Regionalization of flow duration curves

IRINA KRASOVSKAIA[1], LARS GOTTSCHALK[1], ETIENNE LEBLOIS[2] & ALEXIA PACHECO[3]

1 *Department of Geosciences, University of Oslo, PO Box 1047 Blindern, N-0316 Oslo, Norway*
irina.gottschalk@geo.uio.no

2 *Cemagref, 3 bis quai Chauveau, CP 220, F-69336 Lyon Cedex 09, France*

3 *Instituto Costarricense de Electricidad, PO Box 10032-1000, San José, Costa Rica*

Abstract A method for estimating the flow duration curve (FDC) in an ungauged basin (i.e. regionalization of FDCs) is developed. An FDC can be described by the mean value and the coefficient of variation either in terms of empirically determined regional FDCs or as fitted theoretical regional curves. The development of these two first order moments of the FDC along the river network and with the spatial scale, i.e. the basin area, is analysed and a scheme for their interpolation along the river network is elaborated. Daily runoff data records from Costa Rica are used to demonstrate the findings. The estimation errors are highest in relative terms (~30%) for the durations longer than 85%, somewhat smaller for durations less than 20% (~10%) and small for central parts of the FDC (~8%). Differences between the empirical and theoretical curves are minor, although the latter give systematically better results especially in central parts of the FDC.

Key words flow duration curve; regionalization; Costa Rica

INTRODUCTION

The Flow Duration Curve (FDC) represents the relationship between the magnitude and frequency of daily, weekly, monthly (or other time interval) streamflow for a particular river basin, providing an estimate of the percentage of time a given streamflow was equalled or exceeded over a historical period (Vogel & Fennessey, 1994). Foster distinguished (1933) two uses of the FDC: (i) as a probability curve to determine the probability of occurrence of future events; and (ii) as a conventional tool to study the data. The FDC, however, gives a static and incomplete description of a dynamic phenomenon. For a complete description a multivariate distribution, which defines the parent distribution of the data, must be considered, the FDC being its marginal distribution. The FDC is a natural start when analysing streamflow data, as evident from its wide practical application (Vogel & Fenessey, 1995; Holmes *et al*., 2002b).

Usually the FDC is founded on daily records (e.g. Holmes *et al.*, 2002b) although exceptions exist when monthly and 10-day data have been used (e.g. Singh *et al*., 2001). Foster (1933) compared daily, monthly and annual FDCs noting that the differences between the curves for different duration (time scale) change with the type of river basin. He handled the problem by constructing a dimensionless flow duration curve (DFDC), dividing all observations by the mean annual flow and assuming that this curve is transferable to an ungauged site, although with caution.

The DFDC is often a starting point for the regionalization of FDCs, i.e. when predicting the FDC within ungauged basins. While some authors mainly concentrate on the properties of empirical FDCs (quantiles) and map these (e.g. Holmes *et al*., 2002a) other studies assume a theoretical distribution (e.g. LeBoutillier & Waylen, 1993; Singh *et al*., 2001). The lognormal, or the more general Box-Cox transformation, are those used most. For regionalization, two approaches are utilised, *viz.* the delineation of homogenous regions (e.g. by means of cluster analysis) and the use of multiple regression. A DFDC is selected for a homogeneous region, defined by, e.g. physical, climatic and streamflow characteristics of basins. Multiple regression is used to explain the geographic variation of each statistic of the FDC (quantiles, distribution parameters). Size of drainage area and the length of the main river course from the divide of the basin to the site of interest are most often used in regression equations. They implicitly indicate a spatial scale dependence of the FDC, an aspect that will be further developed herein.

In a given river basin with its river network the statistical characteristics are related through contextual rules (Gottschalk *et al.*, 2006). Such rules are commonly not defined in regional hydrology, where river basins are rather seen as independent not interacting objects. Neither multiple regression nor cluster analysis, widely used in hydrology for regionalization, considers different parts of the basin as interacting objects. An approach for regionalization of FDCs presented here, acknowledges that statistical characteristics of runoff develop along the river system in a basin from the headwaters to the outlet. By this the regionalization problem is turned to interpolation of runoff characteristics along this network. The approach for regionalization of FDCs is tested on daily runoff data records from Costa Rica.

EMPIRICAL FLOW DURATION CURVES

Daily runoff series for 74 runoff stations in Costa Rica with record lengths over 40 years have been used. The deviations between the empirical FDC proved to be significant, which contradicts Foster's assumption about transferability. Thus it is necessary to go at least one step further considering not only the mean value (the first order moment) but also the second order moment.

The second order moment in terms of the coefficient of variation (*V*) shows a big variation with a range 0.5–2.2 for the analysed daily data. This can be considered by standardizing DFDC, i.e. a subtraction of 1.0 and division by *V*. These standardized FDC (SFDC) might then be used as "regional" duration curves (cf. Fig. 1). It is seen that the average curve (thick line) represents the series well for the majority of durations. Discrepancies shown by a few stations could be explained by specific geology and climate. Figure 2 shows the DFDC-quantiles as a function of the coefficient of variation, *V*. In general a strong dependence noted allows establishing linear regression relationships for all the quantiles used. The dependence on *V*, for example, is high at the duration level 50%, i.e. the median (*Me*), while at Q25% it is almost constant and independent of *V*.

These two ways of considering the role of *V* are quite similar: both assume a linear dependence on *V*, but in the first case the zero intercept is 1. In both cases there might be a risk that for very high values of *V* the linear relations cross the zero line for values with duration near 100%.

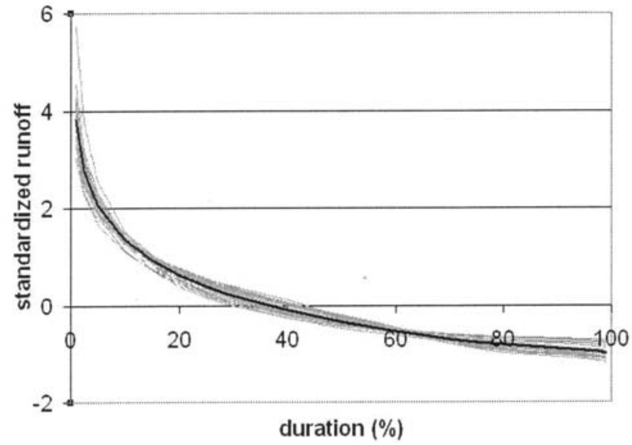

Fig. 1 Regional SFDC (thick line) for the rivers of the Caribbean slope of Costa Rica. DFDCs for individual rivers are given as a background.

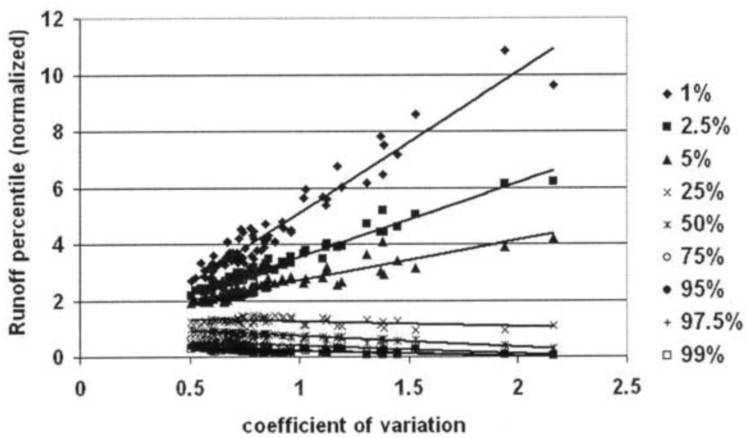

Fig. 2 Relationship between the coefficient of variation (*V*) and normalized runoff (daily values) for Costa Rica.

THEORETICAL FLOW DURATION CURVES

Following Foster (1933), an FDC is a plot of the empirical quantile function Q_p i.e. the pth quantile or percentile of streamflow for a certain duration versus exceedance probability p, where p is defined by:

$$p = 1 - P\{Q \le q\} = 1 - F_Q(q) \tag{1}$$

As noted earlier, many regionalization studies rely on the assumption of a theoretical distribution of FDC. The two most widely used distributions are the 2-parameter gamma and the lognormal. For normalized data both these distributions can be expressed in terms of the coefficient of variation only. For the lognormal we thus have:

$$F_Y(y) = \int_0^y \frac{1}{y\sqrt{2\pi \ln\left(1+V^2\right)}} \exp\left\{-\frac{1}{2}\frac{\left(\ln y + \frac{1}{2}\ln\left(1+V^2\right)\right)^2}{\ln\left(1+V^2\right)}\right\} dy \tag{2}$$

The quantile y_p is determined from $y_p = m_Y + k_p \sigma_Y$ where k_p is the quantile (frequency factor) of the normal distribution. For normalized data (i.e. DFDC) quantile function is:

$$y_p = \exp\left(z_p \sqrt{\ln\left(1+V_Y^2\right)} - \left(\ln\left(1+V_Y^2\right)\right)/2 \right) \tag{3}$$

where z_p is the Gaussian variate of the probability p. For the special case of the median (Me=$y_{0.5}$) for which $z_{0.5}$=0 we find:

$$Me = y_{0.5} = \left(1+V_Y^2\right)^{-1/2} \tag{4}$$

The corresponding cdf of the two parameter gamma distribution for normalized data is:

$$F_X(x) = \int_0^x \frac{1}{V^{2V^{-2}}\,\Gamma\left(V^{-2}\right)} x^{V^{-2}-1} e^{-xV^{-2}} dx \tag{5}$$

The calculation of quantiles x_p must in this case be performed by numerical inversion of the cumulative distribution.

For both distributions the quantiles are thus functions of the coefficient of variation only, as illustrated in Fig 3. The diagrams look very similar in the linear scale with a small difference in the curvature for high values, the gamma having an almost linear growth of the 1% quantile. In a logarithmic scale the difference is more drastic: the gamma distribution declines very quickly towards zero, while the lognormal declines smoothly along concave curves for increasing values of the coefficient of variation.

There is a strong resemblance between the empirical plot in Fig. 2 and the theoretical curves in Fig. 3. For the high flow values (<Q25) the resemblance is highest with the gamma curves while for low flow values (>Q75) the similarity is strongest with the lognormal distribution.

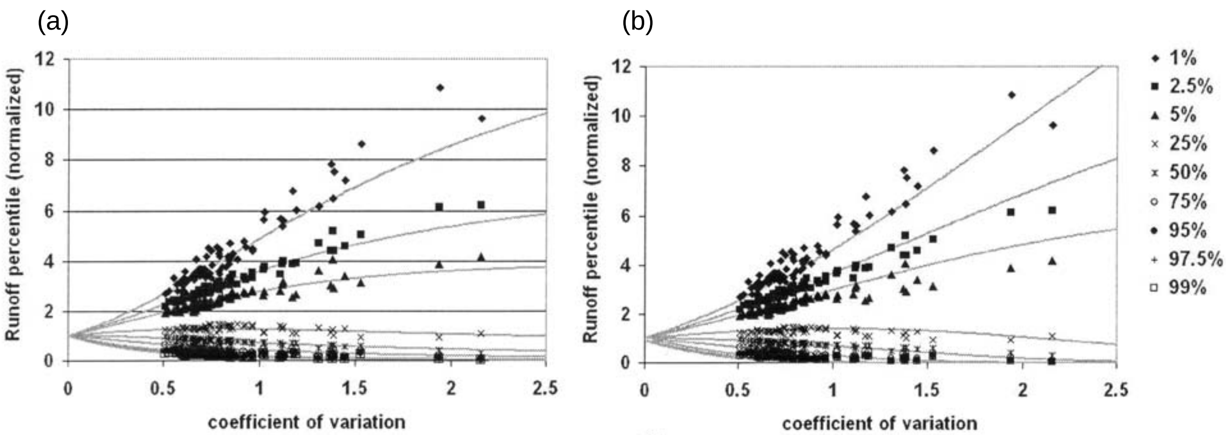

Fig. 3 Quantiles (grey lines) of the lognormal distribution (a) and the gamma distribution (b) as a function of the coefficient of variation.

REGIONALIZATION OF DURATION CURVES

The regionalization problem is here divided into two. The first step is the establishment of an empirical DFDC and/or the parameterization of the DFDC. The analyses referred to earlier indicate a possibility of describing the DFDC through two parameters, namely the long-term annual mean (*MAR*) and the coefficient of variation for daily data (*V*). The second step is to map *MAR* and *V* along rivers. The patterns of spatial variability of the mean and variance of runoff that can be identified from observations are influenced by the fact that observations represent averaged values over drainage basins. To map these quantities, the existence of an instantaneous point runoff (IPR) process is assumed, turning the task into the inverse problem of identifying this process from the observations representing averages in time and space. Then this IPR process is averaged along rivers yielding the desired map.

Mapping the *MAR* value is a rather straightforward task and is basically a problem of stochastic interpolation with local support (or in another vocabulary, block kriging) with a possibility of including a water balance constraint (Gottschalk, 1993b). In the case of Costa Rica, the values at the outlet of the rivers to the ocean are not known and therefore constraints on the water balance were not possible to apply. However, stochastic interpolation with local support is an appropriate method. A complexity in the structure of the covariance/semivariogram for runoff data, representing a mixture of nested and non-nested basins was handled according to the approach by Gottschalk (1993a). The resulting *MAR* map needs to be scrutinized carefully for errors and inconsistencies. Gomez *et al.* (2006) develop this topic further presenting the *MAR* map for Costa Rica.

The variability pattern of the variance identified from observations is still more complex than that for the mean. There are two sources of variability, *viz.* natural variability of the IPR process and variability induced by the variance reduction dependent on the support (i.e. the basin area). This latter part may constitute a significant part of the total variability and it also introduces a dependence on the basin area. The regionalization procedure herein is based on the developments by mainly Gottschalk *et al.* (2006). First, we benefit from the fact that the variation pattern of *MAR* is also reflected in the variance by introducing a new variable, namely the normalized instantaneous point runoff, i.e. IPR divided by the point mean. The covariance function between normalized values $Z(A_1)$ and $Z(A_2)$ over two (basin) averages with areas A_1 and A_2, respectively, is calculated from:

$$\mathrm{Cov}\left[Z(\mathrm{A}_1)Z(A_2)\right] = \frac{1}{A_1 A_2}\int\limits_{A_1}\int\limits_{A_1}\sigma(\mathbf{u}')\sigma(\mathbf{u}'')\rho(\mathbf{u}',\mathbf{u}'')d\mathbf{u}'d\mathbf{u}'' \tag{6}$$

where $\rho(\mathbf{u}',\mathbf{u}'')$ is the point correlation function, $\sigma(\mathrm{u})$ the point standard deviation and with u' representing a point within the basin area A_1 and u'' within the basin area A_2. As we deal with normalized values $Z(A)$ the standard deviation equals the coefficient of variation of the original variable V. The variance is accordingly:

$$\mathrm{V}ar\left[Z(A)\right] = \frac{1}{A^2}\int\limits_{A}\int\limits_{A}\sigma(\mathbf{u}')\sigma(\mathbf{u}'')\rho(\mathbf{u}',\mathbf{u}'')d\mathbf{u}'d\mathbf{u}'' \tag{7}$$

Second, a power relation between *V* and *MAR* of the IPR process is established. Finally, the *MAR* map and the theoretical expression for variance reduction due to local averaging over the basin support (equation 7) is used to produce a map of *V*. The system HydroDem for structuring hydrographical networks (Leblois & Sauquet, 2000) has been used as a background for interpolation. The resulting *V* map (Fig. 4) is consistent with the statistical law for variance reduction and also considers the spatial variability through the dependence of *V* on *MAR*.

APPLICATION

Estimation of the FDC at an ungauged site starts with estimation of the coefficient of variation *V* for this site from the map Fig. 4. Using the dependences illustrated in Figs 2 and 3, either an

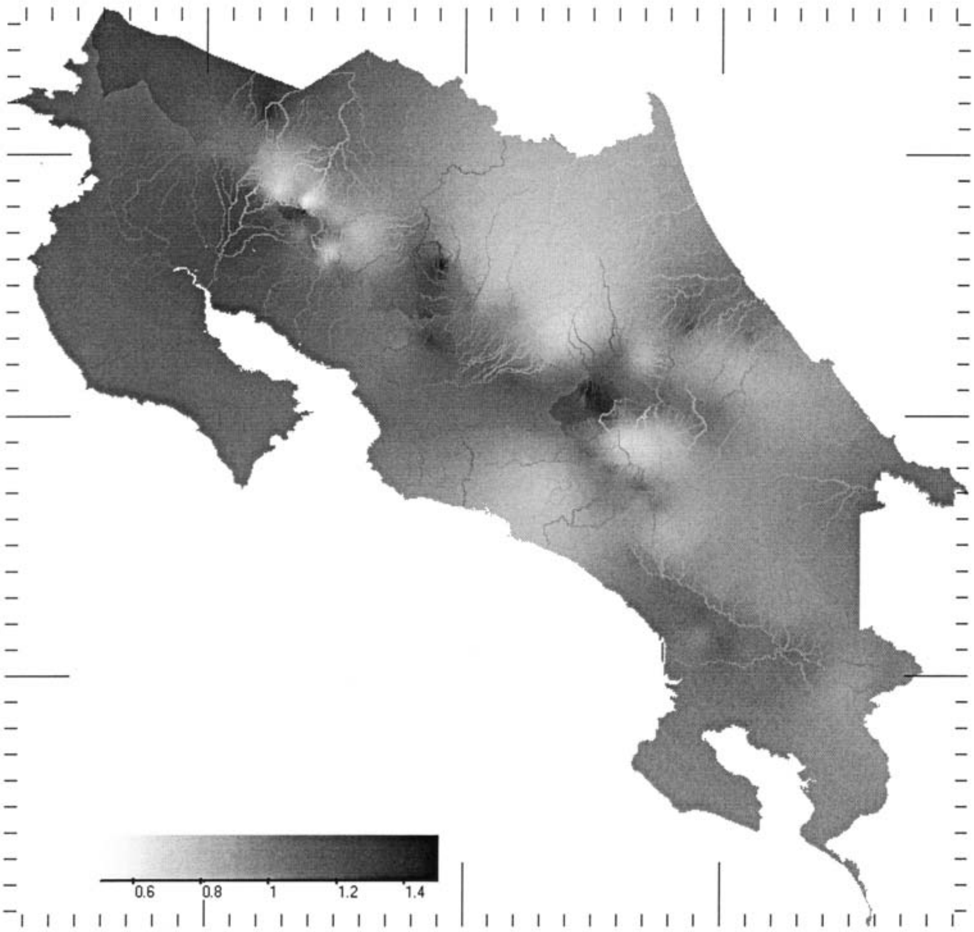

Fig. 4 Map of the coefficient of variation *V* of daily runoff for Costa Rica. The map shows basin integrated values along rivers.

empirical or a theoretical normalized FDC can be estimated. *MAR* is estimated from the corresponding map. Finally, multiplying the normalized FDC by the estimated *MAR* yields FDC at the ungauged site.

Table 1 offers some information about the results. The multiplication by MAR in m^3 or mm $year^{-1}$ yields high correlations between observed and estimated quantiles due to differences in the size of the river basins. Therefore the table shows the normalized values that are more informative with respect to the FDC. As the estimates of SFDC and the theoretical gamma both proved to be inferior, for briefness only estimates of the empirical regional FDC obtained by regression and the theoretical FDC based on the lognormal distribution are compared. The estimation errors are highest in relative terms (~30%) for the percentage points higher than 85%, somewhat smaller for percentage points less than 20% (~10%) and small for the central parts of the FDC (~8%). Differences between the empirical FDC and the lognormal FDC are minor, although the latter gives systematically better results especially in the central parts of the FDC.

Table 1 Estimates of regionalized FDC (normalized values).

Quantile parameter		Q05	Q10	Q25	Q50	Q75	Q90	Q95
Mean		2.53	1.94	1.27	0.77	0.44	0.31	0.28
Standard deviation		0.46	0.25	0.12	0.12	0.12	0.11	0.10
Regional Lognormal FDC	Correlation	0.84	0.59	0.53	0.87	0.68	0.63	0.61
	Abs.std.error	0.25	0.20	0.10	0.06	0.09	0.08	0.08
	Rel.std.error (%)	9.9	10.4	7.9	7.7	20.2	26.0	27.6
Regional emp.FDC, regression	Correlation	0.83	0.53	0.47	0.86	0.64	0.58	0.56
	Abs.std.error	0.26	0.21	0.10	0.06	0.09	0.09	0.08
	Rel.std.error (%)	10.2	10.9	8.4	7.8	21.3	27.3	28.8

CONCLUSIONS

Empirical regional FDC, as well as theoretical FDC based on the lognormal distribution applied here, are standard tools in the regionalization of FDC. A first contribution of this study is the use of not individual but a whole ensemble of FDCs in a basin/region, as shown in Figs 2 and 3. This illustrated that this regional FDC is well described through the two first order moments—the mean annual runoff (*MAR)* and the coefficient of variation (*V*) of daily runoff.

The second contribution of this study is the innovative approach for interpolation of the mean annual runoff (*MAR)* and the coefficient of variation (*V*) of daily runoff that were mapped for all major rivers of Costa Rica. This approach considers the fact that patterns of spatial variability of the mean and variance of runoff reflect observations representing averaged values over drainage basins and furthermore are subordinated through the hierarchical structure of the river network

REFERENCES

Foster, H. A. (1933) Duration curves. *Trans. ASCE* **99**, 1213–1267

Gómez, F., Krasovskaia, I., Gottschalk, L. & Leblois, E. (2006) Interpolation of water balance components for Costa Rica. In: *Climate Variability and Change—Hydrological Impacts* (Proc. Fifth FRIEND World Conf., Havana, Cuba, November 2006). IAHS Publ. 308. IAHS Press, Wallingford, UK (this volume).

Gottschalk, L., Krasovskaia, I., Leblois, E. & Sauquet, E. (2006) Mapping mean and variance of runoff in a river basin. *HESS Discuss*. **3**, 1–35.

Gottschalk, L. (1993a) Correlation and covariance of runoff. *Stochastic Hydrol. Hydraul.* **7**, 85–101.

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

Holmes, M. G. R, Young, A. R., Gustard, A. & Grew, R. (2002a) A new approach to estimate mean flow in the UK. *Hydrol. Earth System Sci*. **6**(4), 709–720.

Holmes, M. G. R, Young, A. R., Gustard, A. & Grew, R. (2002b) A region of influence approach to predicting flow duration curves within ungauged catchments. *Hydrol. Earth System Sci* **6**(4), 721–731.

Leblois, E. & Sauquet, E. (2000) *Grid elevation Models in Hydrology*, Parts 1 & 2. Technical Notes, Cemagref, Lyon, France.

LeBoutellier, D. W. & Waylen, P. R. (1993) A stochastic model of flow duration curves. *Water Resour. Res.* **29**(10), 3535–3541.

Sauquet, E., Gottschalk, L. & Leblois, E. (2000) Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme. *Hydrol. Sci. J.* **45**(6), 799–815.

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

Vogel, R. M. & Fenessey, N. M. (1994) Flow duration curves. I: New interpretation and confidence intervals. *J. Water Resour. Plan. Manage*. **120**(4), 485–504.

Vogel, R. M. & Fenessey, N. M. (1995) Flow duration curves. II: A review of application in water resources planning. *Water Resour. Bull.* **31**(6), 1029–1039.