Estimating mean monthly runoff at ungauged locations: an application to France
Eric Sauquet, Lars Gottschalk and Irina Krasovskaia

ABSTRACT

An approach for estimating mean monthly runoff at ungauged sites is presented. Special attention is paid to include effects of local features such as karst and river regulation by reservoirs. The developments introduced conform with hydrostochastic concepts in that simple physical and statistical laws are inherent in the methods used for mapping. Hence, the approach developed here is consistent with the water balance along the river network. The suggested method combines an application of empirical orthogonal functions and an adapted stochastic interpolation scheme to match the runoff data. The observation data are handled in the frame of a hydrological information system. This allows the display of results either in the form of the change in a statistical parameter along the river branches towards the basin outlet or as a map of the variation of the parameters across the basin or region space. The approach is demonstrated for France.

Key words | empirical orthogonal functions (EOF), karst, kriging, river flow regime, river regulation, runoff mapping

NOMENCLATURE

\( A \) \hspace{1cm} \text{basin/drainage area}
\( L \) \hspace{1cm} \text{number of temporal amplitude functions}
\( M \) \hspace{1cm} \text{number of non-overlapping small target elements}
\( N \) \hspace{1cm} \text{number of observed values}
\( qa(A) \) \hspace{1cm} \text{long-term mean annual runoff generated by drainage area } A
\( qm(A,t) \) \hspace{1cm} \text{monthly runoff generated by drainage area } A \text{ for month } t
\( u \) \hspace{1cm} \text{geographical coordinates}
\( t \) \hspace{1cm} \text{time}
\( X \) \hspace{1cm} \text{basin characteristic}
\( z(A) \) \hspace{1cm} \text{runoff characteristic related to the drainage area } A
\( Z(A,t) \) \hspace{1cm} \text{dimensionless monthly runoff related to drainage area } A \text{ for month } t
\( \alpha \) \hspace{1cm} \text{weight coefficient}
\( \beta \) \hspace{1cm} \text{amplitude function}
\( \varepsilon \) \hspace{1cm} \text{residual}
\( \gamma \) \hspace{1cm} \text{semivariogram}
\( \lambda \) \hspace{1cm} \text{distance between points}
\( E[\lambda] \) \hspace{1cm} \text{average of all possible distances between sub-basins } A_1, A_2
\( f \) \hspace{1cm} \text{probability density function}
\( \mu \) \hspace{1cm} \text{Lagrangian multiplier}

INTRODUCTION

A key issue in applied hydrology is to be able to estimate runoff parameters at ungauged sites. The problem is especially accentuated in Europe today in relation to the application EU Water Framework Directive (http://ec.europa.eu/environment/water/water-framework/index_en.html). Many Directive-related tasks in hydro-ecology and water quality require information about the temporal
variation of runoff at many points on the river network. This necessitates the development of interpolation procedures to estimate flows between measurements at gauging sites.

In practice, two different approaches have been in use for estimating runoff characteristics at ungauged sites. Firstly, a map of the parameter of interest has been constructed from a regional set of estimated values of this parameter from observed data. Long-term mean values of the water balance components, namely precipitation, runoff and evapotranspiration, are the most common parameters for which maps are made (e.g. Lvovich 1973). Other examples of manually mapped runoff parameters are the coefficient of variation and coefficient of skewness of runoff (e.g. Gottschalk et al. 1979). In classical works in hydrology these maps were made by manual procedures creating an isopleth map. However, these manual procedures have been replaced by automatic interpolation procedures (Gottschalk & Krasovskaia 1998 for an overview).

Secondly, runoff parameters have been determined by estimating statistical relations between the parameter of interest and climate and physiographic characteristics, mainly utilizing regression equations. This approach has been applied for estimating mean annual runoff (Solomon et al. 1968; Pentland & Cuthbert 1971; Leibscher 1995; Bishop et al. 1998) but is used more for flood and low flow parameters (Ouarda et al. 2001; Cunderlik & Burn 2002; Tallaksen & van der Lanen 2004; Castellarin et al. 2007). Today, in the era of advanced GIS systems, the technical difference between the two approaches has disappeared, as it is possible to make digital maps of the climatic and physiographic characteristics constituting the independent variables in the regression equation. A map of the dependent variable—the runoff parameter—is then derived by application of simple map algebra. However, from a theoretical point of view there is an important difference. A map relies on geographical proximity criteria, i.e. neighbouring points should be similar in values than more distant ones. However, the regression approach relies on a similarity criteria in catchment attributes i.e. neighbouring points in an attribute space are more similar than more distant ones in that same space.

A critical point in both approaches is that neither guarantees that the physical and statistical laws that are associated with a certain statistical runoff variable are satisfied. That is why a third approach is applied, referred to as the hydrostochastic approach that allows explicit inclusion of such laws in the interpolation procedures.

Several applications based on hydrostochastic concepts to estimate different runoff characteristics along a river network can be found in the literature: e.g. mean annual runoff (Sauquet et al. 2000a), variance/covariance of runoff (Gottschalk et al. 2006), flow duration curves (Krasovskaia et al. 2006) and mean low flow (Pacheco et al. 2006).

Nevertheless, most of the previous works neglected local influences such as karst areas as well as the effect of man-made changes by hydraulic structures. Such local factors induce abrupt changes of the runoff pattern. Indeed, these discontinuities are not suited to interpolation since they are not spatially structured.

Here, an attempt is made to use the basic hydrostochastic concepts to explicitly include such local knowledge. The aim is to predict both natural and influenced seasonal variations of river flow at ungauged locations. For this purpose, the estimation procedure is divided into two consecutive steps. The first step deals with runoff patterns with characteristics behaving as a ‘regionalized variable’ in accordance with the definition of Matheron (1965). A preliminary map is drawn. The second step introduces corrections to represent flow regime altered by karst or man-made influences on affected local sectors.

Developments by Sauquet (2006) are the starting point of this work. The mean annual runoff qa is estimated on a partition, i.e. non-overlapping elements that form the whole study area, and streamflows are obtained by aggregation along the river network. These results are then exploited to derive runoff at finer time resolution by temporal disaggregation of qa. Interpolation will therefore concern variables characterizing the redistribution of the annual volume.

A direct method of estimating the twelve monthly values is to evaluate them independently month by month, in which case a spatio-temporal interpolator needs to be developed. The seasonality induces non-stationarity in the space-time covariance structure, which also needs to be considered by the applied interpolator. Skoien & Blöschl (2007) have developed a spatio-temporal interpolation procedure for daily runoff for the stationary case, which might be extended also to the non-stationary case of mean annual runoff qa.
seasonal variation. The alternative, considered here, is to combine a spatial interpolation procedure with an expansion into Empirical Orthogonal Functions (EOF) of the temporal pattern of variability (Rao & Hsieh 1991; Sauquet et al. 2000b). The non-stationarity is taken into account by this method and dimensionality of the interpolation problem is reduced to a minimum.

In this paper, two theoretical issues are examined. A first section describes how the two procedures suggested by Rao & Hsieh (1991) and Sauquet (2006) can be extended. A second section gives details of how to identify and include thereafter the corrections related to karst, water abstraction and regulation by reservoirs. The study area and data used are then presented. Results are demonstrated using twelve maps of mean monthly runoff. Validation of the interpolation approach is performed for two regions with distinctly different seasonal hydrological patterns. Times series of discharges free from human activities are estimated at the outlet of the main French rivers, then introduced into a simple regime classification and finally mapped. The last section, specifically devoted to river flows influenced by human activities, is followed by a discussion of the results and conclusions.

THE INTERPOLATION PROCEDURE: THEORETICAL ASPECTS

Hydrological variables show large variability across space revealing a non-homogeneous random process. Streamflow has the characteristic of being the result of integration processes over a basin i.e. representing a generalized random process with a local support equal to the basin area. A substantial part of the spatial variability is induced by these differences in support, which thus build in dependence on area. The support influences most parameters such as variance-covariance, skewness of the original data series as well as all statistics of low flow and floods.

When complex variation patterns are concerned, as in the case when runoff is regarded as a space-time random process, a possibility of a direct estimation of the multivariate distribution function describing this process is not tractable. The conventional way of handling this difficulty is to accept a partial characterization (Gottschalk 2005). The most widely used are characterization by distribution function (one-dimensional), second moment characterization (mean, variance-covariance) and empirical orthogonal function (EOF) expansion, i.e. a series representation in terms of random variables and deterministic functions of a random process.

The main hydrometeorological variables (rainfall, evapotranspiration and temperature) observed at the land surface are represented in a three-dimensional (3D) space: the two geographical coordinates \(u\) and time \(t\). Streamflow, on the other hand, is represented in a 2D space i.e. the point along a river \(u\) related to the point with an upstream basin drainage area \(A\) and time \(t\). The most fundamental characteristics of runoff—the long-term mean \(qa(A)\) (which is only a function of the space coordinate, as it is integrated over time)—must allow the runoff values \(qa\), generated by all the non-overlapping sectors falling within the boundaries of the basin \(A\), to sum to the value observed at the outlet \(u\). This is the water balance equation for the lateral flow in a basin or in a statistical sense that the average value for the whole basin should be consistent with averages over its parts.

Mapping the mean value is a rather straightforward task and basically it is a problem of stochastic interpolation with local irregular support (also known as block kriging), with an added water balance constraint. An obstacle might be the complexity in the structure of the covariance/semivariogram of runoff since data represent a mixture of nested and non-nested basins. This problem has been studied by Gottschalk (1993a). Gottschalk (1993b) introduced a method for stochastic interpolation of runoff along the river network with a constraint preserving the water balance, i.e. at each downstream point in the river the runoff is the sum of the upstream inflow. Sauquet et al. (2000a) and Sauquet (2006) developed this methodology further and combined it with a system for structuring hydrographical networks in a hierarchical way. It allows an effective reconstruction of the variation of mean annual runoff (first order moment) along the river network (i.e. as a function of the area \(A\) of the drainage basin) in a basin from streamflow observations and a Digital Elevation Model (DEM). The resolution of the underlying DEM defines the size of computational units (grid cells, sub-basins). It is
A geostatistical framework adapted for mapping runoff features

The first step in an interpolation procedure across space with one realization at hand is to determine a theoretical semivariogram in accordance with runoff data. This approach is widely used in the interpolation of meteorological fields but needs to be adapted for runoff features due to their spatial support. Let us consider the semivariogram \( \gamma(A_1;A_2) \) for a process with local support expressed in terms of the corresponding point process \( z(u) \). \( \gamma(A_1;A_2) \) is given by:

\[
\gamma(A_1;A_2) = \frac{1}{2} \mathbb{E} \left[ \left( \frac{1}{A_1} \int_{A_1} (z(u') - z(u'')) \, du' + \frac{1}{A_2} \int_{A_2} (z(u') - z(u'')) \, du'' \right)^2 \right] 
= \gamma(z(A_1);z(A_2)) - \frac{1}{2} \gamma(z(A_1);z(A_1)) - \frac{1}{2} \gamma(z(A_2);z(A_2)) 
\tag{1}
\]

where

\[
\gamma(z(A_1);z(A_2)) = \frac{1}{A_1 A_2} \int_{A_1 A_2} (z(\mathbf{u}'') - z(\mathbf{u}'')) \, d\mathbf{u}' \, d\mathbf{u}'' 
\tag{2a}
\]

This latter semivariogram with basin support can also be expressed in terms of an expected value (Matérn 1960):

\[
\gamma(z(A_1);z(A_2)) = \int_{\min(\lambda)}^{\max(\lambda)} \gamma(\lambda) f(\lambda) d\lambda = \mathbb{E}[\gamma(\lambda)] 
\tag{2b}
\]

where \( f(\lambda) \) denotes the probability density function of all possible distances \( \lambda \) between points in areas \( A_1 \) and \( A_2 \), respectively. The procedure is to assume a theoretical structure for the instantaneous point process in a continuous space \( u \). An estimated integrated semivariogram value for basins \( A_1 \) and \( A_2 \) is then calculated from Equation (1) and should match the corresponding empirical semivariogram value for these two basins. For applied problems, approximations to \( \gamma \) are usually considered (Gottschalk et al. 2006). Among them, the most common approach consists of fitting one theoretical model as a function of the distance between the centres of gravity (e.g. Huang & Yang 1998; Merz & Blöschl 2004):

\[
\gamma(A_1;A_2) = \gamma(|\mathbf{u}_G(A_1) - \mathbf{u}_G(A_2)|) 
\tag{3}
\]

Sauquet (2006) points out the inconsistency when this distance is used, and suggests an approximate averaged semivariogram:

\[
\gamma(A_1;A_2) \approx \gamma(E(\lambda)) 
\tag{4}
\]

where \( E(\lambda) \) is the average of all possible distances between the two sub-basins \( A_1 \) and \( A_2 \), as suggested by Ghosh (1951):

\[
E(\lambda) = \frac{1}{A_1 A_2} \int_{A_1 A_2} |\mathbf{u}_1 - \mathbf{u}_2| \, d\mathbf{u}_1 \, d\mathbf{u}_2 
\tag{5}
\]

Possible theoretical models of the semivariogram are tested and compared graphically to the empirical semivariogram. The selected function for \( \gamma \) is the one giving the best fit.

The runoff characteristic \( z(A) \) related to the element \( A \) is calculated using the weighted linear combination of \( N \) observed values \( z(A_i), i = 1, \ldots, N \):

\[
z(A) = \sum_{i=1}^{N} \lambda_i z(A_i) 
\tag{6}
\]

where the weights \( \lambda_i, i = 1, \ldots, N \) are found by minimizing the expected error, under an unbiasedness constraint (i.e. an expected bias of zero). Under the assumption that the process is homogeneous, this leads to the resolution of the following linear system (Matheron 1965):

\[
\begin{cases}
\sum_{j=1}^{N} \lambda_j (A_i z(A_j) - \mu(A)) = \gamma(A_i;A), & i = 1, \ldots, N \\
\sum_{i=1}^{N} \lambda_i (A) = 1
\end{cases}
\tag{7}
\]

where \( \mu \) is a Lagrangian multiplier. When spatial homogeneity is rejected, an empirical formula linking the area-related runoff values \( z \) to \( K \) basin characteristics \( X_i, i = 1, \ldots, K \) is fitted:

\[
z'(A) = f(X_i(A)) 
\tag{8}
\]
\( \varepsilon(A) = z(A) - z^*(A) \) is the residual interpolated by kriging under the assumption of a second-order stationarity random field. Including basin descriptors in the empirical formulas is a way of accounting for the fact that streamflow data result from processes operating over the whole basin. The combination of the map of the residuals and the map of \( z^* \) gives an estimate of \( z \) for each area \( A \). In this study, the homogeneity assumption is considered to be valid if no significant relationship between \( z \) and a selection of available basin characteristics can be identified. In fact, the approach applied to runoff characteristics differs above all from the classical types of kriging by the way the semivariogram is computed (Equation (4)).

**The variables under study and the EOF analysis**

As a point of departure, consider monthly runoff generated by areas between two or more gauging stations or by a headwater basin above a single gauging station. Runoff values for areas between gauging stations are calculated by subtracting the streamflow(s) measured upstream from the value observed downstream. From \( N \) gauging stations with drainage basin \( A_i, i = 1, \ldots, N \), \( N \) basins or sub-basins \( A_i, i = 1, \ldots, N \) can be delineated and \( N \) related time series of monthly runoff \( qm(A'_i, t), i = 1, \ldots, N, t = 1, \ldots, 12 \) expressed in mm/month are computed.

In the introductory section, \( qa \) is supposed to be known and a temporal disaggregation of \( qa \) is suggested. A straightforward method of characterizing the way the annual volume is distributed within the year is to consider the twelve runoff values normalized with respect to \( qa \). Consequently, the variables of interest are dimensionless runoff:

\[
Z(A'^*, t) = \frac{qm(A'^*, t)}{qa(A'^*)}
\]

(9)

where \( qa(A'^*) \) is the mean annual runoff expressed in mm/yr. \( Z(A'^*) \) is the average of the twelve long-term monthly flows and an approximation is given by:

\[
\overline{Z}(A'^*) = \frac{qa(A'^*)}{12}
\]

(10)

and

\[
Z(A'^*) = 1/12
\]

(11)

The main advantages of handling \( Z \) are that these twelve coefficients: (1) are dimensionless and thus comparable; and (2) are defined in such a way that the spatial supports correspond to non-overlapping drainage areas. This avoids the problem of redundancy of information from the upper-most parts of the basins when nested basins are used in the spatial analysis.

The \( N \) observed time series of mean normalized monthly runoff are decomposed into Empirical Orthogonal Functions expansion (Holmström 1963). This standard mathematical technique, also known as the Karhunen-Loeve transform, aims at extracting common patterns that represent a large fraction of the variability contained in a dataset. EOF analysis has already been combined with interpolation techniques for mapping the monthly runoff pattern (Rao & Hsieh 1991; Sauquet et al. 2000b). The procedures with respect to the variables under study have been modified in the following way:

1. EOF analysis interprets \( Z(A'_i, t), j = 1, \ldots, N, t = 1, \ldots, 12 \) as linear combinations of \( N \) orthogonal functions \( \beta_j \):

\[
Z(A'_i, t) = \overline{Z}(A'_i) + \sum_{i=1}^{N} \alpha_i(A'_i) \beta_i(t), \quad j = 1, \ldots, N;
\]

\[
t = 1, \ldots, 12
\]

simplified using Equation (11):

\[
Z(A'_i, t) = \frac{1}{12} + \sum_{i=1}^{N} \alpha_i(A'_i) \beta_i(t), \quad j = 1, \ldots, N;
\]

(12)

\[
t = 1, \ldots, 12
\]

The weight coefficients \( \alpha_i(A'_i), i = 1, \ldots, N \) depend on the location and are the eigenvectors of the sample covariance matrix with dimension \( N \). The functions \( \beta_i(t), i = 1, \ldots, N \) in EOF-analysis named amplitude functions, are the products of the transposed matrix of observations \( (Z(A'_i, t) - 1/12), j = 1, \ldots, N, t = 1, \ldots, 12 \) and the eigenvectors \( \alpha_i, i = 1, \ldots, N \). In addition, \( \beta_i(t), i = 1, \ldots, N \) have zero mean.
2. The series expansion is truncated at \( L < N \) terms, retaining \( L \) temporal amplitude functions that explain the largest portion of the variance within the dataset. The contribution of \( \beta_i \) to the total variance in the dataset is equal to the eigenvalue corresponding to the \( i \)th eigenvector \( \alpha_i \) divided by the sum of the eigenvalues. To select the subset of \( L \) temporal functions, a classical procedure consists of sorting the functions \( \beta_i \) by decreasing order according to their related contribution, i.e. from the most informative to the less informative components, and to keep the first \( L \) ordered functions so that the cumulative contribution is higher than a certain threshold in terms of explained variance (e.g. 95%).

3. The weights \( \alpha_i, i = 1, \ldots, L \) of the significant modes are obtained at each of the \( M \) non-overlapping small target elements \( \Delta A_i, i = 1, \ldots, M \) that form a given partition of the study area. The kriging technique suited to runoff characteristics detailed above is applied to the \( L \) variables instead of Equation (12).

4. Equation (12) is used to reconstruct the twelve normalized values of monthly runoff \( Z \). The runoff pattern is estimated for the \( M \) elementary cell \( \Delta A_i, i = 1, \ldots, M \). The results obtained are twelve maps of mean monthly runoff.

Streamflow values are easily derived along the river network at any point \( u \) considered as the outlet of the upstream area using the principle of continuity and neglecting routing lags. The monthly streamflows are the sums of the runoff generated in all fundamental units \( \Delta A_i \) flowing into that location \( u \):

\[
q_m(u, t) = \frac{1}{A} \sum_{\Delta A_i \subset A} q_m(\Delta A_i, t) \Delta A_i, \quad t = 1, \ldots, 12
\]  

(13)

where \( A \) is the drainage area at location \( u \) and streamflows are expressed in mm. The final equation for estimation of runoff at any point in a river is found by inserting Equations (9) and (12) into Equation (13):

\[
q_m(u, t) = \frac{1}{A} \sum_{\Delta A_i \subset A} q(a(\Delta A_i)) \left[ \frac{1}{12} + \sum_{l=1}^{L} \alpha_l(\Delta A_i) \beta_l(t) \right] \Delta A_i, \quad t = 1, \ldots, 12
\]  

(14)

### Accounting for local deviations

At this step, karstic and man-made influences can be accounted for. The procedure suggested here models them as deviations from the estimates derived from the interpolation Equation (14). They are explicitly introduced as additional terms in the water balance equation:

\[
q_m'(u, t) = q_m(u, t) + q_{mk}(u, t) + q_{mr}(u, t),
\]

\[
t = 1, \ldots, 12
\]  

(15)

where \( q_m'(u, t), t = 1, \ldots, 12 \) are the values including the influences; \( q_m(u, t), t = 1, \ldots, 12 \) are given by Equation (14);

\[
q_{mk}(u, t) = \frac{1}{A} \sum_{\Delta A_i \subset A} \left[ \frac{W(\Delta A_i)}{12} + \sum_{l=1}^{L} \alpha_l(\Delta A_i) \beta_l(t) \right] \Delta A_i^k,
\]

\[
t = 1, \ldots, 12
\]

are corrections needed to model karstic influences for areas \( \Delta A_i^k \); and

\[
q_{mr}(u, t) = \frac{1}{A} \sum_{u_i \subset A} \left[ \frac{W(u_i)}{12} + \sum_{l=1}^{L} \alpha_l(u_i) \beta_l(t) \right] \quad t = 1, \ldots, 12
\]

are the corrections determined by the operation of water management at location \( u_i \).

Depending on the sign of the values, \( W(u_i) \) describe the annual volume flowing into \( (W > 0) \) or out of \( (W < 0) \), respectively. \( W(\Delta A_i^k) \) describes the area \( \Delta A_i^k \) under karstic influences and the annual volume diverted to \( (W > 0) \) or from \( (W < 0) \) location \( u_i \). \( q_m(u, t) + q_{mk}(u, t), t = 1, \ldots, 12 \) provide estimates of streamflow from natural conditions including karst. Parameters of the corrections, whatever their origin is, have to be evaluated from observations. It is done by reversing the water budget equation.

The procedure to compute the corrections is iterative. Let us consider \( P \) local features sorted by increasing drainage area. Suppose that the first is a reservoir at location \( u_{r,1} \). The values of \( W(u_{r,1}) \) and \( \alpha_l(u_{r,1}), l = 1, \ldots, L \) are fitted with respect to the observed regulated streamflows \( q_{d,obs} \) and \( q_{m,obs} \) downstream at gauged location \( u_i \) with basin area \( A_i \) both at annual and monthly scales.
The procedure provides estimates of \( qm' \) and \( qa' \):

\[
qm'(u_{1,t}) = qm(u_{1,t}) + \frac{W(u_{1,t})}{12A_1} + \frac{L}{A_1} \sum_{i=1}^{L} \alpha_i(u_{1,t}) \beta_i(t)
\]

(16)

\[
qa'(u_{1,t}) = \sum_{t=1}^{12} qm'(u_{1,t}) = \sum_{t=1}^{12} qm(u_{1,t}) + \sum_{i=1}^{L} \frac{W(u_{1,t})}{12A_1} + \frac{L}{A_1} \sum_{i=1}^{L} \alpha_i(u_{1,t}) \beta_i(t)
\]

(17)

and finally

\[
a_k(u_{1,t}) = \frac{A_1(qm_{obs}(u_{1,t}) - qm(u_{1,t})) \beta_k}{\beta_k \beta_k}
\]

(22)

If the first local feature is a karstic sector \( \Delta A^k \), the correction coefficients \( W(\Delta A^k) \) and \( \alpha_l(\Delta A^k) \) can be derived in the same way:

\[
W(\Delta A^k) = A_1(qa_{obs}(u_{1}) - qa(u_{1})) / \Delta A^k
\]

(23)

The first condition for annual runoff can be written:

\[
qa_{obs}(u_{1}) = qa(u_{1})
\]

(18)

Since \( \beta_i \) has a zero mean, \( \sum_{t=1}^{12} \beta_i(t) = 0 \forall i \in \{1, \ldots, L\}:

\[
\frac{1}{A_1} \sum_{i=1}^{L} \sum_{t=1}^{12} \alpha_i(u_{1,t}) \beta_i(t) = \frac{1}{A_1} \sum_{i=1}^{L} \sum_{t=1}^{12} \alpha_i(u_{1,t}) \beta_i(t)
\]

\[
= \frac{1}{A_1} \sum_{i=1}^{L} \alpha_i(u_{1,t}) \sum_{t=1}^{12} \beta_i(t) = 0
\]

(19)

The first constraint is therefore fulfilled if \( W(u_{1,t}) = A_1(qa_{obs}(u_{1}) - qa(u_{1})) \).

To guarantee the constraints for the monthly values \( qm'(u_{1,t}) = qm_{obs}(u_{1,t}), t = 1, \ldots, 12 \), let us consider the cross-product \( qm(u_{1,t}) \beta_k, k = 1, \ldots, L \):

\[
qm_{obs}(u_{1,t}) \beta_k = qm(u_{1,t}) \beta_k + \frac{W(u_{1,t})}{12A_1} \sum_{t=1}^{12} \beta_k(t)
\]

\[
+ \frac{L}{A_1} \sum_{i=1}^{L} \alpha_i(u_{1,t}) \beta_i \beta_k
\]

(21)

By construction, the \( L \) amplitude functions are orthogonal, i.e. \( \beta_i \beta_i = 0 \) if \( i \neq j \). Thus,

\[
qm_{obs}(u_{1,t}) \beta_k = qm(u_{1,t}) \beta_k + \frac{1}{A_1} \sum_{i=1}^{L} \alpha_i(u_{1,t}) \beta_k \beta_k
\]

and finally

\[
\alpha_l(u_{1,t}) = \frac{A_1(qm_{obs}(u_{1,t}) - qm(u_{1,t})) \beta_k}{\beta_k \beta_k}
\]

\[
= \frac{A_1(qm_{obs}(u_{1,t}) - qm(u_{1,t}))}{\beta_k \beta_k}
\]

(24)

The procedure is applied to gradually obtain the successive local deviations for each feature \( p \) (related to location \( u_{i,p} \) or karstic area \( \Delta A^k \)) from altered streamflows \( qm_{obs} \) and \( qa_{obs} \) observed at location \( u_i \), with basin area \( A_p \) downstream of feature \( p \) and from upstream parameters \( W(\Delta A^k), W(u_{i,j}), j < p, \alpha_l(\Delta A^k) \) and \( \alpha_l(u_{i,j}), j < p, l = 1, \ldots, L \).

If the corrections represent the human influences at location \( u_{i,p} \), then

\[
W(u_{i,p}) = A_p(qa_{obs}(u_{i}) - qa(u_{i})) - \sum_{u_{i,j}} W(u_{i,j})
\]

\[
- \sum_{\Delta A^k, \gamma \neq \gamma} \Delta A^k W(\Delta A^k)
\]

(25)

\[
\alpha_l(u_{i,p}) = \frac{A_p(qm_{obs}(u_{i,p}) - qm(u_{i,p})) \beta_k}{\beta_k \beta_k}
\]

\[
- \sum_{\Delta A^k, \gamma \neq \gamma} \alpha_l(\Delta A^k)
\]

(26)

where \( l = 1, \ldots, L \).
If the corrections model influences of karstic area \( A_p \):

\[
\begin{align*}
\Delta A_p^h W(\Delta A_p^h) &= A_p(qm_{obs}(u_p) - qa(u_p)) \\
&\quad - \sum_{u_{i<j}} W(u_{i,j}) - \sum_{u_{i<j}} \Delta A_p^h W(\Delta A_p^h) \\
\Delta A_p^h \alpha_l(\Delta A_p^h) &= A_p(qm_{obs}(u_p) - qa(u_p)) \beta_l \\
&\quad - \sum_{u_{i<j}} \alpha_l(u_{i,j}) - \sum_{u_{i<j}} \Delta A_p^h \alpha_l(\Delta A_p^h)
\end{align*}
\]

where \( l = 1, \ldots, L \). \( qm_{obs}(u_p, t) - qa(u_p, t) \) and \( qm_{obs}(u_p, t) - qa(u_p, t), t = 1, \ldots, 12 \) are the anomalies of streamflow between observations and estimation without any influence. The last terms in the right-hand sides of Equations (25) to (28) are similar: they quantify all the corrections related to the \( p - 1 \) upstream karstic areas or locations of reservoir, obtained from the previous \( p - 1 \) iterations.

An example is given in Figure 1 with \( P = 3 \) gauging stations at location \( u_i, i = 1, \ldots, 3 \): the first one is downstream of the karstic area \( \Delta A_2^h \) and the second one is influenced by one reservoir at location \( u_1 \). The river flows at location \( u_{1,3} \) are altered by the upstream karstic sector \( \Delta A_3^h \), the reservoir at location \( u_1 \) and an additional reservoir at location \( u_3 \). The grey lines delineate the partition \( \Delta A_i, i = 1, \ldots, M \) of the total drainage area \( A_3 \). In this case, the first anomaly under study is related to the reservoir at location \( u_1 \) using streamflows observed \( (qm_{obs}) \) and estimated \( (qm) \) at location \( u_1 \):

\[
W(u_{1}) = A_1(qm_{obs}(u_1) - qa(u_1))
\]

\[
\alpha_l(u_1) = A_1(qm_{obs}(u_1) - qa(u_1),) \beta_l, \quad l = 1, \ldots, L
\]

The deviation is then computed at location \( u_2 \):

\[
W(\Delta A_2^h) = A_2(qm_{obs}(u_2) - qa(u_2))/\Delta A_2^h
\]

\[
\alpha_l(\Delta A_2^h) = A_2(qm_{obs}(u_2) - qa(u_2)) \beta_l, \quad l = 1, \ldots, L
\]

and lastly at location \( u_3 \):

\[
W(u_{3}) = A_3(qm_{obs}(u_3) - qa(u_3)) - W(u_{1}) - \Delta A_2^h W(\Delta A_2^h)
\]

\[
\alpha_l(u_3) = A_3(qm_{obs}(u_3) - qa(u_3)) \beta_l \\
- \Delta A_2^h \alpha_l(\Delta A_2^h), \quad l = 1, \ldots, L
\]

**APPLICATION**

**The study area and data used**

The study area covers the whole of France. The snowmelt-fed regimes are observed in the mountainous part (high altitude rivers in both the Pyrenees and the Alps) in contrast to the northern and western part under Atlantic climate influences, where rainfall and evaporation drive seasonal variation of runoff.

Runoff estimates were computed for more than 6,100 elements of a partition of the study area \( \Delta A_i, i = 1, \ldots, M \), considered as fundamental units. The median size for \( \Delta A_i \) is equal to 67 km². In addition, France was divided into ten major hydrological sub-areas based on topography (Figure 2), e.g. Region 3 is the Seine river basin; Region 4 comprises Brittany and Normandy; Region 10 includes Corsica, the downstream part of the Rhone river basin and Mediterranean coastal river basins. All sectors apart from Regions 2, 3 and 4 demonstrate pronounced topography.

**Figure 1** Example of one basin with three main sources of influence.
A total of 872 monthly time series free from both human influences and karst influences covering the period 1981–2000 were considered for this application (Figure 2). The dataset was constituted after quality control of all the gauging stations available during the period of interest in the French database HYDRO (http://www.hydro.eaufrance.fr/). These stations have high quality measurements but with some short gaps in their time series. In addition, two groups of locally anomalous stations were constituted: 98 basins with major human impact form reservoirs and 67 basins with a strong control by karst aquifers. Mean annual runoff for the gauged basins ranges from 100 mm to more than 1,500 mm per year, with a median value equal to 370 mm per year. The map for the mean annual runoff was provided by Sauquet (2006).

The interpolation scheme requires a well-defined hierarchical structure of the river network to identify links with runoff variability and to be introduced thereafter in the interpolation procedure. The river network has been extracted from a raster DEM with $1 \times 1$ km cells. Basin attributes are calculated by combining GIS layers with the drainage pattern. At large scale, relief is certainly the most important physiographic factor that influences the temporal runoff distribution within the year, with elevation influencing the air temperature as well as the proportion of snowfall relative to total precipitation. In addition to the mean basin elevation, the coordinates ($XG; YG$) of the centre of gravity are the basin characteristics considered in this study.

The results of the EOF analysis

EOF analysis was performed on the ten sub-datasets of $N$ time series free from any influence. For each region, eleven amplitude functions were identified and arranged in descending order according to their contribution to the explained total variance within the dataset. The choice truncation level $L$ was based on a compromise between the wish to limit the number of weight coefficients to be interpolated (i.e. low value for $L$) and the wish to exhibit perfect agreement between the observed and estimated monthly patterns of streamflow (i.e. high value for $L$). All levels of truncation were tested by calculating the Nash–Sutcliffe coefficient $R^2$ (Nash & Sutcliffe 1970). The statistics considered for selection of truncation level $L$ for each region are: the number of stations ($N > 0.90$) with Nash–Sutcliffe coefficients above 0.90 between estimated and observed monthly runoff values; the regional minimum value of the Nash and Sutcliffe coefficient; and the accumulated explained total variance $\%\text{Var}$.

For all the regions, the first amplitude function $b_1$ describes the most common monthly pattern within the dataset and represents a very large proportion of the explained variance, i.e. from 66% for Region 10 to 97% for Region 4 (Figure 3).

The choices for $L$ are reported in Table 1. One should note that summing the number of stations for each region yields a number exceeding 872. Indeed, basins are used more than once to guarantee continuity on both sides of the boundaries of the regions.

The values of $L$ are in accordance with the dimensionality of the similar types of flow regimes defined for Scandinavia (Krasovskaia et al. 1999). Regions 6, 9 and 10 are certainly the most heterogeneous since these sectors require the highest number of amplitude functions ($L=4$) whereas three amplitude functions are considered for the other sectors. A possible explanation of heterogeneity in Regions 6, 9 and 10 is that these areas encompass a great

Figure 2 | Location of the gauging stations and delineation of the ten regions considered in the interpolation procedure.
climatic variability and consequently numerous river flow patterns. $N_{< 0.90}$ is found for less than 8% of the stations involved in the EOF analysis. A more detailed analysis showed that:

- the low values for the Nash and Sutcliffe coefficient are observed at sub-basins where runoff data demonstrate a smooth pattern or erratic variations within the year (Figure 4);
- the Nash and Sutcliffe criterion is not related to basin size ($R^2 < 0.02$). Comparable errors are therefore expected whatever the size of the drainage area is.

In addition, some of the amplitude functions can be interpreted in terms of river flow regime. This is illustrated by results obtained for two distinct regions (Figures 5 and 6). The climate over Region 3 is under the influence of oceanic conditions, whereas snowmelt-fed regimes and Mediterranean patterns are observed in Region 10.

The first amplitude functions $\beta_1$ displayed in the upper-left panels of Figures 5 and 6 are comparable. This finding can be generalized: the functions $\beta_1$ show a similar pattern from one region to another. When multiplied by a positive coefficient $\alpha_1$, the time series $\alpha_1 \beta_1(t)$ demonstrates high values in winter and low values in summer and could be interpreted as a pattern of rainfall-fed basins in Western Europe. The second amplitude function $\beta_2(t)$ is usually related to the shifts from the average pattern shown by $\beta_1(t)$. For Region 10 (Figure 6), $\beta_2(t)$ displays the maximum in May and a secondary maximum in November. When $\alpha_1$ is close to zero and $\alpha_2 > 0$, $\alpha_2 \beta_2(t)$ is typical of river flow patterns with a main peak in late spring due to snowmelt and a moderate contribution to rainfall runoff in early autumn when air temperature is above zero.

The maps of mean monthly runoff under natural conditions

Empirical relationships $\alpha^*$ introduced in Equation (8) were established between the weight coefficients $\alpha$ and basin characteristics (mean basin elevation and coordinates of the centre of gravity). Both linear and non-linear dependences were investigated. Models were adjusted by the least-square method and, finally, only the 17 significant relationships $\alpha^*$ developed in Table 2 were considered significant and used thereafter for the application.

It is seen in Table 2 that $\alpha$ is highly correlated with elevation in mountainous sectors (e.g. Region 9), while in rather flatter terrain (Regions 1 and 2) the influence of location prevails. Since the seasonality of precipitation is not pronounced and the contribution of snowmelt is weak, the links to the coordinates of the centres of gravity in Region 1 and Region 2 may be related to the effect of geology. In fact, other basin descriptors such as land-use, geology and soil characteristics could be relevant to predict the weight coefficients in areas with lowlands but this kind of variable was not available at an appropriate scale. When correlation is significant (which is the case for 17 of the 33 weight coefficients), residual $\varepsilon$ is the variable $z$ to be interpolated instead of $\alpha$.

<table>
<thead>
<tr>
<th>Number of amplitude functions $K$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Var</td>
<td>0.65</td>
<td>0.75</td>
<td>0.85</td>
<td>0.90</td>
<td>0.95</td>
<td>1.00</td>
<td>0.95</td>
<td>0.90</td>
<td>0.85</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Figure 3 | Proportion of the total variance explained by the $K$ first amplitude functions.

Table 1 | Efficiency coefficients per region when the $L$ first amplitude functions are considered

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
<th>Region 6</th>
<th>Region 7</th>
<th>Region 8</th>
<th>Region 9</th>
<th>Region 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>110</td>
<td>46</td>
<td>127</td>
<td>154</td>
<td>208</td>
<td>205</td>
<td>31</td>
<td>90</td>
<td>65</td>
</tr>
<tr>
<td>$L$</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>%Var</td>
<td>0.981</td>
<td>0.984</td>
<td>0.986</td>
<td>0.994</td>
<td>0.982</td>
<td>0.982</td>
<td>0.984</td>
<td>0.969</td>
<td>0.979</td>
</tr>
<tr>
<td>Nash min.</td>
<td>0.660</td>
<td>0.876</td>
<td>0.660</td>
<td>0.590</td>
<td>0.739</td>
<td>0.826</td>
<td>0.897</td>
<td>0.573</td>
<td>0.898</td>
</tr>
<tr>
<td>$N_{&lt; 0.90}$</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>
Results from the interpolation procedure are illustrated for Region 9 (Figures 7 and 8). Estimations outside of the boundaries of Region 9 are required for rivers with source or tributaries in Switzerland. Significant relationships were found to predict the weight coefficients except $a_4$ (the correlation displayed in Figure 7 is not significant at 10% level). This result confirms the influence of the relief on runoff pattern in the heterogeneous sectors.

Figure 9 exemplifies the successive steps required to map the first weight coefficient $a_1$. Figure 9(a) is a choropleth map of observed values $a_1$. A theoretical model is fitted to the experimental semivariograms related
Table 2: Statistics of the relationships fitted to the weight coefficients where \( H \) denotes the mean elevation (m a.s.l), \((XG, YG)\) define the coordinates of the centre of gravity (Lambert 2 coordinates expressed in km)

<table>
<thead>
<tr>
<th>Region</th>
<th>Coefficient</th>
<th>Empirical relationship</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \alpha_1 )</td>
<td>(-0.1418 \times XG + 0.2184 )</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( 0.8085 \times XG - 0.7082 )</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>( \alpha_1 )</td>
<td>(-0.9845 \times XG - 0.5155 \times YG + 2.7220 )</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>( \alpha_1 )</td>
<td>( 0.0478(H/1000)^2 - 0.0965(H/1000) + 0.0959 )</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( 0.0491 ) if ( H &gt; 250 ) m ( \times H/1000 ) - 0.0326</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 )</td>
<td>( 0.07236 \times (H/1000) - 0.1369 )</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>( \alpha_1 )</td>
<td>( 0.07265 ) if ( H &lt; 920 ) m ( - 0.0725 \times (H/1000) + 0.1392 )</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( -0.0031 ) if ( H &lt; 980 ) m ( + 0.2392 \times (H/1000) - 0.2419 )</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 )</td>
<td>( -0.0979 \times (H/1000) + 0.0521 )</td>
<td>0.38</td>
</tr>
<tr>
<td>7</td>
<td>( \alpha_1 )</td>
<td>( -0.1318 \times (H/1000) + 0.2373 )</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( 0.2730 \times (H/1000) - 0.0878 )</td>
<td>0.86</td>
</tr>
<tr>
<td>9</td>
<td>( \alpha_1 )</td>
<td>( -0.1045 ) if ( H &lt; 880 ) m ( - 0.2043 \times (H/1000) + 0.2838 )</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( 0.0808 \times (H/1000)^2 - 0.2555 \times (H/1000) + 0.0413 )</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 )</td>
<td>( 0.1496 \times (H/1000)^2 - 0.4331 \times (H/1000) + 0.2616 )</td>
<td>0.33</td>
</tr>
<tr>
<td>10</td>
<td>( \alpha_1 )</td>
<td>( 0.0937 ) if ( H &lt; 1,050 ) m ( - 0.1561 \times (H/1000) + 0.2358 )</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( 0.1006 \times (H/1000) - 0.0449 )</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>( \alpha_4 )</td>
<td>( -0.1544 \times (H/1000)^2 + 0.3550 \times (H/1000) - 0.1534 )</td>
<td>0.49</td>
</tr>
</tbody>
</table>
to $a_1$ (Figure 8). Figure 9(d) is a combination of the empirical formula application $a_1^*$ (Figure 9(b)) and the map of residuals $e_1$ (Figure 9(c)). Figure 9(d) is very similar to Figure 9(b) since correlation is strong between $a_1$ and $H$. $a_1$ is close to zero for rivers with a smooth variation of monthly flow whereas a high absolute value of $a_1$ results in a pronounced seasonal variation (Figure 10). This seasonality shows a pluvial regime for high positive $a_1$ values and a nival regime for the opposite case of a high negative value.

The same procedure was applied to the two other weight coefficients $a_2$ and $a_3$ to compute twelve normalized monthly runoff values at ungauged locations for Region 9. Since no empirical relationship between $a_4$ and the basin characteristics was identified, $a_4$ was directly mapped without any intermediate variable. Equation (12) is applied to reconstruct the normalized time series $Z(D_{Ai}, t)$, $t = 1, …, 12$ at each elementary cell of the partition $D_{Ai}$, and the monthly runoff values are obtained by multiplying by $qa(D_{Ai})$ (available from Sauquet 2006).

These steps were repeated for each region. All the results were then combined to create the maps of the long-term mean monthly runoff (Figure 11). Low flows in January are noted both in the Pyrenean and in the Alpine sectors while runoff is abundant in the northwest part of France affected by oceanic rains. High values are found in the mountainous sector with moderate altitude. In July, the air temperature is above zero in both the Pyrenean sector and the Alps. Melting snow at high altitudes in July and August generates high monthly flows; elsewhere evaporation processes are predominant and induce low flow in the rivers. September shows the lowest contrast.

Uncertainty analysis based on the estimated streamflows under natural conditions

A split sample approach was employed to assess the predictive performance for each region to estimate mean monthly streamflows $qm$ at locations unaffected by local regulations and karst. Within each of the regions, 90% of
the stations were used for calibration (EOF analysis, fittings of empirical relationships and theoretical semivariograms) and the remaining stations were used for validation. It is important to note that the additional sites from nearby regions were not involved in this removal procedure. Estimates of monthly streamflow were obtained by aggregation in accordance with Equation (14) and \( q_a \) determined from the map of annual runoff for France (Sauquet 2006).

Basins with karstic influences were not included in this analysis. These steps were repeated ten times for each region, to simulate the ungauged conditions at each station exactly one time. The uncertainty analysis is illustrated for two sectors: Region 4 is representative of homogeneous regions and Region 9 is representative of heterogeneous regions. The performance analysis is based on the differences (Err) between the observed and the estimated normalized monthly runoff \( q_m(u,t)/q_a(u) \), \( t = 1, \ldots, 12 \) at each withdrawn station at location \( u \).

The left panel of Figure 12 summarizes the distributions of errors for each month for all the withheld stations by means of a box plot representation. For comparison purposes, the distributions of observed normalized monthly runoff are displayed in the right panel of Figure 12. Table 3 reports the dispersion of the Nash and Sutcliffe coefficient within sites.

Both Table 3 and Figure 12 suggest that the approach performs relatively well: runoff pattern can be predicted with an acceptable level of confidence. Obviously no significant bias in estimates was detected since interquartile intervals contain the zero value (Figure 12). However, considering the error bars of the box plots, the predictive ability depends on the level of homogeneity of the region, i.e. the more homogeneous the area is, the lower the prediction error. For relatively homogeneous Region 4, 90% of the stations show very good results with a Nash and Sutcliffe coefficient above 0.9. Nevertheless, the score obtained at one station is very poor (–0.99) due to complex groundwater influences that make the observed runoff pattern atypical. In addition, unreliable estimates may be expected in summer when observed values and prediction errors are comparable during this season.

The results are not as good but still acceptable for Region 9; 61% of the sites show a Nash and Sutcliffe coefficient above 0.90. In addition, errors have a wide range for months with a large range of observed values (December and January for Region 4, April to August for Region 9). This effect is more noticeable for Region 9.
River flow regime along the river network, including karst influences

In this section, the natural monthly streamflow pattern is derived from $q_m(u, t) + q_{mk}(u, t)$, $t = 1, \ldots, 12$ and introduced into a river flow regime classification scheme to perform a visual inspection of seasonality.

Numerous classifications of river flow regimes exist in the literature. Some of them are supervised classifications based on the origin of the river flows or/and the occurrence of high and low flows within the year (e.g. Pardé 1955; Lvovich 1973; Krasovskaia & Gottschalk 1992), while others are non-supervised and use some kind of grouping routine such as clustering (e.g. Haines et al. 1988) or Wards algorithm (Krasovskaia 1997).

A simple, but well-adapted classification based on a hierarchical cluster analysis is adopted here. A number of different clustering methods exist. The emphasis of this section is to derive a consistent river flow regime classification for the study area. A simple agglomeration method, i.e. based on Ward’s minimum variance method and Euclidian distance as similarity criteria, is sufficient for this purpose. Observed normalized monthly runoff time series are merged to form $K$ flow regime groups. A representative hydrograph $Z_{ref}(j,t)$, $t = 1, \ldots, 12, j = 1, \ldots, K$ is established for each group $j$. When these groups are formed, the natural normalized monthly runoff time series $Z(u, t)$, $t = 1, \ldots, 12$ at each location $u$ can be assigned to a flow regime group. The Euclidian distances to the $K$ models of flow regime patterns are computed as:

$$
 d(u,j) = \sqrt{\sum_{t=1}^{12} (Z(u,t) - Z_{ref}(j,t))^2} \quad j = 1, \ldots, K
$$

The flow regime at the location is assigned to group $k$, the one having the minimum Euclidian distance among the $K$ possibilities. The classification is based on the 872 normalized monthly runoff time series $Z(A_i, t)$, $i = 1, \ldots, 872, t = 1, \ldots, 12$.

<table>
<thead>
<tr>
<th>Region</th>
<th>Min</th>
<th>1st quartile</th>
<th>Median</th>
<th>3rd quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-0.995</td>
<td>0.960</td>
<td>0.982</td>
<td>0.992</td>
<td>0.998</td>
</tr>
<tr>
<td>9</td>
<td>0.222</td>
<td>0.847</td>
<td>0.935</td>
<td>0.972</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Figure 11 | Maps of mean monthly runoff for France.

Table 3 | Statistics of the Nash and Sutcliffe coefficient calculated between estimated and observed normalized monthly runoff $Z$. 

E. Sauquet et al. | Monthly runoff at ungauged locations 

Hydrology Research | 39.5–6 | 2008
A classical method of fixing the optimal number of groups $K$ is to consider the agglomeration schedule, which quantifies the increase in heterogeneity within groups from one stage to another. Noticeable jumps from one grouping solution to the next indicate that two distinct groups have been combined. Figure 13 suggests solutions with two, four or nine groups. The choice for the solution with nine groups was finally based on hydrological considerations. Indeed, three groups are at least required to make the distinction between the pluvial regimes and the snowmelt-fed regimes and seven groups are required to split the pluvial oceanic patterns and the pluvial Mediterranean patterns.

Figure 14 illustrates the obtained grouping. The dispersion within all the groups are weak, with the exception of Group 9. Few atypical patterns are found in series in Groups 5 and 7. These time series are obtained by subtracting streamflow values measured upstream from those observed downstream, which may result in inconsistencies in the data. The identified nine groups differ by the timing of low and high flows within the year and by the smoothness of the patterns. The classification obtained distinguishes well the various patterns of rainfall-fed river flow regimes due to their prominence in the dataset. On the other hand, too few examples of mountainous river flow regimes are present in the dataset. All of them are contained in Group 9. To allow distinction between types of snowmelt-fed regimes, Group 9 was replaced by four models derived from Pardé’s classification (1955). Together with the other eight groups, it results in twelve characteristic flow regimes in France.
The categorization of the estimated monthly pattern at each point of the map into the twelve river regime groups $Z_{\text{Ref}}(j,t), t = 1, \ldots, 12, j = 1, \ldots, 12$ with Euclidean distance (Equation (29)) as similarity criteria allows a construction of a flow regime map of France (Figure 15). At a national level, pluvial river flow regimes (Groups 1 to 6) dominate. The groups mainly differ by the contrast between the maximum and the minimum of monthly streamflow. Nearly uniform flows through most of the year (Group 1) are found in the northern part of France where large aquifers moderate flows. Almost all hydrographs of Region 4 are allocated in Group 5, which is the confirmation of the strong homogeneity of this region in terms of river flow regime. This runoff pattern is characterized by very low flow in summer, reflecting the lack of deep groundwater storages in the catchment. Group 7 is representative of Mediterranean river flow regimes where small rivers basins experience hot and dry summers and intense rainy events in autumn. Their runoff pattern therefore exhibits severe low flow in summer and high flow in November.

In mountainous areas, uppermost basins display snowmelt-fed regimes (Groups 10, 11 and 12). The lower the outlet is, the lower the contributions of snowmelt to runoff. Groups 8 to 9 are in the transition regime. The seasonal variation of streamflow is affected as much by precipitation timing as by air temperature and topographic influences (on snowpack formation and snowmelt timing). Typically, high flows are observed in spring.

**Estimating river flows affected by human activity**

For this application, the dataset of 98 gauging stations with influenced flows was divided into two subsets.

- **Reference stations** are the closest stations located downstream to the main source of influences. These sites defined the 69 locations $u_{ri}, i = 1, \ldots, 69$ where the parameters $W(u_{ri})$ of Equation (25) and $a_l(u_{ri}), l = 1, \ldots, L$ of Equation (26) were fitted. Here, only stations downstream to artificial reservoirs with capacity
Figure 15 | Map of river flow regime based on the twelve reference hydrographs for France.
larger than 50 hm³ and main abstraction points are considered.

- **Validation stations** are 29 sites used to verify that the monthly pattern is accurately estimated downstream of the reference stations.

**Figure 16** shows results obtained at three reference stations and at three validation stations and allows comparisons between runoff patterns obtained without (grey) and with (black) flow regulation influences. These sites illustrate a wide range of reservoir management objectives: low flow moderation (1 and 4), hydropower generation (3, 5 and 6) and with water withdrawal towards a neighbouring basin (2). Most of them alter the natural flow regime significantly with a reduction of the variability from month-to-month. **Table 4** reveals that both the minimum and the median of the Nash and Sutcliffe criterion calculated between estimated and observed monthly runoff values are higher when corrections are made. The gain in predictive performance is less significant on the validation dataset since the impact of artificial flow alterations is less sensitive when going far downstream from the origin of influences.

**Table 4** | Statistics of the Nash and Sutcliffe coefficient calculated between estimated and observed monthly runoff for the two datasets with or without corrections

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Without correction</th>
<th>Min</th>
<th>1st quartile</th>
<th>Median</th>
<th>3rd quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td></td>
<td>-190.21</td>
<td>-0.80</td>
<td>0.57</td>
<td>0.89</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>With correction</td>
<td>0.06</td>
<td>0.92</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td>-0.93</td>
<td>0.71</td>
<td>0.89</td>
<td>0.94</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>With correction</td>
<td>0.69</td>
<td>0.92</td>
<td>0.95</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>
CONCLUSIONS

An approach to estimate runoff characteristics describing the water resource at ungauged locations based on a hydrostochastic concept has been presented. The procedure is applied in two steps. Firstly, the fraction of runoff without pronounced karstic influences and free from man-made regulation is obtained using objective methods for each element of a partition of the study area. Estimates are obtained by temporal redistribution of annual values $qa$. The twelve mean monthly runoff values divided by $qa$ are first computed at each gauged location. Empirical orthogonal function (EOF) expansion allows interpretation of each normalized runoff pattern as a linear combination of functions defined at the regional scale with the weights that need to be estimated. The interpolation procedure applied is based on geostatistical techniques adapted to account for the related drainage basin supporting areas. This step results in twelve maps of monthly streamflow for the whole of France, but without considering local influences due to karst and man-made regulations.

Secondly, corrections to estimates are made to model local deviations in runoff along the river network. The discontinuities are mainly due to human activities and karstic aquifers, and are not well suited for interpolation since they are not spatially organized. The procedure considers these abrupt changes as corrections. These corrections are deduced from downstream gauged basins under the constraint of water balance and assigned to the location of the hydraulic structures or to the elementary cells supposed to be controlled by karst aquifers. Runoff patterns representative of actual river flow regime are deduced by aggregation along the river network and incorporating the corrections (where relevant); the continuity equation is therefore fulfilled.

This general framework was applied to data from France and uncertainty analysis demonstrated that reliable estimates can be obtained both under natural conditions and when human influences are considered. An objective validation was not possible in case of karstic areas. In principle, the same procedure to map the twelve monthly long-term mean values can be used for deriving the twelve monthly values within a specific year or for a sequence of years. Due to the linearity in the applied methods, the same basic amplitude functions should reappear for individual years and it will only be the weights coefficients that change and need to be estimated. An interesting experiment would therefore be to consider the same temporal functions derived herein on average seasonal variations to predict and map the monthly streamflow for specific years.

The hydrostochastic concepts illustrated here on a monthly scale may be applied to estimate other statistics at ungauged locations. The key step is to identify physical and statistical relationships specific to the variable under study and how that relationship changes along a river network from headwaters to the outlet. To conclude, one should keep in mind that introducing explicitly hydrological properties of the mapped variable in the procedure is a way to guarantee their consistency in space and to expect more reliable estimates.

ACKNOWLEDGEMENTS

This study was partially funded by Electricité de France and the French Ministry of Ecology and Sustainable Development. The authors wish to express their thanks to the French data base HYDRO for supplying the discharge data used in this study. In addition, the two anonymous reviewers are gratefully acknowledged for their valuable comments on the original manuscript.

REFERENCES


First received 8 October 2007; accepted in revised form 23 May 2008