Climate Variability and Change—Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006), IAHS Publ. 308, 2006.

Regionalization of low flow in Costa Rica

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

1 *C.S. Estudios Básicos de Ingeniería, Instituto Costarricense de Electricidad, PO Box 10032-1000, San José, Costa Rica*
apacheco@ice.go.cr

2 *Department of Geosciences, University of Oslo, PO Box 1047 Blindern, N-0316 Oslo, Norway*

Abstract The starting point of this study is the so-called derived distribution function approach for low flows. This approach is extended to apply to PUT (Pit Under Threshold) data from humid tropical conditions. A general theoretical distribution is derived that applies to exponentially distribute dry spells with a nonlinear low flow recession below a threshold level. Regionalization of PUTs is based on a fixed threshold level—the median, which is mapped for the whole country. The low flow data are normalized with respect to the median. The normalized sample distributions are grouped with respect to characteristic features (form of sample distribution, average length of dry spells and intensity of events) and a theoretical regional distribution is estimated for each grouped sample. The methodology has been applied to a set of low flow data from 63 gauging stations in Costa Rica. Important issues are the definition of the common threshold, the relation between dry spells defined from precipitation and runoff records, respectively, characteristic features for grouping of sample distributions as well as principles for parameter estimation of the regional curves of the grouped samples.

Key words regionalization; low flow; dry spells; Costa Rica

INTRODUCTION

The frequency analysis of low flow data in hydrology has by tradition not differed from such analysis of other types of hydrological data. The task has been to find any theoretical distribution function that best fits to a set of observed low flow data. The so-called derived distribution function approach for low flow (Gottschalk & Perzyna, 1989; Gottschalk *et al.*, 1997) offers an alternative. In this case, the choice of theoretical distributions is limited to a family of distributions derived by combining the theory of extreme minimum values with, and basic understanding of, the generating hydrological processes of low flow. Two theoretical principles are exploited in the study of extremes, namely the block method (BM) (Annual Maximum, AMax; Annual Minimum, AMin) and the threshold method (TM) (Embrecht *et al.*, 1999). For maxima, in hydrology the latter method is called "the Peak Over Threshold" (POT). Minimum values are treated less in the literature than maxima. Here the name "Pit Under Threshold" (PUT) is suggested for this type of hydrological data.

The family of derived distributions for low flow referred to earlier are based on the BM principle. Here this approach is extended to PUT data for humid tropics. A general theoretical distribution is derived that applies to exponentially distributed dry spells τ with a nonlinear low flow recession $q_c = f(q_0, \tau)$, where q_0 is a threshold level. The PUT method has the same advantage compared to the BM method for minimum, as the FPOT method has compared to the BM method for maxima, namely it guarantees that the selected events will always be extreme, provided the threshold is well defined. The parallels between PUTs and POTs can, however, not be brought too far, at least for application in hydrology. While maxima above a threshold might be well described as sudden outbursts above this threshold, minima below a threshold is produced by the emptying of a reservoir during a dry spell.

Regionalization of POT data is done by firstly normalizing the individual series by a so-called index flood (mean or median annual flood) and secondly joining all the normalized series into one sample and fitting a regional theoretical curve to this sample. The selection of the critical threshold is usually done individually for each series with the criteria of having equal intensity of events per year λ for all series (say λ = 2–3). Technically this may be an advantage, but at the same time fundamental information is lost about the process compared to the situation of using a common threshold and allowing λ to vary.

Regionalization of PUTs may be carried out in exactly the same manner as for POTs, viz. defining a normalizing "index low flow" and constructing a regional low flow curve. This approach has also been attempted here but yielded a poor result, as it was very difficult to find a regional pattern in the variation index low flow. The alternative developed herein takes advantage of the fact that the derived theoretical distribution gives a direct indication of which processes influence the parameters. The following basic steps are followed:

- a fixed threshold level q_0 is used, determined from the flow duration curve and this threshold also defines the upper bound of the low flow distribution;
- the low flow data are normalized with respect to q_0;
- the normalized sample distributions are grouped with respect to characteristic features (form of sample distribution, average length of dry spells m_τ, intensity of events λ);
- a theoretical regional distribution is estimated for each grouped sample.

This methodology has been applied to a low flow data for 63 gauging stations in Costa Rica, with an average record length over 40 years. The variation patterns of the distribution of dry spells from a sample of 204 precipitation stations for the same observation periods have been jointly analysed. The presentation starts with the derivation of the theoretical PUT distribution. This distribution is then used as a tool to analyse the variation in the pattern of sample low flow distributions across Costa Rica. The key issues in the study are the definition of the common threshold, the relation between dry spells defined from precipitation and runoff records, respectively, characteristic features for grouping sample distributions, as well as principles for parameter estimation of the regional curves of the grouped samples.

A DERIVED DISTRIBUTION FOR LOW FLOW PUTS

A power law distribution for low flow appears when combining an exponential distribution $F_T(\tau) = 1 - \exp(\tau/m_T)$ for the length of dry spells τ with a simple exponential recession curve for low flow $q_c = q_0\ \exp(-\tau/K)$ of the form:

$$F_Q(q) = (q/q_0)^{K/m_T}\ ;\ 0 \le q \le q_0 \tag{1}$$

This distribution will be accepted as a possible basic model for PUTs, although it is limited to the situation of a relatively quick decay towards zero. In situations with a recession with a long tail, i.e. rivers with a large contribution of deep groundwater during dry spells, a nonlinear recession might be more appropriate. The point of departure in this latter case is the discharge q from a non-linear storage S: $q = K^{-1}S^c$ where K and c are parameters. The outflow function in a decay situation (with no inflow) can in this case be calculated as: $q_t = q_0(1 + t/ba)^{-a}$ where $b = K^{\frac{1}{a}-1} q_o - 1/a$ and $a = c/(c - 1)$. The expression for a linear reservoir with an exponentially decaying recession $q_t = q_0 \exp(-t/K)$ is derived as a limiting form for $c \to 1$ and $a \to \infty$, respectively. Assuming an exponential distribution of the length of dry spells as above and introducing the inverse $t = ba\left((q_\tau/q_0)^{-1/a} - 1\right)$ in its expression results in the following theoretical distribution for PUT low flow values:

$$F_Q(q) = \exp\left(-\frac{ba}{m_T}\left((q/q_0)^{-1/a} - 1\right)\right);\ \ 0 \le q \le q_0;\, a < \infty \tag{2}$$

It is important to note that both equations (1) and (2) are truncated distributions. If the number of events observed over a time interval follows the Poisson distribution and the intensity of PUTs per time interval (year) is λ, the general expression for the distribution of annual minima (AMin) Ξ under a threshold is $F_\Xi(\xi) = \exp\{-\lambda F_Q(\xi)\}$. Insertion of $F_Q(q)$ in the latter expression leads to the distribution:

$$F_\Xi(\xi) = \nu^{-1}\left(1 - \exp\left\{-\lambda\left(\frac{\xi}{q_0}\right)^{K/m_T}\right\}\right);\, 0 \le \xi < q_0 \tag{3}$$

This distribution is identified as a truncated Weibull distribution for minima where

$\nu = \left(1 - \exp\{-\lambda\}\right)$ is the level of truncation. It was first introduced by Bernier (1964). It is important to note that the power in the parent distribution is brought forward in the extreme value distribution to determine the behaviour of the tail for the Weibull type of distributions for minima. For the nonlinear case we can proceed accordingly to derive an AMin-distribution:

$$F_{\Xi}(\xi) = \nu^{-1}\left(1 - \exp\left\{-\lambda \exp\left(-\frac{ba}{m_T}\left(\left(\xi/q_0\right)^{-1/a} - 1\right)\right)\right\}\right);\ 0 \le \xi < q_0 \qquad (4)$$

The distribution equation (3) can be compared to the truncated Weibull AMin-distribution proposed by Gottschalk *et al.* (1997) for low flow with origin in Gumbel distributed maximum dry spells with parameters u and α and a linear recession with the recession coefficient K:

$$F_{\Xi}(\xi) = \nu^{-1}\left[1 - \exp\left\{-\left(\frac{\xi \exp(u/K)}{q_0}\right)^{K/\alpha}\right\}\right];\ 0 \le \xi < q_0 \qquad (5)$$

where $\nu = P\{Q > q_0\} = 1 - \exp\left(-\exp(u/\alpha)\right)$, is the level of truncation. Both are truncated Weibull distribution with effectively two parameters but with different interpretations. The difference is in the fact that the latter one is directly formulated for AMin data.

APPLICATION—REGIONAL CURVES OF LOW FOR COSTA RICA

When applying the distribution derived in the previous paragraph with the aim of analysing the patterns of low flow formation across Costa Rica, the four steps indicated in the introduction were followed.

The choice of the threshold level is a very important issue for the selection of extremes. To be in accordance with the developed theory the threshold must be defined so that it not only allows identification of extreme low flow but also the sample of extreme dry spells (duration of recession). Dry spells were also identified from the rainfall records and for each drainage basin an average “rainfall dry spell” was estimated and compared with the dry spell from a corresponding runoff record. Furthermore the sensitivity of the statistical properties (mean and variance) of dry spells, low flow, and recession coefficients to the choice of the threshold level was analysed with a hope that they would stabilize at some level. This was not the case in general. The first criterion became decisive and indicated a rather high level of the duration curve, so the median $q50$ of the daily records was finally accepted as the proper threshold level for the study. The median was mapped for all Costa Rica in accordance with the methodology presented by Krasovskaia *et al.* (2006). The values develop along the river system and are consistent with the statistical laws for how the mean value and the variance change with the increasing support (basin area) when moving downstream.

PUTs for the dry period (“verano”—December to May) were extracted from each of the original daily records including low flow and length of dry spell (recession duration) in accordance with the illustration in Fig. 1. The low flow values were normalized with respect to the median of each daily record. The PUT normalized sample distributions for each station were plotted using the Gringorten plotting position in log–log diagrams. If data follow the theoretical distribution originating from a linear recession curve equation (1) they should plot as a straight line with maximum at zero ($\ln(q/q_0) = 0$ for $q/q_0 = 1$) with the slope (K/m_T). Figure 2 shows examples of sample distributions. It is seen that straight lines are exceptions. The rule is rather a nonlinear recession giving rise to a long tail of the distribution. This indicates a significant contribution of base flow. Even for very long dry spells the majority of rivers do not dry out but are fed by a considerable groundwater discharge. Only seven (Fig. 2, left shows one such station) of the total of 63 showed an approximately linear behaviour with a quick diminishing of low flow into the dry spell. All must be considered as outliers in otherwise quite regular behaviour. Besides these outliers, five record samples showed erroneous behaviour as seen in Fig. 2 (left) for three stations where 1 to 3 of the smallest observations are very close to zero.

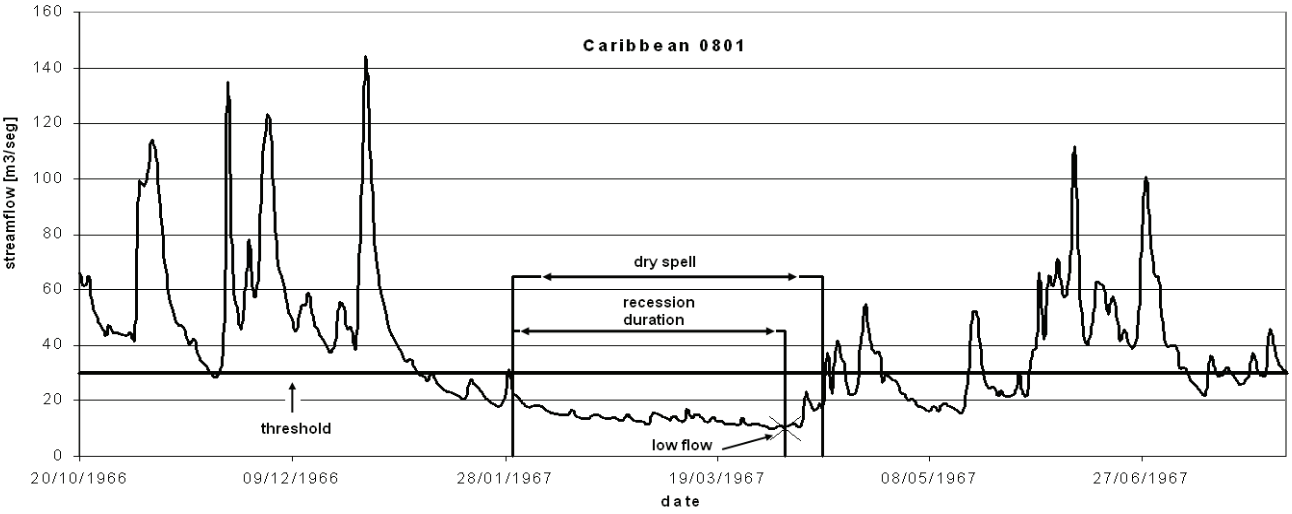

Fig. 1 PUT characteristics extracted from the daily flow record, example for a gauging station in the Pacuare River on the Caribbean slope.

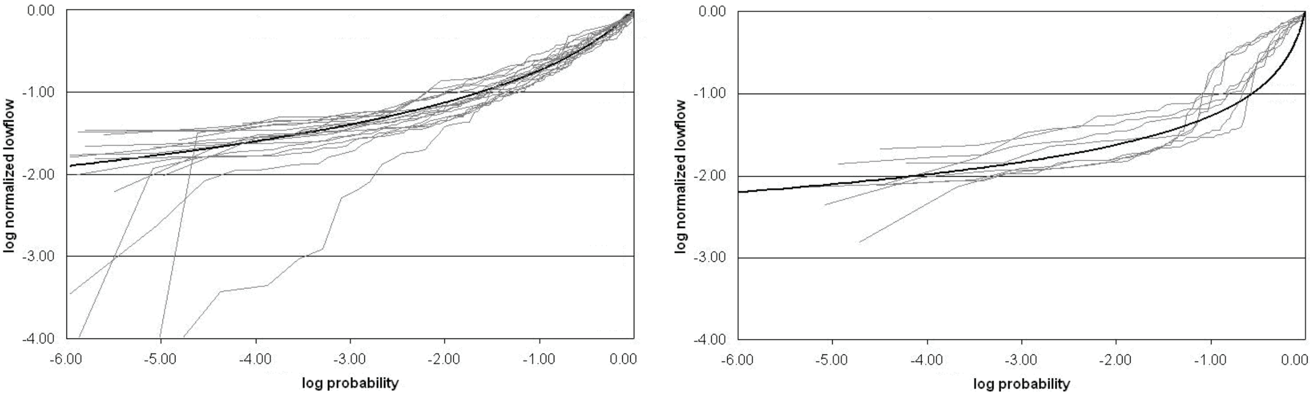

Fig. 2 Sample distributions of PUT low flow for the North Caribbean (left) and the Central Pacific (right). Estimated regional curves are shown by a bold line.

Table 1 Characteristics of regional low flow curves for Costa Rica.

Region	No. stations	No. events	λ	σ_λ	m_T	σ_T	a	$\frac{ba}{m_T}$
S. Caribbean	5	851	6.13	0.89	19.4	6.4	0.713	0.704
N. Caribbean	13	1950	5.51	0.89	26.7	6.5	0.855	0.850
Central Valley	13	1555	3.60	0.72	32.5	13.3	0.813	1.140
N. Pacific A	9	860	3.50	0.95	50.2	16.2	1.349	1.184
N. Pacific B	5	219	2.90	0.87	60.8	14.9	1.269	0.502
Central Pacific	8	529	2.44	0.47	67.3	12.8	0.542	0.195
S. Pacific	10	865	2.77	0.27	57.1	6.3	0.573	0.164
Total	63	6829						

The examples of sample distributions shown in Fig. 2 have already been grouped. Visual inspection of the form of the sample distributions allowed easy manual grouping of stations with similar forms. The statistics of λ and τ were also used as controls. Seven groups were identified in this manner and their characteristics are given in Table 1. The groups are coherent in geographical space and first of all reflect the different character of the dry seasons in Costa Rica, especially between the Caribbean and Pacific slopes. On the Caribbean side short rain episodes are frequent, also during the dry season. This gives rise to several low flow events and relatively short dry spells. The distribution of dry spells fits perfectly with the exponential distribution, which also explains the very smoothly decaying form of the sample distributions shown in Fig. 2 (left) for the North Caribbean region. On the Pacific side, on the other hand, it is typical, with long dry spells up to several months. Figure 2 (right) shows sample distributions from the Central Pacific region with a very sharp bend indicating two populations of dry spells, the shorter ones being of minor interest.

There is a very clear tendency among the identified regions: starting with a high λ and a low m_T for the South Pacific region, then with a gradual decrease in λ and a corresponding increase in m_T when first going north along the Caribbean coast, turning west to the Pacific slope and finally south along the Pacific coast. The regions thus coincide with the climatic regions of Costa Rica. It should be noted that the outlying stations with a linear recessions do have the same dry spell characteristics as the region they are placed in. The, in general, very high baseflow component of runoff far into a dry spell gives evidence of a large contribution from groundwater flow for the majority of stations, while the outlying stations are exceptions in this respect with low baseflow. The outliers are mainly found in the South Caribbean and North Pacific regions, respectively. Consulting a hydrogeological map (SENARA, 2000) of Costa Rica these two regions are partly covered by hydrogeological formation classified as “without potential” while the rest of the country is classified as “high or moderate potential”. The outlying behaviour is interpreted as an effect of the local hydrogeological conditions, but more detailed studies are necessary to confirm this. For the time being the regional curve must be used carefully with the awareness that local factors can have a significant influence.

In the introduction it was mentioned that the PUT (and POT) method guarantees that the selected events always will be extreme provided the threshold is well defined. The threshold used is quite high and this was motivated by the fact that the identified dry spells from runoff records should agree with corresponding dry spells identified from rainfall records. The population of low flow events identified thus includes a wide range of periods of runoff recession, all not really extremes. For the samples from the Caribbean slope the assumption of an exponential distribution of dry spells was very well satisfied. This also results with the fact that the whole population of events, also the not very extreme ones, agrees very well with the theoretical distribution. For the samples from the Pacific slopes this is not the case, especially for those from the Central and South Pacific regions. Here the assumption of exponentially distributed dry spells is not satisfied and it was already commented on the existence of two populations of dry spells, long and short. It is not possible to include both populations in one theoretical model. The problem is solved by using censored samples for parameter estimation for these two regions.

We note that the theoretical distribution equation (3) effectively contains two parameters: q_0, (K/m_T); and the distribution equation (4)–three: q_0, a, ba/m_T. The theoretical moments of the distributions in terms of these parameters have been derived but are not shown here. Expressions for the l-moments, on the other hand, still wait to be investigated. For the time being, we have chosen to apply a least square method to fit the theoretical curves to that of the sample distributions. The same method has been used to fit samples for individual stations and for the regionalized samples. For two regions (CPa and Spa) censored samples were used below an event probability of 0.368 (= exp(−1)), as has already been commented on. The regional parameter values are shown in Table 1.

It is important to note that the graphs in Fig. 2 show event probabilities and not annual probabilities. They thus do not illustrate the variation pattern of the frequency of low flow across Costa Rica, as the number of events for each regional sample is so different. As a final step, the regional curves can be transferred to annual frequencies by applying equation (4). Figure 3 illustrates the seven AMin regional curves confirming that the Northern Pacific is the driest region and that the rivers on the Caribbean slopes have relative high low flow also during the dry period. It is worth mentioning that attempts were made to perform a regionalization directly based on the AMin distribution equation (5) with less successful results.

CONCLUSIONS

Regionalization of low flow has been approached in an innovative way by a joint analysis of information about frequency of low flow and dry spells and the recession. The derived distribution

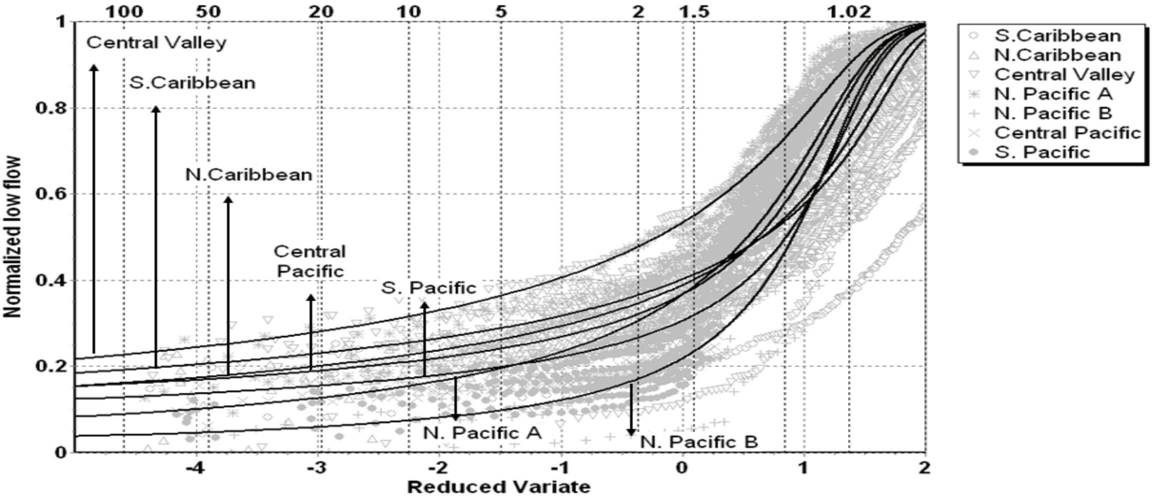

Fig. 3 Regional AMin curves for Costa Rica.

function approach used limits the choice of theoretical distributions to a family of distributions derived by combining the theory of extreme minimum values with a basic understanding of the generating hydrological processes of low flow. The derived theoretical distribution gives a direct indication of which processes influence the parameters.

Unlike many other studies on regionalization, a fixed threshold level determined from the flow duration curve was used, which offered a unique possibility to discern the differences in the processes behind the differences in frequencies of dry spells, called here “PUTs”, i.e. “Pits Under Threshold”. The median $q50$ was adopted as a threshold to be able to match dry spells identified from runoff data and rainfall.

It was possible to group the normalized sample distributions with respect to the form, the average length of dry spells m_τ, and the intensity of events λ. On the Caribbean side, frequent short rain episodes also during the dry season gave rise to many low flow events and relatively short dry spells, perfectly following the exponential distribution. On the Pacific side with several months-long dry spells, two populations of dry spells could be discerned. The obtained regional curves can be transferred to annual frequencies of low flow AMin.

The strength of the suggested approach is in its coherence with the theory of extreme events and possibility to discern the processes behind the occurrence of low flow extremes as well a possibility of consistent mapping of low flow values in a river basin.

REFERENCES

Embrechts, E., Klüppelberg, C. & Mikosch, T. (1999) *Modelling Extremal Events, For Insurance and Finance*, second edn. Springer Verlag, Germany.

Gottschalk, L. & Perzyna, G. (1989) A physically based distribution function for low flow. *Hydrol. Sci. J.* **34**, 559–573.

Gottschalk, L., Tallaksen, L. & Perzyna, G. (1997) Derivation of low flow distribution function using recession curves. *J. Hydrol.* **194**, 239–262.

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 *et al.*) (Proc. Fifth FRIEND World Conference, Havana, Cuba, November 2006). IAHS Publ. 308. IAHS Press, Wallingford, UK (this volume).

SENARA (2000) *Hydrogeological Map of Costa Rica*. SENARA, Costa Rica.