Regional low-flow frequency analysis with a recession parameter from a non-linear reservoir model

Christian Charron[1,]* and Taha B.M.J. Ouarda[1, 2]

[1]Institute Center for Water and Environment (iWATER), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

[2]INRS-ETE, National Institute of Scientific Research, Quebec City (QC), G1K9A9, Canada

*Corresponding author: Email: ccharron@masdar.ac.ae
Tel: +971 2 810 9524

Submitted to Journal of Hydrology
February 2015

Abstract: Several studies have shown that improvements in the regional prediction of low-flow characteristics can be obtained through the inclusion of a parameter characterising catchment baseflow recession. Usually, a linear reservoir model is assumed to define recession characteristics used as predictors in regional models. We propose in this study to adopt instead a non-linear model. Predictors derived from the linear model and the non-linear model are used separately in low-flow regional models along with other predictors representing physiographical and meteorological characteristics. These models are applied to selected gaged catchments. Results show that better performances are obtained with the parameter from the non-linear model. One drawback of using recession parameters for regional estimation is that a streamflow record is required at the site of interest. However, recession parameters can be estimated with short streamflow records. In this study, to simulate the performances obtained at partially gaged catchments, the recession parameters are estimated with very short streamflow records at target sites. Results indicate that, with a streamflow record as short as one year, a model with a recession parameter from the non-linear model leads to better performances than a model with only physiographical and meteorological characteristics.

Keywords: Regional estimation; Low-flows; Recession analysis; Canonical correlation analysis; Non-linear model; Reservoir model.

1. Introduction

It is of major importance to engineers and water managers to properly estimate the frequency of low-flow events, and their spatial and temporal evolution in the region of study (Vogel and Kroll, 1992; Durrans et al., 1999; Smakhtin, 2001; Kroll et al., 2004; Khaliq et al., 2008, 2009; Ouarda et al., 2008b; Fiala et al., 2010). Applications where this information is needed include water supply, hydropower production, dilution of pollution discharge and aquatic wildlife protection. The most commonly used low-flow statistic is the quantile $Q_{d,T}$ defined as the annual minimum average streamflow during d days with a return period of T years. When a sufficient historical streamflow record is available at a given site, low-flow statistics are obtained with a frequency analysis using the observed streamflow data. However, when insufficient or no streamflow record is available, a regional approach needs to be employed (Hamza et al., 2001; Ouarda et al., 2001; Kroll et al., 2004).

For the purpose of regional estimation, regression models are often used to estimate low-flow quantiles with explanatory variables characterising physiographical catchment properties and meteorological conditions. It has been demonstrated that explanatory variables representing geological and hydrogeological characteristics have a strong influence on low-flow regimes (Bingham, 1986; Tallaksen, 1989; Vogel and Kroll, 1992; Smakhtin, 2001; Kroll et al., 2004). However, such variables are hard to establish and difficult to quantify (Demuth and Hagemann, 1994). To address this issue, many authors have used baseflow recession parameters as surrogate to these variables (Bingham, 1986; Tallaksen, 1989; Arihood and Glatfelter, 1991; Vogel and Kroll, 1992, 1996; Demuth and Hagemann, 1994; Kroll et al., 2004; Eng and Milly, 2007). This is justified by the fact that low flows result principally of groundwater discharge into the stream during dry periods.

In studies where recession parameters were used to estimate low-flow statistics, a linear reservoir model is always assumed. This approximation is used for convenience, but a non-linear relation is more accurate in general (Brutsaert and Nieber, 1977; Whittenberg, 1994; Chapman, 2003). In this study, we propose to use a recession parameter assuming the non-linear reservoir model in a regional model along with other physiographical and meteorological characteristics for the estimation of low-flow quantiles. Performances are compared with those with a regional model including instead a parameter that assumes the common linear reservoir model. The regional models are applied to a group of catchments in the province of Quebec (Canada).

A major inconvenience of using recession parameters as predictors is that hydrological data is needed at the site of interest. However, they can be estimated with a short streamflow time series when it is available at the site of interest. A second objective of this paper is to evaluate the potential of this method at such sites (referred to as partially gaged sites). For that, a partially gaged case is simulated at the target site by estimating the recession parameters with only few years of streamflow record data selected randomly from the complete streamflow record.

Regional frequency analysis involves usually two steps: the identification of groups of hydrologically homogeneous catchments and the regional estimation within each individual region. In this study, canonical correlation analysis (CCA) is used to delineate the homogenous regions. This method has been used with success for the regionalization of flood quantiles (Ouarda et al., 2001), low-flow quantiles (Tsakiris et al., 2011) and water quality characteristics (Khalil et al., 2011).

2. Recession curve modeling

Boussinesq (1877) conceptualised the problem of outflow into a penetrating stream channel from an unconfined rectangular aquifer on a horizontal impermeable layer. Brutsaert and Nieber (1977) demonstrated that several solutions to the Boussinesq problem assume the following relation:

$$\frac{dQ}{dt} = -aQ^b \quad (1)$$

where *Q* is the streamflow, *t* is the time, and *a* [$m^{3(1-b)}s^{b-2}$] and *b* [-] are constants.

The linear reservoir model in which $b=1$ is often considered. The solution of Eq. (1) for the outflow is then the simple exponential equation:

$$Q_t = Q_0 e^{-at} \quad (2)$$

where Q_t is the outflow at time *t* and Q_0 the initial outflow. Eq. (2) in its exponential form is often used to describe recession curves because of its simplicity. In that case, *a* characterises the rate of recession. Many authors have used parameters derived from the linear model as predictors in regression models. Vogel and Kroll (1992) and Kroll et al. (2004) used the recession constant $K_b = \exp(-a)$. Eng and Milly (2007) rather used the parameter $\tau = a^{-1}$ to which they referred as the long-term aquifer constant.

Although the linear reservoir model has been largely employed, the power-law model is more appropriate. In several studies where non-linear equations have been fitted to recession discharge data, it was found that the power-law model is more realistic (Moore, 1997; Chapman, 1999, 2003; Wittenberg, 1999). Brutsaert and Nieber (1977) found that *b* takes approximately the value of 1.5 over most of the ranges of low-flow rates. The value of the exponent b ranged from 1.38 to 1.69 for 10 out of 11 catchments in Chapman (2003). In Wittenberg (1994) a mean value of 1.6 was obtained.

We propose in this study to use a parameter derived from the non-linear reservoir model instead of the usually used linear model. Because b is different from 1 in general, it is expected that the performances will be increased by the use of this optimised parameter in regional models. Wittenberg (1999) stated that a value of 1.5 is a typical value for average cases and suggested to calibrate the factor a with b fixed to this value. In this study, a is estimated with a fixed value of b for the whole study area estimated by the average of individual catchment values. To estimate b for a given catchment, a similar approach to Vogel and Kroll (1992) and Brutsaert and Lopez (1998) is used. The relation in Eq. (1) is approximated by:

$$\frac{Q_t - Q_{t-1}}{\Delta t} = -a\left(\frac{Q_t + Q_{t-1}}{2}\right)^b \qquad (3)$$

where Q_t and Q_{t-1} are the streamflow measurements at successive times Δt apart. By taking the logarithm, the following linearised equation is obtained:

$$\ln(Q_{t-1} - Q_t) = \ln(a) + b\ln\left(\frac{Q_t + Q_{t-1}}{2}\right). \qquad (4)$$

The parameter b along with the parameter a in Eq. (4) are estimated using a least square linear regression. Subsequently, given a fixed value of b, the least square estimator of a in Eq. (4) is given by:

$$a = \exp\left\{\frac{1}{d}\sum_{t=1}^{d}\left[\ln(Q_{t-1} - Q_t) - b\ln\left(\frac{Q_t + Q_{t-1}}{2}\right)\right]\right\} \qquad (5)$$

where d is the number of pairs of consecutive streamflow values.

3. Study area

The regional proposed estimation models are applied to a network of 190 gaging stations in the province of Quebec (Canada). Due to the seasonal variations specific to the study area, we consider two distinct low-flow seasons corresponding to the summer and the winter. In this study, we analyse the low-flow quantiles $Q_{7,\,2}$ and $Q_{7,\,10}$ corresponding to return periods of T = 2 and 10 years for a duration d = 7 days, and the low-flow quantile $Q_{30,\,5}$ corresponding to a return period of T = 5 years for a duration d = 30 days for the summer and winter seasons separately. The hydrological, physiographical and meteorological variables used in the present case study came from a low-flow frequency analysis study by Ouarda et al. (2005). The same database has also been used in Ouarda and Shu (2009) for low-flow frequency analysis using artificial neural networks. Only stations with at least 10 years of record data and corresponding to pristine basins were selected. The selected stations passed the Kendall test for stationarity, the Wald-Wolfowitz test of independence, the Wilcoxon test of homogeneity for the mean and the Levene test for homogeneity of the variance. As a result, 127 and 133 stations were selected for the summer season and the winter season respectively.

The locations of the gaging stations are presented in Fig. 1. The stations cover a large area in the southern part of the province of Quebec (Canada) and are located between 45°N and 55°N. The area of the catchments ranges from 0.7 km^2 to 96,600 km^2 with a median value of 3077 km^2. The largest catchments are located in the northern part of the study region. The average flow record size is 32 years of data. Winter mean temperatures for the study area range from -10°C in the south to -21°C in the north and summer mean temperatures range from 20°C in the south to 12°C in the north.

A set of physiographical and meteorological variables are available for each basin. These variables are the basin area (AREA), the latitude of the gaging station (LAT), the mean slope of the drainage area (MSLP), the percentage of the basin area occupied by lakes (PLAKE), the percentage of the basin area covered by forest (PFOR), the mean annual degree days below 0ºC (DDBZ), the mean annual degree days below 0ºC (DDH13), the average annual precipitation (PTMA), the average summer-autumn liquid precipitation (PLMS), the average number of days for which the mean temperature exceeds 27 ºC (NDH27) and the mean curve number (MCN) which is a soil characteristic. These variables are summarised in Table 1.

Low-flow quantiles corresponding to various return periods *T* and durations *d* were estimated. Low-flow data at each station was fitted with an appropriate statistical distribution. The distributions considered include the Generalized Extreme Value, Gumbel, Weibull, two-parameter Lognormal, three-parameter Lognormal, Gamma, Pearson type III, Log-Pearson III and Generalized Pareto distributions. To select the distribution that best fits the hydrological data for each station, the Bayesian information criterion was used. Fig. 2 presents an example of a frequency curve with the observed data on a normal probability plot. The distribution that fits best the observations, the two-parameter Lognormal (LN2), is represented along with the bounds of the 95% confidence interval.

4. Study methodology

4.1. Recession analysis method

The computation of recession characteristics at gaged sites is usually done through a recession analysis. This involves the delineation of baseflow recession segments from hydrographs and subsequently the computation of recession characteristics. In practice, the interpretation of hydrographs is complicated by the fact that, during a recession period, recharge events can often interrupt a recession and produce many recession segments of different lengths. Another interpretative complication comes from the fact that the different streamflow components, that are surface flow, interflow and baseflow, are difficult to quantify at a given time. Given these considerations, various researchers have developed methods to delineate baseflow recession segments from hydrographs.

Traditionally, graphical techniques are used for recession analysis. They are however subjective and applicable only for a few analyses because they are time consuming. For a large database, automated methods are preferred. Several methods have been proposed in the literature. They usually take only decreasing portions of hydrographs in which starting and duration criteria are defined. The minimal length of individual recession segments can usually vary between 4 days and 10 days (Tallaksen, 1995). A portion at the beginning of recession segments can also be removed to avoid the presence of surface flow.

The recession analysis method applied here is based on the procedure proposed by Vogel and Kroll (1992) in which segments of only decreasing 3-day moving average are selected. Only segments with a minimum of 10 days are considered. Furthermore, to minimise surface runoff components, 30 % of the beginning of each segment is subtracted.

The recession parameter a_{nl} is defined for the non-linear reservoir model. It is computed at each catchment with b fixed to b_{opt}, the optimal value for the whole study area. b_{opt} is estimated by averaging the estimated values of b at each basin. The recession variable a_l for the linear reservoir model is also estimated. In that case, b is set to unity (b = 1) for the whole area.

4.2. Delineation of homogenous regions with CCA

Regionalization methods usually involve two steps: defining groups of homogeneous stations and applying an information transfer method over the delineated regions. As in the case of flood regionalization, grouping stations provides generally better estimates because stations in the same group are expected to have similar hydrological responses. Certain delineation methods allow defining geographically contiguous regions. This kind of approach can involve the delineation on the basis of geographic considerations or on the basis of the similitude in residuals obtained by a regression model (Smakhtin, 2001). In reality, two basins can be hydrologically similar without being geographically close. Other methods allow defining groups of catchments that are not necessarily contiguous. Delineation is then made on the basis of the physiographic and climatic characteristics of the catchments. Multivariate statistical analysis methods such as cluster analysis and principal component analysis are then often used (Nathan and McMahon, 1990; Smakhtin, 2001).

Another promising multivariate approach is canonical correlation analysis (CCA). It has been applied in the field of flood regionalization by Ouarda et al. (2000, 2001) and it has been proven to be applicable for low flow regionalization in Tsakiris et al. (2011). This method defines for each target station, a specific set of homogenous stations (neighbourhood). This has the advantage of maximising the similarity between the neighbourhood catchments and the target site. The neighbourhood approach was found in Ouarda et al. (2008a) to be superior to approaches delineating fixed sets of stations for regional flood frequency analysis.

Optimal neighbourhoods need to be delineated for models where a CCA is involved. The jackknife resampling procedure presented in Ouarda et al. (2001) is used. However, when a neighbourhood is defined by the optimisation parameter, it may happen that the number of stations in the neighbourhood is not large enough to be able to carry out the multiple regression. The jackknife resampling procedure is modified in this study to include instead the

s stations with the lowest Mahalanobis distance. This ensures that estimations are obtained at all stations of the study area.

4.3. Regional models

Overall, six regional models are defined depending on which explanatory variables are used and whether neighbourhoods are delineated or not. For three models (ALL, $\mathrm{ALL\_}a_l$ and $\mathrm{ALL\_}a_{nl}$), the information transfer method is used with all the stations of the database without delineation of neighbourhoods. The model ALL includes solely the physiographical and meteorological variables. The $\mathrm{ALL\_}a_l$ and $\mathrm{ALL\_}a_{nl}$ models include in addition the variables a_l and a_{nl} respectively. For three other models (CCA, $\mathrm{CCA\_}a_l$ and $\mathrm{CCA\_}a_{nl}$), neighbourhoods are delineated with CCA prior to the information transfer. The model CCA includes solely the physiographical and meteorological variables and $\mathrm{CCA\_}a_l$ and $\mathrm{CCA\_}a_{nl}$ include in addition the variables a_l and a_{nl} respectively. Models CCA and ALL, for which no hydrological data at the target site is used, represent the ungaged case. Models $\mathrm{CCA\_}a_l$, $\mathrm{CCA\_}a_{nl}$, $\mathrm{ALL\_}a_l$ and $\mathrm{ALL\_}a_{nl}$, for which the hydrological information available at the target site is used, represent the gaged case.

The selection of the variables used for the CCA is based on the previous study of Ouarda et al. (2005) where the CCA method was applied on the same study case for low-flow frequency analysis. The hydrological variables included in the set of response variables are the low-flow quantiles $Q_{7,2}$, $Q_{7,10}$ and $Q_{30,5}$. The physiographical and meteorological variables included in the set of explanatory variables are AREA, PLAKE, NDH27 and MCN for the summer low-flow quantiles and AREA, PLAKE, PFOR and DDBZ for the winter quantiles. To ensure the normality, the low-flow quantiles, AREA and DDBZ are logarithmically transformed and PLAKE is transformed by a square root transformation.

For regional information transfer, the multiple regression model is used. Parameters are estimated with the least square error method. For each multiple regression model, the explanatory variables are selected with a stepwise regression analysis procedure applied to $Q_{7,2}$. The regression models obtained during the summer season for ALL or CCA, ALL_a_l or CCA_a_l, and ALL_a_{nl} or CCA_a_{nl} are respectively given by:

$$\begin{aligned}\log(Q_{d,T}) &= \theta_0 + \theta_1 \log(\mathrm{AREA}) + \theta_2 \log(\mathrm{DDBZ}) + \theta_3 \log(\mathrm{MCN}) \\ &\quad + \theta_4 \log(\mathrm{PTMA}) + \theta_5 \log(\mathrm{NDH\,27}) + \varepsilon\end{aligned}, \tag{6}$$

$$\begin{aligned}\log(Q_{d,T}) &= \theta_0 + \theta_1 \log(\mathrm{AREA}) + \theta_2 \log(a_l) + \theta_3 \log(\mathrm{PLMS}) \\ &\quad + \theta_4 \log(\mathrm{DDH13}) + \theta_5 \log(\mathrm{DDBZ}) + \theta_6 \log(\mathrm{MCN}) + \varepsilon\end{aligned}, \tag{7}$$

$$\begin{aligned}\log(Q_{d,T}) &= \theta_0 + \theta_1 \log(a_{nl}) + \theta_2 \log(\mathrm{AREA}) + \theta_3 \log(\mathrm{PLMS}) \\ &\quad + \theta_4 \log(\mathrm{DDBZ}) + \theta_5 \log(\mathrm{NDH\,27}) + \theta_6 \log(\mathrm{MCN}) + \varepsilon\end{aligned}, \tag{8}$$

where θ_i are the model parameters and ε are the error terms. Similarly, the regression models obtained during the winter season for ALL or CCA, ALL_a_l or CCA_a_l, and ALL_a_{nl} or CCA_a_{nl} are respectively given by:

$$\log(Q_{d,T}) = \theta_0 + \theta_1 \log(\mathrm{AREA}) + \theta_2 \log(\mathrm{DDBZ}) + \theta_3 \log(\mathrm{LAT}) + \theta_4 \log(\mathrm{MCN}) + \varepsilon, \tag{9}$$

$$\log(Q_{d,T}) = \theta_0 + \theta_1 \log(\mathrm{AREA}) + \theta_2 \log(a_l) + \theta_3 \log(\mathrm{LAT}) + \theta_4 \log(\mathrm{DDH13}) + \varepsilon, \tag{10}$$

$$\log(Q_{d,T}) = \theta_0 + \theta_1 \log(\mathrm{AREA}) + \theta_2 \log(a_{nl}) + \theta_3 \log(\mathrm{LAT}) + \theta_4 \log(\mathrm{NDH27}) + \varepsilon. \tag{11}$$

Explanatory variables in Eqs. 6-11 are ordered from the most significant to the least significant one. It can be observed that recession parameters represent very important variables. The recession parameter is generally the most important variable after the basin area and a_{nl} is the most important variable for the summer season.

4.4. Performance criteria

To assess the performances of the regional models, a jacknife resampling procedure is performed. Each gaged site is successively considered ungaged and is removed from the database. A regional model is then applied to obtain an estimate of the quantiles at this target site with the remaining gaged sites. This operation is repeated for all sites of the database. Five indices are used to evaluate the performances (Ouarda and Shu, 2009): the Nash criterion (NASH), the root mean squared error (RMSE), the relative root mean squared error (rRMSE), the mean bias (BIAS), and the relative mean bias (rBIAS).

Performance criteria for models with a recession variable were adapted to represent the performances that can be achieved for the partially gaged case when recession variables are estimated with short streamflow series. The method applied here is similar to the one presented in Eng and Milly (2007). For that, the same jackknife method presented before is applied, but the recession parameter at the target site is estimated with data coming from N years selected randomly through the streamflow record at the target site. This operation is repeated 100 times for each target site and N is varied from 1 to 4 years. The performance indices for the partially gaged cases are defined by the following equations:

$$\mathrm{rRMSE_N} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{100}\sum_{j=1}^{100}\left[(\hat{q}_{ij} - q_i)/q_i\right]^2\right)}, \tag{12}$$

$$\mathrm{rBIAS_N} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{100}\sum_{j=1}^{100}\left[(\hat{q}_{ij} - q_i)/q_i\right]\right) \tag{13}$$

where $\hat{q}_{i,j}$ is the estimate of q_i obtained at site *i* using the sample *j*.

5. Results

The recession analysis method presented in section 4.1 was applied to the gaged catchments of the study area. Fig. 3 presents an example of a hydrograph at the station 020802 for the year 1970. Selected recession segments are identified with the grey areas under the streamflow curve. It can be observed that several recessions occurred during the year. The parameter b was estimated for each catchment. To illustrate the method, $-dQ/dt$ is plotted against Q for the station 030103 on a log-log paper in Fig. 4. The slope of the line estimated with the least-squares method gives an estimate of b for that catchment. The optimal parameter b for the whole study area was obtained by averaging the values obtained at all basins. The value of 1.66 for b_{opt} was obtained for this study area. This result is in agreement with several studies where this parameter was estimated (See section 2.1). Recession variables a_l and a_{nl} were computed for every station with b respectively fixed to 1 and 1.66.

Tables 2 and 3 present the performance indices obtained for the different regional models for summer and winter low-flow quantile estimation respectively. Results show that adding a recession variable to a regression model always improves significantly the performances. In general, better performances are obtained with regression models including a_{nl} instead of a_l. For instance, lower rRMSE and rBIAS are always obtained with a_{nl} instead of a_l. On the other hand, for the summer season, better RMSE and NASH are obtained for $\mathrm{ALL\_}a_l$ compared to $\mathrm{ALL\_}a_{nl}$. Results show also that, in general, the delineation of neighbourhoods with CCA improves the performances for summer quantiles. This is not the case for winter quantiles where performances are generally very similar. This seems to indicate that the overall level of homogeneity in the study region is higher for winter low-flows. Overall best performances are obtained with the model CCA_a_{nl} for the summer season and with ALL_a_{nl} and CCA_a_{nl} for the winter season.

Results in Tables 2 and 3 are obtained under the assumption that the recession variable is available at the target site. In real world cases, the target site is often either ungaged or partially gaged. These results represent a sort of upper bound in terms of regional model performance. Tables 4 and 5 present the performances that are obtained for the simulated partially gaged case. They present the indices $\mathrm{rRMSE}_\mathrm{N}$ and $\mathrm{rBIAS}_\mathrm{N}$ for the regional models when the recession variables at target sites are estimated using a given number N of years. Results indicate that the $\mathrm{rRMSE}_\mathrm{N}$ decreases as N increases and converges to the value obtained with the completely gaged cases for both seasons (see Tables 2 and 3). $\mathrm{rBIAS}_\mathrm{N}$ generally decreases as N increases but occasionally increases instead. This occurs more often for winter quantiles with CCA_a_{nl} and ALL_a_{nl}, although the biases are small in this cases. To assess the improvements obtained by the use of a recession parameter at partially gaged stations, performances of models that include a recession variable (CCA_a_{l}, CCA_a_{nl}, ALL_a_{l} and ALL_a_{nl}) are compared to the performances obtained by their corresponding ungaged models (CCA or ALL). For instance, CCA_a_{nl} is compared with CCA. When a_{nl} is included in the model, $\mathrm{rRMSE}_\mathrm{N}$ and $\mathrm{rBIAS}_\mathrm{N}$ are always better than the ungaged case even when only one year of stramflow data (N=1) is used. On the other side, when a_l is included in the model, more years are required to lead to better performances than the ungaged case. Thus, in general, two years are required to obtain a better $\mathrm{rRMSE}_\mathrm{N}$ for both seasons. To obtain improved $\mathrm{rBIAS_N}$, generally 1 or 2 years are required for both seasons. However, for CCA_a_l and the summer season, 4 years are required for $Q_{30,5}$ and more than 4 years are required for $Q_{7,10}$. These results indicate clearly that it is beneficial to use recession parameters at partially gaged sites in a regional low-flow frequency analysis model. Indeed, quantile estimates are improved even when recession parameters are estimated based on a very limited number of years of streamflow data. These results show also the importance of using the non-linear model instead of the linear one as the performances are improved significantly even with only one year of streamflow data compared to the ungaged case.

6. Conclusions and future work

In this paper, regional low-flow frequency analysis models that include recession parameters as predictors are developed. Two different parameters are considered: the recession coefficient a_l assuming the linear reservoir model and the recession coefficient a_{nl} assuming a non-linear reservoir model where the exponent *b* is fixed to the estimated value of 1.66 for the study area.

The investigation of the appropriate predictors for low-flow statistics is carried out with stepwise regression analysis and leads to the conclusion that the variables from recession parameters are important explanatory variables. The study results clearly indicate that the inclusion of a recession variable in a regional low-flow frequency analysis model improves the performance of the regional estimator. Furthermore, the performances are significantly better with models that include a recession variable from the non-linear reservoir model.

An inconvenience of using recession characteristics is that they can only be obtained for gaged catchments. However, it is possible to estimate these parameters with a limited number of hydrographs. This paper aims also to evaluate the performances obtained with recession parameters estimated from very short streamflow records. Results of the application of regional low-flow frequency models with hydrograph lengths ranging from 1 to 4 years show that performances converge rapidly to those obtained when the parameters are estimated from the complete data record. When the parameter from the non-linear model is included in a regression model, the performances are better than those obtained without recession variable even when only one year of streamflow data is used to estimate the recession parameter. This shows the possibility of combining local hydrological information with regional information at a partially gaged site in a regional model.

These results indicate that it is of interest to dedicate efforts to the development of improved methods for the estimation of recession parameters. Improvements can result principally from the selection of a proper reservoir model and from the recession analysis method. Better reservoir models, in agreement with the real reservoir storage-outflow relationship should be developed. Improved reservoir models, such as the ones that consider various loss and gain sources that affect the streamflow could be used.

Other improvements can result from the development of enhanced recession analysis methods. For instance, recession segments should be representative of baseflow recession discharges, i.e. should represent portions of flow that are free of surface flow and interflow. Stoelzle et al. (2013) compared three different methods for the extraction of recessions and three methods for model fitting. They concluded that the roles of recession extraction procedures and fitting methods for the parameterization of storage–outflow models are complex. They also indicated that the interaction of the recession analysis components has various effects on the derived recession characteristics. These conclusions imply that the results obtained here are strongly associated to the specific recession analysis method used. Future research efforts should focus on the identification of the recession analysis methods that are the best adapted to low flow regionalization.

Other methods for the delineation of homogenous regions should also be considered. For instance, methods based on seasonality characteristics should be very promising (See Cunderlik et al., 2004a, 2004b; Ouarda et al., 2006). Future efforts should also focus on improved modeling of the homogeneity of delineated regions and on the adoption of the multivariate framework (Chebana and Ouarda, 2007, 2008, 2009) for the regional analysis of low-flow characteristics.

References

Arihood, L.D., Glatfelter, D.R., 1991. Method for estimating low-flow characteristics of ungaged streams in Indiana. USGS Water Supply Paper 2372, 22 pp.

Bingham, R.H., 1986. Regionalization of low-flow characteristics of Tennessee streams. Water-Resources Investigations Report 85-4191, 63 pp.

Boussinesq, J., 1877. Essai sur la théorie des eaux courantes, Du mouvement non permanent des eaux souterraines, Acad. Sci. Inst. Fr. 23, 252-260.

Brutsaert, W., Lopez, J.P., 1998. Basin-scale geohydrologic drought flow features of riparian aquifers in the Southern Great Plains. Water Resources Research 34(2), 233-240.

Brutsaert, W., Nieber, J.L., 1977. Regionalized drought flow hydrographs from a mature glaciated plateau. Water Resources Research 13(3), 637-643.

Chapman, T., 1999. A comparison of algorithms for stream flow recession and baseflow separation. Hydrological Processes 13(5), 701-714.

Chapman, T.G., 2003. Modelling stream recession flows. Environmental Modelling & Software 18(8–9), 683-692.

Chebana, F., Ouarda, T.B.J.M., 2007. Multivariate L-moment homogeneity test. Water Resources Research 43(8), W08406.

Chebana, F., Ouarda, T.B.M.J., 2008. Depth and homogeneity in regional flood frequency analysis. Water Resources Research 44, W11422.

Chebana, F., Ouarda, T.B.J.M., 2009. Index flood–based multivariate regional frequency analysis. Water Resources Research 45, W10435.

Cunderlik, J.M., Ouarda, T., Bobée, B., 2004a. Determination of flood seasonality from hydrological records. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 49(3), 511-526.

Cunderlik, J.M., Ouarda, T.B.M.J., Bobée, B., 2004b. On the objective identification of flood seasons. Water Resources Research 40(1), W01520.

Demuth, S., Hagemann, I., 1994. Estimation of flow parameters applying hydrogeological area information. IAHS Publications 221, 151-157.

Durrans, S., Ouarda, T., Rasmussen, P., Bobée, B., 1999. Treatment of Zeroes in Tail Modeling of Low Flows. Journal of Hydrologic Engineering 4(1), 19-27.

Eng, K., Milly, P.C.D., 2007. Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges. Water Resources Research 43(1), W01201.

Fiala, T., Ouarda, T.B.M.J., Hladný, J., 2010. Evolution of low flows in the Czech Republic. Journal of Hydrology 393(3–4), 206-218.

Hamza, A., Ouarda, T.B.M.J., Durrans, S.R., Bobée, B., 2001. Développement de modèles de queues et d'invariance d'échelle pour l'estimation régionale des débits d'étiage. Canadian Journal of Civil Engineering 28(2), 291-304.

Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. Journal of Hydrology 405(3–4), 277-287.

Khaliq, M.N., Ouarda, T.B.M.J., Gachon, P., 2009. Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers: The effect of short- and long-term persistence. Journal of Hydrology 369(1-2), 183-197.

Khaliq, M.N., Ouarda, T.B.M.J., Gachon, P., Sushama, L., 2008. Temporal evolution of low-flow regimes in Canadian rivers. Water Resources Research 44(8), W08436.

Kroll, C.N., Luz, J.G., Allen, T.B., Vogel, R.M., 2004. Developing a watershed characteristics database to improve low streamflow prediction. Journal of Hydrologic Engineering, ASCE 9(2), 116-125.

Moore, R.D., 1997. Storage-outflow modelling of streamflow recessions, with application to a shallow-soil forested catchment. Journal of Hydrology 198(1–4), 260-270.

Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalisation. Journal of Hydrology 121(1–4), 217-238.

Ouarda, T.B.M.J., Jourdain, V., Gignac, N., Gingras, H., Herrera, E., Bobée, B., 2005. Développement d'un modèle hydrologique visant l'estimation des débits d'étiage pour le Québec habité. INRS-ETE, Rapport de recherche No. R-684-f1, 174 pp.

Ouarda, T.B.M.J., Ba, K.M., Diaz-Delgado, C., Carsteanu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., Bobée, B., 2008a. Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. Journal of Hydrology 348(1-2), 40-58.

Ouarda, T.B.M.J., Charron, C., St-Hilaire, A., 2008b. Statistical Models and the Estimation of Low Flows. Canadian Water Resources Journal 33(2), 195-206.

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. Journal of Hydrology 330(1-2), 329-339.

Ouarda, T.B.M.J., Girard, C., Cavadias, G.S., Bobée, B., 2001. Regional flood frequency estimation with canonical correlation analysis. Journal of Hydrology 254(1-4), 157-173.

Ouarda, T.B.M.J., Haché, M., Bruneau, P., Bobée, B., 2000. Regional flood peak and volume estimation in northern Canadian basin. Journal of cold regions engineering 14(4), 176-191.

Ouarda, T.B.M.J., Shu, C., 2009. Regional low-flow frequency analysis using single and ensemble artificial neural networks. Water Resources Research 45(11), W11428.

Smakhtin, V.U., 2001. Low flow hydrology: a review. Journal of Hydrology 240(3–4), 147-186.

Stoelzle, M., Stahl, K., Weiler, M., 2013. Are streamflow recession characteristics really characteristic? Hydrology and Earth System Sciences 17(2), 817-828.

Tallaksen, L.M., 1989. Analysis of time variability in recession. IAHS Publications 187, 85-96.

Tallaksen, L.M., 1995. A review of baseflow recession analysis. Journal of Hydrology 165(1–4), 349-370.

Tsakiris, G., Nalbantis, I., Cavadias, G., 2011. Regionalization of low flows based on Canonical Correlation Analysis. Advances in Water Resources 34(7), 865-872.

Vogel, R.M., Kroll, C.N., 1992. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. Water Resources Research 28(9), 2451-2458.

Vogel, R.M., Kroll, C.N., 1996. Estimation of baseflow recession constants. Water Resources Management 10(4), 303-320.

Wittenberg, H., 1994. Nonlinear analysis of flow recession curves. IAHS Publication 221, 61-67.

Wittenberg, H., 1999. Baseflow recession and recharge as nonlinear storage processes. Hydrological Processes 13(5), 715-726.

Table 1. Explanatory variables available for the study area.

Explanatory variable	Units	Notation
Drainage area	km^2	AREA
Mean slope of the drainage area	degree	MSLP
Percentage of forested area	%	PFOR
Percentage of lakes	%	PLAKE
Mean annual precipitation	mm	PTMA
Mean summer-fall precipitation	mm	PLMS
Mean annual degree-days < 0 ºC	degree day	DDBZ
Mean annual degree-days > 13 ºC	degree day	DDH13
Number of days where temperatures > 27 ºC	day	NDH27
Mean curve number	-	MCN
Latitude of the gaging station	degree	LAT
Recession parameter (linear case, b=1)	s^{-1}	a_l
Recession parameter (non-linear case, $b=b_{opt}$)	$m^{3(1-b)}s^{b-2}$	a_{nl}

Table 2. Performances of regional models for summer low-flow quantile estimation.

Statistic	Quantile	Regional model					
		CCA	CCA_a_l	CCA_a_{nl}	ALL	ALL _a_l	ALL_a_{nl}
NASH	$Q_{30, 5}$	0.940	0.960	**0.969**	0.905	0.945	0.930
	$Q_{7, 2}$	0.936	0.978	**0.982**	0.904	0.961	0.939
	$Q_{7, 10}$	0.917	0.961	**0.978**	0.878	0.951	0.939
RMSE	$Q_{30, 5}$	41.64	34.05	**29.90**	52.41	39.83	44.99
	$Q_{7, 2}$	51.71	30.46	**27.75**	63.37	40.35	50.33
	$Q_{7, 10}$	41.34	28.47	**21.27**	50.16	31.79	35.46
rRMSE(%)	$Q_{30, 5}$	43.84	35.12	**30.25**	48.31	43.20	34.06
	$Q_{7, 2}$	44.57	30.88	**24.35**	48.18	37.72	26.81
	$Q_{7, 10}$	51.81	45.28	**38.66**	57.71	50.83	40.13
BIAS	$Q_{30, 5}$	-3.96	**-0.40**	-2.46	-4.08	-0.93	1.94
	$Q_{7, 2}$	-4.17	**-0.59**	-1.77	-3.97	-0.64	3.02
	$Q_{7, 10}$	-3.54	**-0.85**	-1.97	-3.60	-1.04	1.86
rBIAS(%)	$Q_{30, 5}$	8.24	7.47	**3.90**	9.29	6.67	4.79
	$Q_{7, 2}$	8.27	5.76	**2.62**	9.52	5.45	3.27
	$Q_{7, 10}$	11.30	11.37	**6.23**	12.81	9.11	6.40

Bold values correspond to best performances.

Table 3. Performances of regional models for winter low-flow quantile estimation.

Statistic	Quantile	Regional model					
		CCA	CCA_a_l	CCA_a_{nl}	ALL	ALL _a_l	ALL_a_{nl}
NASH	$Q_{30, 5}$	0.877	0.940	0.957	0.872	0.938	**0.959**
	$Q_{7, 2}$	0.877	0.949	0.964	0.873	0.947	**0.967**
	$Q_{7, 10}$	0.874	0.957	0.966	0.868	0.957	**0.969**
RMSE	$Q_{30, 5}$	19.62	13.65	11.61	20.01	13.88	**11.29**
	$Q_{7, 2}$	21.86	14.14	11.82	22.23	14.40	**11.27**
	$Q_{7, 10}$	18.08	10.50	9.37	18.47	10.56	**8.92**
rRMSE(%)	$Q_{30, 5}$	35.97	29.84	**22.94**	37.11	30.72	23.50
	$Q_{7, 2}$	32.92	25.35	**18.54**	33.53	25.86	19.16
	$Q_{7, 10}$	42.17	36.41	30.13	43.40	37.18	**28.23**
BIAS	$Q_{30, 5}$	-1.87	-1.71	**-1.11**	-2.61	-2.56	-1.26
	$Q_{7, 2}$	-1.92	-1.79	-1.26	-2.77	-2.64	**-1.15**
	$Q_{7, 10}$	-1.62	-1.25	-0.80	-1.97	-1.89	**-0.44**
rBIAS(%)	$Q_{30, 5}$	6.19	4.56	3.30	5.83	4.00	**2.47**
	$Q_{7, 2}$	5.67	3.70	2.43	5.20	3.08	**1.74**
	$Q_{7, 10}$	7.59	5.67	4.24	7.50	5.29	**3.27**

Bold values correspond to best performances.

Table 4. Performances of regional models for partially gaged basins (Summer low-flow quantile estimation).

Quantile	N	CCA_a_l		CCA_a_{nl}		ALL_a_l		ALL_a_{nl}	
		rRMSE$_N$	rBIAS$_N$	rRMSE$_N$	rBIAS$_N$	rRMSE$_N$	rBIAS$_N$	rRMSE$_N$	rBIAS$_N$
$Q_{30,5}$	1	46.22	9.40	**37.45**	**6.37**	53.05	**8.76**	**41.39**	**8.03**
	2	**40.16**	8.59	34.41	6.38	**47.29**	7.48	38.28	7.45
	3	38.39	8.29	33.05	6.07	46.32	7.63	37.43	7.41
	4	37.89	**8.20**	32.68	5.87	45.96	7.48	36.87	7.49
$Q_{7,2}$	1	46.11	8.73	**34.28**	**6.10**	53.41	**8.83**	**38.47**	**7.73**
	2	**37.32**	**6.94**	29.46	5.38	**44.63**	6.92	32.81	6.67
	3	35.97	6.98	28.15	5.33	42.45	6.67	30.88	6.32
	4	34.54	6.89	26.88	4.84	41.79	6.45	30.52	6.44
$Q_{7,10}$	1	64.73	15.87	**49.48**	**10.55**	65.77	12.44	**50.93**	**11.14**
	2	53.41	13.37	44.65	9.51	57.76	**11.31**	46.37	9.99
	3	**51.21**	12.94	43.26	9.13	**55.87**	10.66	44.72	9.90
	4	50.29	12.99	42.12	8.87	54.99	10.43	43.69	9.60

Bold values correspond to performances surpassing the corresponding ungaged case model.

Table 5. Performances of regional models for partially gaged basins (Winter low-flow quantile estimation).

Quantile	N	CCA_a_l		CCA_a_{nl}		ALL_a_l		ALL_a_{nl}	
		$rRMSE_N$	$rBIAS_N$	$rRMSE_N$	$rBIAS_N$	$rRMSE_N$	$rBIAS_N$	$rRMSE_N$	$rBIAS_N$
$Q_{30,5}$	1	36.65	**5.58**	**27.35**	**0.21**	38.76	**5.43**	**28.94**	**-1.05**
	2	**32.44**	4.91	25.10	-0.10	**34.76**	4.85	25.46	-1.88
	3	31.58	4.76	24.26	-0.28	33.14	4.30	24.95	-2.08
	4	31.38	4.85	23.76	-0.32	32.54	4.31	24.66	-2.09
$Q_{7,2}$	1	33.18	**4.93**	**24.34**	**-0.71**	35.17	**4.67**	**24.62**	**-2.41**
	2	**28.87**	4.30	21.18	-1.25	**30.53**	3.87	21.99	-2.85
	3	27.52	3.87	20.53	-1.28	28.67	3.47	20.93	-2.90
	4	27.05	3.93	19.94	-1.40	28.11	3.38	20.66	-3.05
$Q_{7,10}$	1	45.79	7.60	**36.80**	**0.63**	47.92	**7.41**	**34.02**	**-1.29**
	2	**41.08**	**6.53**	33.41	0.13	**42.16**	6.10	30.97	-1.79
	3	39.27	6.21	31.82	-0.14	40.84	5.94	30.29	-2.02
	4	38.41	5.88	31.65	-0.28	39.83	5.67	29.81	-2.07

Bold values correspond to performances surpassing the corresponding ungaged case model.

Figure captions

Figure 1. Hydrometric stations of the study area.

Figure 2. Normal probability plot for the station 023402.

Figure 3. Streamflow for the year 1970 at station 020802. Recession segments are identified with grey areas.

Figure 4. Plot of $-dQ/dt$ versus Q for the station 030103.

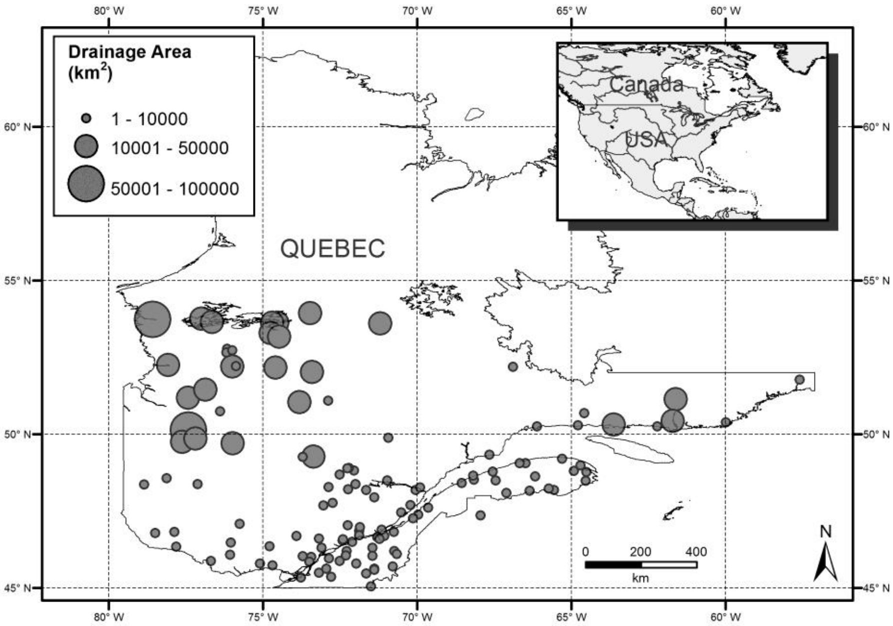
Drainage Area
(km2)
1 - 10000
10001 - 50000
50001 - 100000
QUEBEC
Canada
USA
N
0
200
400
km
80° W
75° W
70° W
65° W
60° W
60° N
55° N
50° N
45° N

Figure 1.

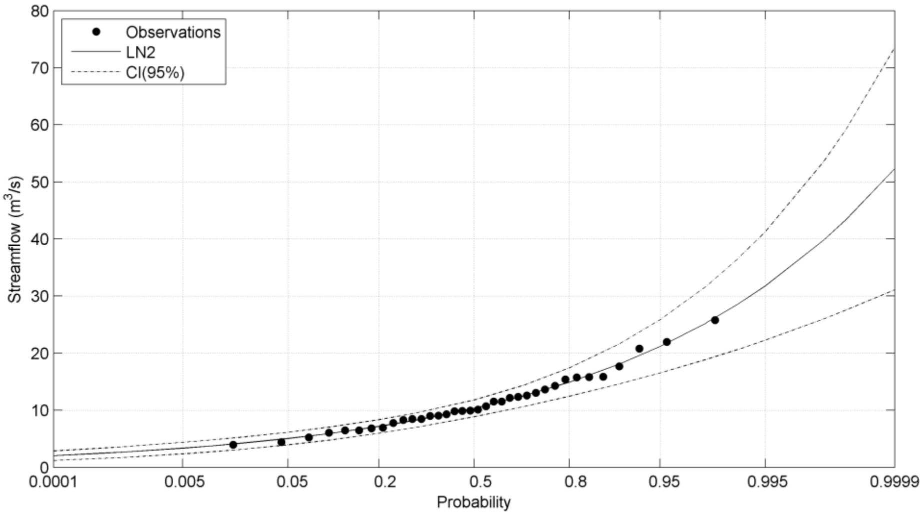
Observations
LN2
CI(95%)
Streamflow (m3/s)
Probability
0
10
20
30
40
50
60
70
80
0.0001
0.005
0.05
0.2
0.5
0.8
0.95
0.995
0.9999

Figure 2.

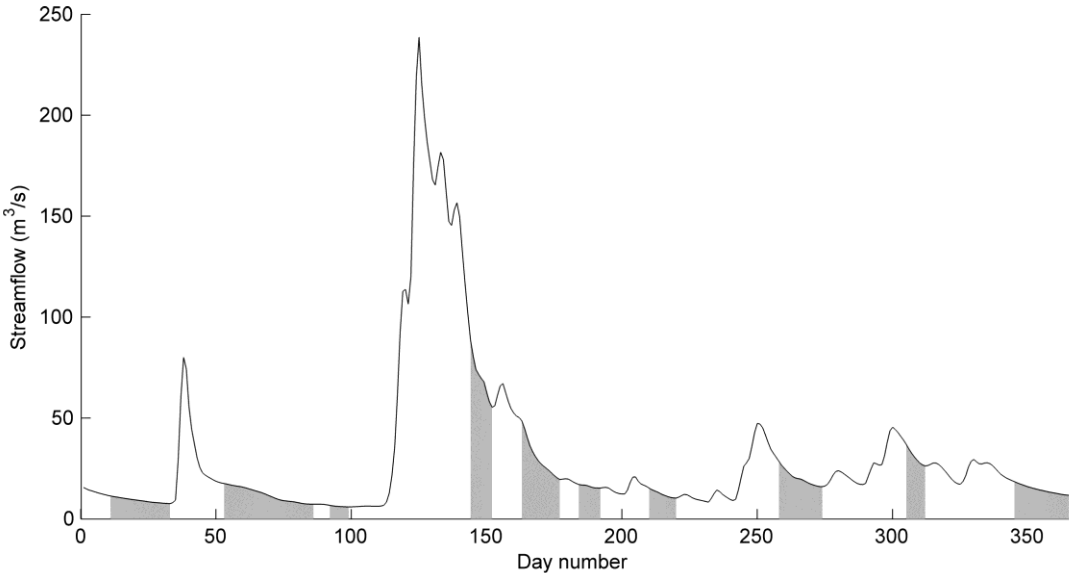
Streamflow (m3/s)
250
200
150
100
50
0
Day number
0
50
100
150
200
250
300
350

Figure 3.

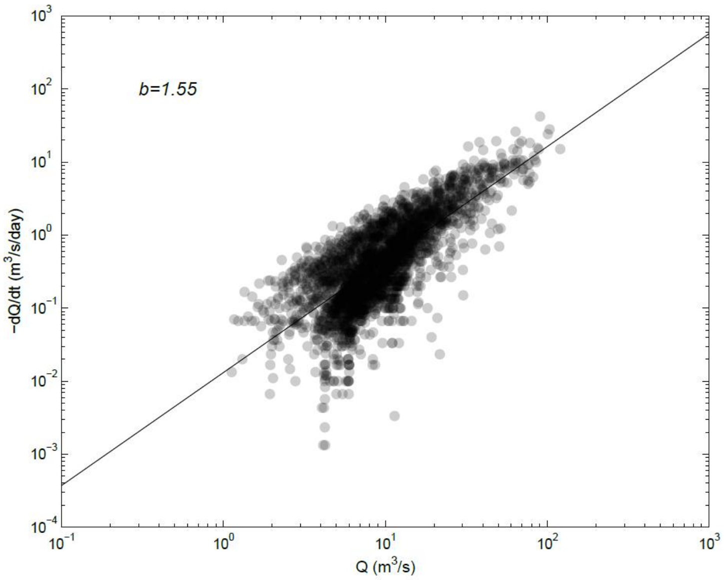

Figure 4.