

GAGAM v1.2: An Improvement on Peak Labeling and Genomic Annotated Gene Activity Matrix Construction

Original

GAGAM v1.2: An Improvement on Peak Labeling and Genomic Annotated Gene Activity Matrix Construction / Martini, Lorenzo; Bardini, Roberta; Savino, Alessandro; DI CARLO, Stefano. - In: GENES. - ISSN 2073-4425. - ELETTRONICO. - 14:1(2023). [10.3390/genes14010115]

Availability:

This version is available at: 11583/2974377 since: 2023-01-05T16:34:05Z

Publisher:

MDPI

Published

DOI:10.3390/genes14010115

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)



Spatially-smooth regionalization of flow duration curves in non-pristine basins

Daniele Ganora, Francesco Laio, Alessandro Masoero, and Pierluigi Claps

Politecnico di Torino, Department of Environment, Land and Infrastructure Engineering, Torino, Italy

Correspondence to: Daniele Ganora (daniele.ganora@polito.it)

Published: 12 May 2016

Abstract. The flow duration curve (FDC) is a fundamental signature of the hydrological cycle to support water management strategies. Despite many studies on this topic, its estimation in ungauged basins is still a relevant issue as the FDC is controlled by different types of processes at different time-space scales, thus resulting quite sensitive to the specific case study.

In this work, a regional spatially-smooth procedure to evaluate the annual FDC in ungauged basins is proposed, based on the estimation of the L -moments (mean, L -CV and L -skewness) through regression models valid for the whole case study area. In this approach, homogeneous regions are no longer required and the L -moments are allowed to continuously vary along the river network, thus providing a final FDC smoothly evolving for different locations on the river. Regressions are based on a set of topographic, climatic, land use and vegetation descriptors at the basin scale. Moreover, the model ensures that the mean annual runoff is preserved at the river confluences, i.e. the sum of annual flows of the upstream reaches is equal to the predicted annual downstream flow.

The proposed model is adapted to incorporate different “sub-models” to account for local information within the regional framework, where man-induced alterations are known, as common in non-pristine catchments. In particular, we propose a module to consider the impact of existing/designed water withdrawals on the L -moments of the FDC.

The procedure has been applied to a dataset of daily observation of about 120 gauged basins on the upper Po river basin in North-Western Italy.

1 Introduction

Flow duration curves (FDC) are widely used tools to represent water availability in a river basin and are thus considered for many water resources planning and management purposes, like concessions for water uses and the planning of new hydropower plants. The FDC represents the percentage of time a certain value of discharge (usually at the daily scale) is equaled or exceeded in a river section over a specified a period of time. A FDC computed for a single year is called “annual”; if the observations of multiple years are merged together, the FDC is usually referred to as “period-of-record” or “total”. The empirical FDC can be easily built up by plotting the sorted the observations versus their frequency of non-exceedance computed with the Weibull plotting position, although the FDC is more frequently represented with respect to the exceedance frequency, thus resulting as a de-

creasing function. The evaluation of the FDC in ungauged basins is still a major issue in hydrological modeling, despite a large body of literature available on this topic, as recently reviewed by Castellarin et al. (2013).

In this work, we develop a regional model for prediction of the FDC in ungauged basins, developed and applied in the upper Po river basin (in North-Western Italy, an area mainly characterized by alpine and piedmont environments). The statistical framework developed is able to take advantage of local information about some types of anthropic effects. The aim of the work is to provide a regional tool to estimate the mean annual FDC in a generic watershed based on morphometric and climatic descriptors.

2 Methods

The regional spatially-smooth (SS) statistical estimation method proposed in this paper is based on a framework developed by Laio et al. (2011) in the context of regional flood frequency analysis. Analogies between the procedures are obtained by representing the mean annual FDC by means of its L -moments, i.e. distribution-free statistics which describe the mean, the variability and the skewness of the empirical FDC curve. L -moments are linear combinations of order statistics and are widely used in statistical hydrology; here we will refer to as μ for the mean value, while the dimensionless coefficients τ (L -CV) and τ_3 (L -skewness) respectively describe the variability and the skewness of the curve (see e.g., Hosking and Wallis, 1997, and Grimaldi et al., 2011, for more details on L -moments).

In the SS method the L -moments are the variables to be regionalized, i.e. the variables that will be predicted in ungauged basins. This is not so common in the literature as most of the models use as regionalization variables directly the FDC quantiles or the parameters of an analytical function fitted to the FDC. Then, the complete FDC can be calculated by fitting a probability distribution function to the set of predicted regional L -moments. In this way, the errors introduced by the preliminary choice of a distribution to fit the empirical curve do not enter in the regionalization procedure. The issue of the choice of a suitable distribution to represent the estimated FDC represents the last step in the regionalization process and will be discussed in the results section.

The present approach is referred to as spatially-smooth because it does not require the delineation of homogeneous regions, i.e. groups of basins sharing the same statistical characteristics (and thus the same L -moments τ and τ_3 within the region). Rather, L -moments are allowed to continuously vary in the space, so the FDC can vary accordingly. Following Wagener et al. (2013), the SS approach belongs to the “mapping function” paradigms in which the set of relationship between the L -moments and the descriptors (and not the L -moments themselves) remain valid for the whole area of interest.

Ultimately, this approach can be considered an extension of the index-flow approach, as the dimensionless FDC is not constant in a region but is allowed to change site by site.

To build the regional model, the sample L -moments of the mean annual FDC must be computed at each available gauging station; estimation in ungauged sites is then obtained by building multiple linear regression models based on some basins descriptors (i.e. morphologic, climatic, geological, etc characteristics of the catchments), also known as descriptors. For a generic variable y to be regionalized, two different model structures have been considered:

$$y = a_0 + a_1 \cdot x_1 + \dots + a_d \cdot x_d + \varepsilon \quad (1)$$

$$y = a_0 \cdot x_1^{a_1} \cdot \dots \cdot x_d^{a_d} \cdot \varepsilon \quad (2)$$

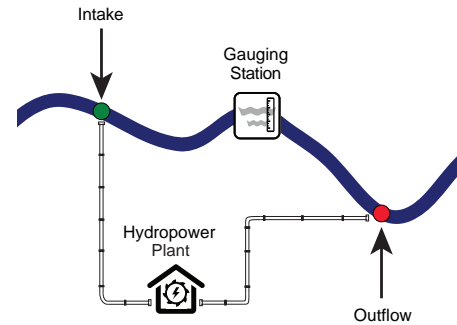


Figure 1. Example of gauging station bypassed by a water flow.

where x_d represents the generic basin descriptor, a_d are the coefficients to be estimated, and ε is the residual to be minimized. Equation (1) can be solved by the Least Squares method (LS), while Eq. (2) can be still solved within the LS framework after the log-transformation of both side of the equation. After the selection of a set of suitable regression models, the prediction of μ , τ and τ_3 in a generic ungauged basin can be performed.

To provide reliable results, observed data should be natural flow observations, i.e. with no alterations due to upstream water uses. However, many gauging stations are located on rivers affected by man-induced alterations: in such cases, the observations should be previously processed to provide “naturalized” values. Where the actual amount of daily derived flow is known, corrections are straightforward, but this information is seldom available. This is the case (see Fig. 1) of run-on-river power plants where the gauging station is located between the intake and the outflow of the system and no information are available about the actual flow derived by the plant.

To overcome this problem, and exploiting also the data relative to such stations, a new methodology, developed by Ganora et al. (2013), has been used to obtain the L -moments of the “natural”. In essence, the L -moments computed on the measurements affected by changes (i.e., downstream the intake) are converted to a set of naturalized L -moments subsequently used in the analysis. The method only requires the maximum discharge that can be withdrawn (said ΔQ) and the affected (observed) L -moments and does not need to assess the complete time series of the natural flow. Note that ΔQ is a known characteristic of the plants, available, in our case, from a regional database of water infrastructures.

The correction method uses the parameter:

$$K = \frac{\Delta Q}{\mu^D}, \quad (3)$$

where μ^D is the mean observed discharge (i.e. downstream the intake; superscript D) to correct the L -moments by means of the equations

Table 1. Observed and corrected L -moments for gauging stations affected by relevant water intakes.

Basin	K	μ obs.	μ corr.	τ obs.	τ corr.	τ_3 obs.	τ_3 corr.
Devero at Baceno	6.069	1.73	4.51	0.385	0.218	0.374	0.161
Germanasca at Perrero	0.637	3.14	4.6	0.708	0.528	0.627	0.465
Isorno at Pontetto	4	0.75	1.93	0.485	0.285	0.677	0.315
Po at Castiglione Torinese	2.016	54.57	117.5	0.585	0.371	0.607	0.33
S.Bernardino at S.Bernardino Santino	0.414	4.35	5.82	0.728	0.576	0.67	0.55
Stura di Demonte at Vinadio	3.053	2.62	5.45	0.482	0.291	0.536	0.258
Stura di Viù at Germagnano	0.734	4.77	7.26	0.596	0.435	0.568	0.406
Toce at Domodossola	3.391	12.21	29.34	0.375	0.224	0.469	0.226
Varaita at Torrette	8.824	1.36	3.2	0.383	0.212	0.541	0.221

$$\begin{cases} \mu^U = \mu^D \cdot \Psi[K] \\ \tau^U = \tau^D \cdot \frac{0.5}{L_2^*[K]} \\ \tau_3^U = \tau_3^D \cdot \frac{1}{3} \frac{L_2^*[K]}{L_3^*[K]} \end{cases} \quad (4)$$

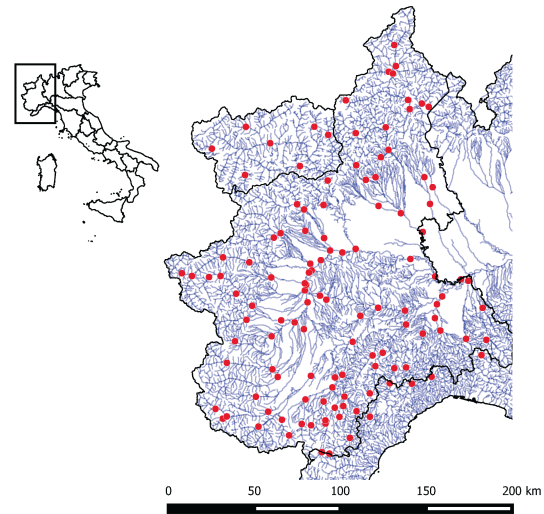
In Eq. (4) the superscript D indicates the downstream (observed) value, while U is the upstream (naturalized) statistic. Functions $\Psi[K]$, $L_2^*[K]$ and $L_3^*[K]$ give the correction factor for each L -moment and are reported in the Appendix A. Details about the correction method are discussed in Ganora et al. (2013).

3 Data availability

The available hydrological dataset refers to North-West of Italy (see Fig. 2) and includes 129 gauging stations with a total of 1438 station-year of daily observations, that have been selected after different quality controls. Only years with no more than 3 missing values have been included in the dataset. The available records have a mean length of 11 years, with actual durations ranging between 2 and 52 years. Many stations are quite recent and their data do not completely overlap with longer records, however they allow a larger spatial coverage of the model. More details about the dataset can be found in Ganora et al. (2013).

For each station, the annual FDCs have been computed based on the daily discharge values; the $N = 365$ values have been sorted and associated to the non-exceedance frequency $F_i = \frac{i}{N+1}$, with $i = 1 \dots N$ being the index of the sorted sample (leap years curves have been resampled on the frequency domain to obtain 365 values). Finally, the mean annual FDC at each station has been computed by taking the average of each annual FDC quantile (i.e., the average frequency-by-frequency of all the annual values).

The database encompasses a number of basins affected by different water withdrawals for different uses and with different intake systems. In many cases, for our purposes, such anthropic effects can be neglected (e.g., negligible intake with

**Figure 2.** Location of gauging stations used in the analysis.

respect to the average flow; high-elevation reservoir affecting only the upper part of the river and with negligible effect on the long-term statistics). In other cases alterations can be corrected with measurements of real intake (e.g. large irrigation canals). However, other cases require a correction of the observed streamflow statistics to provide reliable “natural” values to be used in the regional analysis (with the method described in the previous section). This is the case of 9 gauging stations located downstream the intake of run-of-the-river hydropower plants and upstream their water return. For these intakes no withdrawal data were available, as no gauging recording system was installed. Table 1 reports the main characteristics of the above mentioned stations with man-induced alterations, as well as the corrected (“naturalized”) L -moments.

4 Spatially-Smooth regional estimation

Sample L -moments, i.e. μ , τ and τ_3 , have been computed from the available average FDCs with the necessary corrections to obtain “naturalized” values in the 9 cases described

above. To support the implementation of the regional model, a number of catchment descriptors (about 120), including morphologic and climatic characteristics of the basin, soil and vegetation type, land use, etc., have been computed for each basin. The most relevant descriptors have been also reported in an “Atlas of basin characteristics” (Gallo et al., 2013). The descriptors act as independent variable in the regression models and some of them are expected to provide a statistical dependence with each L -moments.

Concerning the mean annual flow, to implement the regression model we considered the transformed variable

$$h = (\mu \cdot 31\,536)/A, \quad (5)$$

which represents the flow in mm when μ in $\text{m}^3 \text{s}^{-1}$ and A the catchment area in km^2 . Using h in the additive model structure (Eq. 1) and considering only descriptors that represents area-averaged values (e.g. mean basin elevation, spatial average of annual rainfall, etc.), the model guarantees the congruence of the mean discharge along the stream, i.e. the mean annual flow downstream of a confluence is equal to the sum of the two mean annual flow values computed upstream. This is a notable feature of the method that allows to keep the estimates of mean annual flow congruent at the confluences.

In order to select the most appropriate regional model, we performed all the possible regressions combining 1 to 4 descriptors as independent variables to estimate h ; each regression was tested for significance with the t Student test (5 % level of significance) and for multicollinearity with the VIF test (limit value equal to 5, see e.g. Montgomery et al., 2001). Models passing the tests were finally sorted according to their adjusted coefficient of determination and the best-performing ones were further evaluated by visual check of diagnostic error plots. Regressions have been solved using the Weighted Least Square method (e.g. Montgomery et al., 2001) considering the inverse of the square root of record length as weighting coefficients.

Concerning the L -moments τ and τ_3 , an analogous procedure has been performed but without requiring any congruence at the confluences, and thus without restricting the set of possible descriptors. In this case, both the additive and the multiplicative model structures have been considered, thus considering a larger number of possible models (about 10^6 models have been checked for each L -moment and each model structure).

The final models, applicable to the whole area of interests are:

$$h = -7.3605 \times 10^2 + 1.2527 \times \text{MAP} + 3.2569 \times 10^{-1} \times z_m + 5.2674 \times \text{fourier}_{B1} - 6.7185 \times \text{clc}_2 \quad (6)$$

$$\begin{aligned} \tau = & -2.896 \times 10^{-1} - 2.688 \times 10^{-3} \times \text{clc}_3 + 9.643 \\ & \times 10^{-5} \times a_{75} + 1.688 \times 10^{-4} \times \text{MAP} + 2.941 \\ & \times 10 \times c_{\text{int}}, \end{aligned} \quad (7)$$

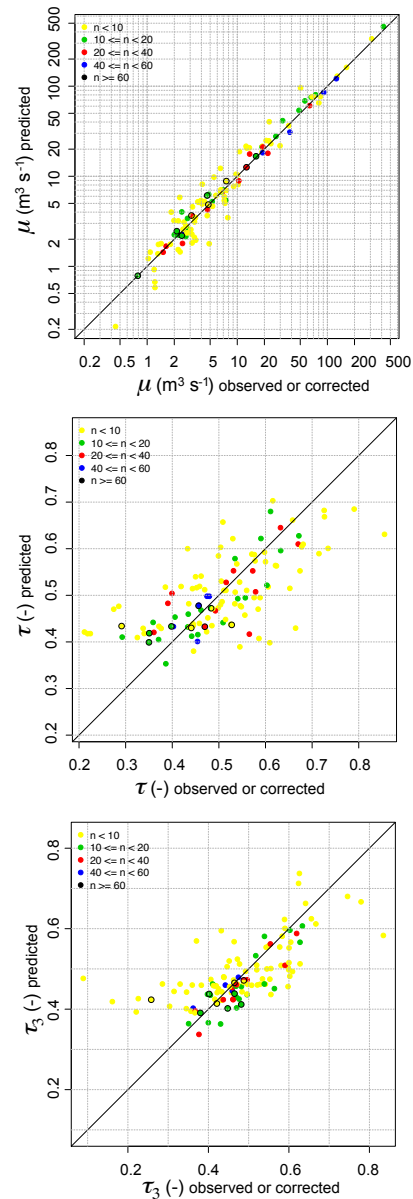


Figure 3. Natural values of the L -moments (observed or corrected) compared with regional predictions. Colors represent record lengths; circled points are the corrected values.

$$\tau_3 = 4.755 \times z_{\text{max}}^{-0.2702} \times \text{SD}(\text{IDF}_a)^{0.06869} \times \text{CV}_{\text{rp}}^{0.2106},$$

where MAP is the mean annual precipitation, z_m and z_{max} are the mean and maximum basin elevation, fourier_{B1} and CV_{rp} are rainfall regime parameters, clc_2 and clc_3 are land use parameters, a_{75} is the 75th percentile of the hypsographic curve, c_{int} and $\text{SD}(\text{IDF}_a)$ are extreme-rainfall statistics. All the parameters are computed at the basin scale; details are reported in Gallo et al. (2013).

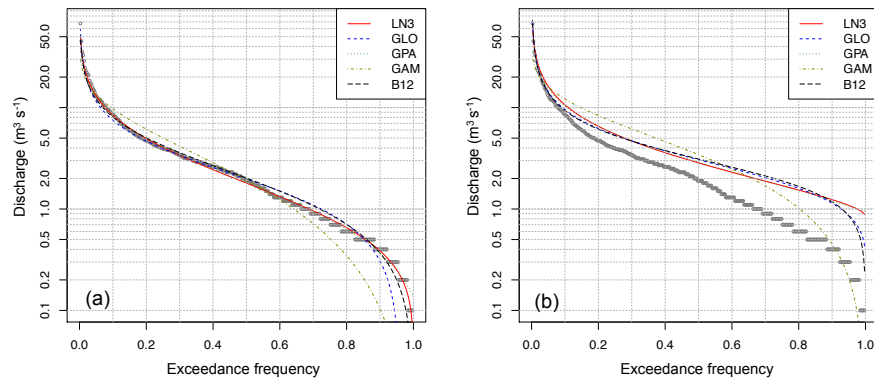


Figure 4. Empirical (points) and analytical (lines) flow duration curve for the Agogna at Momo river: (a) fitting on sample L -moments; (b) fitting on regional L -moments. Probability distributions are: log-Normal (LN3), Generalized Logistic (GLO), Generalized Pareto (GPA), Gamma (GAM) and Burr XII (B12).

Performances of the selected regression models can be examined in Fig. 3 where predictions in cross-validation are reported. The mean annual discharge prediction presents little scatter, while L -CV and L -skewness tend to be more dispersed. Worse performances in higher-order L -moments prediction were in fact expected, but it is worth noting that the points with larger errors are in general characterized by shorter time series and thus by larger sample uncertainty. Further diagnostics on residuals (not reported) have been performed to check possible heteroskedasticity and residual correlation.

5 Selection of FDC analytical form

The regional statistical framework presented is related to the estimation of the L -moments, similarly to Laio et al. (2011). The L -moments can be used to fit different probability distribution functions to represent the FDC and the choice of the analytical form of the FDC can be done as the last step of the procedure. To suggest a suitable distribution to represent the FDC in ungauged basins, different distributions have been investigated, all with 3 parameters: the log-Normal (LN3) the Generalized Logistic (GLO), the Generalized Pareto (GPA) and the Gamma (GAM) (details can be found in Hosking and Wallis, 1997), and the Burr XII (also called B12) described, for example, in Ganora and Laio (2015).

A first series of test of fitting was performed with the more common distributions (LN3, GLO, GPA, GAM). The main drawback observed was that in many cases negative stream-flow values in the lower tail were obtained. This occurs when the distributions have a variable lower bound that can allow (depending on the parameter set) negative predictions for high exceedance frequencies. Further constraints to the parameters of the distribution can be added to keep the result consistent, but such controls seem hardly applicable in a regional context, for two reasons: first, the adjustments in general require numerical optimization that become difficult to

be implemented in large-scale models; second, it is in general preferable to avoid complex (and likely unstable) estimation methods when the model will be used by practitioners that may not have confidence with numerical techniques.

To overcome the problem of negative quantiles without optimizing the distribution's parameters, we introduce the use of the Burr probability distribution, well known in different scientific communities but rarely used in the hydrological field (Ganora and Laio, 2015). In particular, the proposed function is the Burr XII distribution, originally proposed by Burr (1942) as a two-parameter function, and later extended in the three-parameter form (cumulative frequency function):

$$F(x) = 1 - \left[1 - k \left(\frac{x}{\lambda} \right)^c \right]^{\frac{1}{k}} \quad k \neq 0$$

$$= 1 - \exp \left[- \left(\frac{x}{\lambda} \right)^c \right] \quad k = 0, \quad (8)$$

where k and c are the shape parameters and λ is the scale parameter. The parameter k can be any real number ($-\infty < k < \infty$), while $c > -k$ and $\lambda > 0$. The domain is always positive, being $0 \leq x \leq \infty$ for $k \leq 0$ and $0 \leq x \leq \lambda k^{-1/c}$ otherwise; for this kind of application we considered only non-positive values of k in order to use only unbounded (from above) curves. The condition $k = 0$ is the limiting case for which $F(x)$ becomes the two-parameter Weibull distribution. Another limiting case appears for $k \rightarrow -\infty$ (with $c \rightarrow \infty$ simultaneously) and lead to the Pareto distribution:

$$F_P(x) = 1 - \left(\frac{x_0}{x} \right)^\gamma \quad x > x_0$$

$$= 0 \quad \text{otherwise,} \quad (9)$$

where $-\infty/k = \gamma$ and $x_0 = \lambda(-k)^{-1/c}$.

The Burr function results quite “flexible”, and thus suitable to fit many different shapes of empirical data, thanks to the wide range of skewness and kurtosis values it covers; nevertheless it has only three parameters, thus avoiding over-fitting problems. On the other hand, despite the simple analytical form of the Burr XII, the estimation of the two shape

parameters is not straightforward as it requires the joint inversion of two nonlinear equations (not shown here). For details about the estimation of the parameters with the method of L -moments as well for other characteristics of the distribution the reader is referred to Ganora and Laio (2015) and references therein.

To proceed with the model selection, a comparison between the different distributions is in order. This comparison has been performed by fitting the analytical curves to 365 quantiles corresponding to the non-exceedance probabilities $F_i = i/366$ with $i = 1 \dots 365$. For each distribution, the difference between the empirical and the estimated quantiles has been calculated (on both untransformed and log-transformed discharge values). The RMSE provided a error statistic for each considered station.

Results show that LN3 and GPA have generally good fitting performances; GLO and B12 have slightly larger errors, but comparable with LN3 and GPA, while the GAM does not provide reliable results. However, the LN3 provides negative values in 21 % of the stations, the GLO in 51 % and the GPA in 11 %. The average duration (in days per year) affected by negative quantiles is 10, 14 and 19 days for the LN3, GLO and GPA respectively (the average is computed considering only the stations with at least one negative value). In the context of the data set analysed here the Burr distribution results the best choice, as it provides adequate fitting capabilities without the need of further optimization to avoid negative or inconsistent quantile predictions. An example of fitting is reported in Fig. 4a where the studied analytical functions have been superimposed on the empirical FDC.

Panel b of Fig. 4 shows the same record, but compared to the analytical FDCs computed from the regional L -moments. An overestimation of the mid-right part of the FDC is evident, while the left side of the curve seems properly recognized. In this case, although the shift could be due to an inaccurate estimation of the L -moments, it can also be caused by the presence of unknown (or underestimated) water intakes upstream the gauging station. Under this perspective, the regional model can be also used as a diagnostic tool to identify basins in which an improvement of the knowledge on the actual water use is recommended. This kind of analysis is currently under development.

6 Conclusions

A regional model for the estimation of the mean annual flow duration curve (FDC) has been presented in the paper. The model is made of three regression-based equations to estimate the L -moments (i.e., the mean, the L -CV and the L -skewness) of the FDC at any ungauged basin, which can be subsequently used to fit the whole curve. In particular, concerning the mean annual flow, the river network structure is explicitly accounted for: at any confluence, the mean annual flow in the downstream section is the sum of the mean values in the two upstream river reaches, thus ensuring the congruence of the prediction.

Moreover, the method allows the use in the calibration phase of records affected by water derivations, even if the actual derived flow is not known or hardly retrievable. This is possible thanks to a parsimonious corrections method that provide “naturalized” L -moments considering only simple “geometrical” information about the water intake and does not require the reconstruction of the whole natural time series.

Finally, among different analytical representations of the FDC, the three-parameter Burr XII distribution has been proved to be the most suitable for the case study. In fact, while it provides good fitting performances, it does not allow the estimation of negative discharge values, which frequently appear by applying other analytical forms. This is advantageous as it is not necessary to optimize the parameters of a distribution to force the predicted quantiles to positive values only, thus making the model suitable for large-scale unsupervised applications.

Due to its flexibility, the present modeling framework is easily adaptable to further extensions like, for instance, the inclusion of observations from new gauging stations (by combining local and regional L -moments), while its use as a diagnostic tools to investigate the actual effect of existing water derivations is already in progress.

Appendix A

The correction method developed by Ganora et al. (2013) is based on the hypothesis that the upstream FDC L -moments can be obtained by correcting the observed ones through Eq. (4) in a parsimonious way, i.e. using just simple information about the intake system. This is necessary when scarce information are available about the plant, while more sophisticated models could be adopted where the available information is more detailed.

The method assumes the intake flow is exactly corresponding to the actual available flow below the threshold ΔQ . In this work we also consider the plant is always in exercise, even though the original method allows to account for a number of days of stop during the year. Correction coefficients are obtained by studying the effect of water withdrawal on a virtual FDC defined by an exponential distribution; in this way the problem can be treated analytically and solutions are based on a single parameter.

Applicability of correction factors has been tested with real data (i.e. also with non-exponential FDC) showing that μ and τ can be very efficiently reconstructed while, as expected, corrections for τ_3 provide larger errors, but still useful for the scope of the analysis.

To allow the implementation of Eq. (4), from Ganora et al. (2013), we define:

$d^* = \text{Exp}[W[K]]$ where $W[K]$ is the Lambert W -function,

i.e. the inverse of equation $K = We^W$;

$\psi[K] = K / W[K]$;

$L_2 = \psi[K] \cdot S_2[d^*] + k \cdot (d^{*2} - d^*)$;

$L_3 = \psi[K] \cdot S_3[d^*] + k \cdot (-2d^{*3} + 3d^{*2} - d^*)$;

$S_2[d^*] = \frac{1}{2}d^* (2 - d^* + 2(d^* - 1) \cdot \text{Ln}[d^*])$;

$S_3[d^*] = \frac{1}{6}d^* (6 + d^* (4d^* - 9) - 6(d^* - 1)(2d^* - 1) \cdot \text{Ln}[d^*])$.

Acknowledgements. Funding from the Italian Ministry of Education, Universities and Research (FIRB-RBFR12BA3Y project), and from the European Union – Horizon 2020 ERC-2014-CoG (“CWASI” project, ref. 647473) are acknowledged.

References

- Burr, I. W.: Cumulative frequency functions, *Ann. Math. Stat.*, 13, 215–232, 1942.
- 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., and Vogel, R. M.: Prediction of flow duration curves in ungauged basins, in: *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, edited by: Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., Cambridge University Press, 2013.
- Gallo, E., Ganora, D., Laio, F., Masoero, A., and Claps, P.: Atlante dei bacini imbriferi piemontesi ISBN:978-88-96046-06-7, available at: http://www.idrologia.polito.it/web2/open-data/Renerfor/atlante_bacini_piemontesi_LR.pdf (last access: 4 April 2016), 2013 (in Italian).
- Ganora, D. and Laio, F.: Hydrological Applications of the Burr Distribution: Practical Method for Parameter Estimation, *J. Hydrol. Eng.*, 20, 11, doi:10.1061/(ASCE)HE.1943-5584.0001203, 2015.
- Ganora, D., Gallo, E., Laio, F., Masoero, A., and Claps, P.: Analisi idrologiche e valutazioni del potenziale idroelettrico dei bacini piemontesi ISBN:978-88-96046-07-4, available at: http://www.idrologia.polito.it/web2/open-data/Renerfor/analisi_idrologiche_LR.pdf (last access: 4 April 2016), 2013 (in Italian).
- Grimaldi, S., Kao, S. C., Castellarin, A., Papalexiou, S. M., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A.: Statistical hydrology, in: *Treatise on Water Science, Vol. 2: The Science of Hydrology*, edited by: Wilderer, P., Chap. 2.18, 479–517, Elsevier, doi:10.1016/B978-0-444-53199-5.00046-4, 2011.
- Hosking, J. and Wallis, J.: *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, 1997.
- Laio, F., Ganora, D., Claps, P., and Galeati, G.: Spatially smooth regional estimation of the flood frequency curve (with uncertainty), *J. Hydrol.*, 408, 67–77, doi:10.1016/j.jhydrol.2011.07.022, 2011.
- Montgomery, D., Peck, E., and Vining, G.: *Introduction to linear regression analysis*, in: *Regional flood frequency*, edited by: Ouarda, T., Girard, C., Cavadias, G., and Bobee, B., Wiley Series Probability and Statistics, 3rd Edn., Wiley, New York, 2001.
- Wagener, T., Bloeschl, G., Goodrich, D. C., Gupta, H. V., Sivapalan, M., Tachilawa, Y., Troch, P. A., and Weiler, M.: A synthesis framework for runoff prediction in ungauged basins, in: *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, edited by: Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., Cambridge University Press, 2013.