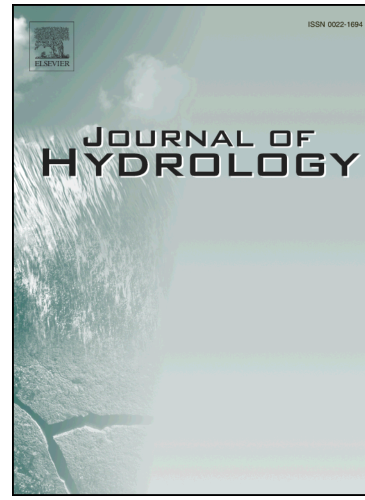
Accepted Manuscript



Research papers

Low-flow frequency analysis at ungauged sites based on regionally estimated streamflows

Ana I. Requena, Taha B.M.J. Ouarda, Fateh Chebana

PII: S0022-1694(18)30438-4
DOI: https://doi.org/10.1016/j.jhydrol.2018.06.016
Reference: HYDROL 22865

To appear in: *Journal of Hydrology*

Received Date: 26 July 2017
Revised Date: 2 April 2018
Accepted Date: 6 June 2018

Please cite this article as: Requena, A.I., Ouarda, T.B.M., Chebana, F., Low-flow frequency analysis at ungauged sites based on regionally estimated streamflows, *Journal of Hydrology* (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.06.016

Low-flow frequency analysis at ungauged sites based on regionally estimated streamflows

Ana I. Requena [1,*], Taha B. M. J. Ouarda [1] and Fateh Chebana [1]

[1] INRS-ETE, 490 rue de la Couronne, Quebec, QC, Canada. (requena.ana.isa@gmail.com; fateh.chebana@ete.inrs.ca; taha.ouarda@ete.inrs.ca).

Corresponding author: Ana I. Requena (requena.ana.isa@gmail.com)

Abstract

Estimation of low-flow quantiles or indices at ungauged sites is traditionally done through regional low-flow frequency analysis. However, traditional methods imply a prior aggregation of the regional information due to the usual focus on a given quantile. This leads to loss of information for estimating additional quantiles or performing additional applications. In the present study, the recently proposed regional streamflow based frequency analysis (RSBFA) approach is evaluated for the estimation of low-flow quantiles. The approach, originally applied to floods, is based on regionally estimating the daily streamflow series at the ungauged site to later obtain hydrological quantiles through a local frequency analysis. In the present study, the RSBFA approach is applied to low flows through a case study in the province of Quebec, Canada, considered in prior traditional regional low-flow studies. Although the RSBFA approach does not systematically lead to the best results, its relevance resides in the fact that the whole regionally estimated daily streamflow series is provided at the ungauged site. This allows an easy estimation of low-flow quantiles associated to particular durations, such as the 7-day (30-day) average low flows corresponding to a return period of 2 or 10 years (5 years). Furthermore, a large range of additional low-flow quantiles, such as $q_{95\%}$ (streamflow expected to be exceeded 95% of the time divided by the catchment area), may be obtained from the regionally estimated streamflow series with a very good performance. Therefore, any specific or absolute, annual or seasonal low-flow quantile may be easily obtained by a local low-flow frequency analysis without repeating the regional procedure. The approach does not require complex models, and it may also allow combining regional and local information in a straight forward manner.
Keywords: Regional frequency analysis; low-flow event; ungauged site; flow duration curve; daily streamflow series; quantile.

1. Introduction

Estimation of low-flow quantiles, also called low-flow indices, is essential for water resource management, water quality regulation and habitat protection (e.g. Pyrce 2004). This estimation is obtained at ungauged sites by regional low-flow frequency analysis. Most low-flow frequency analysis approaches reported in the literature are based on regional (or global) regression (e.g. Ouarda et al. 2005; Laaha and Blöschl 2006; Tsakiris et al. 2011). They establish a relation between low-flow quantiles and catchment descriptors. More complex procedures for characterising this relation, such as by artificial neuronal networks, are also studied (e.g. Ouarda and Shu 2009). Other approaches can be classified as geostatistical methods (e.g. Castiglioni et al. 2009). They account for spatial correlation between low flows at nearby sites; and a dense gauging network is required for their application. One example is the baseflow correlation technique (e.g. Stedinger and Thomas 1985; Reilly and Kroll 2003). The procedure requires the availability of a number of streamflow measurements during the baseflow conditions for the same days at both gauged and "ungauged" sites; and a high correlation between sites. A low-flow hydrology review is presented in Smakhtin (2001), and a broad comparison among different regional low-flow procedures can be found in Salinas et al. (2013).

Most studies on regional low flows focused on the estimation of a given quantile or of a limited number of them. Examples of the former case include $q_{95\%}$ defined as the streamflow expected to be exceeded 95% of the time divided by the catchment area (e.g. Laaha and Blöschl 2006); $Q_{10,7}$ defined as the 7-day average low flow corresponding to a return period of 10 years (e.g. Reilly and Kroll 2003); and Q_{355} defined as the streamflow exceeded on average 355 days in a year (e.g. Castiglioni et al. 2011). A number of studies presented results for several low-flow quantiles, such as $Q_{2,7}$ and $Q_{10,7}$ (e.g. Vogel and Kroll 1992; Tsakiris et al. 2011) or $Q_{2,7}$, $Q_{10,7}$ and $Q_{5,30}$ (e.g. Ouarda and Shu 2009).

Alternatively, low-flow quantiles may be estimated by a regional flow duration curve (FDC) or by a regional low-flow frequency curve built from gauged sites (e.g. see Fennessey and Vogel 1990; Smakhtin 2001; Castellarin et al. 2004; Castellarin et al. 2013). However, only partial information is also obtained by these approaches due to the focus on a given type of low flows. For instance, a given regional FDC that allows estimating $Q_{95\%}$ defined as the streamflow expected to be exceeded 95% of the time, cannot be directly used to estimate other low-flow quantiles such as $Q_{2,7}$ or $Q_{5,30.}$ In addition, a given regional low-flow frequency curve cannot simultaneously estimate $Q_{2,7}$ and $Q_{5,30}$, for which the regional procedure needs to be repeated. Although some of these quantiles may be correlated (e.g. see Fennessey and Vogel 1990; Salinas et al. 2013), their direct estimation may be valuable for addressing different purposes of a low-flow analysis. In this regard, Smakhtin (2001) underlined that despite correlations, different low-flow quantiles should be directly estimated from observed streamflow series when the quality of the record is suitable. In relation to ungauged sites, it was indicated that several low-flow quantiles may be extracted from streamflow series generated at a target site when a suitable model is considered for such estimation.

An alternative approach to construct FDCs was introduced by Vogel and Fennessey (1994) with the aim of providing a median annual FDC that is less sensitive to the period of record considered, and which may have an impact on the low-flow range of the FDC if such a period is not long enough. This approach also allows obtaining regional annual FDCs related to a given probability, which may in turn allow estimating low-flow (or flood) quantiles (Castellarin et al. 2007). However, similar to what happens in the aforementioned approaches, annual FDCs would need to be properly recalculated if low-flow or flood quantiles related to aggregated data, such as $Q_{2,7}$, need to be estimated (e.g. Claps and Fiorentino 1997).

In the present study, a recently proposed approach for estimating hydrological quantiles at ungauged sites is adapted and evaluated in the case of low flows. The approach, called regional streamflow estimation based frequency analysis (RSBFA), is introduced and applied to floods in Requena et al. (2017). The notion behind the RSBFA approach is the regional estimation of daily streamflow series at the ungauged site, to later perform a local frequency analysis to estimate desired quantiles (Ouarda 2016). An overview of the path adopted by RSBFA in comparison to traditional low-flow regional frequency analysis approaches is presented in Fig. 1. This alternative approach has a number of advantages in comparison to traditional regional methods. It provides the complete daily streamflow series at the ungauged site, where all the regional information is included. Therefore, it avoids the prior aggregation step involved in traditional regional frequency analysis procedures focused on a given quantile or given application (e.g. low flows or floods), and allows using all the regional information just once.

In particular, the application of the RSBFA approach to carry out a regional low-flow frequency analysis implies that, for instance, any low-flow quantile related to any average annual (or seasonal) d-day minimum flow series may be easily obtained from the regionally estimated streamflow series. Therefore, a whole set of low-flow quantiles may be estimated without the need to repeat the regional analysis if an additional quantile is required. It also facilitates the straightforward estimation of both specific and absolute quantiles. The approach is very flexible due to the fact that regionally estimated streamflow series may be in turn used for performing a number of applications such as stationary or non-stationary frequency analysis, and univariate or multivariate frequency analysis studies on low flows or floods. Furthermore, the approach does not specifically need a dense gauging network or streamflow measurements at the ungauged site for streamflow estimation, and does not rely on complex statistical models. Finally, the RSBFA approach allows easily combining local and regional information in the case of partially gauged sites, without the need to resort to complex statistical models such as empirical Bayes estimation (Fill and Stedinger 1998) or parametric Bayesian approaches (Seidou et al. 2006). Indeed, if a short streamflow record is available at the target site, it can simply be combined with the regionally estimated streamflow record into a single streamflow series. At-site frequency analysis can then be applied to the combined streamflow series in a straightforward manner. It is important to note that the RSBFA approach also presents the advantage of allowing the direct use of daily streamflow estimates that are often already available at the ungauged sites from previous regional transfer studies. The considered RSBFA approach is adapted and applied to low flows through a case study in the province of Quebec, Canada, where several regional low-flow frequency analysis approaches have been previously applied in the literature (Sect. 3). The methodology is described in Sect. 2. Results are illustrated, compared to the ones obtained by traditional approaches, and discussed in Sect. 4. Conclusions are presented in Sect. 5.

2. Methodology

The regional approach considered in the present study for estimating low-flow quantiles at ungauged sites is the RSBFA approach presented in Requena et al. (2017). It consists of two main steps:

- **(i)** Regional estimation of daily streamflow series at the destination (or ungauged) site; which is in turn divided into:
  - **a)** Regional estimation of the FDC streamflow quantiles at the destination site;

b) Transfer of daily streamflow series from source sites to the destination site;

(ii) Estimation of low-flow quantiles at the destination site by a local frequency analysis.

Note that step (i) is often carried out in advance, and hence the results of the estimation of streamflow series at the destination site are often already available. The regional FDC streamflow quantile estimation in step (i.a) is generally carried out for a limited number of streamflow quantiles, which are obtained from the whole daily streamflow information. This regional estimation allows obtaining the daily streamflow series at the ungauged site by the transfer procedure in step (i.b). A given minimum flow series may then be extracted from the regionally estimated streamflow series at the destination site, with the aim of obtaining any desired low-flow quantile by a local frequency analysis (step (ii)). Therefore, the low-flow quantiles finally estimated in step (ii) are separately obtained from the streamflow quantiles regionally estimated in step (i.a), and they do not need to be identified during the regional procedure.

Step (i) is common for the application of the procedure to any type of frequency analysis, such as floods or low flows, and is summarised in Sect. 2.1. However, step (ii) needs to be slightly adapted to the type of analysis to be performed. The adaptation is related to which series need to be extracted from the regionally estimated streamflow series in order to apply the local frequency analysis. The adaptation to low flows is presented in Sect. 2.2. Assessment measures for evaluating the performance of the approach when applied to low flows are presented in Sect. 2.3.

2.1. Regional estimation of daily streamflow series

The first step of the methodology consists in the regional estimation of the daily streamflow series at the destination site. A summary of the procedure is presented herein.

a) **Regional estimation of the FDC streamflow quantiles:** The estimation of the FDC streamflow quantiles at the destination site is carried out through the regression-based logarithmic interpolation method presented in Shu and Ouarda (2012). The modifications introduced by Requena et al. (2017) for applying a more objective procedure to ensure decreasing monotonicity over the FDC are also considered. First, a point-wise FDC is built at each gauged site from the observed daily streamflow series. This is achieved by identifying the streamflow quantiles associated with given percentile points (or probabilities of exceedance) p. The following 17 un-evenly fixed percentile points are considered as suitable for properly characterising the FDC (Shu and Ouarda 2012): p = 0.01%, 0.1%, 0.5%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, and 99.99%. Second, each of these streamflow quantiles is estimated at the destination site by a regional regression equation in order to build its point-wise FDC:

$$\ln(\mathrm{Q_p}) = \ln(\mathrm{b_0}) + \mathrm{b_1}\ln(\mathrm{V_1}) + \mathrm{b_2}\ln(\mathrm{V_2}) + \cdots + \mathrm{b_k}\ln(\mathrm{V_k}), \quad (1)$$

where Q_p are the streamflow quantile values at the gauged sites for a given percentile point p, and $(V_1, V_2, \dots, V_k)$ are the descriptors selected by stepwise regression for characterising Q_p. The values of the regional parameters $(b_0, b_1, \dots, b_k)$ are obtained from Eq. (1) by using information from gauged sites. They are then used for estimating Q_p at the destination site by replacing its catchment descriptor values in Eq. (1).

By definition, a FDC must present a decreasing monotonicity over percentile points. Hence, if the estimated point-wise FDC does not initially preserve this condition, a smoothing curve is fitted to ensure it (Ouarda et al. 2010):

$$Q_p = \frac{ap+b}{p^c+e}\,, \qquad a < 0,\ b > 0 \text{ and } c \in [0,1] \tag{2}$$

where a, b, c and e are parameters estimated by a least square approach. Finally, intermediate points of the FDC are estimated by logarithmic interpolation, or by the smoothing curve (if it was previously fitted to ensure decreasing monotonicity), in order to supply values required during the streamflow transfer process.

b) **Transfer of daily streamflow series:** Transfer of daily streamflows from source sites to a destination site is based on the nonlinear spatial interpolation method (e.g. Hughes and Smakhtin 1996). The latter assumes the same probability of exceedance p of the streamflow of a given day for the source and destination sites. The influence of several source sites may be considered in the procedure by weighting the streamflow estimated by each source site. Based on the findings of Shu and Ouarda (2012), four source sites are considered in the present study, and their selection is based on spatial proximity (e.g. Hughes and Smakhtin 1996). Note that in the literature, from three to five source sites are usually considered for streamflow transfer (e.g. Patil and Stieglitz 2012; Ergen and Kentel 2016).

2.2. Estimation of low-flow quantiles by a local frequency analysis

The second step of the RSBFA approach is the estimation of quantiles at the destination site by a local frequency analysis, which in this study is focused on estimating low-flow quantiles. The procedure is described below. Once the daily streamflow series at the ungauged site is regionally estimated through the procedure described in Sect. 2.1, it is used to extract annual or seasonal minimum flow series. In this study, winter and summer events are considered (Herrera 2008). The average seasonal d-day minimum flow for summer and winter events is obtained for each year through a moving window. This is done for both d = 7 and 30 days (see Sect. 3 for details).

Distributions commonly used in low-flow frequency analysis are considered for fitting each minimum flow series (e.g. see Smakhtin 2001; Ouarda and Shu 2009). Candidate distributions in this study are the Gumbel (G), Weibull (W2), two-parameter lognormal (LN2), three-parameter lognormal (LN3), generalised extreme value (GEV), gamma (GA), Pearson type III (P3), log-Pearson type III (LP3) and generalised Pareto (GP). Several parameter estimation methods are considered in the procedure, such as the method of moments, the method of L-moments or the maximum likelihood approach. The well-known Bayesian information criterion (BIC) (Schwarz 1978) is applied to identify the best model for fitting the data by considering parsimony (see Laio et al. 2009 for details). The desired low-flow quantiles are then estimated by the selected distribution. Goodness-of-fit tests and graphical representations may be useful in identifying the best model; yet only the BIC is considered in this study for more objective comparison purposes (see Sect. 3).

2.3. Evaluation of the approach

Error assessment measures used for evaluating traditional regional low-flow frequency analysis approaches are considered for assessing the performance of the low-flow quantiles estimated in this study. Their application is based on a jackknife (or leave-one-out) procedure, which entails estimating low-flow quantiles at a gauged site as if it were ungauged. This is a procedure commonly applied to assess the performance of regional approaches, and their use in this study allows obtained results to be compared with those of traditional approaches applied to the same case study. In the present study, this implies that the daily streamflow series regionally estimated at a given site (Sect. 2.1) is obtained without considering its available information. The probability distribution selected for fitting the extracted data series (Sect. 2.2) is then used for estimating the desired low-flow quantiles. The procedure is performed for each site. In particular, the assessment measures considered are the Nash criterion (NASH), root mean squared error (RMSE), relative root mean squared error (RRMSE), mean bias (BIAS) and relative mean bias (RBIAS):

$$\mathrm{NASH} = 1 - \frac{\sum_{i=1}^{N_t}(q_i - \hat{q}_i)^2}{\sum_{i=1}^{N_t}(q_i - \bar{q})^2} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{t}}}\sum_{i=1}^{N_t}(q_i - \hat{q}_i)^2} \tag{4}$$

$$\mathrm{RRMSE} = \sqrt{\frac{1}{N_{\mathrm{t}}}\sum_{i=1}^{N_t}\left(\frac{q_i - \hat{q}_i}{q_i}\right)^2} \tag{5}$$

$$\mathrm{BIAS} = \frac{1}{N_{\mathrm{t}}}\sum_{i=1}^{N_t}(q_i - \hat{q}_i) \tag{6}$$

$$\mathrm{RBIAS} = \frac{1}{N_{\mathrm{t}}}\sum_{i=1}^{N_t}\left(\frac{q_i - \hat{q}_i}{q_i}\right), \tag{7}$$

where $\hat{q}_i$ is the low-flow quantile regionally estimated through the jackknife procedure at the site i, q_i is the local low-flow quantile at the site i, N_{t} is the total number of sites in the study region, and $\bar{q}$ is the mean of the local low-flow quantile over the N_{t} sites. The local low-flow quantile is estimated from observed streamflow series at the given site. It is important to mention that NASH is equivalent to the coefficient of determination used in evaluating low-flow approaches in some other studies (Viglione et al. 2013). The overall performance of a given approach is usually considered as suitable if NASH is equal to or larger than 0.8 (e.g. Chokmani and Ouarda 2004). RMSE indicates the precision in the regionally estimated low-flow quantiles regarding the local quantiles; whereas relative RMSE reflects the average precision. Relative RMSE eliminates the scale effect that is present in RMSE. Consequently, RMSE may be dominated by the error of a few large basins, while this is not the case for relative RMSE. BIAS indicates if the regionally estimated low-flow quantiles tend to underestimate/overestimate the local quantiles, and relative BIAS reflects the average tendency for an underestimation/overestimation. Note that error measures such as RMSE and BIAS may be large if local low-flow quantiles are large; whereas relative RMSE and BIAS may be large if local low-flow quantiles are small (see Eqs. 5 and 7). Therefore, different error measures should be considered for a complete and proper performance assessment.

3. Case study

The case study considered for the application of the RSBFA approach to low flows is the hydrometric station network of the southern part of the province of Quebec, Canada. For comparison purposes, the definitions of low-flow events and quantiles, as well as the sites used in the regional low-flow frequency analysis approach adopted in the present study are the same as the ones used in previous low-flow studies dealing with the same database. Three low-flow studies exist in this regard. Low-flow quantiles were regionally estimated based on their relation with catchment descriptors by multiple regression (MR) in Ouarda et al. (2005) and by artificial neuronal networks (ANN) in Ouarda and Shu (2009). The effect of considering a recession parameter based on a non-linear reservoir model as predictor in low-flow quantile estimation was investigated in Charron and Ouarda (2015). Due to the specific focus of the latter study, different information from the one used by the former two studies was considered. Hence, only low-flow quantiles obtained in Ouarda et al. (2005) and Ouarda and Shu (2009) are considered for comparison purposes in the present study.

In the aforementioned studies, winter and summer low-flow events are studied separately to account for the strong differences in seasonal low-flow regimes. The low-flow quantiles ($Q_{T,d}$) for the return periods T = 2, 5, 10, and durations d = 7 and 30 days ($Q_{5,30}$, $Q_{2,7}$ and $Q_{10,7}$) are then analysed for the winter and summer seasons. These low-flow quantiles are commonly used for water quality control and habitat maintenance (e.g. see Pyrce 2004). Note that the d-day T-year low-flow quantile ($Q_{T,d}$) is overall defined as the average annual (or seasonal) d-day minimum flow expected to be exceeded on average T-1 out of every T years (e.g. Reilly and Kroll 2003) or with a non-exceedance probability of 1/T (e.g. Tsakiris et al. 2011).

Hydrological data is available for the 190 hydrometric stations managed by the Ministry of the Environment of the province of Quebec. As a result of a data quality assessment for low-flow analysis, 134 and 129 sites were respectively considered when estimating $Q_{T,30}$ and $Q_{T,7}$ for the summer season. For the winter season, 135 and 133 sites were respectively considered when estimating $Q_{T,30}$ and $Q_{T,7}$. The sites have natural flow records of at least 10 years, and passed stationarity (Kendall 1975) and independence (Wald and Wolfowitz 1943) tests. These four sets of sites are also considered in Ouarda et al. (2005) and Ouarda and Shu (2009). Commonly used distributions in low-flow frequency analysis were considered in the model selection process performed by the BIC. Corresponding $Q_{T,30}$ and $Q_{T,7}$ local low-flow quantiles are available from a low-flow frequency analysis performed at each of the studied sites (Ouarda et al. 2005). A summary of the descriptive statistics of the local low-flow quantiles is presented in Table 1.

Physiographical and meteorological descriptors are also available at each site. The following descriptors were selected by stepwise regression to characterise low-flow quantiles in Ouarda et al. (2005), and were also considered in Ouarda and Shu (2009). These descriptors are the catchment area (BV), fraction of the catchment controlled by lakes (PLAC), fraction of the catchment occupied by forest (PFOR), annual mean degree-days below 0°C (DDBZ), summer mean liquid precipitation (PLME), curve number (CN) and average number of days with a mean temperature exceeding 27°C (NJH27). A summary of their descriptive statistics is also displayed in Table 1. In particular, winter low-flow quantiles were characterised by BV, PLAC, PFOR and DDBZ; and summer low-flow quantiles were characterised by BV, PLAC, PLME, NJH27 and CN. A log transformation was used for BV and DDBZ, and a square root transformation for PLAC. They are also used in this study for comparison purposes.

4. Results and discussion

The described methodology is applied to the case study for the four low-flow quantiles being considered: winter $Q_{T,30}$, winter $Q_{T,7}$, summer $Q_{T,30}$ and summer $Q_{T,7}$. Since each low-flow quantile is based on a different set of sites, the procedure is applied four times. As a result, a 30-year regionally estimated daily streamflow series is obtained each time, and the corresponding minimum flow series is extracted: the winter 7-day, winter 30-day, summer 7-day and summer 30-day minimum flow series. Note that the procedure could be done just once if a unique set of sites was considered. In such a case, the four minimum flow series would have been obtained from the same daily streamflow series regionally estimated at the ungauged site. Nevertheless, the four sets of sites are used in the present study for comparison purposes with results obtained by traditional regional low-flow frequency analysis approaches.

The coefficients of determination corresponding to the regressions used to estimate the FDC streamflow quantiles at ungauged sites are very high for the four low-flow quantiles, with values equal to or larger than 0.91. These values of the coefficient of determination are obtained through a jackknife procedure, as explained in Sect. 2.3. Note that if low values of the coefficient of determination were obtained, additional descriptors could be needed in the regional regression. Otherwise, the performance of the regression-based RSBFA approach should be carefully assessed. The performance of regionally estimated daily streamflow series is in the same order of magnitude as in Shu and Ouarda (2012), where a different study period and different sites were considered. The distributions mostly selected for fitting the minimum flow series extracted from the daily streamflow series regionally estimated at each of the studied sites are the G, GA, LN2, LN3 and W2 distributions. For the winter 7-day minimum flow series, the G, GA, LN2, W2 and LN3 distributions were found to provide the best fit for 39%, 23%, 18%, 17% and 2% of the sites, respectively. For the winter 30-day minimum flow series, the G, LN2, GA, LN3 and W2 were selected for 39%, 27%, 21%, 6% and 4% of the sites. For the summer 7-day minimum flow series, the LN2, G, GA, W2 and LN3 were selected for the 34%, 20%, 19%, 13% and 12% of the sites. And for the summer 30-day minimum flow series, the LN2, GA, W2, G and LN3 distributions were selected for 42%, 16%, 14%, 13% and 10% of the sites.

Low-flow quantiles regionally estimated by applying the RSBFA approach are visually compared with local low-flow quantiles in Fig. 2. These Q-Q plots show regionally estimated quantiles to be similar to local quantiles, although several exceptions exist. Winter low-flow quantiles show two clear outliers, sites ‘081007’ and ‘081002’. These two sites were also found as outliers in traditional approaches, and the reason for their low-flow underestimation was attributed to having a very large PLAC and DDBZ values (Ouarda and Shu 2009).

Site ‘081002’ is found as a less noticeable outlier for summer low-flow quantiles (Fig. 2). In the summer season, traditional approaches identified site ‘090601’ as a noticeable outlier. This site was also found as an outlier for the RSBFA approach. The reason of its underestimation was attributed to a very large PFOR, low PLAC and low CN compared with sites of a similar size. The MR approach also identified site ‘076601’ as an outlier, but the site was correctly estimated for ANN-based approaches (Ouarda and Shu 2009). The reason of its overestimation was attributed to a low NJH27 and a low CN compared to sites of a similar size. This site was also found as an outlier in the present study. Overall, low-flow Q-Q plots for the RSBFA approach seem to be closer to MR results than to ANN-based results (see Ouarda and Shu 2009, Figures 4 to 9). Error $(q_i - \hat{q}_i)$ and relative error $\left(\frac{q_i - \hat{q}_i}{q_i}\right)$ results for each site are respectively displayed in Fig. 3 and Fig. 4. Aforementioned sites ‘081007’, ‘081002’ and ‘090601’ present a large error (Fig. 3), but their associated relative errors are not especially large (Fig. 4). The exception is site ‘076601’, due to the association of a small local low-flow quantile (see Fig. 2). Without taking into account site ‘076601’, sites with a large relative error are usually associated to a small error due to presenting a small local low-flow quantile.

Performance of low-flow quantiles estimated by the RSBFA approach is quantitatively compared with performance of traditional low-flow regional approaches. Assessment criteria results for winter and summer low-flow quantiles are displayed in Fig. 5 and Fig. 6, respectively. As seen in both figures, all NASH values are larger than 0.8; and hence, all approaches may be considered as suitable. Winter low-flow quantile performance results in Fig. 5 are diverse. The advanced Ensemble ANN (EANN) approach obtains the best results for all criteria except for BIAS. The worst RMSE and RBIAS values are obtained by the RSBFA approach. The worst

RRMSE values are given by the Single ANN (SANN) approach, except for $Q_{10,7}$. The worst BIAS values are presented by the MR approach.

Summer low-flow quantile performance results in Fig. 6 show a more defined situation. ANN-based approaches always lead to better results than the MR or RSBFA approaches, both of which make use of regional regression. Recall that the RSBFA approach uses regional regression for estimating the point-wise FDC at the ungauged site. This better behaviour of ANN-based approaches is probably due to their ability to represent a non-linear relationship between summer low-flow quantiles and catchment descriptors. In particular, the EANN approach always lead to the best results. By analysing summer results obtained by the RSBFA and by the MR approach; it is found that the RSBFA approach leads to lower RMSE and BIAS, but larger RRMSE and RBIAS values.

The larger non-linearity presented by summer low-flow quantiles in comparison to winter low-flow quantiles for the case study may also be supported based on the fact that the non-linear SANN and EANN approaches have smaller RRMSE and RBIAS values for summer than for winter results; whereas the opposite happens for the MR or RSBFA approaches (see Fig. 5 and Fig. 6). In this regard, it is important to mention that winter results obtained for the RSBFA approach by considering a larger number of available descriptors, led to an overall performance improvement; yet the opposite occurred with summer results (not shown). Therefore, the use of additional descriptors in the (linear) regression-based approach only improves results for winter low-flow quantiles, which supports the argument of summer low-flow quantiles having a non-linear behaviour. Thus, approaches that are able to account for non-linearity are expected to lead to better results in this case. Note that a number of regional approaches considering non-linearity have been recently introduced in hydrology; yet they are usually applied to floods (e.g. Chebana et al. 2014; Durocher et al. 2015, Ouali et al. 2017).

As could be expected, results indicated the EANN approach to be an advanced and notable approach able to adequately estimate winter and summer low-flow quantiles. The other approaches may also be considered as suitable, as they have large NASH values. However, they present a different performance depending on the additional assessment criteria to be considered, and on the type of event being studied. In this regard, it is important to underline that although results obtained by the RSBFA approach are not the best ones, they are overall comparable to results obtained by traditional approaches. Furthermore, these results are achieved although, unlike traditional approaches, the RSBFA approach is not specifically applied to estimate a given low-flow quantile and does not use complex models. Also note that the whole procedure used to estimate $Q_{2,7}$ by the EANN, SANN or MR approaches has to be repeated to obtain any other low-flow quantile such as $Q_{5,30}$. While on the contrary, in the case of the RSBFA, $Q_{5,30}$ is obtained from the regionally estimated streamflow series at the ungauged site. Hence, in this case, estimating the low-flow quantile only requires extracting the associated minimum flow series and fitting a probability distribution. Moreover, when the desired quantile is $Q_{10,7}$, it is directly obtained by the fitted distribution already used to estimate $Q_{2,7}$. Therefore, the RSBFA approach has the advantage of being able to directly obtain additional low-flow quantiles from the corresponding fitted probability distribution or from the existing regionally estimated streamflow series.

A further example of additional low-flow quantiles directly obtained from the regionally estimated series is given below. For this case, the well-known $q_{95\%}$ is considered, which is the streamflow expected to be exceeded 95% of the time divided by the catchment area. Jackknife-based assessment results are obtained by calculating this specific low-flow quantile $q_{95\%}$ from the regionally estimated streamflow series at the site considered as ungauged, and from the observed streamflow series at the given site. Performance results are shown in Table 2, and the corresponding Q-Q plot is displayed in Fig. 7. Note that four sets of sites were considered in the present study for comparing *d*-day *T*-year low-flow quantiles obtained by traditional approaches. As illustration for this particular analysis, results are shown for the set of sites considered when estimating winter $Q_{T,7}$. As seen in Table 2 and also in Fig. 7, $q_{95\%}$ results show a very good performance, with a very large NASH and very small values for the other assessment criteria. The same holds when obtaining results by considering any of the other sets of sites.

The absolute low-flow quantile $Q_{95\%}$ as estimated when obtaining the point-wise FDC at the ungauged site by regression equations using information from all gauged sites (result of step (i.a)), is also compared with the value of $Q_{95\%}$ calculated from the regionally estimated streamflow series obtained by the transfer procedure (result of step (i.b)). As shown in Table 2 and Fig. 7, $Q_{95\%}$ results are largely improved when obtained from the regionally estimated streamflows. This is likely due to the positive effect of the four source sites in the transfer procedure. The same holds when obtaining results by considering any of the other sets of sites. Note that $q_{95\%}$ results regarding the FDC are not shown due to the fact that the FDC estimation is not based on specific quantiles. Also note that RRMSE and RBIAS values are the same for $Q_{95\%}$ and $q_{95\%}$ obtained from the regionally estimated streamflow series, as $q_{95\%}$ is a specific quantile obtained by dividing the estimated $Q_{95\%}$ by the catchment area.

Additionally, any kind of local frequency analysis could be performed at the ungauged site, thanks to the availability of the whole daily streamflow series. In this regard, the proposed approach may facilitate the development of multivariate low-flow frequency analysis to properly deal with the marked multivariate nature of low flows (Ouarda et al. 2008). Furthermore, the RSBFA approach allows easily combining local and regional information, which is a tedious matter usually addressed in the literature by proposing complicated methods such as Bayesian procedures (e.g. Fill and Stedinger 1998; Seidou et al. 2006). All the aforementioned benefits support the use of the RSBFA approach to conduct regional frequency analysis in practice.

Further research could consist in the application of the RSBFA approach by considering different procedures for estimating the FDC at the ungauged site. Furthermore, the FDC estimation at the ungauged site could be performed by using a subset of sites selected based on similarity and homogeneity instead of by using all available sites (e.g. Castellarin et al. 2004; Mendicino and Senatore 2013).

5. Conclusions

A recently proposed approach for conducting regional hydrological frequency analysis is applied to low flows in this study. The approach, called regional streamflow estimation based frequency analysis (RSBFA), was originally applied on a flood case study when it was introduced. It consists in the regional estimation of the daily streamflows at the ungauged site, with the aim of later performing a local frequency analysis to obtain hydrologic quantiles at the ungauged site. A suitable performance is observed for the RSBFA approach for the case study. Although the RSBFA approach does not systematically provide the best results, in part due to its lack of focus on the direct estimation of particular low-flow quantiles (as opposite to traditional approaches), it possesses a number of practical benefits. The method consists in a simple approach that provides the complete daily streamflow series at the ungauged site where all the regional information is included. Practically, this turns the ungauged site into a gauged site, which allows easily obtaining any additional low-flow quantiles, as well as performing any local (e.g. seasonal or annual, univariate or multivariate, stationary or non-stationary, etc.) frequency analysis on the target site without repeating the regional procedure. A strong advantage of the RSBFA procedure is represented by the easy estimation of low-flow quantiles associated with particular durations (e.g. 7 or 30 days) from the regionally estimated daily streamflow series. Therefore, in practice, traditional low-flow frequency analysis approaches could be preferred if few low-flow quantiles are of interest; whereas the RSBFA approach would be preferred if the focus is on the estimation of a large number and/or different types of low-flow quantiles. The RSBFA approach would be also relevant to estimate both flood and low-flow quantiles, as well as to easily combine local and regional streamflow series.

Acknowledgements

The financial support provided by the Merit scholarship program for foreign students – Postdoctoral research fellowship of the *Ministère de l'Éducation et de l'Enseignement Supérieur du Québec* managed by the *Fonds de recherche du Québec – Nature et technologies* is gratefully acknowledged. The authors also thank the financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would also like to thank the Editor, András Bárdossy, the Associate Editor, Félix Francés, and two anonymous reviewers who help improve the quality of the present paper.

Acronym list

Acronym	Definition
ANN	Artificial neuronal networks
BIC	Bayesian information criterion
EANN	Ensemble artificial neuronal networks
FDC	Flow duration curve
MR	Multiple regression
RSBFA	Regional streamflow based frequency analysis
SANN	Single artificial neuronal networks

References

Castellarin, A., Botter, G., Hughes, D.A., Liu, S., Ouarda, T.B.M.J., Parajka, J., Post, D.A., Sivapalan, M., Spence, C., Viglione, A., Vogel, R.M., 2013. Prediction of flow duration curves in ungauged basins, in "Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales", edited by G. Bloschl et al., University Press, Cambridge. 135-162.

Castellarin, A., Camorani, G., Brath, A., 2007. Predicting annual and long-term flow-duration curves in ungauged basins. Adv. Water Resour. 30, 937-953.

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

Castiglioni, S., Castellarin, A., Montanari, A., 2009. Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. Journal of Hydrology. 378, 272-280.

Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J., Laaha, G., Blöschl, G., 2011. Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging. Hydrology and Earth System Sciences. 15, 715-727.

Charron, C., Ouarda, T.B.M.J., 2015. Regional low-flow frequency analysis with a recession parameter from a non-linear reservoir model. Journal of Hydrology. 524, 468-475.

Chebana, F., Charron, C., Ouarda, T.B.M.J., 2014. Regional frequency analysis at ungauged sites with the generalized additive model. Journal of Hydrometeorology. 15(6), 2418-2428.

Chokmani, K., Ouarda, T. B. M. J., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water Resour. Res. 40, W12514.

Claps, P., Fiorentino, M., 1997. Probabilistic flow duration curves for use in environmental planning and management. In Integrated Approach to Environmental Data Management Systems. Springer Netherlands. 255-266.

Durocher, M., Chebana, F. Ouarda, T.B.M.J., 2015. A nonlinear approach to regional flood frequency analysis using projection pursuit regression. Journal of Hydrometeorology. 16(4), 1561-1574.

Ergen, K., Kentel, E., 2016. An integrated map correlation method and multiple-source sites drainage-area ratio method for estimating streamflows at ungauged catchments: A case study of the western Black Sea region, Turkey. J. Environ. Manage. 166, 309–320.

Fennessey, N., Vogel, R.M., 1990. Regional flow-duration curves for ungauged sites in Massachusetts. J. Water Resour. Plann. Manage. 116, 530-549.

Fill, H., D., Stedinger, J., R., 1998. Using regional regression within index flood procedures and an empirical Bayes estimator. J. Hydrol. 210, 128-145.

Herrera, E., 2008. Développement d'une méthodologie hydrologique/statistique pour estimer les débits d'étiage au Québec habité. Québec, Université du Québec, Institut national de la recherche scientifique. 296 pp.

Hughes, D., Smakhtin, V., 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. Hydrological Sciences Journal. 41, 851-871.

Kendall, M. G., 1975. Rank Correlation Methods. Griffin, London.

Laaha, G., Blöschl, G., 2006. A comparison of low flow regionalisation methods–catchment grouping. Journal of Hydrology. 323, 193-214.

Laio, F., Di Baldassarre, G., Montanari, A., 2009. Model selection techniques for the frequency analysis of hydrological extremes. Water Resour. Res., 45, W07416.

Mendicino, G., and Senatore, A., 2013. Evaluation of parametric and statistical approaches for the regionalization of flow duration curves in intermittent regimes. J. Hydrol., 480, 19–32.

Ouali, D., Chebana, F., Ouarda, T.B.M.J., 2017. Fully nonlinear statistical and machine learning approaches for hydrological frequency estimation at ungauged sites. Journal of Advances in Modeling Earth Systems (accepted for publication).

Ouarda T.B.M.J., 2016. Regional Flood Frequency Modeling. Chapter 77. In Handbook of Applied Hydrology (V.P. Singh, Editor). McGraw-Hill, New York.

Ouarda, T., Shu, C., 2009. Regional low-flow frequency analysis using single and ensemble artificial neural networks. Water Resour. Res. 45, W11428.

Ouarda, T. B. M. J., Jourdain, V., Gignac, N., Gingras, H., Herrera, E., Bobée, B., 2005. Development of a hydrological model for the regional estimation of low-flows in the province of Quebec (in French). Res. Rep. R-684-f1, 174 pp., Eau, Terre, et Environ., Inst. Nat. de Rech. Sci., Sainte-Foy, Que., Canada.

Ouarda, T.B.M.J, Charron, C., St-Hilaire, A., 2008. Statistical models and the estimation of low flows. Canadian Water Resources Journal. 33(2), 195-206.

Ouarda, T.B.M.J., Charron, C., Martel, B. 2010. User guide for the software UNGAUGED for daily streamflow estimation in the Great Lakes System. Report for the International Joint Commission (IJC), 21 pp.

Patil, S., Stieglitz, M., 2012. Controls on hydrologic similarity: Role of nearby gauged catchments for prediction at an ungauged catchment. Hydrol. Earth Syst. Sci. 16, 551–562.

Pyrce, R., 2004. Hydrological low flow indices and their uses. Watershed Science Centre, (WSC) Report.

Reilly, C.F., Kroll, C.N., 2003. Estimation of 7-day, 10-year low-streamflow statistics using baseflow correlation. Water Resour. Res. 39(9), 1236.

Requena, A.I, Ouarda, T., Chebana, F., 2017. Flood frequency analysis at ungauged sites based on regionally estimated streamflows. Journal of Hydrometeorology 18(9), 2521-2539.

Salinas, J., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., et al., 2013. Comparative assessment of predictions in ungauged basins–Part 2: Flood and low flow studies. Hydrology and Earth System Sciences. 17, 2637-2652.

Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics. 6, 461-464.

Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., Bobée, B., 2006. A parametric Bayesian combination of local and regional information in flood frequency analysis. Water Resources Research. 42, W11408, 1-21.

Shu, C., Ouarda, T.B., 2012. Improved methods for daily streamflow estimates at ungauged sites. Water Resour. Res. 48, W02523.

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

Stedinger, J.R., Thomas Jr, W.O., 1985. Low-flow frequency estimation using base-flow measurements. 85-95. US Geological Survey.

Tsakiris, G., Nalbantis, I., Cavadias, G., 2011. Regionalization of low flows based on canonical correlation analysis. Adv. Water Resour. 34, 865-872.

Viglione, A., Parajka, J., Rogger, M., Salinas, J., Laaha, G., Sivapalan, M., et al., 2013. Comparative assessment of predictions in ungauged basins-Part 3: Runoff signatures in Austria. Hydrology and Earth System Sciences. 17, 2263.

Vogel, R.M., Fennessey, N.M., 1994. Flow-duration curves. I: New interpretation and confidence intervals. J. Water Resour. Plann. Manage. 120, 485-504.

Vogel, R.M., Kroll, C.N., 1992. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. Water Resour. Res. 28, 2451-2458.

Wald, A., Wolfowitz, J., 1943. An exact test for randomness in the non-parametric case based on serial correlation. The Annals of Mathematical Statistics. 14, 378-388.

Table 1. Descriptive statistics of local low-flow quantiles and catchment descriptors.

Variable		Unit	Notation	Min	Mean	Max	Std. dev
30-day 5-year low-flow quantile	Winter	$m^3\ s^{-1}$	$Q_{5,30}$	0.00	26.27	369.00	55.52
7-day 2-year low-flow quantile			$Q_{2,7}$	0.00	28.71	406.00	62.18
7-day 10-year low-flow quantile			$Q_{10,7}$	0.00	22.68	341.00	50.75
30-day 5-year low-flow quantile	Summer	$m^3\ s^{-1}$	$Q_{5,30}$	0.01	70.44	1280.00	167.27
7-day 2-year low-flow quantile			$Q_{2,7}$	0.00	85.62	1560.00	203.89
7-day 10-year low-flow quantile			$Q_{10,7}$	0.00	58.91	1080.00	143.49
Catchment area		km^2	BV	0.69	5655.52	96600	11685.7
Fraction of the catchment controlled by lakes		%	PLAC	0.00	6.33	32.00	6.57
Fraction of the catchment occupied by forest		%	PFOR	6.50	85.78	100.00	15.97
Annual mean degree-days below 0°C		degree-day	DDBZ	920.60	1635.15	2963.10	529.29
Summer mean liquid precipitation		mm	PLME	306.00	464.51	664.00	77.40
Curve number		-	CN	21.00	45.08	78.20	12.56
Average number of days with mean temperature > 27°C		-	NJH27	0.80	12.28	36.60	7.57

Table 2. Jackknife assessment results for the specific and absolute low-flow quantiles $q_{95\%}$ and $Q_{95\%}$ for the sites used when estimating winter $Q_{T,7}$.

Quantile	Estimate from	NASH	RMSE	RRMSE (%)	BIAS	RBIAS (%)
$q_{95\%}$	Regional streamflows	0.988	0.00	4.49	0.00	0.97
$Q_{95\%}$	Regional streamflows	0.999	1.79	4.49	0.50	0.97
	Regional FDC	0.882	21.09	48.62	2.27	-7.78

Figure caption list

Fig. 1 Illustration of the paths adopted by the RSBFA approach and traditional low-flow regional frequency analysis techniques.

Fig. 2 Low-flow Q-Q plot for the RSBFA approach. Particular sites are marked.

Fig. 3 Error of the low-flow quantile regionally estimated by the RSBFA approach. Particular sites are marked.

Fig. 4 Relative error of the low-flow quantile regionally estimated by the RSBFA approach. Particular sites are marked.

Fig. 5 Assessment criteria results for the winter season.

Fig. 6 Assessment criteria results for the summer season.

Fig. 7 Example of Q-Q plots for the specific and absolute low-flow quantiles q95% and Q95%. Regional estimation obtained by considering the sites used when estimating winter $Q_{T,7}$.

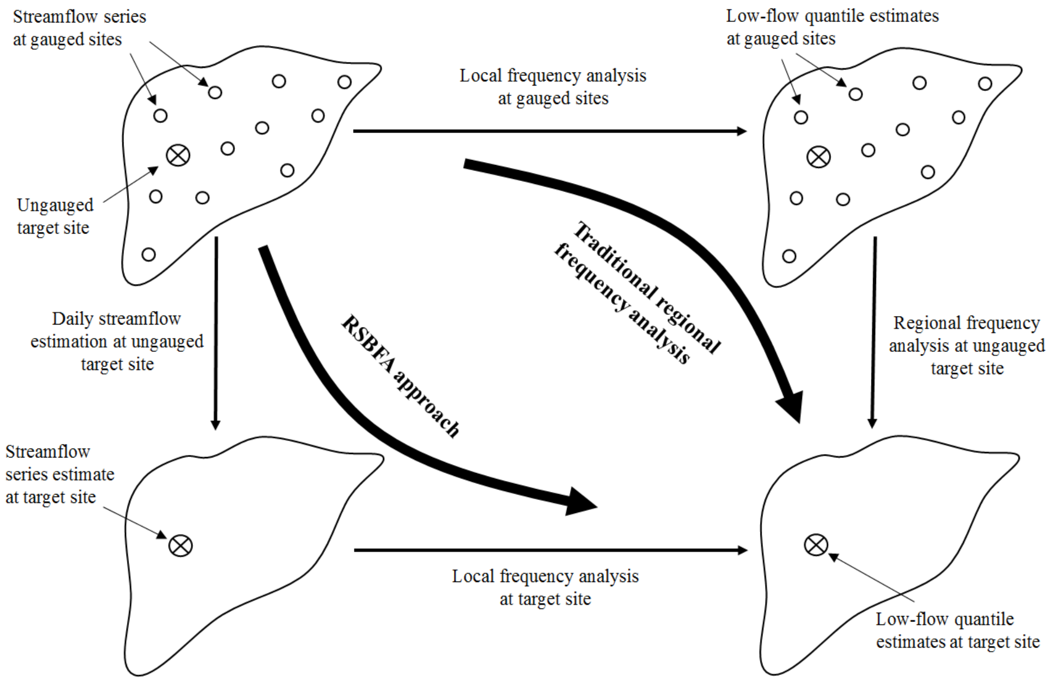


Fig. 1 Illustration of the paths adopted by the RSBFA approach and traditional low-flow regional frequency analysis techniques.

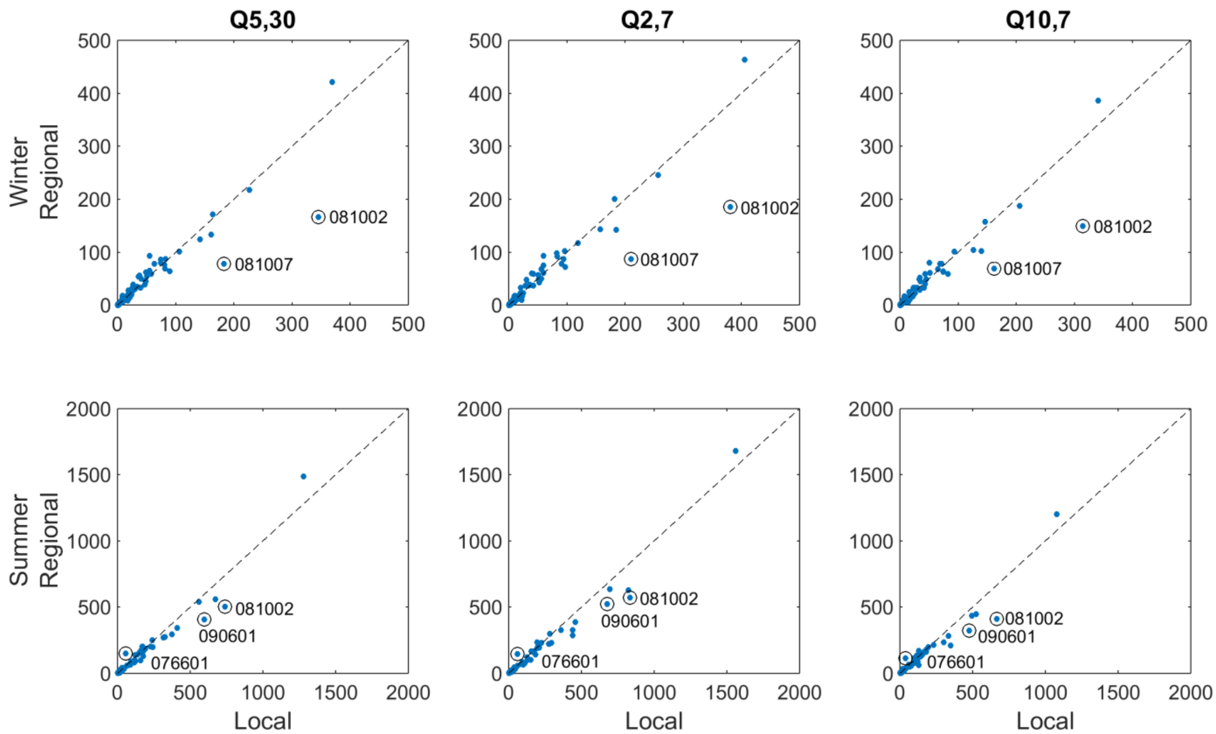


Fig. 2 Low-flow Q-Q plot for the RSBFA approach. Particular sites are marked.

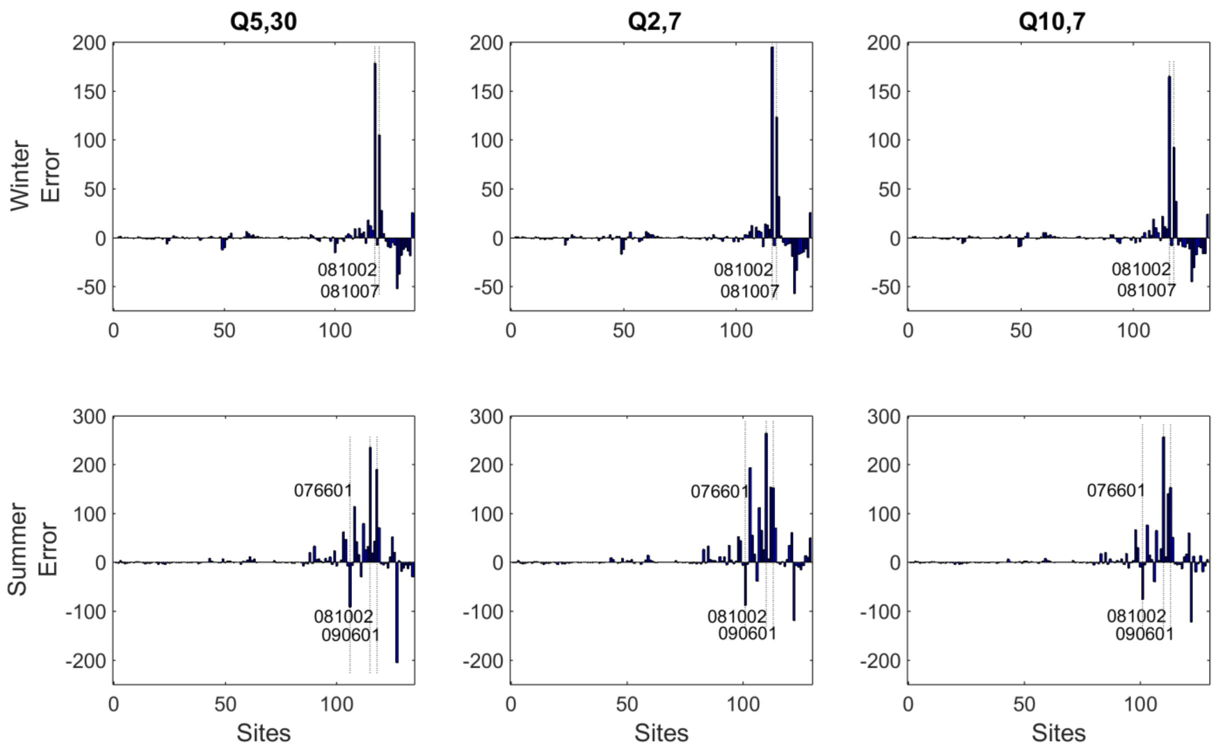


Fig. 3 Error of the low-flow quantile regionally estimated by the RSBFA approach. Particular sites are marked.

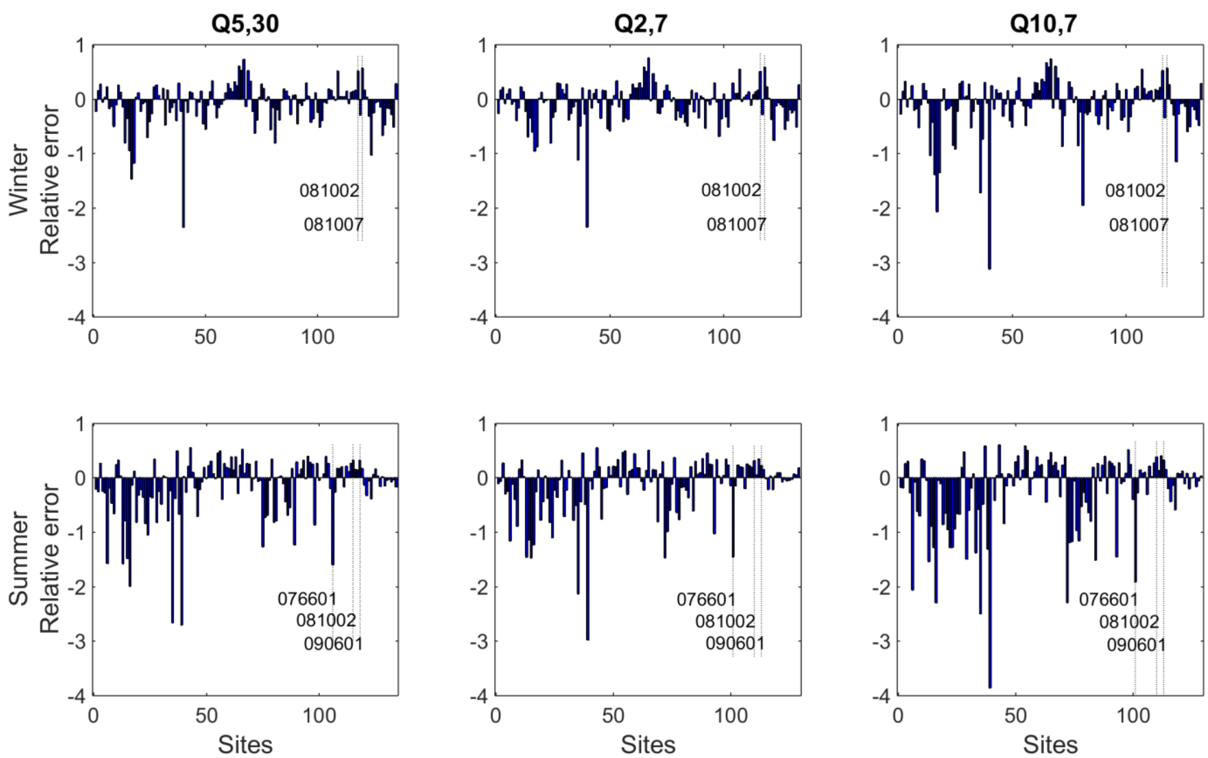


Fig. 4 Relative error of the low-flow quantile regionally estimated by the RSBFA approach. Particular sites are marked.

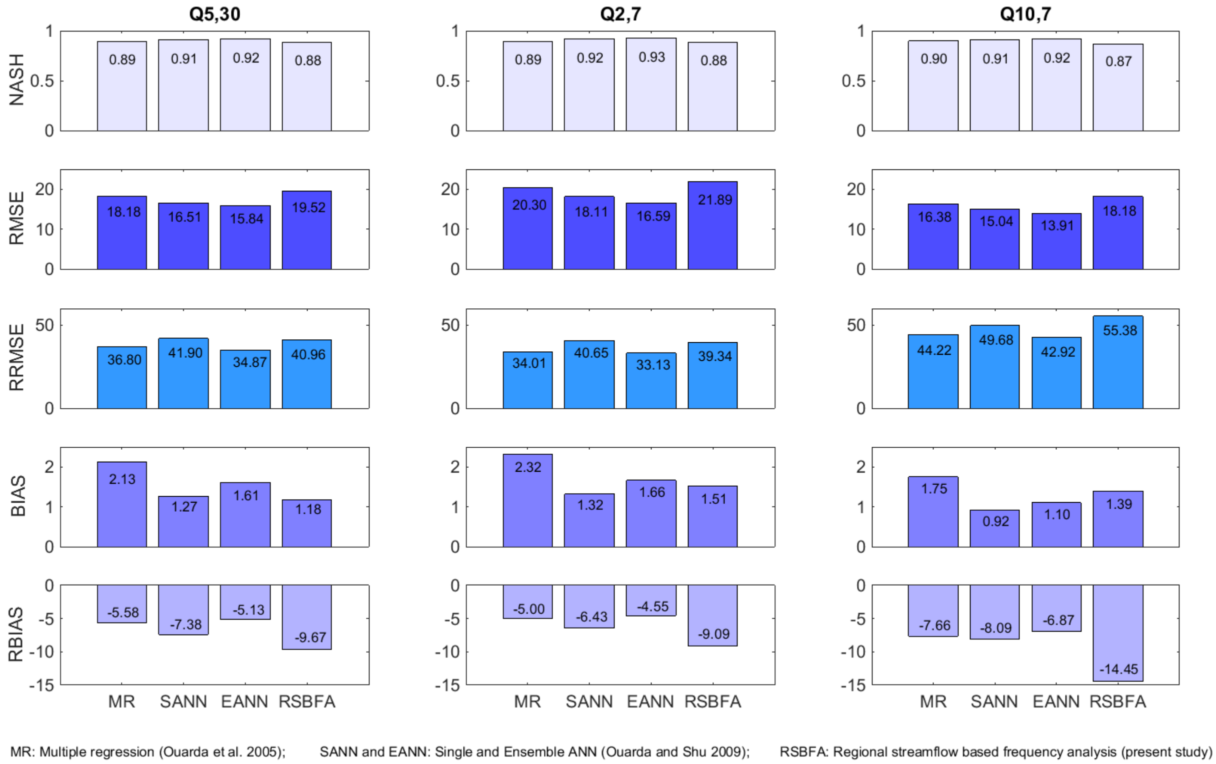


Fig. 5 Assessment criteria results for the winter season.

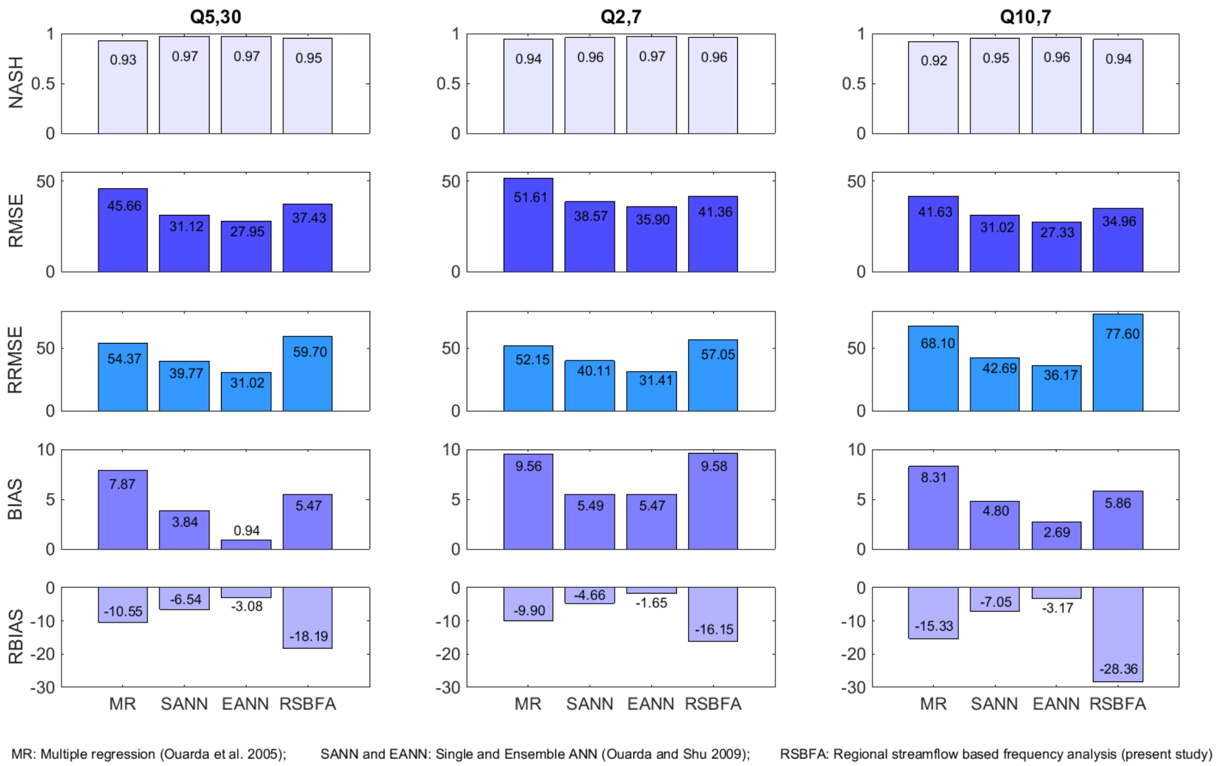


Fig. 6 Assessment criteria results for the summer season.

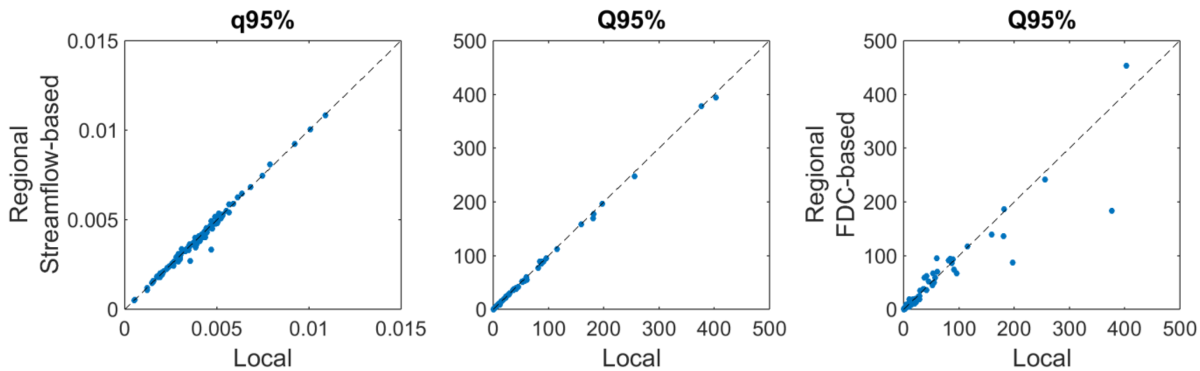


Fig. 7 Example of Q-Q plots for the specific and absolute low-flow quantiles $q_{95\%}$ and $Q_{95\%}$. Regional estimation obtained by considering the sites used when estimating winter $Q_{T,7}$.

Highlights

- The regional streamflow based frequency analysis approach is applied to low flows.
- Results are compared to traditional methods for a case study in Quebec, Canada.
- The performance and advantages of the proposed approach are discussed.
- The approach estimates different quantiles without redoing the regional procedure.