Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX. IAHS Publ. 307, 2006.

Predicting river flow statistics at ungauged locations—a hydrostochastic approach

ERIC SAUQUET[1], LARS GOTTSCHALK[2], IRINA KRASOVSKAIA[2] & ETIENNE LEBLOIS[1]

1 *Cemagref, Hydrology - Hydraulics Research Unit, 3 bis quai Chauveau CP220, F-69336 Lyon cedex 09, France*
eric.sauquet@cemagref.fr

2 *Department of Geosciences, University of Oslo, PO Box 1022 Blindern, N-0315 Oslo, Norway*

Abstract The specific objective of this paper is to provide an overview of the collaborative projects between the University of Oslo (Norway) and Cemagref (France). The research undertaken concentrated on the development of methods for estimating statistical properties of the water balance components (especially runoff) at ungauged locations. Probability and Statistical Theory is used to estimate runoff characteristics related to various scales in time and space. A river basin was studied as a system consisting of interacting entities without splitting it into these entities—looking for patterns rather than isolated steps of causality, i.e. letting the data speak for themselves. Links between the water balance statistics and the way they develop in time and/or in space, i.e. along the river network, are analysed and these dependences are introduced explicitly in interpolation procedures. Thus the approach preserves the statistical properties in terms of covariances and semi-variograms at a basin scale as well as satisfying water balance constraints. A review of the basic concepts is given first. Second, applications of the concepts are presented with examples of their application to different extensive datasets (France and Costa Rica) and for different hydrological variables. Mean annual and monthly discharges and flow duration curves are investigated in this article.

key words empirical orthogonal function; flow duration curve; runoff; scale; stochastic interpolation

INTRODUCTION

A fundamental problem in applied hydrology is to estimate runoff parameters at an ungauged location. Various methods have been developed for this purpose and the results are often summarized by maps of runoff characteristics over the studied region. Maps produced manually (Gannett, 1912) were published first, since computation facilities were limited at the time they were made. The procedures involved in producing these maps needed lots of attention and were highly time consuming and usually subjective. Empirical relationships among average streamflow, land use, geomorphology and climate have received wide attention for several decades (Solomon *et al.*, 1968; Liebscher, 1972; Hawley & McCuen, 1982; Gustard *et al.*, 1989; Vogel *et al.*, 1999). They have usually been established by multivariate regional regression. Drainage area and precipitation are by far the most explanatory variables. Other basin characteristics may be incorporated, but their inclusion in relationships is not warranted. One critical point is that when values are required at ungauged locations, each gauged basin has the same weight in the calculation. Our experience is that geographical proximity criteria give more reliable estimates and this is the background for the approach developed herein.

Runoff is, by definition, an integrator of a random spatially distributed process. The spatial variability of runoff characteristics is mostly governed by the river network. From upstream to downstream, the basins collect water from tributaries with sometimes different river flow regimes. One consequence is that the hydrological information cannot be propagated in all directions but in a preferred direction along the river network. The size of the basin as well as the topology of the river network influence most parameters like variance-covariance, skewness of the original data series, and all statistics of low flow and floods (Gottschalk, 1993).

The objective of this paper is to present an approach that we will call the "hydrostochastic" approach. This approach attempts to incorporate the previously mentioned properties of runoff characteristics in an interpolation framework:

– a river basin is studied as a system consisting of interacting blocks without decomposing it into blocks—looking for patterns rather than identifying steps of causality, i.e. letting data speak for themselves;
– a river basin is studied in its context looking at it from outside in search of contextual rules of its functioning (water balance, consistency with statistical laws);
– the theoretical background is given by Probability and Statistical Theory to analyse and model results of integration in space over a drainage basin and in time over an interval following the structure of the river network.

These concepts have been applied in a number of studies (Gottschalk & Krasovskaia, 1998; Sauquet *et al*. 2000a,b; Skøien *et al*., 2005). The following sections briefly describe the developments. Applications of these concepts to map monthly discharges and to estimate flow duration curves are presented using different extensive datasets in France and in Costa Rica.

MAPPING MEAN ANNUAL RUNOFF

Starting with the most fundamental characteristic of runoff—the long-term mean $qa(A)$ for a basin A—it must be so that the values over M sub-basins ΔA_i, $i = 1, \ldots, M$ should sum up to the value at the outlet for total basin A. This is the water balance equation for the lateral flow in a basin or in a statistical sense 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 support (or block kriging) with an added water balance constraint. Gottschalk (1993) 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*. (2000) developed this methodology further and combined it with a system for structuring hydrographical networks in a hierarchical way called HydroDem (Leblois & Sauquet, 2000). It allows an effective reconstruction of the variation of mean annual runoff (first order moment) along the river network in a basin from discharge observations and a DEM. The resolution of the underlying DEM defines the size of computational units (grid cells, sub-basins). The applications below use maps of the long-term annual mean derived in this manner.

MAPPING MEAN MONTHLY RUNOFF PATTERN

Variables under study

Traditionally, the variable considered is runoff observed at the outlet at gauging stations. When measurements at a gauging station are directly used in the interpolation scheme, redundant information is introduced due to partial overlapping drainage areas resulting in bias in the spatial analysis. Here, runoff generated by each portion of the basin between two or more gauging stations or uppermost headwaters are considered instead. These runoff values are calculated by subtracting the discharge(s) measured upstream from the value observed downstream. If N gauging stations are available, N basins or sub-basins A_i ,i = 1, ..., N can be defined and N related runoff values computed. To eliminate scale effects within the dataset due to the size of the basin, mean annual runoff $qa(A_i)$ and the 12 monthly discharges $qm(A_i, t)$, t = 1, …, 12 are expressed in mm year^{-1} and mm month^{-1}, respectively.

The pattern of monthly values is estimated following a procedure that accounts for the dependence between consecutive monthly discharges imposed by groundwater storage, flow routing in addition to scale effect, and links to topological river patterns. Temporal disaggregation of the mean annual discharge is considered to ensure consistency with the map of mean annual runoff. The interpolation will deal with dimensionless discharges, i.e. the 12 coefficients:

$$Z(A,t) = \frac{qm(A,t)}{qa(A)} \tag{1}$$

where $qa(A)$ is the mean annual runoff. $\overline{qm}(A)$ is the average of the 12 long-term monthly flows and an approximation is given by:

$$\overline{qm}(A) = qa(A)/12 \tag{2}$$

and:

$$\overline{Z}(A) \approx 1/12 \tag{3}$$

To keep the month-to-month dependence within the flow pattern, the time-series of Z is interpreted as linear combinations of L temporal functions β_i invariant in space:

$$Z(A,t) = \overline{Z}(A) + \sum_{i=1}^{L} \alpha_i(A)\,\beta_i(t), t = 1, ..., 12 \tag{4}$$

where $L \leq 12$ and the weight coefficients $\alpha_i(A)$ vary in space but are constant in time. The Empirical Orthogonal Function (EOF) analysis (Holmström, 1963) is a convenient method to extract the L significant amplitude functions from a dataset of observed times-series. α_i, i = 1, …, L are the eigenvectors of the covariance matrix between all the sub-basins. The number of variables to be interpolated is reduced from 12 to L.

An adaptation of kriging

With kriging the estimated value $z(\boldsymbol{u})$ at location $\boldsymbol{u}$ is a linear combination of observed values $z(\boldsymbol{u}_i)$, $i = 1, \ldots, N$ located in the neighbourhood of $\boldsymbol{u}$:

$$z(\boldsymbol{u}) = \sum_{i=1}^{N} \lambda_i z(\boldsymbol{u}_i) \tag{5}$$

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

$$\begin{cases} \sum_{j=1}^{N} \lambda_j(\boldsymbol{u})\gamma\left(\left|\boldsymbol{u}_i - \boldsymbol{u}_j\right|\right) - \mu(\boldsymbol{u}) = \gamma\left(\left|\boldsymbol{u}_i - \boldsymbol{u}\right|\right), i = 1, \ldots, N \\ \sum_{j=1}^{N} \lambda_j(\boldsymbol{u}) = 1 \end{cases} \tag{6}$$

where: $\left|\mathbf{u}_i - \mathbf{u}_j\right|$ denotes the Euclidian distance between $\mathbf{u}_i$ and $\mathbf{u}_j$, μ is a Lagrangian multiplier and γ is the theoretical model function to the distance fitted to the empirical semi-variogram. This approach is widely used in the interpolation of meteorological fields (e. g. Creutin & Obled, 1982) but needs to be adapted for runoff features z_q that are related to areas. In particular, a relevant distance between pairs of basins has to be defined. Huang & Yang (1998) used the distance between the centres of gravity. The main drawback in this approach is that several basins may have the same centre of gravity. That is why the average of all possible distances between two sub-basins A and B, as suggested by Ghosh (1951), is used here instead. This distance is determined as:

$$d(A,B) = \frac{1}{AB} \iint_{A,B} \left|\boldsymbol{u}_A - \boldsymbol{u}_B\right| \boldsymbol{du}_A \boldsymbol{du}_B \tag{7}$$

Possible theoretical models of the semi-variogram are tested and compared graphically to the empirical semi-variogram. The selected function for γ is the one giving the best fit. The runoff characteristic $z_q(A)$ related to the element A is calculated using the weighted linear combination of N observed values $z_q(A_i)$, $i = 1, \ldots, N$:

$$z_q(A) = \sum_{i=1}^{N} \lambda_i z_q(A_i) \tag{8}$$

When spatial homogeneity is rejected, an empirical formula linking the area-related runoff values z_q to K basin characteristics X_i, $i = 1, \ldots, K$ is fitted:

$$z_q{}^*(A) = f(X_i(A)) \tag{9}$$

$\varepsilon_q(A) = z_q(A) - z_q{}^*(A)$ 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_q* gives an estimate of z_q for each area ΔA_i, i = 1, …, M. A special case is when runoff is estimated on elements of a partition of the study area ΔA_i, i = 1, …, M. Estimated runoff values can be provided along the river network at any point **u** considered as the outlet of the upstream area. The annual and monthly discharges are the sums of the runoffs generated in all fundamental units ΔA_i flowing into that location **u**.

$$qa(\boldsymbol{u}) = \frac{1}{A} \sum_{\Delta A_i \subset A} qa(\Delta A_i)\Delta A_i \tag{10}$$

$$qm(\boldsymbol{u},t) = \frac{1}{A} \sum_{\Delta A_i \subset A} qm(\Delta A_i)\Delta A_i \text{ , } t = 1, ..., 12 \tag{11}$$

where A is the drainage area at location **u** and discharge are expressed in mm. Equation (11) can be developed using equation (3) and equation (4):

$$qm(\boldsymbol{u},t) = \frac{1}{A} \sum_{\Delta A_i \subset A} qa(\Delta A_i) Z(\Delta A_i,t)\Delta A_i = \frac{1}{A} \sum_{\Delta A_i \subset A} qa(\Delta A_i)\left[\frac{1}{12} + \sum_{i=1}^{L} \alpha_i(A)\beta_i(t)\right]\Delta A_i \tag{12}$$

REGIONALIZATION OF FLOW DURATION CURVES

Flow Duration Curves (FDC) are widely used in applied hydrology. A common practice for regionalization of FDCs is to derive a dimensionless FDC (DFDC) from observed FDCs valid within a homogeneous region, assuming that this DFDC can be transferred to a site with no data. This DFDC is obtained by dividing all observations by the mean annual flow (Yu & Yang, 1996, Singh *et al.*, 2001). Here the same standardization procedure is used, but the regionalization is supported by scaling properties, i.e. a relationship between the parameters of the DFDC and the first moments of the time series. First we consider the normalized daily runoff, in the same way as before (equation (1)):

$$Z(A,t) = q(A,t) / qa(A) \tag{13}$$

and then model the covariance between the values $Z(A_1)$ and $Z(A_2)$ over two (basin) averages with areas A_1 and A_2, respectively, by a theoretical function:

$$\mathrm{Cov}[Z(A_1); Z(A_2)] = \frac{1}{A_1 A_2} \int_{A_1}\int_{A_1} Cov(\boldsymbol{u}',\boldsymbol{u}'')d\boldsymbol{u}'d\boldsymbol{u}'' = \frac{1}{A_1 A_2} \int_{A_1}\int_{A_1} \sigma(\boldsymbol{u}')\sigma(\boldsymbol{u}'')\rho(\boldsymbol{u}',\boldsymbol{u}'')d\boldsymbol{u}'d\boldsymbol{u}'' \tag{14}$$

where $Cov(\mathbf{u}',\mathbf{u}'')$ and $\rho(\boldsymbol{u}',\boldsymbol{u}'')$ is the point covariance and correlation functions, $\sigma(\mathbf{u})$ the point standard deviation and with **u′** representing a point within the basin area A_1 and **u″** within the basin area A_2. The variance is formalized by:

$$\mathrm{V}ar[Z(A)] = \frac{1}{A^2} \int_{A}\int_{A} Cov(\boldsymbol{u}',\boldsymbol{u}'')d\boldsymbol{u}'d\boldsymbol{u}'' = \frac{1}{A^2} \int_{A}\int_{A} \sigma(\boldsymbol{u}')\sigma(\boldsymbol{u}'')\rho(\boldsymbol{u}',\boldsymbol{u}'')d\boldsymbol{u}'d\boldsymbol{u}'' \tag{15}$$

Aggregated values were assigned to each couple of gauged basins applying equation (15). The choice of the best point function is based on a comparison between theoretical values and empirical values. As we are dealing with normalized values $Z(A)$ the standard deviation equals the coefficient of variation of the original variable $q(A)$. These calculations allow mapping of the coefficient of variation along the river network. The next step is to establish the dependency between the DFDCs and use them to calculate percentiles of the normalized flow duration curve. The interpolated mean annual runoff (obtained by the procedure detailed in the previous section) is introduced to finally calculate the percentiles of the FDC (for more details see Krasovskaia *et al*., 2005).

APPLICATION

The study areas

The study areas cover Costa Rica and France, except for the rivers flowing into the Mediterranean Sea. In France, the snowmelt-fed regimes are found in the mountainous part (high altitude rivers in the Pyreneans) in contrast to the northern and western part of the study area under Atlantic climate influences, where the pluvial regime is governed by rainfall and evaporation dominates (alluvial plains of the Seine basin and rivers from Brittany). In Costa Rica, tropical river flow regimes dominate with high flows in autumn and a temporal variability closely connected to the El Nino Southern Oscillation Index. The French and Costa Rican data sets will be used to illustrate mapping mean monthly runoff and flow duration curves, respectively.

Database

A total of 726 French stations and 70 Costa Rican stations with minor human impact were considered for this application. Mean annual runoff for the gauged basins ranges from 100 to more than 1500 mm in the French area. Mean annual runoff can be higher than 5 m year^{-1} in the centre of Costa Rica, varying on average around 2000–4000 mm year^{-1}.

The interpolation scheme requires a well-defined hierarchical structure of the river network to identify links with runoff variability and to be utilized thereafter in the interpolation procedure. The river network has been extracted from a raster digital elevation model DEM with 1 × 1 km cells. Basin attributes are calculated by combining GIS layers with the drainage pattern. The mean basin elevation was a single basin characteristic utilized as at large scale relief, is the most important physiographic factor. The French study area was divided into five major hydrological sub-areas based on topography. In Costa Rica the hydrological conditions are different at either side of the water divider between the Pacific and the Caribbean basins. Separate calculations have therefore been made for the Caribbean and Pacific regions.

Mapping mean monthly runoff, an example from France

Results are detailed for one region which includes the Dordogne basin, the Garonne basin and coastal rivers in the southwest of France. Runoff estimates were computed for more than 5100 elements of a partition of the study area ΔA_i, i = 1, …, M, considered as fundamental units (i.e. M non-overlapping small target elements that form the whole study area).

EOF analysis was achieved using a sub dataset of 205 gauging stations. Eleven independent amplitude functions were identified and arranged in descending order according to their contribution to the explained variance. The first four of them $\beta_i, i = 1, ..., 4$ (Fig. 1) explained more than 98% of the total variance within the dataset, and consequently four weight coefficients were considered for the mapping. The amplitude function β_1 stands for the largest portion of the explained variance (84%) and describes the most common monthly pattern within the dataset.

The regionalization procedure is illustrated on the example of the mean annual runoff qa and the first weight coefficient α_1. Empirical relationships were established between runoff characteristics and H, the mean elevation:

$$qa^* = -0.0423H + 261.73 + 0.0735 \ (R^2 = 0.40) \tag{16}$$

$$\alpha_1^* = -0.0263(H/1000)^2 + 0.0144(H/1000) + 0.0735 \ (R^2 = 0.55) \tag{17}$$

The same procedure is applied to the three other weight coefficients α_i, i = 2, ..., 4. Residuals are calculated at the gauging stations and an exponential model is fitted to the experimental variogram (Fig. 2). Figure 3 displays the final map of the first weight coefficient α_1, which is a combination of the map of residuals ε_1 and the empirical formula application α_1^*. α_1 is close to zero for rivers with an even flow regime whereas a high positive value for α_1 shows a contrasted regime, with pronounced seasonal variation.

All the results from the interpolation were combined to estimate the long-term mean monthly pattern (Fig. 4). Low flows in January are noted in the Pyrenean sector

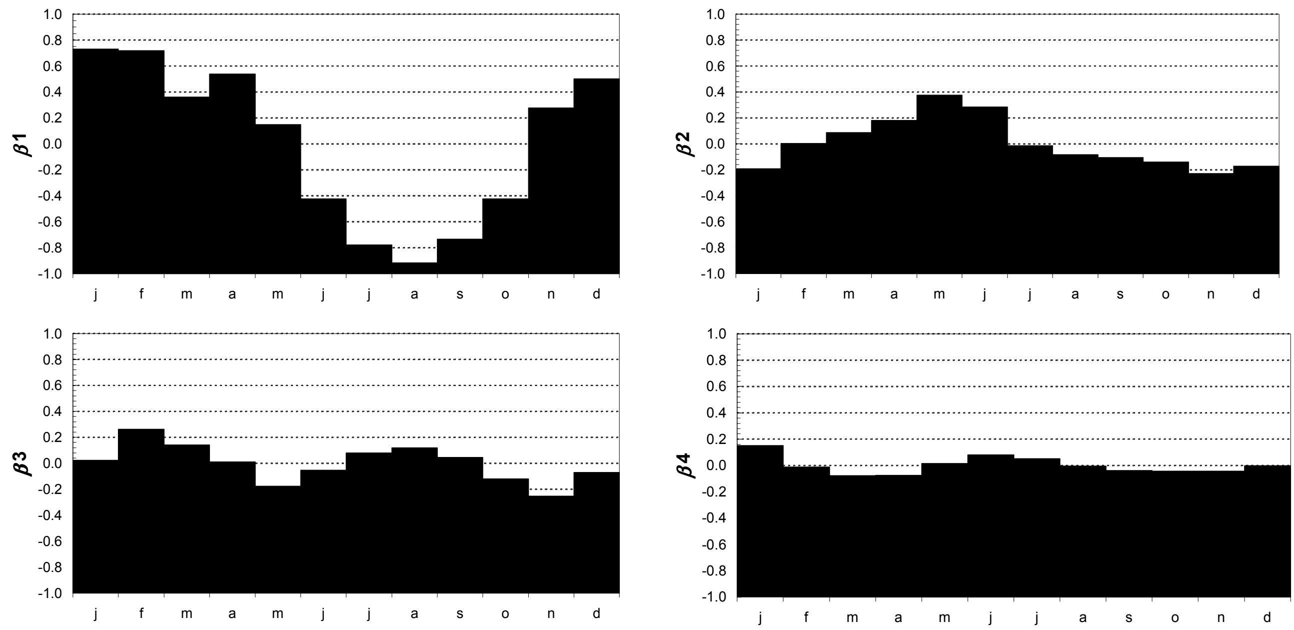

Fig. 1 Amplitude functions for southwestern France.

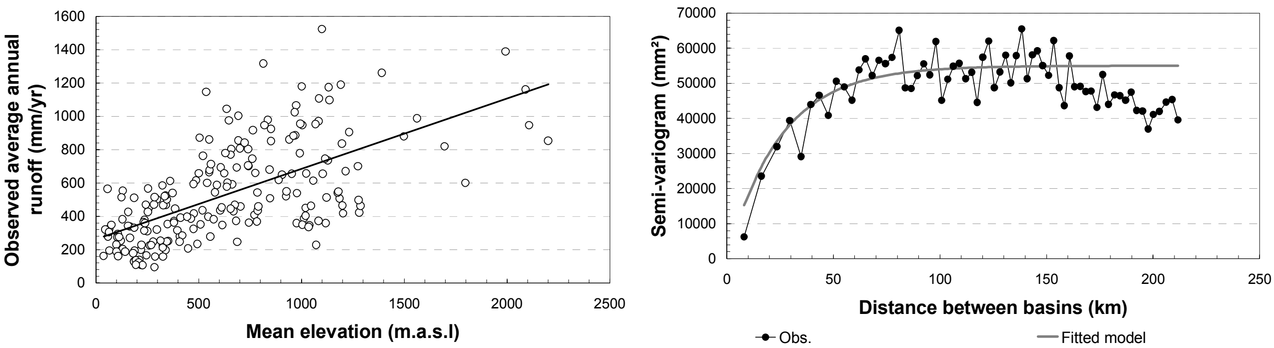

Fig. 2 Empirical relationship between mean annual runoff and mean elevation and semi-variograms for the residual of this relationship for the southwestern part of France.

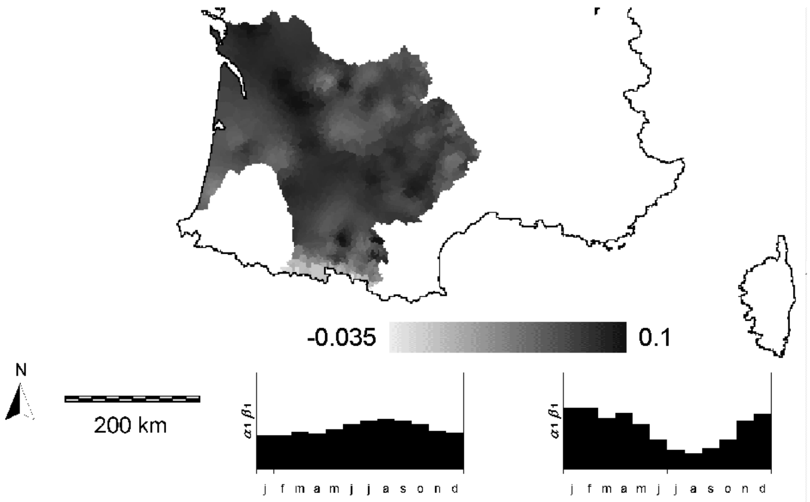

Fig. 3 Map of the first spatial component for the southwestern part of France.

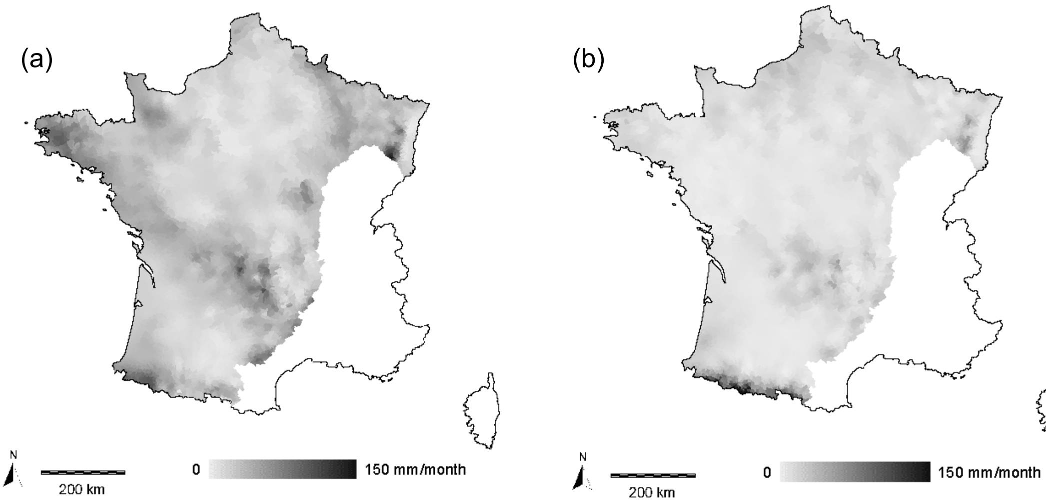

Fig. 4 Map of mean runoff for (a) January and (b) July for France.

while water is abundant in the northwestern 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 the Pyrenean sector. Melting snow at high altitude in the Pyreneans in July generates monthly flow peaks, elsewhere evaporation processes are predominant and cause low flow in the rivers.

Regionalization of flow duration curves, an example from Costa Rica

The DFDC were mapped along the river net using the established strong dependency between the percentiles of the normalized runoff and the coefficient of variation, shown in Fig. 5, together with the interpolation procedures described in the previous chapter for interpolating the mean annual runoff and the coefficient of variation along the river net.

The dependence shown in Fig. 5 allows establishing a linear regression between the values of the variation coefficient and percentiles of the normalized runoff. As example, the relationship between the normalized percentile p = 1% and the coefficient of variation is linear:

$$Z_{p=1\%} = 4.99Cv + 0.12 \quad (R^2=0.93) \tag{18}$$

It is interesting to note that at the duration level of 80% there is almost no dependence on the coefficient of variation, while it is high for, for example, 50%, i.e. the median.

The map of the coefficient of variation over Costa Rica is shown in Fig. 6. This map is consistent with the statistical law for variance reduction of the mean annual runoff (not shown in this paper). Figure 7 shows the Dimensionless Flow Duration Curves (DFDC) for the Caribbean region. It can be noticed that almost all DFDCs

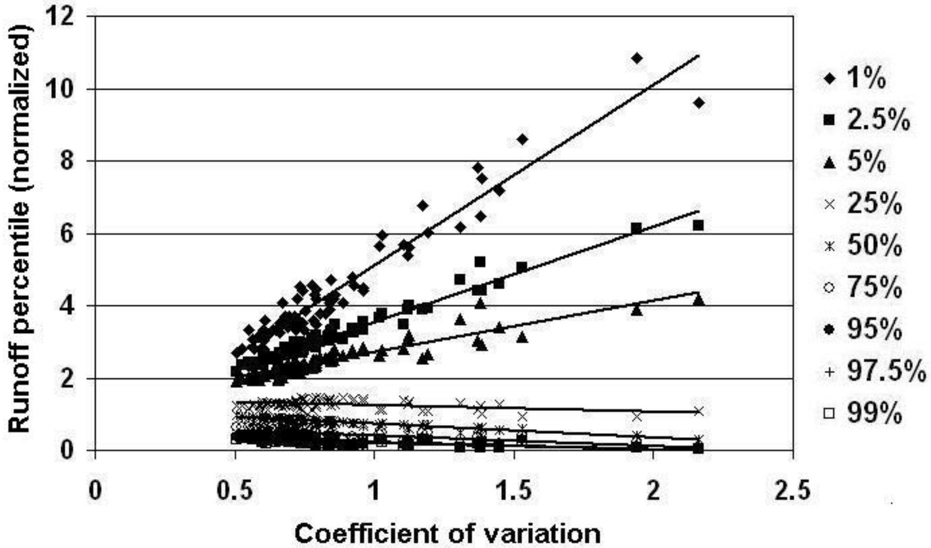

Fig. 5 Relationship between the coefficient of variation and normalized percentiles in Costa Rica.

have a value equal to one when the duration is close to 30%. This property is also observed under other climate conditions.

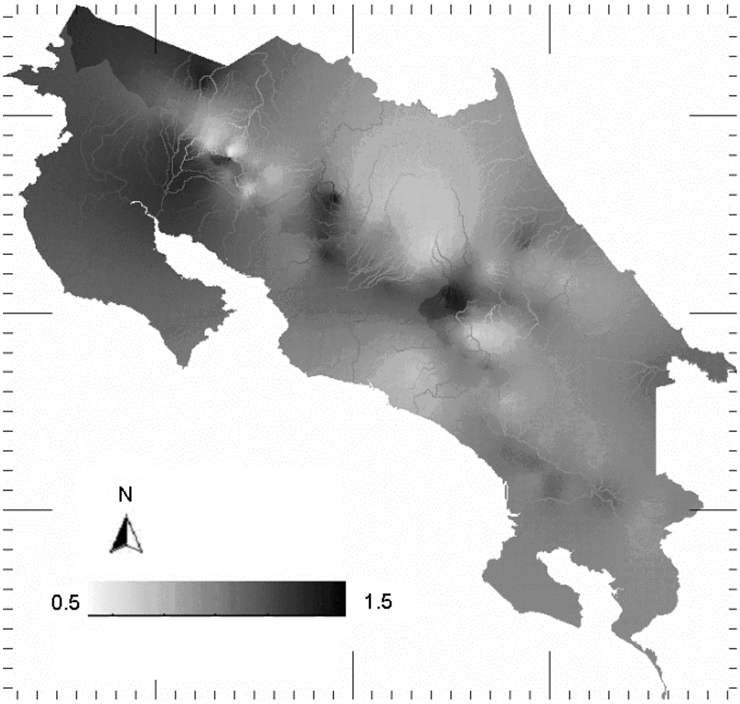

Fig. 6 Map of the coefficient of variation of daily runoff for Costa Rica. The map shows basin values along rivers.

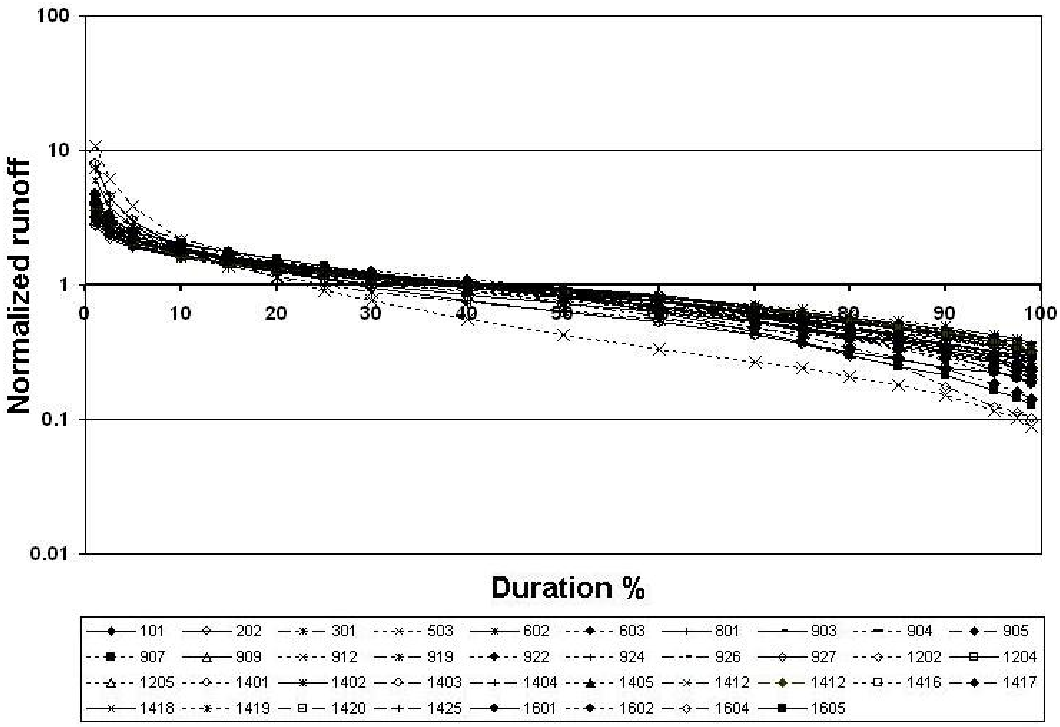

Fig. 7 Observed Dimensionless Flow Duration Curves (Caribbean area, Costa Rica).

Thus, the procedure for estimating the daily runoff corresponding to the percentiles *p*, is as follows: extract the value of the coefficient of variation and the mean annual runoff from the map, apply the relationships between the coefficient of variation and the normalized percentiles and multiply the obtained value by the extracted mean annual runoff value.

CONCLUSIONS

An approach to estimate runoff characteristics at ungauged locations based on a hydrostochastic concept is presented. Runoff is the specific focus for the approach with due consideration that this variable represents an integrated value over a basin, i.e. in theoretical terms a generalized random process in time and space. Furthermore the approach benefits from the fact that different statistical parameters are internally linked by statistical and physical laws. The basic parameters are the long-term mean and the variance. The developments have been demonstrated on two examples from France and Costa Rica. These developments have one common point: the map of mean annual runoff from which monthly pattern and DFDC can be straightforwardly derived. In practice, the same normalized variable is studied at different time scale and its spatial properties are introduced to estimate monthly runoff and percentiles.

The hydrostochastic approach is a “top-down” alternative to “bottom-up” rainfall–runoff modelling and even complements the latter for validation purposes, as rainfall is not involved in the interpolation framework. Comparison of maps derived by each method, respectively, is a convenient way to identify errors in measurements as well as in interpolation or modelling. Noticeable differences between the interpolated values and those deduced from the models may be a sign of deficiency of one method or of inconsistency in the data. Modelling should benefit from these developments, as one major difficulty is to improve the estimate of the parameters of lumped conceptual models.

Acknowledgements The authors wish to express their thanks to the French data base HYDRO and Instituto Costarricense de Electricidad for supplying the discharge data used in this study.

REFERENCES

Creutin, J. D. & Obled C. (1982) Objective analysis and mapping techniques for rainfall fields an objective comparison. *Water Resour. Res*. **18**, 413–431.

Gannett, H. (1912) Map of the United States showing mean annual runoff. In surface water supply of the United States, 1911. *US Geol. Survey Water Supply Papers* 301–312, Government Printing Office, Washington, DC, USA, pt. II.

Ghosh, B. (1951) Random distances within a rectangle and between two rectangles. *Bull. Calcutta Math. Soc*. **43,** 17–24.

Gottschalk, L. & Krasovskaia, I (1998) Development of grid-related estimates of hydrological variables. *Report of the WCP-Water Project B.3, Geneva, WCP/WCA, February 1998.*

Gottschalk, L. (1993) Interpolation of runoff applying objective methods. *Stochast. Hydrol. Hydraul*. **7**, 269–281.

Gustard, A, Roald, L. A., Demuth, S., Lumadjeng, H. S. & Gross R. (1989) *Flow Regimes from Experimental and Network Data (FREND)*, Vol. 1. UNESCO, Paris, France.

Hawley, E. M. & MacCuen, R.H. (1982) Water yield estimation in western United States. *J. Irrig. Drain. Div*. **108,** 25–34.

Holmström, I. (1963) On a method for parametric representation of the state of the atmosphere. *Tellus* **15**, 127–149.

Huang, W. C. & Yang, F. T. (1998) Streamflow estimation using Kriging. *Water Resour. Res*. **34**, 1599–1608.

Krasovskaia, I., Gottschalk, L, Leblois, E. & Pacheco, A. (2006) Regionalization of flow duration curves. In: *Climate Variability and Change—Hydrological Impacts* (ed. by S. Demuth, A. Gustard, E. Planos, F. Scatena & E. Servat (Proc. Fifth FRIEND World Conference held at Havana, Cuba, November 2006), 105–110. IAHS Publ. 308. IAHS Press, Wallingford, UK.

Leblois, E. & Sauquet, E. (2000) Grid elevation models in hydrology—Part 1: Principles and literature review; Part 2: HydroDEM. User's manual. Cemagref, Technical notes, Lyon, France.

Liebscher, H. (1972) A method for runoff-mapping from precipitation and air temperature data. In: *Proc. Symp. World Water Balance* (Reading, 1970), vol. 1, 115–121. IAHS Publ. 92. IAHS Press, Wallingford, UK.

Matheron, G. (1965) Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature. Ed. Masson, Paris, France (in French).

Singh, R. D., Mishra, S. K. & Chowdhary, H. (2001) Regional flow-duration models for large number of ungauged Himalayan catchments for planning microhydro projects. *J. Hydrol. Engng* **6**(4), 310–316.

Sauquet, E., Gottschalk, L. & Leblois, E. (2000a) Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme. *Hydrol. Sci. J.* **45**, 799–816.

Sauquet E., Krasovskaïa, I. & Leblois, E. (2000b) Mapping mean monthly runoff pattern using EOF analysis. *Hydrol. Earth System Sci*. **4**, 79–93.

Skøien, J. O., Merz, R. & Blöschl G. (2005) Top-kriging geostatistics on stream networks. *Hydrol. Earth System Sci.*, Discussions, **2**, 2253–2286.

Solomon, S. I., Denouvilliez, J. P., Chart, E. J., Woolley, J. A. & Cadou C. (1968) The use of a square grid system for computer estimation of precipitation, temperature, and runoff. *Water Resour. Res*. **4**, 919–929.

Vogel, R. M., Wilson, I. & Daly, C. (1999) Regional regression models of annual streamflow for the United States. *J. Irrig. Drain. Engng*, **125**, 148–157.

Yu, P. S. & Yang, T. C. (1996) Synthetic regional flow duration curve for southern Taiwan. *Hydrol. Processes* **10**(3), 373–391.