Chapter 5

Modelling Characteristics of Low-Flows with Time-Dependent Data

5.1 Introduction

An accurate analysis of low-flow regimes occurring in a given stream or river is of crucial importance in watershed management because of the following reasons. Firstly, low-flows will constrain the design of key infrastructure facilities such as water supply and irrigation systems, river navigation networks, and hydroelectric power plants. Secondly, they will indicate to the water-manager the maximum levels of BOD[1] and the maximum allowable concentrations of other pollutants (e.g. Hg, Pb, Zn, P, N, Rn) that should not be reached in a given stream so that its ecosystem will not be jeopardized or damaged during a drought period.

In general, longer low-flow periods will increase investment costs of a given infrastructure facility in a non-linear way. Additionally, an erroneous estimation of such regimes will cause substantial economic losses for a region since the water shortage will hamper production processes.

In order to better understand this phenomenon, it is necessary to determine the most likely period of the year when it may occur. In other words, this means that the temporal distribution of discharge and precipitation should be determined within a given domain (i.e. a basin) for different time intervals during a water year, say months. By knowing these two observables and assuming that the annual change of underground storage is insignificant, the basic form of the water budget for a given spatial unit Ω_i during a given time interval t can be determined as

$$\left\langle P_i^t \right\rangle - \left\langle Q_i^t \right\rangle - \left\langle \mathcal{V}_i^t \right\rangle - \cancel{\left\langle \Delta S_i^t \right\rangle}^{\approx 0} = 0\,, \tag{5.1}$$

where the variables P, Q, $\mathcal{V}$, and ΔS stand for precipitation, discharge, evapotranspiration, and change of underground storage. The operator $\langle \cdot \rangle$ represents the integral of a given variable over the spatial domain Ω_i and/or during the time interval t (e.g. one month). This equation must hold everywhere because it represents the principle of conservation of mass within the system. The results of (5.1) can then be averaged in order to have an unbiased estimator for each variable at a given time

---

1 Biochemical Oxygen Demand (BOD) refers to the amount of oxygen that would be consumed if all the organics in one litre of water were oxidized by bacteria and protozoa (ReVelle and ReVelle, 1988).

interval (e.g. for January). The result of such a procedure for the Study Area is depicted in Figure 5.1. This graphical presentation shows that the most probable low-flow spells would take place during summer (M, J, J, A, S, O), in which the evapotranspiration will increase because of higher air temperature; which in turn will reduce the river discharge although the precipitation has increased within its basin.

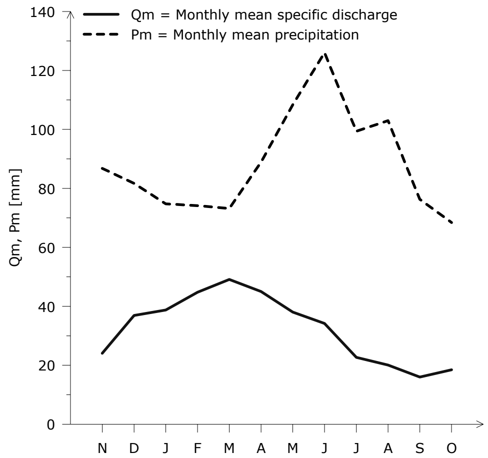

Figure 5.1 Annual water balance of the Study Area. Each value is computed over the period 1961 to 1993.

Because of this fact, the present study will only consider low-flow spells that happen during summer. Having defined the time span for the study of low-flows, the following question can be formulated in connection with the general aim of the present research: are the land cover changes that have taken place within the study area influencing in some way the probabilities of occurrence and/or the total duration of low-flow events?

In order to answer this question, the available information will be first described.

Table 5.1 Correlation matrix $[\mathbf{R}]$ among explained variables Q_{13}, Q_{14}, Q_{15}, and Q_{16} and some climatic explanatory variables in summer. The sample size is equal to 860.

	Q_{13}	Q_{14}	Q_{15}	Q_{16}	x_{25}	x_{31}	x_{37}	x_{38}
Q_{13}	1					Symmetric		
Q_{14}	0.771	1						
Q_{15}	0.291	0.345	1					
Q_{16}	0.633	0.711	0.633	1				
x_{25}	-0.335	-0.587	-0.113	-0.251	1			
x_{31}	0.103	0.246	0.036	0.130	-0.312	1		
x_{37}	0.143	0.252	0.060	0.142	-0.297	0.846	1.	
x_{38}	0.664	0.849	0.297	0.587	-0.569	0.228	0.241	1.

5.2 Description of Time-Dependent Variables

The time series depicted in Figure 5.2 for catchments No. 11 and No. 13 as well as the correlation matrix shown in Table 5.1 point out the same fact: the explained variables $\{Q_l \ \forall\, l = 13,\ldots,16\}$ are mutually correlated.

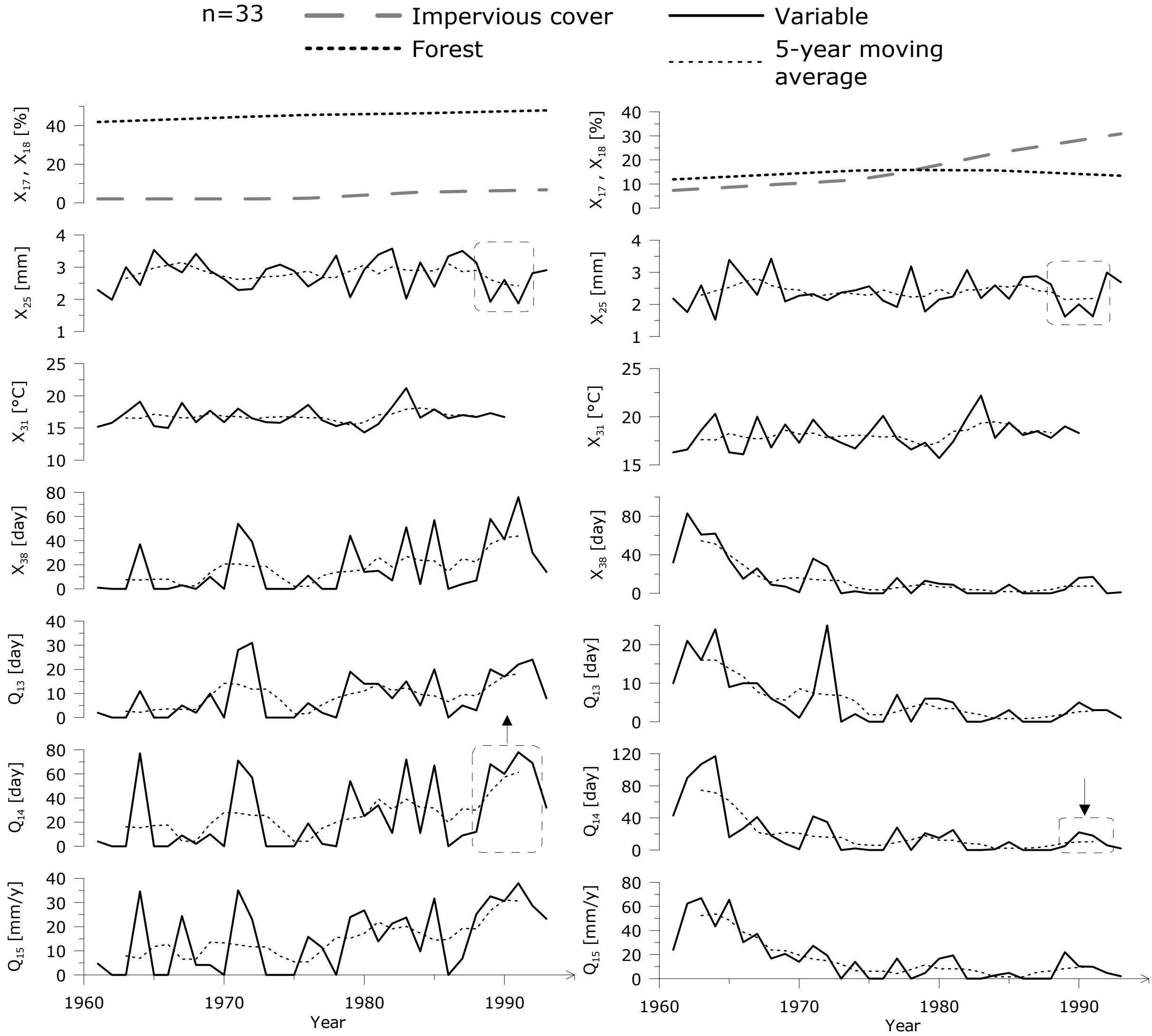

Figure 5.2 Time series showing the trends (by means of a 5-year running average) and the actual values for variables Q_{13}, Q_{14}, and Q_{15} as well as for some explanatory variables including land cover for two catchments of approximately the same size ($A \cong 125\,[\mathrm{km}^2]$). On the left panel is catchment No. 11 with growing shares of forest and impervious cover; whereas on the right panel is catchment No. 13 which has endured a steady land use transition from grassland (permeable land cover) to settlement (impervious land cover) and a steady decline of forest since the mid 70s. The data shown here correspond to the period from 1961 to 1993.

Hence, it would be sufficient to model one of them in order to give an answer to the previous question. An evident selection will be the total drought duration Q_{14} since it exhibits the higher correlations not only with all potential predictors but also with the rest of the explanatory variables.

The following remarks can be stated based on the time series shown in Figure 5.2.

- Firstly, that both the daily mean precipitation (for summer) x_{25} and the daily mean air temperatures in July x_{31} do not reveal any significant trend;
- Secondly, that the composed variable x_{38}, which accounts for days with dry circulation patterns and a decreasing antecedent precipitation index exhibits a completely different behaviour

depending on the shares of land cover while almost constant seasonal-mean precipitation and temperature have been observed;

- Thirdly, that the variable x_{38} is highly correlated with total drought durations (Q_{14}), and;
- Finally, that the land cover variables seem to have played a significant role in the duration of low-flows at mesoscale level, especially the shares of forest and/or impervious areas. As is shown in Figure 5.2, a combination of a rapid growth of impervious cover accompanied by a decline of forest may have led to a rapid shrink of the total drought durations; conversely, slightly growing shares of forest and impervious areas may have led to an increase of total drought durations. In other words, the explained variable Q_{14} has been attenuated by land cover variables. The rectangles with dashed line shown in Figure 5.2 illustrate this fact, i.e. the same climatic phenomenon (a dry year) may produce different outcomes depending on the land cover situation within the catchment, as well as on its morphology.

The distribution of the explained variable Q_{14} is positively skewed (skewness = 1.45) with its mode and median occurring at 0 and 6 [day] respectively. Hence, three positively skewed theoretical distributions from the exponential family, specifically the exponential, the gamma, and the Weibull distributions, were fitted to the observations using the maximum likelihood method. The χ^2 test statistic obtained for each fit was 1.2, 2.5, and 1.05 respectively. Based on this statistical test, whose $p\text{-value} = 0.31$, it is possible to assume that the data can be modelled using a Weibull distribution with a shape factor $\mathrm{a} = 1.035$ and a scale factor $\mathrm{b} = 29.708$. The sample also indicates that the explained variable is heteroscedastic with regard to one of its predictors, namely x_{38}. Additional statistics of the explained variable can be found in the Appendix 4.

5.3 Total Drought Duration

Using the procedures described in Chapter 4 it was found that the most significant variables to explain Q_{14} are $\{x_j \; j = 1,7,8,9,10,13,16,17,18,19,31,38\}$. In this case, forest and permeable cover have been evaluated at basin level whereas impervious cover has been evaluated within the floodplains and buffer zones of the stream network, i.e. $\{x_j \; \forall j = 17,19 \; \wedge \mathcal{L}_i \equiv \Omega_i \;\; \forall i = 1,\ldots,46\}$ and $\{x_j \; \forall j = 18 \; \wedge \mathcal{L}_i \equiv \mathcal{B}_i \subset \Omega_i \;\; \forall i = 1,\ldots,46\}$.

Using these twelve variables and a sample with 752 observations, all possible combinations of predictors have been calculated using the following model with three variants (predictors). Explicitly it can be written as

$$
\begin{aligned}
& Q_{i14}^t \sim \mathrm{Weibull}(\mathrm{a}, \mathrm{b}_i^t) \quad Q_{i14}^t > 0 \quad \forall i,t \\
& E\left(Q_{i14}^t\right) = \begin{cases} 0 & x_{i38}^t = 0 \\ \mu_i^t & x_{i38}^t > 0 \end{cases} \\
& \mathrm{var}\left(Q_{i14}^t\right) = \begin{cases} 0 & x_{i38}^t = 0 \\ \kappa\left(\mu_i^t\right)^2 & x_{i38}^t > 0 \end{cases}
\end{aligned}
\tag{5.2}
$$

where

$$\begin{aligned} \mu_i^t &= \mathrm{b}_i^t \Gamma\left(1+\mathrm{a}^{-1}\right) \\ \kappa &= \Gamma\left(1+2\mathrm{a}^{-1}\right)-\Gamma^2\left(1+\mathrm{a}^{-1}\right) \end{aligned} . \tag{5.3}$$

Three predictors that are functions of the explanatory variables will be used. They will be called (POT), and multi-linear potential 1 and 2 (MLP1, MLP2) respectively. They can be written as follows

$$\eta_i^t = \beta_0 \prod_j \left(x_{ij}^t\right)^{\beta_j} , \tag{5.4}$$

$$\eta_i^t = \beta_0 + \sum_{j \in \mathbf{G} \cup \mathbf{U}} \beta_j x_{ij}^t + \beta_{J^*} \prod_{j \in \mathbf{M}} \beta_j \left(x_{ij}^t\right)^{\beta_j} , \tag{5.5}$$

and

$$\eta_i^t = \beta_0 + \sum_{j \in \mathbf{U}} \beta_j x_{ij}^t + \beta_{J^*} \prod_{\substack{j \\ j \notin \mathbf{U}}} \left(x_{ij}^t\right)^{\beta_j} \tag{5.6}$$

respectively. In all cases the link function will be the identity one, so that

$$\eta_i^t = \mu_i^t \qquad \forall i, t . \tag{5.7}$$

Where

$$\begin{aligned} \mathbf{U} &= \left\{x_j \;\; j = 17,18,19\right\} \\ \mathbf{G} &= \left\{x_j \;\; j = 1,7,8,9,10,13,16\right\} \\ \mathbf{M} &= \left\{x_j \;\; j = 31,38\right\} \\ \beta_0, \beta_j, \beta_{J^*} &= \text{coefficients to be optimised.} \end{aligned}$$

Table 5.2 shows the summarised results of applying the proposed method (see Chapters 3 and 4) to the present dataset.

Table 5.2 Robust models for total drought duration in summer (1 = a variable is included in the model, otherwise it is omitted). The estimated deviance as well as results for the cross validation statistic and the Akaike's information criterion is presented. The most robust models are highlighted with the symbol ⌘ . All values are dimensionless.

Summer: $Q_{i14}^t \sim \mathrm{Weibull}(\mathrm{a}, \mathrm{b}_i^t)$

Model No.	x_1	x_7	x_8	x_9	x_{10}	x_{13}	x_{16}	x_{17}	x_{18}	x_{19}	x_{31}	x_{38}	Predictor	Link	κ	AIC	θ	Obs.
3502		1	1		1		1	1		1	1	1	POT	identity	0.611	6293.9	20.42	
3149				1			1	1	1		1	1	MLP1	identity	0.811	6217.8	24.16	
2964	1	1	1			1		1				1	MLP2	identity	0.794	5958.8	7.18	⌘

Table 5.3 Parameter estimates and results of the permutation test (the Monte Carlo p-values with R=500) obtained for the selected model MLP2 No. 2964.

Parameter	β_0	β_{17}	β_{J*}	β_1	β_7	β_8	β_{13}	β_{38}
Estimates	-0.269	0.071	14.658	-0.075	-0.711	-2.126	0.236	0.869
p-value	-	0.045	-	$\simeq 0$	$\simeq 0$	$\simeq 0$	0.016	$\simeq 0$

Table 5.4 Additional quality measures for the selected robust model.

Model No.	Type	Season	E_1 [day]	E_2 [day^2]	E_3 [day]	E_4 [$-$]	E_5 [day]	E_6 [$-$]	E_7 [$-$]
2964	MLP2	Summer	0.00	175.83	13.26	0.45	9.26	0.31	0.86

During the process of selection of predictors and for a given model type, it has been observed that some variables appear always or very often as elements of the subset of the best models. This fact is illustrated in Table 5.2 where variables x_{17} and x_{38} have been always present. The selected model whose variables are all significant at 5% not only indicates that the total drought duration within a catchment primarily depends on the macroclimatic conditions represented by the variable x_{38}, but also that the morphology of the catchment and the land cover will play an important role. This evidence also provides valuable support to the remarks presented above in Section 5.2. Hence, variables such as mean terrain slope in the buffer zones of streams x_7, drainage density x_8, and share of forest cover x_{17} should be taken into account when watershed management plans are carried out.

The behaviour of the water system concerning the total drought duration appears to have complex and non-linear relationships with the observables or predictors. The following reasons help to corroborate this statement. On the one hand, the variable x_{38} has a nonlinear relationship with the explained variable (see Table 5.2), which not only depends on the macrocirculation patterns but also on the antecedent precipitation index. The latter, which is an indicator of the soil moisture, is, in turn, directly related to the share of forest within a catchment and inversely related to the share of impervious areas. On the other hand, the share of forest appears as a linear predictor of the explained variable, too (see Table 5.2). Such a complex relationship makes the analysis of low flows more complicated to model.

Fortunately, using the proposed method, a model composed of six predictors out of twelve potential ones has been found, i.e. model MLP2 No. 2964. It has not only a correlation coefficient of 0.86 between the observed and calculated total drought duration, but it also exhibits the smallest Jackknife statistic (7.18) compared with other potential robust models (see Table 5.2). In addition to that, the model's output largely supports the presumption that the explained variable has been drawn from a Weibull distribution, as can be seen in the Q-Q plot of Figure 5.3, although deviations are accounted at the right tail of the distribution.

A Q-Q plot is a scatterplot, in which '*each coordinate pair consists of a data value and a corresponding theoretical estimate for that data value derived from the empirical cumulative probability estimate*' (Wilks, 1995). The empirical cumulative probability estimate for the i^{th} smallest data value will be assumed to be equal to $p(x_{(i)}) = i/(n_0+1) \approx \Pr\{X \leq x_{(i)}\}$, where n_0 is the sample size. Hence, the i^{th} coordinate pair of the Q-Q plot in the present case is given by $[x_{(i)}, F^{-1}(p(x_{(i)}))]$, where $F^{-1}()$ represents the fitted inverse Weibull CDF with parameters a and b given above.

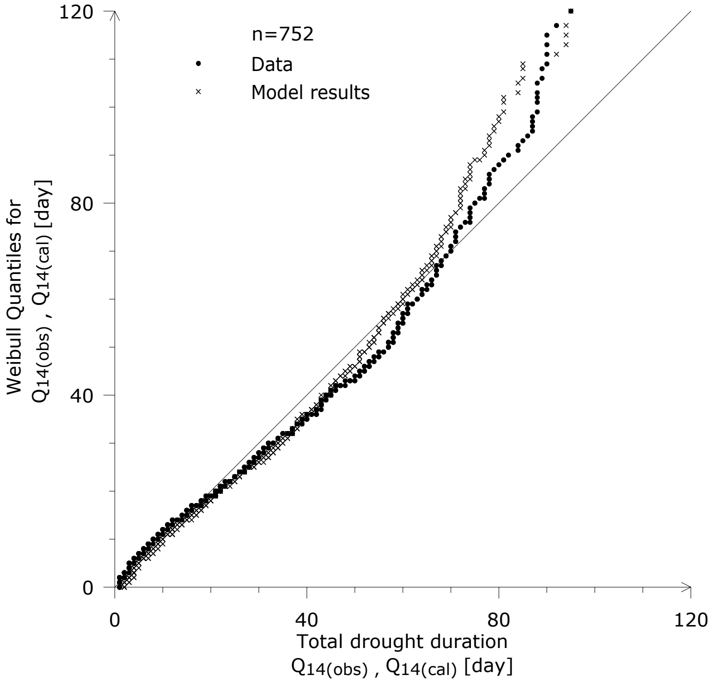

Figure 5.3 Q-Q plot showing the fit of a Weibull(a, b) distribution to both the observed and the calculated total drought duration that have occurred in the case Study Area during the period of 1961 to 1993. The calculated values are the output of model type MLP2 No. 2964. A perfect fit would have all points falling on the 1:1 line.

The Q-Q plot depicted in Figure 5.3 shows two facts. Firstly, it illustrates how the fitted Weibull distribution has been able to reproduce the empirical distribution of the data up to values of a total drought duration of about 88 [day], which corresponds to the 97th percentile. In other words, the fit works satisfactorily with the exception of the right tail, which exhibits larger differences because the Weibull distribution allocates too much probability to the few observations with values greater than 88 [day], which are too few in the given sample. This is why the Weibull quintiles are above the 1:1 line. Secondly, this plot depicts clearly how closely the selected model has been able to reproduce the empirical distribution function of the data up to about 75 [day].