WATER RESOURCES MANAGEMENT IN SCARCITY CONDITIONS: FDC VARIABILITY ANALYSIS IMPLICATION ON ENVIRONMENTAL STREAMFLOW REQUIREMENTS

A. LONGOBARDI and P. VILLANI

1 Department of Civil Engineering, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: alongobardi@unisa.it

EXTENDED ABSTRACT

The management of water resources, more and more limited and poor in quality, represents a present key issue in hydrology. The development of a community is highly related to the management of available water resources and there is a need, for this reason, to rationalize them, to plan their conscious use and to preserve their quality and related wildlife. Environmental streamflow requirements represent then a particular aspect of water resources management.

Environmental flow has been defined in many different ways but, on a general perspective, it can be defined as the river discharge that allows the sustainable maintenance of a river environment and, therefore, should allow protection of the natural river ecosystem. The scientific environmental flow estimating methodologies, perhaps different from country to country, are mainly based on analysis of historical flow data or on investigation of relationships between a particular aspect of the habitat and river discharge. It is possible to distinguish between theoretical and experimental methodologies. Examples of theoretical methods are the so known hydrological methods that estimate environmental flow as a function of catchment properties, such as drainage area, mean monthly discharge, mean annual discharge and so on. Among hydrological methods, statistical methods are based on the FDC flow duration curve estimation. Environmental flow is then defined as the discharge for a particular duration; generally large duration values are considered, corresponding to the lower tail of the FDC.

FDCs have a broad use in hydrological and engineering application, which has been sometimes criticized because their representation, and therefore interpretation, tightly depends on the algorithm used to compute it. It is indeed possible to compute the FDC for the whole period of observation but it is also possible to derive a mean annual FDC. If n is the number of recorded years, it is possible to consider n different FDCs, each corresponding to a particular year. It is rather intuitive to imagine that the n FDCs would have different shapes between them, due to the well-known climate variability from year to year. This variability, in extreme climatic conditions makes questionable the definition of an average FDC. The empirical FDCs uncertainty analysis inevitably produces an uncertainty in the definition of the environmental streamflow requirements, which estimation should embed the considered elements.

An application will be presented for a river basin located in a particular region of Southern Italy, experiencing a typical Mediterranean climate.

Keywords: water resources and river basin management, flow duration curve, environmental streamflow requirement, Mediterranean area.

1. INTRODUCTION

Environmental flow has been defined in many different ways but on a general perspective it can be defined as a river discharge that allows the sustainable maintenance of a river environment and therefore should allow protection of the natural river ecosystem. The scientific environmental flow estimating methodology differs from country to country and
is mainly based on analysis of historical flow data or on investigation of relationships between a particular aspect of the habitat and river discharge. It is possible to distinguish between theoretical and experimental methodologies. Some of the most simple methods relates it to a percentage of mean annual runoff (Tennant, 1976), some other include more physical and climate properties, such as the drainage area, percentage of pervious area, mean monthly discharge, annual rainfall, average elevation (Kobold and Brilly, 1994; Castellarin et al., 2007; Longobardi and Villani, in review). In this way the streamflow variability is however not considered and no information can be given about the frequency of occurrences. Some other methods, the so called theoretical-hydrological, environmental streamflow requirements indices can be derived for empirical FDC, as a discharge value for a particular duration. FDCs are variable on an annual base, because of the year to year climate variability and the use of indicators relying on their calculation also allows the possibility to assess how important is the impact climate conditions on environmental streamflow requirements. To measure the frequency of the occurrences of particular low flow values, it is possible to perform a frequency analysis, where a particular probability distribution is fitted to the sample data and then use, as an example, to associate a return period to a particular low flow value. Gottschalk et al. (2013) indicate that most of the widely used distributions are the Gumbel, Fréchet and Weibull as well as particular cases of the Generalised Extreme Value (GEV) distribution. But to predict the magnitude that a particular low flow value can assume, regression analysis can be performed where year by year, or for streamflow value of a particular return period, the low flow statistics is related to physical and catchment properties also frequently placed within a regionalization context (Eng and Milly, 2007; Mohaomud, 2008; Longobardi and Villani, 2008; Cheng et al., 2012).

An application is presented in the following for a river basin located in a particular region of Southern Italy, experiencing a typical Mediterranean climate.

2. THE CASE STUDY

The studied catchment is the Tanagro at Polla gauging station, a 660 km² drainage area catchment located in Southern Italy. The available data is a 65 years of natural daily streamflow time series. According to the Thornthwaite index, on an annual base, it experiences a humid climate, with a mean annual precipitation of about 1243 mm and a mean annual potential evapotranspiration of about 700 mm.

![Figure 1. Climate characterization (left side) and annual flow duration curve envelope for the Tanagro@Polla (mean annual FDC is red marked).](image)

Average monthly precipitation plot is showed in Figure 1 – right side. The minimum amount of precipitation is observed from June to September and the maximum from November to February. Inversely, average monthly potential evapotranspiration is maximum during the period from June to September and minimum during the period from November to February. The climate is then typically Mediterranean, with very marked dry
and wet seasons recognizable within the year. The Tanagro river basin is a moderately permeable catchment where about the 45% of total streamflow is the outcome of deep groundwater systems. Its geological properties and the large seasonal rainfall variability are perhaps responsible of the Tanagro river basin large variability of the hydrological regimes, with a quite steep streamflow duration curve (Figure 1 – left side).

3. Q 347 FREQUENCY ANALYSIS

In this first section, a frequency analysis application is given for the variable Q347, assumed as the indicator for the minimum environmental flow. Once annual FDCs have been computed for the observed discharge time series, the discharge value corresponding to the duration of 347 days has been selected for each FDCs, and a frequency analysis has been performed for the sample vector data Q347 (Figure 2).

![Figure 2. Daily discharge values for the Tanagro@Poll (blu line) and annual Q347 occurrences (red circle).](image)

The empirical Q347 distribution has then been fitted with reference to some of the most used frequency distribution in hydrological applications, the Normal (1) and the Generalized Extreme Value GEV distribution (2):

\[
F(x; \mu, \sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx 
\]  

(1)

\[
F(x; \mu; \sigma; \xi) = \exp\left\{-\left[1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} 
\]  

(2)

The choice for the Normal distribution lies in the fact that the hypothesis of normality is generally in regression analysis, later performed. The Jarque Brera and the \(\chi^2\) tests will be applied to verify the normality hypothesis. The choice instead for the GEV distribution lies in the fact that the Generalised Extreme Value distribution are the most suited for extreme values analysis (Gottschalk et al., 2013). N and GEV probability distribution parameters estimates, along with the probability density and cumulative functions fittings, are given in Figure 2. The GEV shape parameter \(\xi\) estimates is negative and the GEV distribution correspond in this case with the type III extreme value distribution or the Weibull distribution (3):

\[
F(x; \mu; \sigma; \xi) = \begin{cases} 
1 & x \geq \mu \\
\exp\left\{-\left[1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & x < \mu 
\end{cases} 
\]  

(3)
Visual inspection through Normal and Weibull probability plots (Figure 4) and hypothesis tests performances, indicate the goodness of fit of the mentioned distributions. Both the Jarque Brera, which critical value has been calculated with an intensive Monte Carlo simulation given the small sample size, and the \( \chi^2 \) tests accepted the null hypothesis that the sample data Q347 comes from a Normal distribution, at the 5% significance level. At the same level of significance the Kolmogorov-Smirnov test rejected instead the null hypothesis that the sample data Q347 come either from a Normal nor for a Weibull distribution, for given parameters.

Fitting are further compared in terms of main statistics and key values frequency occurrences. Main statistics for the empirical, Normal and GEV probability distribution are given in Table 1. The sample average Q347 of about 2.14 mc/s, exceeded for the 48% record in the sample data, has the same return period for both N and GEV distributions. Maximum Q347 observed value of about 4.5 mc/s, has a return period of about 100 years and 50 years, respectively for the N and GEV distribution.

**Figure 3.** Empirical and fitted pdf (left side) and cdf (right side) distribution for the sample vector data Q347.

**Figure 4.** Normal and Weibull probability plot for the sample vector data Q347.

**Table 1:** Main statistics for the empirical, Normal and GEV probability distributions.

<table>
<thead>
<tr>
<th>distribution</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std</th>
<th>skew</th>
<th>kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>1.194</td>
<td>4.617</td>
<td>2.464</td>
<td>0.695</td>
<td>0.520</td>
<td>3.163</td>
</tr>
<tr>
<td>Normal</td>
<td>0.549</td>
<td>4.932</td>
<td>2.433</td>
<td>0.680</td>
<td>0.0902</td>
<td>2.734</td>
</tr>
<tr>
<td>GEV</td>
<td>0.661</td>
<td>5.482</td>
<td>2.518</td>
<td>0.797</td>
<td>0.714</td>
<td>3.628</td>
</tr>
</tbody>
</table>
4. Q 347 REGRESSION ANALYSIS

In the following sections, the sample vector Q347 is explored in a regression analysis framework with the aim to identify a linear regression model, from simple to multiple dimension, to predict Q347 from independent variables. Main variables accounted for in the regression analysis are the annual precipitation (P), the annual temperature (T), the annual specific discharge (Qs) and the annual BFI index. As a particular discharge values, the annual Q347 would indeed depend on both climate and catchment properties. Among catchment properties, a major role is played by the BFI index, as demonstrated, among others, by the authors in an empirical study involving, at the regional scale, the currently investigated river basin (Longobardi and Villani, in review).

In a first attempt, index regression analysis, the Q347 variable is directly regressed against the mean climate and catchment properties. In a second attempt, the annual empirical FDC is considered as a distribution function, an analytical function is fitted to the empirical one and analytical function parameters are then regressed against the climate and catchment properties. Q347 values are then computed for each regressed annual FDC and compared to the observations.

4.1. Index regression analysis

Significant independent variables have been selected through a step-wise linear multiple regressive approach, which resulted in the specification of the following four linear regressive method to predict annual Q347 (Table 2).

Table 2: Simple and multiple regression models for annual Q347 prediction.

<table>
<thead>
<tr>
<th>Regression model</th>
<th>Model equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod 1</td>
<td>Q347 = +6.41*BFI - 0.356</td>
</tr>
<tr>
<td>Mod 2</td>
<td>Q347 = -0.009<em>BFI + 5.96</em>Qs + 0.234</td>
</tr>
<tr>
<td>Mod 3</td>
<td>Q347 = -0.0003<em>BFI - 0.019</em>Qs + 5.96*P + 0.246</td>
</tr>
<tr>
<td>Mod 4</td>
<td>Q347 = -0.032<em>BFI + 0.0002</em>Qs - 0.019<em>P + 5.84</em>T + 0.765</td>
</tr>
</tbody>
</table>

Models have been compared in terms of conventional statistics, such as the standard deviation (std), the bias, the root mean squared error (rmse), and the Pearson correlation coefficient (r2cor), corrected for the number of model parameters:

\[ r_{2cor} = 1 - (1 - r_{cor})^{\frac{k-1}{k-p}} \]  \hspace{1cm} (4)

where p is the number of model parameters and k is the sample length. Results are given in Table 3.

The same table also provides jackknife errors estimates derived from the application of the jackknife resampling procedure, to assess different sources of error such as mean total true error, mean apparent error, a measure of goodness of fit, and mean expected excess error, a measure of model robustness (Efron, 1982).

The four regression prediction models performances are comparable both in terms of conventional statistic and jackknife error estimates, which provide an average absolute error of about 20% on Q347 prediction.

Table 3: Regression prediction models performances (std, bias, rmse, r2cor), and jackknife errors estimates (mt, ma, me).

<table>
<thead>
<tr>
<th>Regression model</th>
<th>std</th>
<th>bias</th>
<th>rmse</th>
<th>r2cor</th>
<th>mt</th>
<th>ma</th>
<th>me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod 1</td>
<td>0.414</td>
<td>0.372</td>
<td>0.556</td>
<td>0.356</td>
<td>0.198</td>
<td>0.192</td>
<td>0.007</td>
</tr>
<tr>
<td>Mod 2</td>
<td>0.421</td>
<td>0.355</td>
<td>0.551</td>
<td>0.369</td>
<td>0.199</td>
<td>0.189</td>
<td>0.009</td>
</tr>
<tr>
<td>Mod 3</td>
<td>0.423</td>
<td>0.351</td>
<td>0.550</td>
<td>0.371</td>
<td>0.209</td>
<td>0.188</td>
<td>0.020</td>
</tr>
<tr>
<td>Mod 4</td>
<td>0.426</td>
<td>0.343</td>
<td>0.547</td>
<td>0.378</td>
<td>0.206</td>
<td>0.187</td>
<td>0.019</td>
</tr>
</tbody>
</table>
4.2. Analytical distribution parameters regression analysis

Longobardi and Villani (in review) have demonstrated that, for the region under investigation, annual FDCs can analytically be described through the use of a lognormal distribution. In fact, if the duration \( d \) is replaced by the frequency \( F \), expressed as \( F = d/(365+1) \), the three-parameter lognormal distribution can be adopted:

\[
    z = a \log(q - q_0) + b
\]

(4)

where \( q \) is the daily normalized streamflow for a given frequency (%), \( a, b \) and \( q_0 \) the model parameters and \( z \) is the normal reduced variate representing the frequency of the discharge. The parameter \( q_0 \) represents the streamflow distribution lower limit (%), whereas, if the normal standardized variate \( z \) is defined as:

\[
    z = \frac{\log(q - q_0) - \mu[\log(q - q_0)]}{\sigma[\log(q - q_0)]}
\]

(5)

the parameter \( a, b \) and \( q_0 \) correspond to:

\[
    a = \frac{1}{\sigma[\log(q - q_0)]} \quad \quad b = -\frac{\mu[\log(q - q_0)]}{\sigma[\log(q - q_0)]} \quad \quad q_0 = \mu(q) - \frac{\sigma(q)}{t}
\]

(6)

where \( t \) is a function of the third order momentum. Lognormal distribution parameters \( a, b \) and \( q_0 \) are estimated for each of the annual FDCs and simple to multiple regression models have been identified to predict each of the parameters through climate and catchment properties independent variables. A step-wise regression approach, as in the case of the quantile regression analysis, suggested that for both \( a, b \) and \( q_0 \), the better model performances are associated to a regressive model type 4, where all of the selected independent variables are accounted for. Regressed annual FDCs are derived and for each of them the Q347 is selected and compared to the observed value. Conventional statistics are summarized in Table 4.

<table>
<thead>
<tr>
<th>Regression model</th>
<th>std</th>
<th>bias</th>
<th>rmse</th>
<th>r2cor</th>
<th>mt</th>
<th>ma</th>
<th>me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod 4</td>
<td>0.499</td>
<td>0.304</td>
<td>0.584</td>
<td>0.420</td>
<td>0.244</td>
<td>0.219</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Regression model performances achieved in this second attempt, based on the identification of a multiple regressive model to predict the lognormal FDC parameters, are almost similar to the one resulted in the direct quantile regression analysis.

5. CONCLUSIONS

The presented paper has reported on different analysis concerning an indicator of the environmental streamflow requirements. This index is assumed to be the streamflow discharge for a particular duration, Q347, and it is calculated, on an annual base, from empirical annual FDC. FDCs are indeed characterized by an annual variability, which is mainly driven by climate conditions which differs from year to year: the Mediterranean climate, that is experienced by the studied catchment, is particularly extreme on this side. As a possibility to describe the occurrences frequency of the environmental streamflow requirements index, Q347 sample data has been approximated, for the case study, with a Normal and a Weibull probability distributions. The fitting of a probability distribution has enabled the possibility to associate a value of the return period to each particular value of annual Q347, and eventually to understand how severe are for the case study, the extreme Q347 statistics. As a possibility to identify a predictive model for Q347, different step-wise regression equations have been calibrated and assessed on the base of a jackknife validation technique. A regression analysis based on the sole index Q347 and a
regression analysis of the whole FDC have been performed. Independent variable assumed are the BFI, the annual specific discharge, the annual precipitation and the annual temperature. Regression models performance are almost similar, with an absolute relative error of about 20% and a moderate Pearson correlation coefficient of about 0.45. among all of the considered independent variable, the BFI is however the index that on its side is able to explain the 35% of the explained variance.

REFERENCES