Parry et al., 2007). However, the purpose of this study was to compare the ability of various approaches to characterise differences in flow regimes between sites across New Zealand given current climatic and land cover conditions rather than to characterise differences through time. For empirically-based methods it was therefore assumed that differences in hydrological regimes between sites far exceeded any differences in hydrological regimes that may have occurred due to differences in observation periods (which were different for each observed time-series) despite some evidence for inter-decadal patterns in some, but not all, indices for particular regions of New Zealand but not others (e.g. McKerchar and Henderson, 2003; Booker, in press).

b. Observed hydrological indices

Several hydrological indices were calculated for each observed flow time-series (Table 2). These indices were chosen because they represent a range of hydrological conditions including floods and droughts, can be used to estimate water resource availability, and are used in environmental flow setting procedures. Mean flow, Q_{bar}, represents total potential water availability, is used for scaling of dimensionless metrics such as standardised flow duration curves (e.g. Booker and Snelder, 2012) and may be used when comparing sites for ecological studies (e.g. Leathwick et al., 2005). The proportion of flow in each month may be of interest when investigating seasonality of flow. The proportion of flow in February, Q_{Feb}, was chosen as an example because the mid-summer month of February represents a generally dry month in which both irrigation demand (the largest consumptive water use in New Zealand) and ecological stress are likely to be high. The 7-day mean annual low flow, Q_{MALF}, is often used as an indicator of low flow in ecological studies (e.g. Caruso, 2002; Suren and Jowett, 2006) and to represent one component of the flow regime in environmental flow assessments (e.g. Richter et al., 1997; Poff et al., 1997). Since limits to water resource use may be expressed as proportions of Q_{MALF}, this index is of particular interest in New Zealand (MFE, 2008). Mean annual flood, Q_F, may be used for flood risk assessment and flood design, but may also be used as a surrogate for physical disturbance (e.g. Poff and Ward, 1989; Poff, 1996) especially when compared to geomorphological characteristics such as sediment grain size and channel slope (Clausen and Plew, 2004). All four of these hydrological indices may also be used for data driven environmental classifications (e.g. Snelder and Booker, 2012). Many further hydrological indices could have been compared, but it was desirable to provide an expedient analysis and there is known to be a high degree of covariance within sets of these indices (Clausen and Biggs, 1997; Olden and Poff, 2003).

In order to minimise the likelihood of low flow periods crossing years, each day in each observed time-series was assigned to a water year starting on the 1st of October. Water years with more than 30 days of missing data were excluded from the analysis. Calculations of (Q_{MALF}), and mean annual flood (Q_F) were based on water years. Q_{MALF} was calculated as being the mean of the 7-day running average annual low flow in each water year.

Many hydrological indices are scale-dependent; bigger catchments have larger values of Q_5, Q_{MALF}, Q_F and Q_{bar} than smaller catchments. The values for these indices were therefore standardised by dividing by catchment area. Further transformations were then applied in order to more closely approximate normal distributions (Table 2).

c. Flow duration curves

FDCs represent the relationship between magnitude and frequency of flow by defining the proportion of time for which any discharge is equalled or exceeded (Vogel and Fennessey, 1994; Vogel and Fennessey, 1995). Flow duration curves are a useful tool for quantifying flow regimes for both resource availability (Snelder et al., 2011) and for departure from a reference state (Vogel et al., 2007). For each flow time-series two observed FDCs were calculated from mean daily flows. FDCs were calculated from: a) mean daily flows in all months of the year; and b) mean daily flows in February. These two FDCs represent the probability distribution of flow over all-time and the probability distribution of flow for the month of February over all years. As above, February was chosen to represent a dry month in which both irrigation demand and ecological stress are likely to be high.

For calculation of each FDC, mean daily flows for each gauging station were sorted lowest to highest and then interpolated onto percentile values from 0 to 100 in intervals of 1 to determine the proportion of the time that each flow was not exceeded. Each FDC was therefore characterised using the same number of data points (101), providing for a balanced study design in further statistical analysis. All daily flows were divided by catchment area to allow modelling of differences in mean flow whilst standardising for differences in catchment size. This was in contrast to the method of Booker and Snelder (2012) which investigated only the shapes of FDCs after having standardised by Q_{bar}.

d. Catchment characteristics

A GIS representation of the New Zealand river network comprising 550,000 segments, their unique upstream catchments and an associated database of catchment characteristics were used to provide information for each gauging station. The catchment characteristics include a range of categorical and continuous variables (Snelder and Biggs, 2002; Snelder et al., 2004; Leathwick et al., 2011). The GIS river network and associated databases have previously been used to define a hierarchical classification of New Zealand's rivers called the River Environment Classification (REC; Snelder and Biggs 2002). These databases provide inventories for river resource analysis and management purposes (Snelder and Hughey, 2005; Leathwick et al., 2011; Clapcott et al., 2010; Clapcott et al., 2011). They have also been used to create nationwide models for estimating flow statistics such as flood flows (Pearson and McKerchar, 1989), low flows (Pearson, 1995), mean flow (Woods et al., 2006) and shapes of FDCs (Booker and Snelder, 2012) at ungauged sites using relationships between these hydrological metrics and catchment characteristics. Snelder et al. (2005) showed that grouping river segments by nested categorical subdivisions of climate and topography, known as the Source-of-Flow grouping factor (Table 3), provided an a priori hydrological regionalisation.

3. Estimation methods

For this study four methods for calculating hydrological indices and FDCs at ungauged locations were compared (Figure 2). Method 1 used a physically-based approach. Method 2 used a data-driven empirical approach that was informed by hydrological theory to estimate each hydrological index separately. Method 2 can be classified as being a hybrid metric-conceptual approach under the classification proposed by (Pechlivanidis et al., 2011). Method 2 was named after a sequence of projects collectively known as the Hydrology of Ungauged Catchments (HUC) projects. Method 3 used an empirically-based regression approach. Method 4 combined a physically-based and empirically-based approach. All methods were able to produce estimates for all reaches that comprise the NZ river network and were therefore applicable to ungauged sites across New Zealand.

a. Method 1 TopNet

Topnet is a spatially distributed time-stepping hydrological model which combines TOPMODEL concepts of sub-surface storage controlling the dynamics of the saturated contributing area and baseflow recession (Beven and Kirkby, 1979; Beven et al., 1995) with submodels for snow and plant canopies, and a kinematic wave channel routing algorithm (Goring, 1994). See McMillan et al. (2013) for further detailed description and Clark et al. (2008) for complete model equations.

TopNet has two fundamental components: (i) simulating the water balance over sub-catchments throughout a river basin, and (ii) routing streamflow from each sub-catchment to the basin outlet. The water balance model includes simulating the storages and fluxes of water in the canopy, snowpack, unsaturated and saturated soil zone. TopNet also accounts for time delay due to flow routing within each sub-basin. Runoff from each sub-basin flows into a digital stream network and is routed through the river network. For this application TopNet models used daily precipitation and temperature data from the New Zealand Virtual Climate

Station Network (Tait, 2008, Tait et al., 2006), which was then disaggregated to hourly resolution using stochastic disaggregation for precipitation (Rupp et al., 2009). Additional model boundary conditions were estimated directly from GIS data on topography, soil and vegetation (Clark et al., 2008; McMillan et al., 2013).

For catchment specific applications TopNet parameters can be calibrated to optimise model performance (e.g. Bandaragoda et al., 2004; McMillan et al. 2013). However, in this case uncalibrated national TopNet models of New Zealand (Henderson et al., 2011) were run using an hourly timestep over the period 1973-2010. Two different versions of TopNet were available. National TopNet Version 0 was discretised using Strahler-1 sub-catchments from the REC. The typical catchment area of a Strahler-1 catchment is 0.7 km^2. This version had a spatially uniform value for the parameter, *f*, which represents the decline in saturated hydraulic conductivity of the soil with depth (Clark et al., 2008). This parameter effectively controls responsiveness of river flow to rainfall. National TopNet Version 1 was discretised using Strahler-3 sub-catchments from the REC. This version had a spatially distributed set of values for *f*. The *f* parameter took different values according to the hydrological regionalisation described by Toebes and Palmer (1969), ranging from values more than 8 m^{-1} for steep catchments in the Southern Alps to less than 1 m^{-1} in flat catchments on the volcanic plateau in the central North Island (see Figure 1 for place names). Where flow time-series were required for Strahler-1 and Strahler-2 catchments flow data were downscaled by multiplying flows from the nearest available Strahler-3 node in the REC network by the ratio of the catchment area of the required location with that of the substitute location. For both Version 0 and Version 1 hourly data for the river reach in which each gauging station was located were averaged over each calendar day to obtained mean daily flow time-series. Hydrological indices were then calculated using the same algorithms as were applied to the observed flow time-series.

Ideally both observed and estimated time-series would be available for a very long period (e.g. 100 years). However, the available observed flow time-series did not all cover the same period, and TopNet data were available for a uniform time period (1973-2010). This provided the opportunity to test the sensitivity of correspondence between observed and estimated hydrological indices to synchronisation of the observed and TopNet estimated time-series. Observed and TopNet Version 1 estimated indices were compared using two different procedures. For the first procedure, indices calculated from all available observed flows (5 years or more) were compared with those calculated from all available TopNet Version 1 estimated flows (1973-2010). Essentially this procedure assumed that, when averaged over time, both the observed and TopNet estimated time-series represented the long term hydrological conditions (i.e. that both observed and TopNet estimated time-series were stationary and that records were sufficiently long to characterise long term conditions). For the second procedure only the time period for which both observed flows and TopNet estimated flows were available was identified for each gauging station. Observed indices for this period were then compared with TopNet Version 1 estimated indices for the same period at each gauged location. Better fit between synchronised observed and estimated values (the second procedure) in comparison to non-synchronised (the first procedure) would indicate non-stationarities in the observed hydrological regimes that were detectable in the TopNet time-series. Some observed time-series fell completely outside of the TopNet time-series. This reduced the number of time-series available for the second procedure compared to the first.

b. Method 2 HUC

The approach used to estimate Q_{bar} for Method 2 (HUC) is described in Woods et al. (2006). Woods et al. (2006) evaluated four simple models of mean annual runoff throughout New Zealand, predominantly based on precipitation information and estimated evapotranspiration. Model results were compared to observed data and synthesised estimates of catchment runoff. The preferred model of Woods et al. (2006) subtracts an estimate of annual actual evapotranspiration from a precipitation surface. Annual actual evapotranspiration is estimated according to the ratios of potential evapotranspiration with annual precipitation, and a single water balance parameter which is estimated by independent calibration. This method applies a regional bias correction to the results of a previously uncorrected model.

The approach used to estimate Q_{Feb} for Method 2 was to employ a regionalisation of Q_{Feb} based on Source-of-Flow groupings in the REC and New Zealand island (i.e. North Island or South Island, Figure 1), where Source-of-Flow is a combination of the climate and topography classes of a catchment (Table 3). For each region Q_{Feb} was the mean of the Q_{Feb} for all observed flow records that belong to that class in that island. For cases where no measured flow was available, expert judgement was applied to make use of data from other classes.

The approach used to estimate Q_{MALF} for Method 2 is described in Henderson et al. (2004). Figure 3 shows a schematic description of the model and its parameters. These fall into three categories: a) climate parameters (T the average length of a dry season, N the number of rain events in that season, P the amount of rain in the dry season); b) flow parameters (Q_{mean} the mean flow, Q_0 the average flow at the start of the dry season, α the fraction of that rain that affects the streamflow); and c) catchment parameters that describe the way in which water is released from catchments during the dry season (b and T*). Estimates of all these input parameters have previously been developed for all of New Zealand (Henderson et al, 2004). The parameter Q_0 corresponds to the average flow at the start of the dry season. The predictions are most sensitive to the value of the b parameter, which describes the type of river flow recession. For example, catchments in dry catchments typically have b values near 1, hill country catchments typically have b values near 2, and catchments with volcanic geology typically have b values of 3 or larger.

The approach used to estimate Q_F for Method 2 is described in Pearson and McKerchar (1989) and McKerchar and Pearson (1989). Essentially, these estimates are gained from interpolation onto ungauged sites from a contour map of Q_F which was itself derived from a spatial interpolation of observed data. Since this approach used instantaneous flow data to calculate Q_F, rather than mean daily values, it was anticipated that the approach would overestimate Q_F in comparison to observed values derived from mean daily values. However, the estimates were still included in the analysis.

The approach used to estimate FDCs for Method 2 was to assume a log-normal probability distribution as a model of the flow duration curves. This is a log transformation of

$$g(x,\vartheta) = \left(1/\sqrt{2\pi}\vartheta_2\right)\exp\left[-1/2\left((x-\vartheta_1)/\vartheta_2\right)^2\right], \qquad \text{Equation 1}$$

which has two parameters, θ_1 and θ_2. It was further assumed that θ_1 could be estimated as the mean flow (Q_{bar} from Method 2) and that θ_2 would be estimated as a linear function of the b parameter, which was also used to calculate Q_{MALF} for Method 2. The approach used to estimate FDC_{Feb} was to scale the estimated FDC for Method 2 by the estimated Q_{Feb} for Method 2.

c. Method 3 Random Forests

A regression technique called Random Forests was used to apply a regression of each observed hydrological index (Table 2) and each of the three parameters describing a GEV distribution of the all-time FDC and the FDC for February as a function of available catchment characteristics (Table 4). This method uses machine-learning by combining many regression trees into an ensemble to produce more accurate regressions by drawing several bootstrap samples from the original training data and fitting a tree to each sample (Breiman, 2001; Cutler et al., 2007). Random forest models fitted using catchment characteristics have previously been shown to be able to explain variation in hydrological patterns such as parameters describing FDCs (Booker and Snelder, 2012), the frequency of events that exceed three time the median flow (Booker, in press) and various other hydrological indices (Snelder and Booker, 2012). Each random forest was developed by growing 500 trees. As the number of trees (k) increases the generalisation error always converges and it was assumed that use of 500 trees was sufficiently high to ensure convergence.

The predictions from random forest models were tested using a leave-one-out cross validation procedure referred to here as jack-knifing (Efron, 1982; Booker and Snelder, 2012). This cross-validation procedure was applied by leaving out all data associated with each of the 485 sites and then estimating each hydrological index for the left-out site from all remaining sites. The results from this procedure produced estimates as if each site were ungauged (Ganora *et al*., 2009). Comparison between observed and jack-knifed values allowed an assessment of both the robustness and reliability for estimation at ungauged sites (Castellarin et al., 2004).

For each time-series, the parameters describing a GEV distribution,

$$G(x,\vartheta)=\exp\left[-\left(1-\left(\vartheta_3\left(x-\vartheta_1\right)\right)/\vartheta_2\right)^{1/\vartheta_3}\right], \qquad \text{Equation 2}$$

were fitted to all observed mean daily flows and all observed mean daily flows in February. In both cases observed mean daily flows were divided by catchment area for each gauging station prior to fitting the GEV parameters. The GEV distribution is described by three parameters and has shown to represent the range of shapes of standardised FDCs found across New Zealand. See Booker and Snelder (2012) for further discussion of estimating standardised FDCs at ungauged sites across New Zealand using various statistical techniques to generalise parameters describing various probability distributions.

d. Method 4 TopNet Corrected

FDCs calculated using the jack-knifed Random Forests method represent a unique FDC at any location in the New Zealand river network as if each location were ungauged. This provided the opportunity to correct for bias in the TopNet estimated FDCs using the Random Forests estimated FDC at each site as if it were an observed FDC. Therefore the jack-knifed Random Forests FDCs were used to calculate a correction factor for each percentile, *i*, of the TopNet FDC for each site, *j*.

$$\text{TopNet Corrected}_{ij} = \text{TopNet}_{ij} * (\text{Random Forest}_{ij} / \text{TopNet}_{ij}) \qquad \text{Equation 3}$$

Since the exceedance percentile of each datum in each TopNet time-series was known, these corrections could also be applied to each TopNet time-series. This allowed re-calculation of each hydrological index from each corrected time-series. This procedure was repeated separately for TopNet Version 0 FDCs and TopNet Version 1 FDCs.

4. Observed versus predicted values

Scatterplots of observed versus predicted values after having standardised and transformed each index (Table 2) were plotted for each index for each method. These scatterplots were overlaid with a linear regression with observed values on the y-axis as recommended by Piñeiro et al. (2008). Following the suggestion of Moriasi et al. (2007), three model performance metrics were calculated for each set of observed versus predicted values: Nash-Sutcliffe efficiency (NSE); percent bias (pbias); and ratio of the root mean square error to the standard deviation of observed data (RSR). NSE is a dimensionless metric that determines the relative magnitude of the residual variance (“noise”) compared to the observed data variance (“information”) (Nash and Sutcliffe, 1970). NSE values of 1 indicate a perfect match between estimates and observations, whereas values of 0 indicate performance equal to estimating the mean observed value across all observations. pbias measures the average tendency of the simulated data to be larger or smaller than their observed counterparts (Gupta et al., 1999). Negative pbias values represent overestimation and positive values indicate underestimation. RSR standardises RMSE using the observations standard deviation, and it combines both an error index and the additional information recommended by Legates and McCabe (1999). Lower RSR values indicate better model performance, with 0 indicating perfect correspondence between estimates and observations. See Moriasi et al. (2007) and references therein for full details of these performance evaluation metrics. The same metrics were applied to 101 points representing log specific (flow per unit catchment area) FDCs for each site for each method for the February and all-time FDCs separately.

5. Results

a. Hydrological indices

Synchronisation of TopNet Version 1 with the observed time-series made little impact on the performance metrics (NSE, RSR and pbias) when compared to using the full TopNet time-series (Table 5). This was especially the case for Q_{bar}, Q_{MALF} and Q_F. For Q_{bar}, synchronisation marginally reduced an overestimation bias, but also resulted in a small reduction in performance in terms of NSE and RSR (reduced NSE, increased RSR). For Q_{MALF}, synchronisation resulted in increased overprediction bias, but marginally improved performance in terms of NSE and RSR. The process of synchronisation did alter performance for Q_F as synchronisation improved performance in terms of NES and RSR, but substituted an overprediction bias with an underprediction bias of the same magnitude. These results indicate that it was not the case that there were non-stationarities in observed hydrological regimes that were generally detectable in the TopNet time-series for Q_{bar}, Q_{MALF} or Q_F. This may not have been the case for Q_{Feb}. This is an understandable result as Q_{bar}, Q_{MALF} and Q_F will be less sensitive to inter-annual variability than Q_{Feb}. This is because Q_{bar} is an average calculated over all the record, and both Q_{MALF} and Q_F are both averages of indices calculated for each year of record, whereas Q_{Feb} is calculated over a smaller time-window in each year of record.

Overall there was more difference in performance between TopNet Version 0 and TopNet Version 1 than there were differences between synchronisation and non-synchronisation of TopNet Version 1. This indicates that TopNet results are more sensitive to changes to the TopNet *f* parameter than to either the assumption that the 1973-2010 time-series represent the long-term flow regime, or any non-stationarities combined with relatively short records in the observed time-series.

When compared to TopNet Version 0, TopNet Version 1 reduced an overestimation of Q_{bar}, but reduced performance in terms of NSE and RSR. For Q_{Feb}, TopNet Version 1 marginally improved NSE, reduced an overestimation pbias, but increased RSR. For Q_{MALF}, TopNet Version 1 dramatically improved NSE, improved RSR and replaced a large overestimation with an underestimation of lesser magnitude. For Q_F, TopNet Version 1 reduced performance of all metrics when compared to TopNet Version 0. This indicates that high flows were not better predicted following the regionalisation of the TopNet *f* parameter. However, over all four indices there were greater differences between methods (TopNet, HUC and Random Forests) than there was between the two TopNet versions (Table 5, Figure 4).

The TopNet time-series was corrected using the jack-knifed Random Forests FDC estimates and then used to estimate the hydrological indices. For all indices and both TopNet versions, corrected estimates improved performance in terms of NSE and RSR when compared to the uncorrected TopNet estimates. Corrected estimates produced less bias as indicated by smaller magnitude pbias when compared to uncorrected estimates from both TopNet versions for all indices except Q_{Feb} for Version 1 and Q_F for version 0. Correction of TopNet Version 1 caused an increase in overprediction of Q_{Feb}. Correction of TopNet

Version 0 caused an overprediction to change to an underprediction of greater magnitude. Overall, correction greatly reduced differences in performance between the two TopNet versions (Table 5, Figure 4).

For Q_{bar} and Q_{Feb} there was more difference between TopNet Version 0 and TopNet Version 1 than there was between TopNet Version 1 and TopNet 1 Corrected. After correction, the performance of Q_{bar} estimated from both TopNet versions matched the performance of those estimated using Random Forests. This was because the correction procedure forced the TopNet corrected estimated FDCs to match jack-knifed Random Forests estimated FDCs and therefore TopNet corrected Q_{bar} matched jack-knifed Random Forests estimated Q_{bar}.

NSE was positive (negative values indicate that the mean observed value is a better predictor than the simulated value) for all indices for all methods except Q_F for Method 2 HUC (Table 5). This indicates that, except for Q_F from the HUC method, all methods provided some degree of useful information about patterns in the estimated values. In this comparison HUC estimates of instantaneous Q_F were compared with observed Q_F calculated from mean daily flow data. Poor performance and, in particular, overestimation of Q_F for Method 2 HUC was therefore not surprising. In fact, McKerchar and Pearson (1989) previously showed that the method was able to explain a substantial fraction of the observed variation in Q_F when compared to observed values calculated from instantaneous flow data.

For Q_{bar} the HUC method performed best in terms of both NSE and RSR. This is the method already recommended by Woods, et al. (2006). For Q_{MALF}, Q_F and Q_{Feb} the Random Forests method performed best in terms of both NSE and RSR. The Random Forests method also gave the lowest magnitude pbias for Q_F and Q_{Feb} but not for Q_{MALF} (Table 5). These findings correspond well with visual inspection of observed against predicted values, which indicated that the Random Forests method reduced scatter and produced unbaised estimates for all four indices but was out-performed by Method 2 HUC for Q_{bar} (Figure 4).

b. Flow duration curves

More sites had better performance as indicated by higher NSE values, lower RSR values and lower magnitude pbias for all-time FDCs compared to February FDCs regardless of estimation method (Figure 5). This indicates greater uncertainties associated with estimation of February FDCs compared to all-time FDCs. More sites had better performance in terms of NSE, RSR and pbias for TopNet Version 1 in comparison to TopNet Version 0 for the all-time FDC and the February FDC in particular. Negative pbias values for many TopNet Version 0 estimated February FDCs indicated consistent underestimation. This consistent underestimation was not present for TopNet Version 1, which showed an equal likelihood for either underestimation or overestimation of the February FDC. This indicated that regionalisation of the TopNet *f* parameter improved flow estimation, particularly in February.

Both the HUC and the Random Forests methods performed better than either of the uncorrected TopNet methods for both the all-time and February FDCs. Both all-time and February FDCs had more sites with higher NSE, lower RSR and lower magnitude pbias when estimated using the Random Forests method compared to the other methods. Since the TopNet 1 Corrected estimated all-time FDC was corrected using the jack-knifed Random Forests estimated FDC, performance of the TopNet 1 Corrected estimated all-time FDC was the same as the jack-knifed Random Forests estimated FDC.

c. National estimates for New Zealand

All methods were able to provide predictions for ungauged sites across New Zealand which reproduced the major regional variations in observed Q_{MALF} (Figure 6). These geographical patterns included a strong east-west gradient in the South Island as well as the influence of the Southern Alps (see Figure 1 for place names). As they cross the eastern plains of the South Island, large mountain-fed rivers with markedly higher Q_{MALF} stand out against a background of comparatively lower-yielding lowland streams. To the northeast of the central North Island, the rivers draining a volcanic plateau have relatively high $Q_{MALF,}$ with large storage capacity in the thick pumice and ash layers sustaining low flows (Mosley and Pearson, 1997). Both Random Forests (Figure 6c) and TopNet (Figure 6d) predicted lower values of Q_{MALF} than HUC (Figure 6b) for the south west coast of the South Island, but predicted slightly higher Q_{MALF} for most other locations in comparison with HUC. It should be noted that none of the methods were designed to take account of large engineering schemes such as those currently in place on several of New Zealand's large rivers (e.g. the Waikato, Rangitata, Waitaki, Clutha and Waiau rivers).

6. Discussion

A limited set of hydrological indices along with both the all-time and February FDCs were investigated (Table 2). This set of hydrological indices included those representing both high and low flow extremes as well as an aspect of seasonality. These indices are commonly used for water resource planning in New Zealand, however not all aspects of the flow regime, such as the frequency of mid-range flows, were represented. This aspect of the flow regime could have been included by calculating various additional indices such as the number of events exceeding three times the long-term median flow (FRE3; Biggs 2000), but no HUC method was available for estimating this index. National estimates of FRE3 using random forests, including comparison with observed values, were calculated and compared with observations by Booker (in press).

For the Random Forests method FDCs were described using the three parameter GEV distribution. Other distributions could have been used including log Pearson Type III (LP3; Ganora et al., 2009) or a mixed gamma distribution (Cheng et al., 2012). Booker and Snelder

(2012) showed that, although the LP3 distribution may provide better fits to observed FDCs when standardised by mean flow, uncertainties in generalising the LP3 parameters from catchment characteristics meant that a method using the GEV distribution to parameterise the shape of the FDC gave better performance for prediction at ungauged locations.

The same set of independent variables was used to model all four hydrological indices. Procedures designed to optimise the set of independent variables such as the Model Improvement Ratio (Murphy et al. 2010) were not employed to optimise the predictor data set. This approach may not have provided optimal Random Forest models in all cases as one would expect different sets and different numbers of independent variables to best predict each dependent variable. For example, summer temperature might be expected to be related to low flows, but not flood flows. Despite this the Random Forests method still outperformed the other methods even when a leave-one-out cross validation procedure was applied to allow for independent assessment of estimation performance against observed data.

Although many performance metrics are available to assess model performance, NSE, RSR and pbias were used as recommended by Moriasi et al., (2007). Although these three metrics are designed to quantify different aspects of model performance, they often gave consistent information regarding model performance.

The aim of this work was to assess the ability of various methods to estimate hydrological conditions for ungauged catchments in the absence of major hydrological alterations such as that caused by abstraction, storage or diversion. The ability to estimate the effects of either climate change or land cover change were not assessed. It may be necessary to assess the potential effects of climate change (Zemansky et al., 2012; Earman and Dettinger 2011), land use change (Scanlon et al., 2007) or their combined effects (Brekke et al., 2004) on flow regimes to develop rational management strategies. Both TopNet and the Random Forests models described above have inputs that could be changed to assess the impacts of climate change. However, the validity of this approach was not tested here. It should be noted that there are several issues relating to model structure and parameterisation that would need to be resolved when using physically-based models to predict the hydrological impacts of environmental change (Wagener, 2007). Similarly, when using flexible empirically-based models such as Random Forests to predict outside of the fitted model domain it is important to understand how the algorithms perform when projected into the new environmental conditions (Elith and Graham, 2009).

These results indicate that Random Forests outperformed both TopNet versions for all four hydrological indices as well as for FDCs. This finding corresponds well with the findings of others. For example, Parkin et al. (1996) found that streamflow predictions from an a priori parameterised physically-based model contained considerable uncertainty. It should be noted that, although TopNet Version 1 arguably represents the best currently available physically-based approach for application to ungauged sites across New Zealand, this method was uncalibrated. It is known that calibration of TopNet parameters can significantly improve estimation performance by optimising model performance against observed flows (e.g. Bandaragoda et al., 2004; McMillan et al., 2013). Calibration procedures are only possible for catchment specific applications with available flow data. It is possible to transfer calibrated parameter sets to ungauged sites (e.g. Yu and Yang, 2000) given a suitable regionalisation procedure (McDonnell and Woods, 2004; Li et al., 2010; Olden et al., 2012; Coopersmith et al., 2012). Although calibration procedures have been applied to TopNet for several catchments (Bandaragoda et al., 2004; Clark et al., 2008; McMillan et al. 2013), a procedure to regionalise the calibrated parameter values is not currently available. Such procedures can be hampered by issues such as equifinality within the calibration parameter sets (Beven 2006; Bárdossy, 2007).

The Random Forests method can be used to estimate a unique FDC at any location in the New Zealand river network. These estimated FDCs could be used to provide a more reliable regionalisation than would be the case using data from observed locations alone because they represent variability across all of New Zealand rather than a sample of observed FDCs (Snelder and Booker, 2012). Furthermore, the Random Forests estimated FDC's at ungauged locations could provide the opportunity to calibrate TopNet parameters against an estimated FDC for ungauged locations in the New Zealand river network. This would require a method that allowed calibration against an observed (or estimated) FDC (e.g. Yu and Yang 2000; Yadav et al., 2007; Westerberg et al., 2011). Such a method may be developed as part of future work. However, considerable improvements in performance were gained when both TopNet versions were corrected using the jack-knifed estimated FDCs from Random Forests. This indicates that TopNet performance can be increased considerably without automated parameter set calibration procedures (Yu and Yang, 2000) or increased understanding of hydrological processes controlling variability of FDCs across catchments (Yaeger et al., 2012). Furthermore, the correction procedure reduced differences in performance between TopNet Version 0 and TopNet Version 1.

The TopNet correction procedure tested here represents one relatively crude method of combining a process-based approach with a data-based approach. The procedure provides estimates calculated using a data-based approach to correct for bias within FDCs calculated using a process-based approach. This contrasts with alternative approaches which have augmented stochastic approaches with more process-based approaches by incorporating different components of catchment dynamic responses into stochastic models (e.g. Botter et al., 2007a, 2007b, 2009; Muneepeerakul et al., 2010; Cheng et al., 2012) or by applying a water balance modelling framework to divide the FDC into three parts (Yokoo and Sivapalan, 2011).

The TopNet correction procedure provided results that matched the performance of Random Forests for Q_{bar} and the all-time FDC, but not for Q_{Feb}, Q_{MALF} or Q_F. It should be noted that this procedure allowed improved estimation of the entire time-series of flows using both TopNet versions. This method has a major advantage over the Random Forest method because any required hydrological indices can be calculated from the estimated time-series. In contrast, the Random Forests method requires fitting of new models to any newly calculated indices prior to estimation of these new indices at ungauged sites.

7. Conclusion

Results showed the Random Forests method provided the best estimates of both FDCs and all four hydrological indices except mean flow. Mean flow was best estimated using the already published HUC method (Woods et al., 2006). Results also showed that considerable gains in estimation performance can be made by correcting estimates calculated using physically-based models with estimated values calculated using empirically-based models.

8. Acknowledgements

Many thanks to Emily Walker, Dale Hansen (both NRC), Gillian Crowcroft, Clive Coleman (both AC), Bevan Jenkins (WRC), Mike Thompson (GWRC) and Tony Gray (ECan) for assistance in providing data. Thanks to Kathy Walter and Jani Diettrich for help collating and extracting flow data. Thanks to Shailesh Singh and Jani Diettirch for running and extracting data from national TopNet models. Thanks to Roddy Henderson for providing internal review and to Ton Snelder for help with model assessment algorithms. This research was funded by NIWA under Freshwater and Estuaries Programme 2, Sustainable Water Allocation and Freshwater and Estuaries Programme 1, Water Resources (2012/13 SCI).

9. References

Bandaragoda, C., Tarboton, D.G., Woods, R. 2004. Application of TOPNET in the distributed model comparison project. Journal of Hydrology 298, 178– 201.

Bárdossy, A. 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences Discussions 11, 703–710.

Beven, K.J. 1997. TOPMODEL: a critique. Hydrological Processes 11, 1069–1085.

Beven, K.J. 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320, 18–36.

Beven, K.J., Binley, A.M. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6, 279–298.

Beven, K.J., Kirkby, M. J. 1979. A physically based variable contributing area model of basin hydrology, Hydrol. Sci. Bull. 24, 43–69.

Beven K.J. 1989. Changing ideas in hydrology: the case of physically-based models. Journal of Hydrology 105, 157–172

Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., Freer, J. 1995. TOPMODEL, in: Singh, V. P. (Ed.), Computer Models of Watershed Hydrology. Water Resour. Publ., Highlands Ranch, Colorado, pp. 627–668.

Biggs, B.J.F. 2000: Eutrophication of streams and rivers: dissolved nutrient chlorophyll relationships for benthic algae. Journal of the North American Benthological Society 19: 17–31.

Bloschl, G., Sivapalan, M. 1995. Scale issue in hydrological modelling – a review, Hydrological Processes 9, 251–290.

Booker, D.J. (in press) Spatial and temporal patterns in the frequency of events exceeding three times the median flow (FRE3) across New Zealand. Journal of Hydrology (NZ).

Booker, D.J., Snelder, T.H. 2012. Comparing methods for estimating flow duration curves at ungauged sites. Journal of Hydrology 434–435, 78–94.

Botter, G., Porporato, A., Rodriguez-Iturbe, I., Rinaldo, A. 2007a. Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. Water Resour. Res. 43, W02417, doi:10.1029/2006WR005043.

Botter, G., Porporato, A., Daly, E., Rodriguez-Iturbe, I., Rinaldo, A. 2007b. Probabilistic characterization of base flows in river basins: Roles of soil, vegetation, and geomorphology. Water Resour. Res. 43, W06404, doi:10.1029/2006WR005397.

Botter, G., Porporato, A., Rodriguez-Iturbe, I., Rinaldo, A. 2009. Nonlinear storage-discharge relations and catchment streamflow regimes. Water Resour. Res. 45, W10427, doi:10.1029/2008WR007658.

Breiman, L. 2001: Random forests. Machine Learning 45, 15–32.

Brekke, L.D., Miller, N.L., Bashford, K.E., Quinn, N.W.T., Dracup, J.A. 2004. Climate change impacts uncertainty for water resources in the San Joaquin River basin, California. Journal of the American Water Resources Association 40, 149–164.

Caruso, B.S. 2002. Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago, New Zealand. Journal of Hydrology 257, 115–133.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A. 2004: Regional flow-duration curves: reliability for ungauged basins. Advances in Water Resources 27, 953–965.

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., Sivapalan, M. 2012. Exploring the physical controls of regional patterns of flow duration curves – Part 1: Insights from statistical analyses. Hydrol. Earth Syst. Sci. 16, 4435–4446.

Chow, V.T., Maidment, D.R., Mays, L.W. 1988. Applied Hydrology. McGraw-Hill, USA.

Clapcott, J.E., Collier, K.J., Death, R.G., Goodwin, E.O., Harding, J.S., Kelly, D., Leathwick, J.R., Young, R.G. 2011. Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. Freshwater Biology, 10.1111/j.1365-2427.2011.02696.x

Clapcott, J.E., Young, R.G., Goodwin, E.O., Leathwick, J.R. 2010. Exploring the response of functional indicators of stream health to land-use gradients. Freshwater Biology 55, 2181–2199.

Clark, M.P., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Rupp, D.E., Schmidt, J., Uddstrom, M.J. 2008. Hydrological data assimilation with the Ensemble Kalman Filter: use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources 31, 1309–1324.

Clausen, B., Biggs, B.J.F. 1997. Relationships between benthic biota and hydrological indices in New Zealand streams. Freshwater Biology 38, 327-342.

Clausen, B., Plew, D. 2004. How high are bed-moving flows in New Zealand rivers? Journal of Hydrology (NZ) 43, 19–37.

Coopersmith, E., Yaeger, M.A., Ye, S., Cheng L., Sivapalan, M. 2012. Exploring the physical controls of regional patterns of flow duration curves – Part 3: A catchment classification system based on regime curve indicators. Hydrology and Earth System Sciences 16, 4467–4482.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

Earman, S., Dettinger, M. 2011. Potential impacts of climate change on groundwater resources-a global review. Journal of Water and Climate Change 24, 213–229.

EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council of October 23 2000 Establishing a Framework for Community Action in the Field of Water Policy. Official Journal of the European Communities, L327/1eL327/72. 22.12.2000.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, Pa. 92p.

Elith, J., Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77.

Engeland, K., Hisdal, H. 2009. A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model. Water Resources Management 23, 2567–2586

Fahey, B. Jackson, R. 1997. Hydrological impacts of converting native forests and grasslands to pine plantations, South Island, New Zealand. Agricultural and Forest Meteorology 84, 69–82.

Fenicia, F., McDonnell, J.J., Savenije, H.H.G. 2008. Learning from model improvement: on the contribution of complementary data to process understanding, Water Resources. Research 44, W06419, doi:10.1029/2007WR006386.

Ganora, D., Claps P., Laio, F., Viglione A. 2009. An approach to estimate nonparametric flow duration curves in ungauged basins. Water Resources Research 45, W10418, doi:10.1029/2008WR007472.

Goring, D.G. 1994. Kinematic shocks and monoclinal waves in the Waimakariri, a steep, braided, gravel-bed river, in: Proceedings of the International Symposium on waves: Physical and numerical modelling, University of British Columbia, Vancouver, Canada, 336–345.

Gupta, H.V., Beven, K.J., Wagener, T. 2006. Model Calibration and Uncertainty Estimation. Encyclopedia of Hydrological Sciences, in: Anderson, M.G. (Ed.). Encyclopedia of Hydrological Sciences, John Wiley & Sons Ltd: Chichester, pp. 1–17.

Gupta, H.V., Sorooshian, S., Yapo. P.O. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrologic Eng. 4(2): 135–143.

Hamilton, J.D. 1994. Time Series Analysis. Princeton University Press, Princeton.

Henderson, R.D., Woods, R.A., Schmidt, J. 2004. A new low flow model for New Zealand - Part 3. New Zealand Hydrological Society 2004 Conference abstract, Queenstown, November, 2004.

Henderson, R. D., Woods, R. A., Singh, S.K., Zammit, C.L. 2011. Surface Water Components of New Zealand's National Water Accounts. NIWA Client Report. Christchurch, New Zealand, NIWA. CHC2011-051 (SNZ11501) for Statistics New Zealand: 45pp.

Jakeman, A.J., Hornberger, G.M. 1993. How much complexity is warranted in a rainfall-runoff model? Water Resources Research 29, 2637–2649.

Kiem, A.S., Franks, S.W., Kuczera, G. 2003. Multi-decadal variability of flood risk. Geophysical Research Letters 30, DOI:10.1029/2002GL015992.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., Hastie, T. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshwater Biology 50, 2034–2052.

Leathwick, J.R., Snelder, T., Chadderton, W.L., Elith, J., Julian, K., Ferrier, S., 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. Freshwater Biology 56, 21–38.

Legates, D. R., McCabe, G. J. 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233-241.

Li, M., Shao, Q., Zhang, L., Chiew, F.H.S. 2010. A new regionalization approach and its application to predict flow duration curve in ungauged basins. Journal of Hydrology 389, 137–145.

McDonnell, J., Woods, R. 2004. On the need for catchment classification. Journal of Hydrology 299, 2–3.

McKerchar, A.I., Henderson, R.D. 2003. Shifts in flood and low-flow regimes in New Zealand due to interdecadal climate variations. Hydrological Sciences 48, 637–654.

McKerchar, A.I., Pearson, C.P. 1989. Flood Frequency in New Zealand. Hydrology Centre Publication No. 20, DSIR Division of Water Sciences, Christchurch, N.Z. 87p.

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M. J. 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. Hydrol. Earth Syst. Sci. 17, 21–38,

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer R.J. 2008. Climate change: stationarity is dead: whither water management? *Science 319*, 573–574.

MFE 2008. Proposed national environmental standard on ecological flows and water levels: discussion document. Wellington, New Zealand: New Zealand Ministry for the Environment. 61p.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the American Society of Agricultural and Biological Engineers 50, 885–900.

Mosley, M.P., Pearson, C.P. 1997. Floods and droughts: the New Zealand experience. New Zealand Hydrological Society, Wellington, New Zealand, 206p.

Muneepeerakul, R., Azaele, S., Botter, G., Rinaldo, A., Rodriguez-Iturbe, I. 2010. Daily streamflow analysis based on a two scaled gamma pulse model, Water Resour. Res. 46, W11546, doi:10.1029/2010WR009286,

Murphy, M.A., Evans, J.S., Storfer, A. 2010. Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology 91, 252–261

Nash, J.E., Sutcliffe, J.V. 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles. Journal of Hydrology 10, 282–290.

New Zealand Government, 2011. National Policy Statement for Freshwater Management 2011. New Zealand.

Olden, J., Kennard, M., Pusey, B. 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. Ecohydrology 5, 503–518.

Olden, J.D., Poff, N.L. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications 19, 101–121.

Parkin, G., O'Donnell, G., Ewen, J., Bathurst, J.C., O'Connell, P.E., Lavabre, J. 1996. Validation of catchment models for predicting land-use and climate change impacts. 2. Case study for a Mediterranean catchment. Journal of Hydrology 175, 595–613.

Parry, O.F., Canziani, J.P., van den Linden, P.J., Hanson, C.E. 2007. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ Press, Cambridge, UK.

Pearson, C.P. 1995. Regional frequency analysis of low flows in New Zealand rivers. Journal of Hydrology (NZ) 30, 53–64.

Pearson, C.P., McKerchar, A.I. 1989. Flood estimation - a revised procedure. Transactions of the Institution of Professional Engineers New Zealand 16(2/CE): 59–65.

Pechlivanidis, I.G., Jackson, B.M., McIntyre, N.R., Wheater, H.S. 2011. Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global NEST Journal 13, 193–214.

Piñeiro, G., Perelman, P., Guerschman, J.P., Paruelo, J.M. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216, 316–322.

Poff, N.L. 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. Freshwater Biology 36, 71–91.

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J. R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C. 1997. The natural flow regime. A paradigm for river conservation and restoration. Bioscience 47, 769–784.

Poff, N.L., Richter, B.D., Arthington, A.H., Bunn, S.E., Naiman, R.J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B.P., Freeman, M.C., Henriksen, J., Jacobson, R.B., Kennen, J.G., Merritt, D.M., O'Keefe, J.H., Olden, J.D., Rogers, K., Tharme, R.E., Warner, A. 2010. The Ecological Limits of Hydrologic Alteration (ELOHA): A new framework for developing regional environmental flow standards. Freshwater Biology 55, 147–170.

Poff, N.L., Ward J.V. 1989. Implications of streamflow variability and predictability for lotic community structure: a regional analysis of streamflow patterns. Canadian Journal of Fisheries and Aquatic Sciences 46, 1805–1818.

Richter, B.D., Sparks, R.E., Stromberg, J.C. 1997. The natural flow regime. A paradigm for river conservation and restoration. Bioscience 47, 769–784.

Rupp, D. E., R. F. Keim, M. Ossiander, M. Brugnach and J. S. Selker, 2009. Time scale and intensity dependency in multiplicative cascades for temporal rainfall disaggregation, Water Resources Research, 45, W07409, doi:10.1029/2008WR007321.

Scanlon, B.R., Jolly, I., Sophocleous, M., Zhang, L. 2007. Global impacts of conversions from natural to agricultural ecosystems on water resources: Quantity versus quality. Water Resources Research 43, doi:10.1029/2006WR005486.

Singh, V.P. 1995. Watershed modelling, in: Singh, V.P. (Ed.) Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, pp. 1–22.

Singh V.P., Frevert D. 2006. Watershed models. Boca Raton, Taylor & Francis.

Singh, V., Woolhiser, D. 2002. Mathematical Modeling of Watershed Hydrology. Journal of Hydrologic Engineering 7, 270–292.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe E. 2003. IAHS decade on predictions in ungauged basins (PUB), 2003-2012: shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal 48, 857–880.

Smakhtin, V.U. 2001. Low flow hydrology: a review. Journal of Hydrology 240, 147–186.

Snelder, T.H., Biggs, B.J.F. 2002. Multi-scale river environment classification for water resources management. Journal of the American Water Resources Association 38, 1225–1240.

Snelder, T.H., Booker, D.J. 2012. Natural flow regime classifications are sensitive to definition procedures. River Research and Applications. DOI: 10.1002/rra.2581.

Snelder, T.H., Booker, D.J., Lamouroux, N. 2011. A method to assess and define environmental flow rules for large jurisdictional regions. International Journal of River Basin Management 1-13, DOI: 10.1111/j.1752-1688.2011.00556.x

Snelder, T.H., Cattaneo, F., Suren, A.M., Biggs, B.J.F., 2004. Is the river environment classification an improved landscape-scale classification of rivers? J. North Am. Benthol. Soc. 23, 580–598.

Snelder, T.H., Hughey, K.F.D., 2005. On the use of an ecological classification to improve water resource planning in New Zealand. Environ. Manage. 36, 741–756.

Snelder, T.H., Rouse, H.L., Franklin, P.A., Booker, D.J., Norton, N., Diettrich, J. 2013. The role of science in setting water resource use limits: a case study from New Zealand, Hydrological Sciences Journal, DOI:10.1080/02626667.2013.793799

Snelder, T.H., Woods, R., Biggs, B.J.F. 2005. Improved eco-hydrological classification of rivers. River Research and Applications 21, 609–628.

Suren, A.M., Jowett, I.G. 2006. Effects of floods versus low flows on invertebrates in a New Zealand gravel-bed river. Freshwater Biology 51, 2207–2227.

Tait, A. (2008). Future projections of growing degree days and frost in New Zealand and some implications for grape growing. Weather and Climate 28, 17–36.

Tait, A., Henderson, R.D., Turner, R., Zheng, X. 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. International Journal of Climatology 26, 2097–2115.

Toebes, C., Palmer, B.R., 1969. Hydrological regions of New Zealand. New Zealand Ministry of Works Miscellaneous Hydrological Publication No. 4. 45p.

Vogel, R.M., Sieber, J. Archfield, S.A., Smith, M.P., Apse, C.D., Huber-Lee, A. 2007. Relations among storage, yield, and instream flow. Water Resources Research 43, W05403, doi:10.1029/2006WR005226.

Vogel, R.M., Fennessey, N.M., 1994. Flow duration curves I: new interpretation and confidence intervals. J. Water Resour. Plann. Manage. (ASCE) 120, 485–504.

Vogel, R.M., Fennessey, N.M., 1995. Flow duration curves II: a review of applications in water resources planning. Journal of the American Water Resources Association 31, 1029–1039.

Vörösmarty, C.J., McIntyre, P., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Reidy Liermann, C., Davies, P.M. 2010. Global threats to human water security and river biodiversity. Nature 467, 555–561.

Wagener T., 2007. Can we model the hydrological impacts of environmental change? Hydrological Processes 21, 3233–3236.

Westerberg, I.K., Guerrero, J.-L., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J.E, Xu, C.-Y. 2011. Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences 15, 2205–2227.

Woods, R.A., Hendrikx, J., Henderson, R.D., Tait, A.B., 2006. Estimating mean flow of New Zealand rivers. Journal of Hydrology (NZ) 45, 95–110.

Yadav, M., Wagener, T. and Gupta, H.V. 2007. Regionalization of constraints on expected watershed response behavior. Advances in Water Resources, 30, 1756-1774. doi:10.1016/j.advwatres.2007.01.005

Yaeger, M.A., Coopersmith, E., Ye, S., Cheng L., Viglione, A., Sivapalan, M. 2012. Exploring the physical controls of regional patterns of flow duration curves – Part 4: A synthesis of empirical analysis, process modelling and catchment classification. Hydrology and Earth System Sciences 16, 4483–4498.

Yokoo, Y., Sivapalan, M. 2011. Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis, Hydrol. Earth Syst. Sci. 15, 2805–2819, doi:10.5194/hess-15-2805-2011.

Yu, P.S., Yang,T.C. 2000. Using synthetic flow duration curves for rainfall–runoff model calibration at ungauged sites. Hydrological Processes, 14, 117–133.

Zemansky, G., Hong, Y-S.T, Rose, J., Song, S.H., Thomas, J. 2012. Assessing the effects of climate change on water resources: the Waimea Plains. Journal of Hydrology (NZ) 51, 45–62.

Tables

Table 1. Codes, descriptions and numbers of sites used in the analysis. See Snelder and Biggs (2002) and Snelder and Hughey (2005) for full descriptions of codes.

Table 2. Hydrological Indices derived from observed mean daily flows.

Table 3. Summary of the defining characteristics, categories and category membership criteria that combine to define Source-of-Flow groupings within the REC.

Table 4. Codes and descriptions of independent variables used to fit regression models. See Leathwick et al., (2011) for full descriptions.

Table 5. Various metrics quantifying correspondence between observed and predicted values for four hydrological indices (Table 2) using various estimation methods.

Figures

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

Figure 2. Schematic showing different methods used to estimate hydrological indices and flow duration curves (FDCs).

Figure 3. Hydrology of ungauged catchments (HUC) low flow model and parameters.

Figure 4. Observed against calculated values for each index for each method (n = 485). Grey dashed line is linear regression. Black line is 1:1 such that x-limits are equal to y-limits for all plots. Qbar is mean flow. QFeb is proportion of flow in February. QMALF is 7-day mean annual low flow. QF is mean annual flood.

Figure 5. Box and whisker plots of Nash-Sutcliffe efficiency, RSR (ratio of the root mean square error to the standard deviation of observed data) and pbias (average tendency of the calculated data to be larger or smaller than their observed counterparts) at each site for all-time and February flow duration curves for each method (n = 101 points at each of 485sites).

Figure 6. All observations and for each method predictions of 7-day mean annual low flow (MALF) for all rivers of Strahler order greater than three. TopNet results are for uncorrected TopNet Version 1.

Figure

Figures

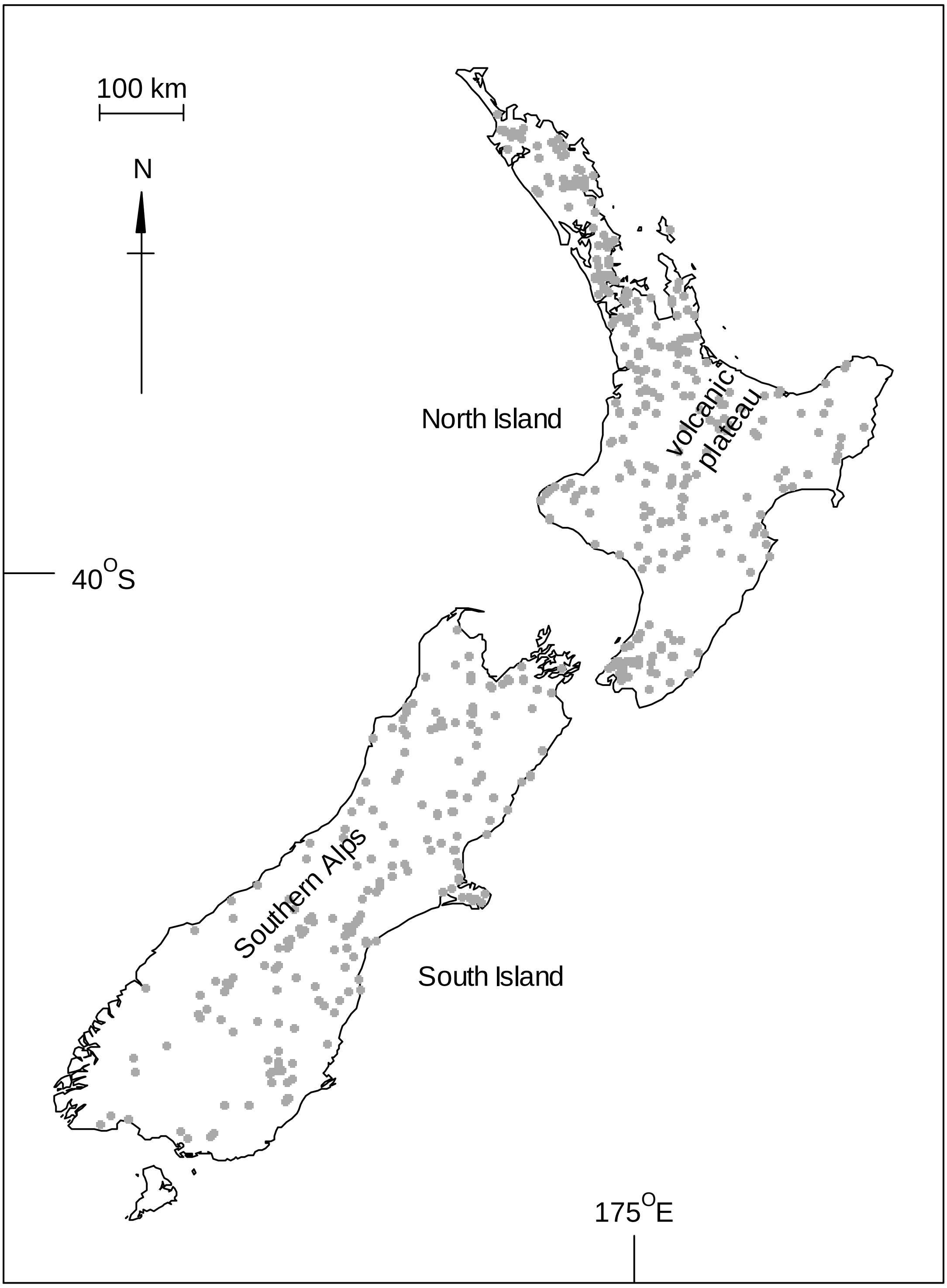

Figure 1. Map showing the locations of the gauging stations used in this study.

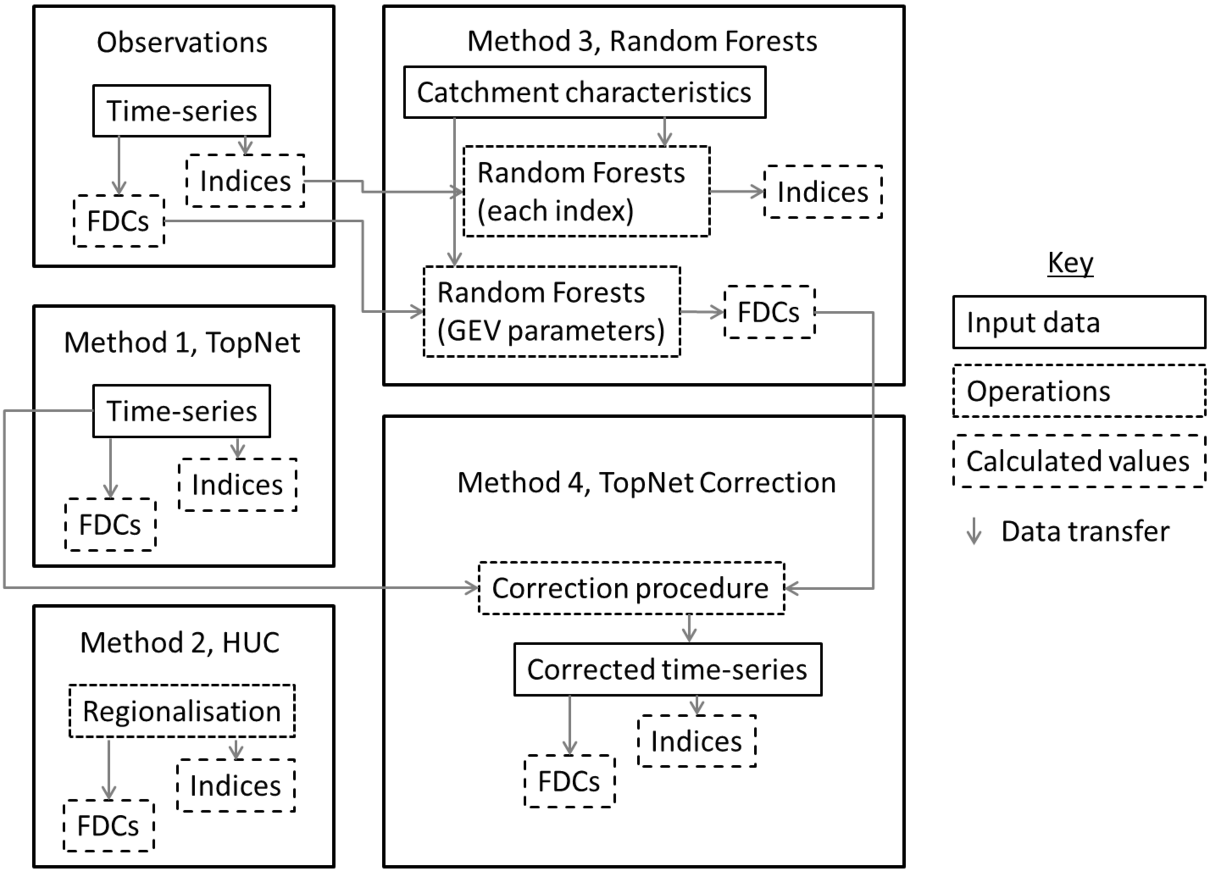

Figure 2. Schematic showing different methods used to estimate hydrological indices and flow duration curves (FDCs).

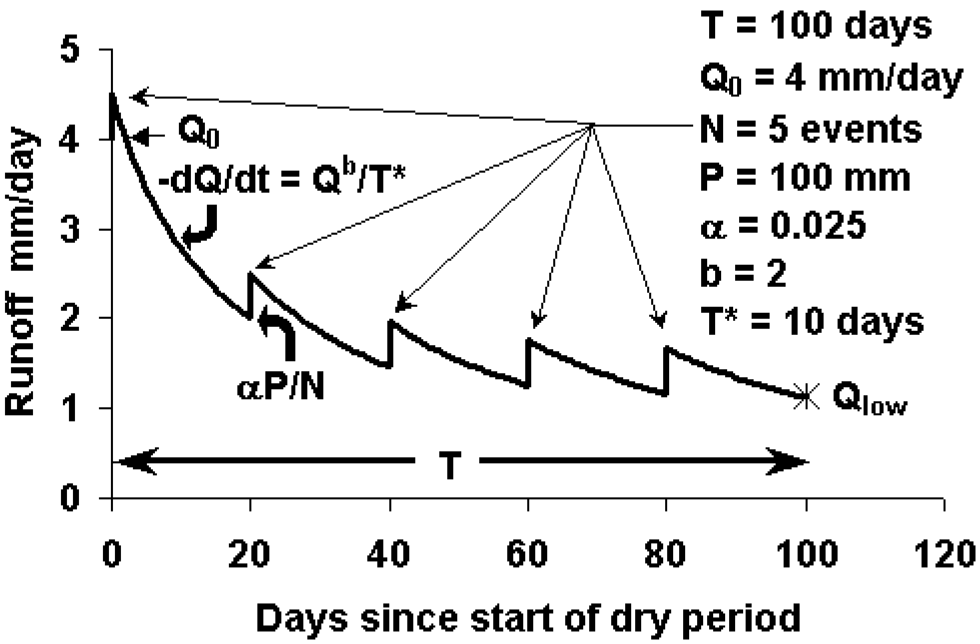

Figure 3. Hydrology of ungauged catchments (HUC) low flow model and parameters.

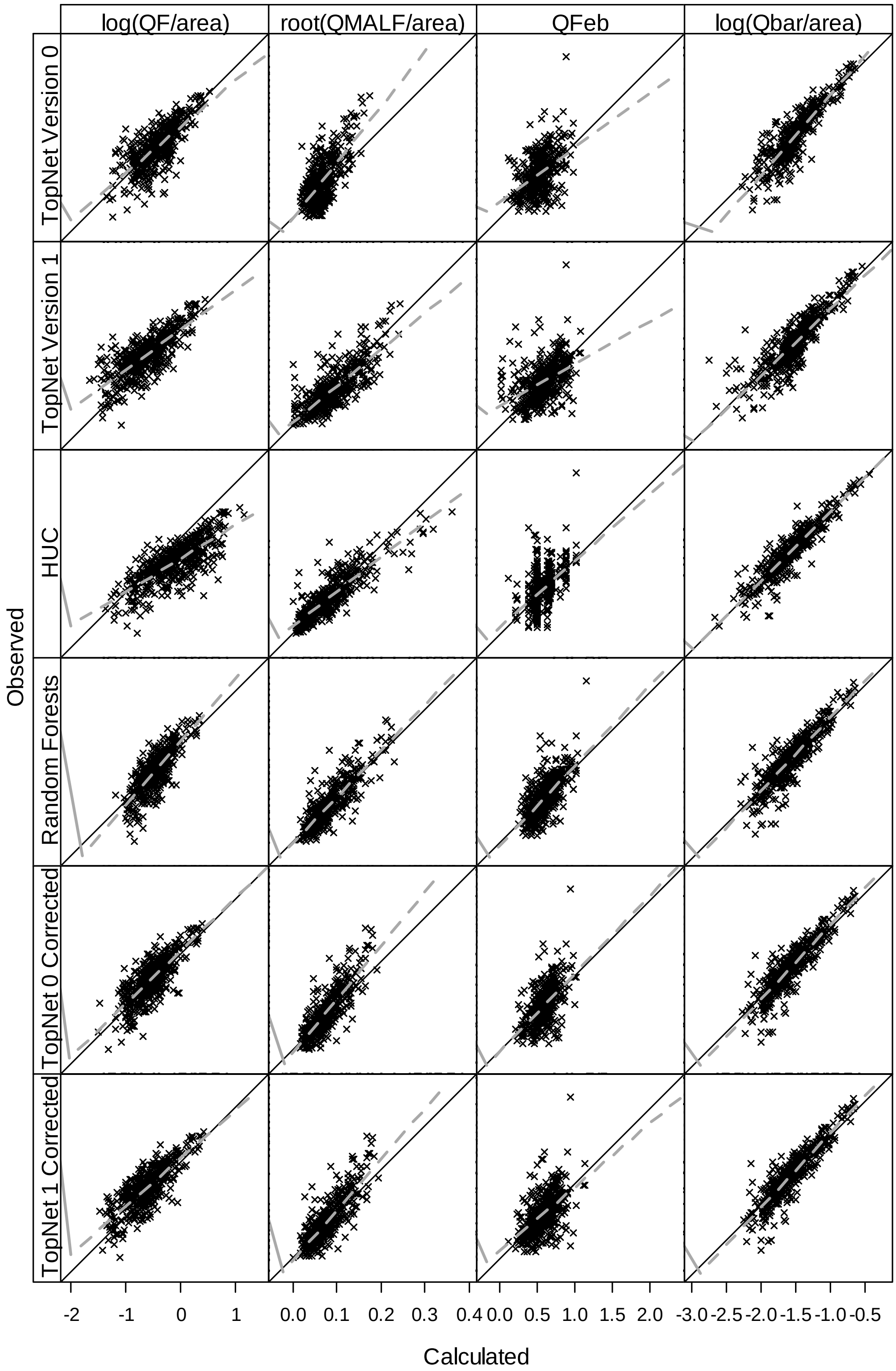

Figure 4. Observed against calculated values for each index for each method (n = 485). Grey dashed line is linear regression. Black line is 1:1 such that x-limits are equal to y-limits for all plots. Qbar is mean flow. QFeb is proportion of flow in February. QMALF is 7-day mean annual low flow. QF is mean annual flood.

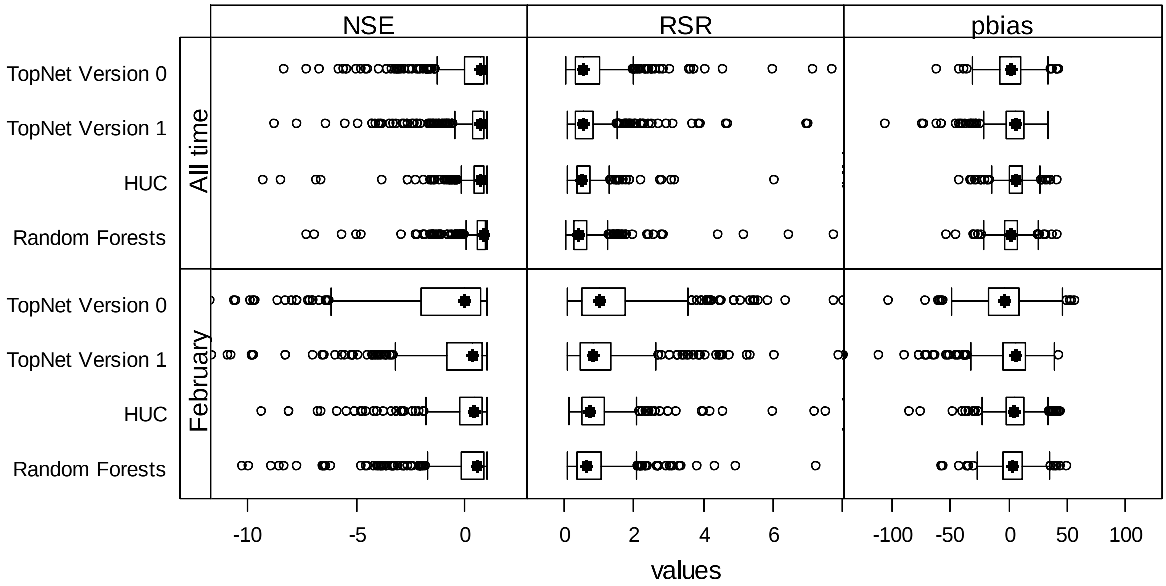

Figure 5. Box and whisker plots of Nash-Sutcliffe efficiency, RSR (ratio of the root mean square error to the standard deviation of observed data) and pbias (average tendency of the calculated data to be larger or smaller than their observed counterparts) at each site for all-time and February flow duration curves for each method (n = 101 points at each of 485sites).

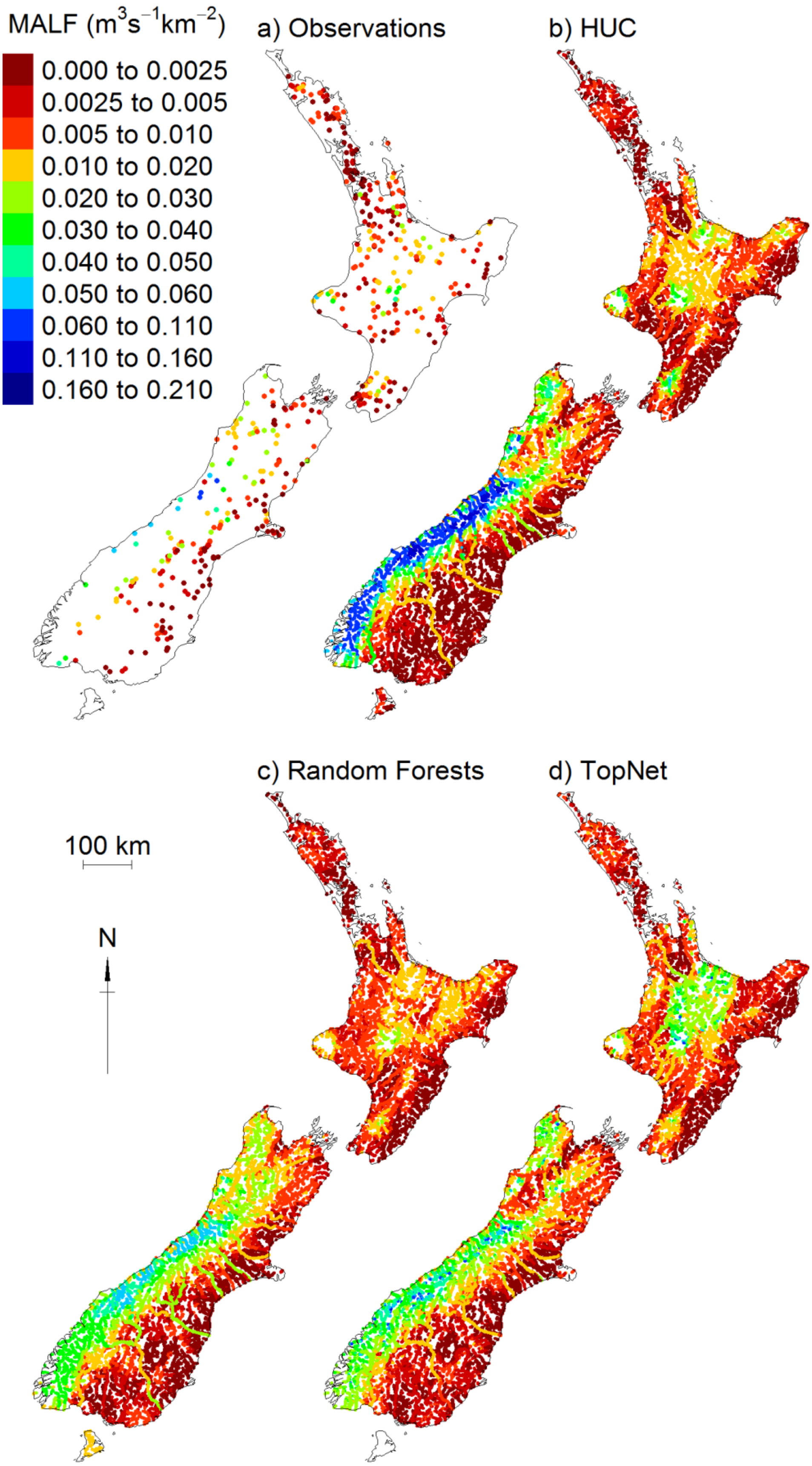

Figure 6. All observations and for each method predictions of 7-day mean annual low flow (MALF) for all rivers of Strahler order greater than three. TopNet results are for uncorrected TopNet Version 1.

Tables

Table 1. Codes, descriptions and numbers of sites used in the analysis. See Snelder and Biggs (2002) and Snelder and Hughey (2005) for full descriptions of codes.

Code	Description	Number of sites, total
Island		
N	North Island	289
S	South Island	196
Climate		
WD	Warm-dry	18
WW	Warm-wet	152
WX	Warm-extremely wet	4
CD	Cool-dry	75
CW	Cool-wet	154
CX	Cool-extremely wet	82
Topographic source of flow		
GM	Glacial mountain	10
H	Hill	167
L	Low elevation	241
Lk	Lake	19
M	Mountain	48
Land cover		
B	Bare	16
EF	Exotic-Forest	22
IF	Indigenous-Forest	105
P	Pastoral	247
S	Scrub	17
T	Tussock	63
U	Urban	15

Table 2. Hydrological Indices derived from observed mean daily flows.

Index	Description	Calculation	Standardisation	Transformation
Q_{bar}	Mean flow over all time	Mean of all daily flows	Divide by catchment area to get specific mean flow ($m^3\ s^{-1}\ km^{-2}$)	Log base 10
Q_{Feb}	Proportion of flow in February	Mean of all daily flows for each calendar month after having divided by the overall mean flow	Divide by mean flow over entire record to get proportion of flow in February (unitless)	None
Q_{MALF}	Mean of minimum 7-day flow in each year	Mean of minimum flow for each water year after having applied a running 7-day mean to the daily flows	Divide by catchment area to get specific QMALF ($m^3\ s^{-1}\ km^{-2}$)	Square root
Q_F	Mean of maximum flow in each year	Mean of maximum flow for each water year	Divide by catchment area to get specific QF ($m^3\ s^{-1}\ km^{-2}$)	Log base 10
FDC	Probability distribution of daily flow	Interpolation of the cumulative frequency distribution of daily flows on to 101 points (0 to 100 in steps of 1)	Divide by catchment area to get specific FDC ($m^3\ s^{-1}\ km^{-2}$)	Log base 10
FDC_{Feb}	Probability distribution of daily flow for February	Interpolation of the cumulative frequency distribution of daily flows for each calendar month on to 101 points (0 to 100 in steps of 1)	Divide by catchment area to get specific FDC ($m^3\ s^{-1}\ km^{-2}$)	Log base 10

Table 3. Summary of the defining characteristics, categories and category membership criteria that combine to define Source-of-Flow groupings within the REC.

Defining characteristic	Categories	Notation	Category membership criteria
Climate	Warm-extremely-wet	WX	Warm: mean annual temperature ≥ 12°C
	Warm-wet	WW	Cool: mean annual temperature < 12°C
	Warm-dry	WD	Extremely Wet: mean annual effective precipitation[a] ≥ 1500 mm
	Cool-extremely-wet	CX	Wet: mean annual effective precipitation > 500 and < 1500 mm
	Cool-wet	CW	Dry: mean annual effective precipitation ≤ 500 mm
	Cool-dry	CD	
Topography	Glacial-mountain	GM	GM: M and % permanent ice > 1.5%
	Mountain	M	M: > 50% annual rainfall volume above 1000 m ASL
	Hill	H	H: 50% rainfall volume between 400 and 1000 m ASL
	Low-elevation	L	L: 50% rainfall below 400 m ASL
	Lake	Lk	Lk: Lake influence index[b] > 0.033

a. Effective precipitation = annual rainfall – annual potential evapotranspiration

b. See Snelder and Biggs (2002) for a description.

Table 4. Codes and descriptions of independent variables used to fit regression models. See Leathwick et al., (2011) for full descriptions.

Variable name	Description
usPET_Q	Annual potential evapotranspiration of catchment (mm)
usRainDays10_Q	Catchment rain days, greater than 10 mm/month (days/year)
usAnRainVar_Q	Coefficient of variation of annual catchment rainfall (m)
usSteep_Q	% annual runoff volume from area of catchment with slope > 30° (%)
usCatElev	Average elevation in the upstream catchment (m)
usParticleSize_Q	Catchment average of particle size (ordinal scale)

Table 5. Various metrics quantifying correspondence between observed and predicted values for four hydrological indices (Table 2) using various estimation methods.

Index	Method	n	NSE	pbias	RSR
$\log(Q_{bar}/area)$					
	TopNet_0	485	0.73	4.050	0.523
	TopNet_1 Sync	456	0.70	3.138	0.552
	TopNet_1	485	0.71	3.469	0.537
	HUC	485	0.87	0.298	0.363
	RFjacked	485	0.80	-0.241	0.446
	TopNet_0 Corrected	485	0.80	-0.410	0.447
	TopNet_1 Corrected	485	0.80	-0.433	0.447
Q_{Feb}					
	TopNet_0	485	0.09	11.733	0.955
	TopNet_1 Sync	456	0.29	-2.420	0.843
	TopNet_1	485	0.08	2.499	0.960
	HUC	485	0.22	5.354	0.884
	RFjacked	485	0.44	0.216	0.748
	TopNet_0 Corrected	485	0.31	2.872	0.828
	TopNet_1 Corrected	485	0.27	3.020	0.853
$\mathrm{root}(Q_{MALF}/area)$					
	TopNet_0	485	0.36	17.496	0.797
	TopNet_1 Sync	454	0.59	-11.031	0.643
	TopNet_1	485	0.58	-10.739	0.646
	HUC	485	0.71	-0.506	0.540
	RFjacked	485	0.75	0.157	0.499
	TopNet_0 Corrected	485	0.66	9.132	0.587
	TopNet_1 Corrected	485	0.67	5.923	0.571
$\log(Q_F/area)$					
	TopNet_0	485	0.50	7.523	0.704
	TopNet_1 Sync	456	0.30	-36.797	0.837
	TopNet_1	485	0.31	-34.958	0.832
	HUC*	485	-0.45	73.012	1.206
	RFjacked	485	0.63	-0.674	0.609
	TopNet_0 Corrected	485	0.55	-16.521	0.668
	TopNet_1 Corrected	485	0.46	-31.733	0.734

* Table footnote: In this comparison HUC estimates of instantaneous Q_F were compared with observed Q_F calculated from mean daily flow data.