Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Estimation of extreme flow quantiles and quantile uncertainty for ungauged catchments

DONALD H. BURN[1], TAHA B. M. J. OUARDA[2] & CHANG SHU[2]

1 *Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada*
dhburn@civmail.uwaterloo.ca

2 *INRS-ETE, University of Québec, 490 rue de la Couronne, Québec City, Quebec G1K 9A9, Canada*

Abstract Pooled frequency analysis is used to estimate extreme event quantiles at catchments where the data record is either short or not available. This can be accomplished by combining (pooling) information from hydrologically similar sites to increase the available information for estimating the required extreme event quantiles. This paper compares methods for estimating extreme event quantiles at ungauged catchments and determines the uncertainty associated with the estimated extreme event quantiles. The techniques are demonstrated and evaluated using data from a collection of catchments in the Canadian province of Ontario. A geographic nearest neighbour index flood-based approach resulted in the lowest mean squared error value.

Key words frequency analysis; quantile estimation; site-focused pooling; uncertainty; ungauged sites

INTRODUCTION

Pooled flood frequency analysis is used when at-site data are insufficient to provide a reliable estimate of extreme flood quantiles. The pooling process involves combining extreme flow information from gauged sites to enhance the estimation of quantiles at sites that are either ungauged or have a short data record. Two main approaches to pooled flood frequency analysis are the index flood method (Dalrymple, 1960) and quantile regression. In the index flood method, the distributions of flood peaks at different sites within a pooling group are assumed to be the same except for a scale parameter, which is the index flood for the site. In quantile regression, catchment characteristics of a site are related to the flood quantile of interest using a regression-based model.

The delineation of pooling groups was traditionally based on geographical boundaries leading to spatially contiguous pooling groups (regions). However, geographical closeness does not necessarily guarantee hydrological similarity. Acreman & Wiltshire (1989) suggested a pooling approach that involves dispensing with fixed groups. This idea was further developed by Burn (1990) into a focused pooling scheme, wherein a homogeneous group is specifically designed for a site of interest (the target site).

Most pooling schemes are based on grouping catchments according to physiographic similarity. At-site flood statistics are then used to test the homogeneity of the pooling groups. A pooling approach that does not include flood magnitude data, and has recently gained popularity, is based on flood seasonality (Magilligan & Graber, 1996; Burn, 1997; Castellarin *et al.*, 2001; Cunderlik & Burn, 2002; Cunderlik *et al.*, 2004; Ouarda *et al.*, 2006). The main advantage of this approach is that flood seasonality is described using flood date data, which are practically error-free.

This paper addresses issues associated with estimating extreme event quantiles at ungauged catchments using focused pooling approaches. A variety of methods for estimating extreme event quantiles are evaluated in terms of the accuracy of the extreme quantile estimates and the amount of uncertainty associated with the estimates. The options that are evaluated and compared include several nearest neighbour approaches and an artificial neural network ensemble approach. The techniques are demonstrated and evaluated using data from catchments in the Canadian province of Ontario.

METHODOLOGY

Seasonality measures

The timing and regularity of flood events can be described in terms of directional statistics (Fisher, 1993). Dates of flood occurrence are defined as a directional statistic by converting the Julian date to an angular value (θ_{i}) using (Bayliss & Jones, 1993):

$$\theta_{\mathrm{i}} = (JulianDate)_{\mathrm{i}} \frac{2\pi}{365} \tag{1}$$

A flood date can be interpreted as a vector with unit magnitude and a direction given by θ_{i}. The mean direction ($\overline{\theta}$) is calculated as the addition of unit vectors; the resultant vector has the direction of the mean direction of the individual unit vectors. A measure of the variability of flood occurrences about the mean date, $\bar{r}$, can be derived as the mean resultant length calculated from coordinates of individual dates of flood occurrence:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \; ; \bar{r} \in \langle 0,1 \rangle \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_{\mathrm{i}}) \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n}\sin(\theta_{\mathrm{i}}) \tag{2}$$

where n is the number of flood samples and $\bar{x}$ and $\bar{y}$ give the Cartesian coordinates of the catchment in seasonality space. The parameter $\bar{r}$ provides a dimensionless measure of the spread of the data. A value of unity indicates that all floods in the sample occurred on the same day of the year while values closer to zero indicate that there is greater variability in the date of occurrence of flood events for a given catchment (Bayliss & Jones, 1993). The dissimilarity between two catchments, i and j, can be defined using the Euclidean distance:

$$DIST_{ij} = \sqrt{\left(\bar{x}_i - \bar{x}_j\right)^2 + \left(\bar{y}_i - \bar{y}_j\right)^2} \tag{3}$$

A focused pooling group for a site of interest can then be formed by sequentially adding the next most similar catchment to the pooling group for the target site. This process continues until the target number of station years of record have been added to the pooling group or other stopping criteria are satisfied.

Index flood-based approaches for ungauged catchments

Three index flood-based approaches are evaluated for estimating flood quantiles at ungauged catchments. The first approach involves using similarity in catchment characteristic space as the basis for forming a pooling group. This method will be referred to as CC. The second involves using catchment characteristics to define the closest gauged catchment to the target site and then using the seasonality pooling group for the nearest neighbour (NN) site to represent the pooling group for the target site. This option is referred to as CC–NN. The third approach involves using the seasonality pooling group for the geographically closest gauged site as the pooling group for the target ungauged site. This option is referred to as G–NN. For each option, catchments are added to the pooling group until 500 station-years of record are included. This implies sufficient information in the pooling group to accurately estimate the 100-year event, based on the 5T guideline (Jakob *et al.*, 1999).

Information from stations in the pooling group is combined using a weighted average of the L-moment ratios for the stations in the pooling group where the weighting depends on the number of years of record for a station and the similarity of the station to the target site. The weighted L-moments ratios are then used to estimate quantiles for the growth curve at the target site based on the specified distribution function (for further details, see Burn, 2003).

The uncertainty associated with the quantile estimates for each of the index flood-based options is estimated through a balanced resampling process. Resampling approaches involve creating new samples from the original sample by a bootstrapping process that involves randomly selecting data points, with replacement, from the original sample. In balanced resampling, each data point appears the same number of times in the union of the resampled data sets. To preserve the spatial correlation structure of the data in the pooling group, a vector bootstrap approach is used. In vector bootstrapping, resampling is done on years such that selecting a year implies that all sites with a data value for that year have the corresponding data value included in the bootstrap sample. This approach ensures that the spatial correlation structure in the original data set is preserved in the resampled data sets. The use of vector balanced resampling implies that all years from the collection of years with data will appear the same number of times in the union of the resampled data sets. Once the pooled data set has been assembled, quantiles can be estimated in accordance with the pooling method employed and confidence intervals calculated from the empirical distribution of the T-year quantiles (Burn, 2003).

Quantile regression based approaches for ungauged catchments

Quantile regression was implemented using an artificial neural network (ANN) ensemble method. An ANN is an information processing system with massive parallelism and high connectivity (Haykin, 1994) and may be treated as a universal approximator. The nonlinear nature of the activation function enables ANNs to form highly nonlinear functional relationships from observed data.

An ANN ensemble is used to improve the generalization capability and stability of a single ANN. An ANN ensemble consists of a number of ANNs that are trained for the same purpose; the results produced by these individual networks are combined to generate a unique output. Shu & Burn (2004) introduced ANN ensemble methods to estimate the index flood and flood quantiles at ungauged sites. There are two steps for generating an ANN ensemble: (1) generate individual ensemble members; and then (2) combine the multiple member outputs to produce the output of the ensemble. Ensemble members can be formed by training the different networks on different subsets of the training set using the bagging or boosting algorithms. Averaging the results across the individual networks is a frequently used approach for combining the results of different networks. A more detailed review of different ensemble techniques is provided in Shu & Burn (2004).

In this paper, the selected network type is the Multilayer Perceptron (MLP). The MLP adopted in this paper consists of an input layer, one hidden layer, and an output layer. The input layer accepts values of the input variables, which are the available catchment characteristics. The output layer provides the estimation of an extreme event quantile. Layers between the input and output layer are called hidden layers. Five neurons are used in the hidden layer of the ANN. The tan-sigmoid transfer function is used in the hidden layer and the linear transfer function is used in the output layer. Input and output data are standardized before being provided to the ANNs. The Levenberg-Marquardt algorithm is used for ANN training.

The bootstrap aggregation (bagging) approach is used to generate the individual ANNs. The bagging algorithm was introduced by Breiman (1996) to improve the accuracy of predictions. The algorithm is based on the bootstrap statistical resampling technique (Efron & Tibshirani, 1993). Suppose a bootstrap sample is drawn from the training set with replacement, and the sample size is set to the same size as the size of the training set. An ANN is then trained by using the bootstrap sample and the process is repeated a number of times until the desired number of member networks is generated (an ensemble size of 20 is used here). The ensemble output is derived by averaging the outputs from the member networks. Finally, a confidence interval can be estimated using the mean and standard deviation of the 20 member networks and assuming that the Normal distribution applies. This estimation option is referred to as ANN.

APPLICATION

Description of case study area

The ungauged quantile estimation options were applied to a collection of 85 catchments in the province of Ontario (Canada). Figure 1 illustrates the location of the catchments. The sites cover three ecozones and two climatic zones and represent a range of basin sizes from 40 km^2 to 8900 km^2 (median basin size is 340 km^2). The annual maximum daily flow values for the 85 sites have a median record length of 38 years (range of 23 to 90 years). For each catchment, four catchment descriptors were used: catchment area (in km^2); slope of the principal watercourse (in m/km); fraction of the area containing lakes and marshes (dimensionless); and mean annual precipitation (in mm). The catchment descriptors were taken from Birikundavyi *et al.* (1997). Figure 1 also highlights the location of a subset of 15 catchments used to

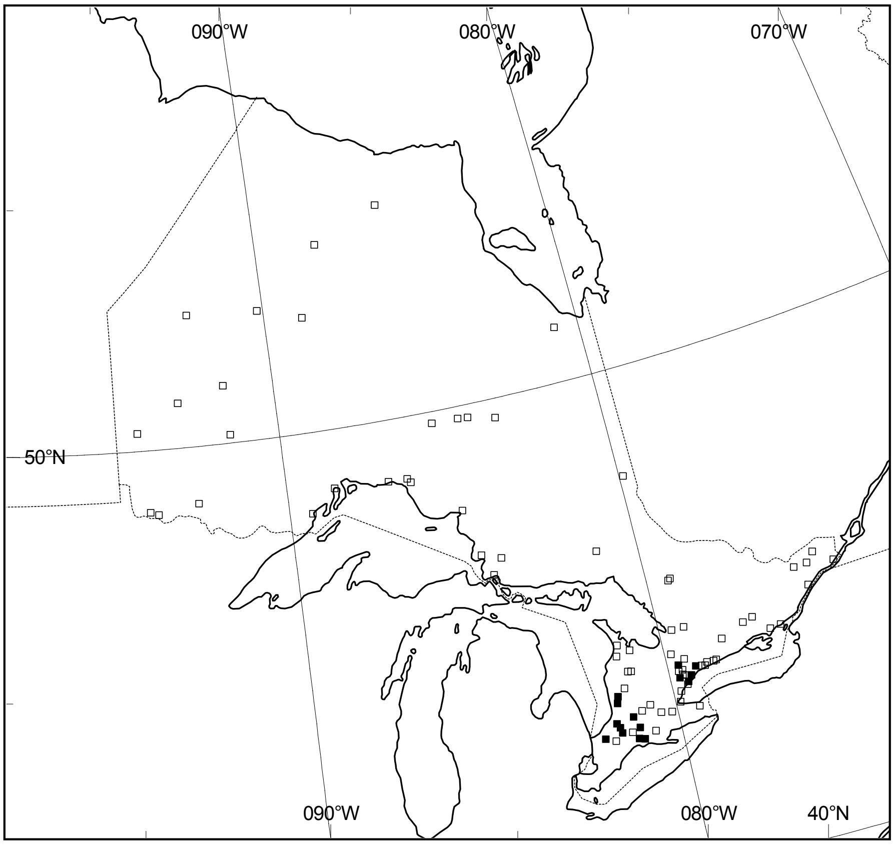

Fig. 1 Location of the 85 catchments in the province of Ontario. Solid symbols denote the location of the 15 sites used to evaluate the methods.

evaluate the methods. Each of these 15 sites was sequentially considered to be ungauged and extreme flow quantiles were estimated for the sites using information from the remaining sites.

Results

The first step in the analysis was to estimate a pooling group for each site based on seasonality measures using the distance measure defined in equation (3). The 85 catchments are shown in Fig. 2 in seasonality-space. The angle between a line from the origin to the catchment and the horizontal axis labelled "Jan" indicates the mean date of flood occurrence and the distance of the catchment from the origin indicates the value of $\bar{r}$ (catchments close to the origin have a low value for $\bar{r}$ and those closer to the unit circle have a higher value). As with Fig. 1, the locations of the 15 evaluation catchments are highlighted. The seasonality-based pooling groups were constructed to contain 500 station-years of record and the homogeneity of the pooling groups was checked using the Hosking & Wallis homogeneity test (Hosking & Wallis, 1997). If necessary, revisions were made to the pooling groups to improve the homogeneity. The typical size for a pooling group created using seasonality measures was approximately 13 sites.

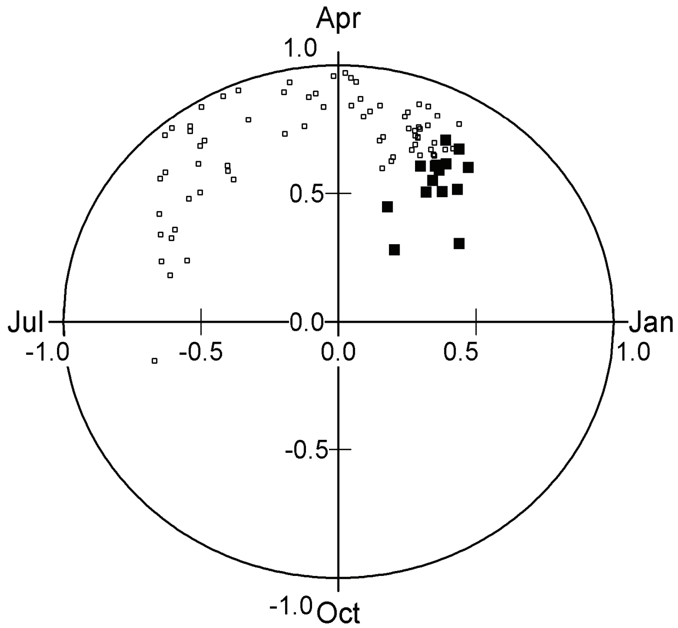

Fig. 2 Seasonality–space representation of the 85 catchments. Solid symbols denote the 15 sites used in the evaluation of the methods.

The performance of the ungauged methods was evaluated by comparing estimated growth curves calculated using each of the methods with actual growth curves, where the actual (or "true") value was taken as the at-site value for the 10- and 20-year events and was taken as the pooled estimate using seasonality measures for the 50-, 100- and 200-year events. A different basis of comparison is used for the shorter *versus* the longer return period events since reliable at-site estimates should be obtained from the available record for 10- and 20-year events (median record length of 38 years), while pooled information is required for reliable estimates for the longer return periods. The mean squared error (MSE) was calculated for each of the methods for estimating quantiles at ungauged catchments. The results are summarized in Table 1 with MSE values calculated separately for the 10- and 20-year return periods and the 50-, 100-, and 200-year return periods. The results reveal that the G–NN option gives the lowest MSE for both sets of return periods. The ANN quantile regression approach results in the largest MSE for the estimation of the short return period events, but is competitive with the geographic nearest neighbour approach for the longer return period events.

Table 1 also presents the average standardized 95% confidence interval widths for the quantile estimates. The width of the 95% confidence interval is standardized by

Table 1 Summary of MSE values and confidence interval widths.

Option	10 and 20 year results	50, 100, and 200 year results	Average standardized confidence interval width
CC	0.031	0.114	0.84
CC–NN	0.040	0.129	0.92
G–NN	0.025	0.033	1.01
ANN	0.044	0.046	0.14

dividing by the quantile estimate and then averaged over all return periods examined and averaged over all 15 evaluation sites. The results reveal markedly different average width values for the ANN *versus* the index flood-based approaches, reflecting the differing approaches taken to estimate the confidence intervals. It is clear that the ANN approach results in very narrow confidence intervals; the CC approach provides the narrowest average confidence limit of the index flood approaches while the widest confidence intervals are obtained with the G–NN approach.

CONCLUSIONS

A geographic nearest neighbour index flood-based approach was found to provide the lowest mean square error values for estimating the growth curve at ungauged catchments. The three index flood-based approaches provided similar results in terms of the prediction of the 10- and 20-year growth curve values, but the geographic nearest neighbour was superior for the longer return period events. The ANN quantile regression approach was not as efficient for the estimation of the shorter return period events, but was competitive with the geographic nearest neighbour approach for the longer return period events. The index flood-based approaches all resulted in considerably wider confidence limits than were obtained with the ANN ensemble approach. Further work is required to properly interpret the confidence limits from the ANN ensemble approach in comparison to those from the index flood-based approaches, as the methods used for confidence limit estimation are very different. Future work will evaluate other quantile regression approaches and will explore K-Nearest Neighbour approaches.

Acknowledgements This research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). This support is gratefully acknowledged.

REFERENCES

Acreman, M. C. & Wiltshire, S. E. (1989) The regions are dead; long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. In: *FRIENDS in Hydrology* (ed. by L. Roald, K. Nordseth & K. A. Hassel), 175–188. IAHS Publ. 187. IAHS Press, Wallingford, UK.

Bayliss, A. C. & Jones, R. C. (1993) *Peaks-over-threshold Flood Database: Summary Statistics and Seasonality.* Report no. 121, Institute of Hydrology, Wallingford, UK.

Birikundavyi, S., Rouselle, J. & Nguyen, V.-T.-V. (1997) Estimation régionale de quantiles de crues par l'analyse des correspondances. *Can. J. Civil Engng* **24**(3), 438–447.

Breiman, L. (1996) Bagging predictors. *Machine Learning* **26**(2), 123–140.

Burn, D. H. (1990) Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.* **26**(10), 2257–2265.

Burn, D. H. (1997) Catchment similarity for regional flood frequency analysis using seasonality measures. *J. Hydrol.* **202**, 212–230.

Burn, D. H. (2003). The use of resampling for estimating confidence intervals for single site and pooled frequency analysis. *Hydrol. Sci. J.* **48**(1), 25–38.

Castellarin, A., Burn, D. H. & Brath, A. (2001) Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *J. Hydrol.* **241**, 270–285.

Cunderlik, J. M. & Burn, D. H. (2002) The use of flood regime information in regional flood frequency analysis. *Hydrol. Sci. J.* **47**(1), 77–92.

Cunderlik, J. M., Ouarda, T. B. M. J. & Bobée, B. (2004) On the objective identification of flood seasons. *Water Resour. Res.* **40**(1), W01520, doi:10.1029/2003WR002295.

Dalrymple, T. (1960) Flood-frequency analyses. *USGS Water-Supply paper 1543-A*.

Efron, B. & Tibshirani, T. J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York, USA.

Fisher, N. I. (1993) *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, UK.

Haykin, S. (1994) *Neural Networks – A Comprehensive Foundation*. Macmillan College Publishing, New York, USA.

Hosking, J. R. M. & Wallis, J. R. (1997) *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, Cambridge, UK.

Jakob, D., Reed, D. W. & Robson, A. J. (1999) Choosing a pooling-group. In: *Flood Estimation Handbook*, vol. 3. Institute of Hydrology, Wallingford, UK.

Magilligan, F. J. & Graber, B. E. (1996) Hydroclimatological and geomorphic controls on the timing and spatial variability of floods in New England, USA. *J. Hydrol.* **178**, 159–180.

Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P. & Bobée, B. (2006) Data-based comparison of seasonality-based regional flood frequency methods. *J. Hydrol.* **330**, 329–339.

Shu, C. & Burn D. H. (2004) Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour. Res.* **40** (9), W09301, doi:10.1029/2003WR002816.