A simple model for low flow forecasting in Mediterranean streams

K. Risva[1], D. Nikolopoulos[2], A. Efstratiadis[2] and I. Nalbantis[1*]
[1] *School of Rural and Surveying Engineering, National Technical University of Athens, Greece*
[2] *School of Civil Engineering, National Technical University of Athens, Greece*
[*] *e-mail: nalbant@central.ntua.gr*

Abstract: Low flows commonly occur in rivers during dry seasons within each year. They often concur with increased water demand which creates numerous water resources management problems. This paper seeks for simple yet efficient tools for low-flow forecasting, which are easy to implement, based on the adoption of an exponential decay model for the flow recession curve. A statistical attribute of flows preceding the start of the dry period is used as the starting flow, as for example the minimum flow of early April. On the other hand, the decay rate (recession parameter) is assumed as a linear function of the starting flow. The two parameters of that function are time-invariant, and they are optimised over a reference time series representing the low flow component of the observed hydrographs. The methodology is tested in the basins of Achelous, Greece, Xeros and Peristerona, Cyprus, and Salso, Italy. Raw data are filtered by simple signal processing techniques which remove the effect of flood events occurring in dry periods, thus allowing the preservation of the decaying form of the flow recession curve. Results indicate that satisfactory low flow forecasts are possible for Mediterranean basins of different hydrological behaviour.

Key words: dry period; recession rate; Savitzky-Golay filter; benchmark function; Monte Carlo calibration; uncertainty

1. INTRODUCTION

The World Meteorological Organization (WMO) (1974) defines low flow as the river discharge during prolonged dry weather. Actually, low flows are periodic phenomena and integral components of the river regime. Particularly, in Mediterranean catchments exhibiting significant variability across seasons, baseflow is the major, and occasionally the sole component of river hydrographs during dry periods. Therefore, low-flow estimations, in terms of magnitude, frequency and duration, are essential premise in solving numerous water engineering and management problems.

The literature reports multiple attempts for low-flow modelling, including hydrograph recession techniques, stochastic autoregressive functions, physically-based models, and data-driven approaches (Smakhtin, 2001). A typical concept is the linear reservoir model that represents the recession limb of a hydrograph as the outflow from a tank of infinite storage capacity, implementing the groundwater storage. The model is linear since outflow is expressed as a constant fraction of storage, thus resulting to an exponential decay function.

This article investigates the use of the above approach as a forecasting tool for the dry-period recession, using daily flow data retrieved at the first half of April. Key objective is providing a parsimonious model of acceptable predictive capacity, which can be implemented for operational purposes (e.g., annual irrigation planning). The model is evaluated at four rivers with different low-flow dynamics, namely Achelous (Greece), Salso (Italy), and Xeros and Peristerona (Cyprus).

2. OUTLINE OF METHODOLOGY

2.1 Model assumptions

Generally, a hydrograph is separated into rising limbs, reflecting increases in discharge from

precipitation events, and recession limbs, representing delayed flow due to processes in the saturated and unsaturated zone. Particularly, low flows during the dry period are mostly attributed to groundwater responses, which can be well represented as outflows from a linear reservoir. Under this assumption, the low flow component of the dry-period hydrograph of a specific year j can be modelled by an exponential decay:

$$q_{jt} = q_{0j} \exp(-k_j t) \tag{1}$$

where q_{0j} is the low flow at the beginning of the dry period of year j, t is the day index, and k_j is a recession parameter, associated with the macroscopic hydrological behaviour of the aquifer. Here, eq. (1) is applied in discrete form, using a daily time interval, flows are expressed in m^3/s, and the time is considered in days, while k_j is expressed in inverse daily units. The start and end of the dry period, the initial flow q_{0j}, and the recession parameter k_j is varying across hydrological years. In the remaining text the index j is omitted when unnecessary.

For convenience, we consider a *reference time horizon* of six months, from April 15th to October 15th, which is a reasonable assumption for the maximum duration of the dry period in Mediterranean rivers. Moreover, after preliminary investigations, we decided taking as initial discharge, q_{0j}, the minimum flow value of the first two weeks of April. Thus, for each year j, q_{0j} is a priori determined based on observed flow data, in contrast to the recession coefficient k_j, which is inferred through calibration, i.e. by fitting eq. (1) to the observed dry-period hydrograph.

2.2 Derivation of adjusted low flow data

For a given record of daily discharge data between April 15th and October 15th, and a given q_{0j}, the estimation of k_j is far beyond a typical calibration problem, since the duration of the dry period is not constant, while the dry-period hydrograph contains both rising and recession limbs, as well as individual peaks. Therefore, for each year, it is essential to extract the low flow component from the total hydrograph; these data are next referred to as *adjusted low flows*. In order to construct the adjusted data from observed flows, while also determining the effective length of the dry period, conventionally starting at April 15th, we employ a multi-step processing procedure, as illustrated in Fig. 1 and outlined next.

Step (a): The dry-period hydrograph exhibits both large- and small-scale fluctuations; the former are due to flood events whereas the latter are generally induced by random errors of the monitoring instruments (e.g., stage recorders), also resulting to local minima that do not have physical meaning. In order to remove such effects and obtain a smoothed hydrograph, we employ the numerical filter introduced by Savitzky and Golay (1964), which allows increasing the signal-to-noise ratio without greatly distorting the signal. This is achieved by replacing raw values by new ones, which are computed from a moving polynomial fitted to $2n + 1$ neighbouring points, with n being at least equal to the order of the polynomial. This is handled similarly to a weighted moving average, since the coefficients of the smoothing procedure remain constant.

Step (b): We define the beginning and end of the dry period, and its last flow value, q_{jn}. First, using as cut-off threshold the initial flow, q_{0j}, we seek for the last flow value being greater than q_{0j}; the date with that value denotes the actual beginning of the dry period, whereas observed flows for previous dates are ignored within model fitting. Next, we calculate the average flows over 15-day intervals and their first differences until October 15th. Moving backwards in time, we examine whether the flow trend changes from negative to positive, which marks the end of the dry period.

Step (c): We remove all flow values above the line joining q_{0j} with q_{jn}, i.e.

$$q_{jt} = q_{0j} - (q_{0j} - q_{nj}) t / n_j \tag{2}$$

since the negatively exponential low flow model (1) is by construction below this line. In eq. (2), n_j is the number of days between April 15th and the end of the dry period of year j.

Step (d): We reduce the flow peaks remaining below the theoretical upper threshold (2), considering small-scale flood events of typical duration up to two days. In this context, we apply twice a cut-off procedure comprising the identification of all local flow maxima over the dry period, which are next set equal to the value of the previous day. The remaining flows between q_{0j} with q_{nj} comprise the adjusted low flow sample, used in fitting eq. (1). This is generally non-continuous and contains much less values than the length n_j, since all important local flood events are removed.
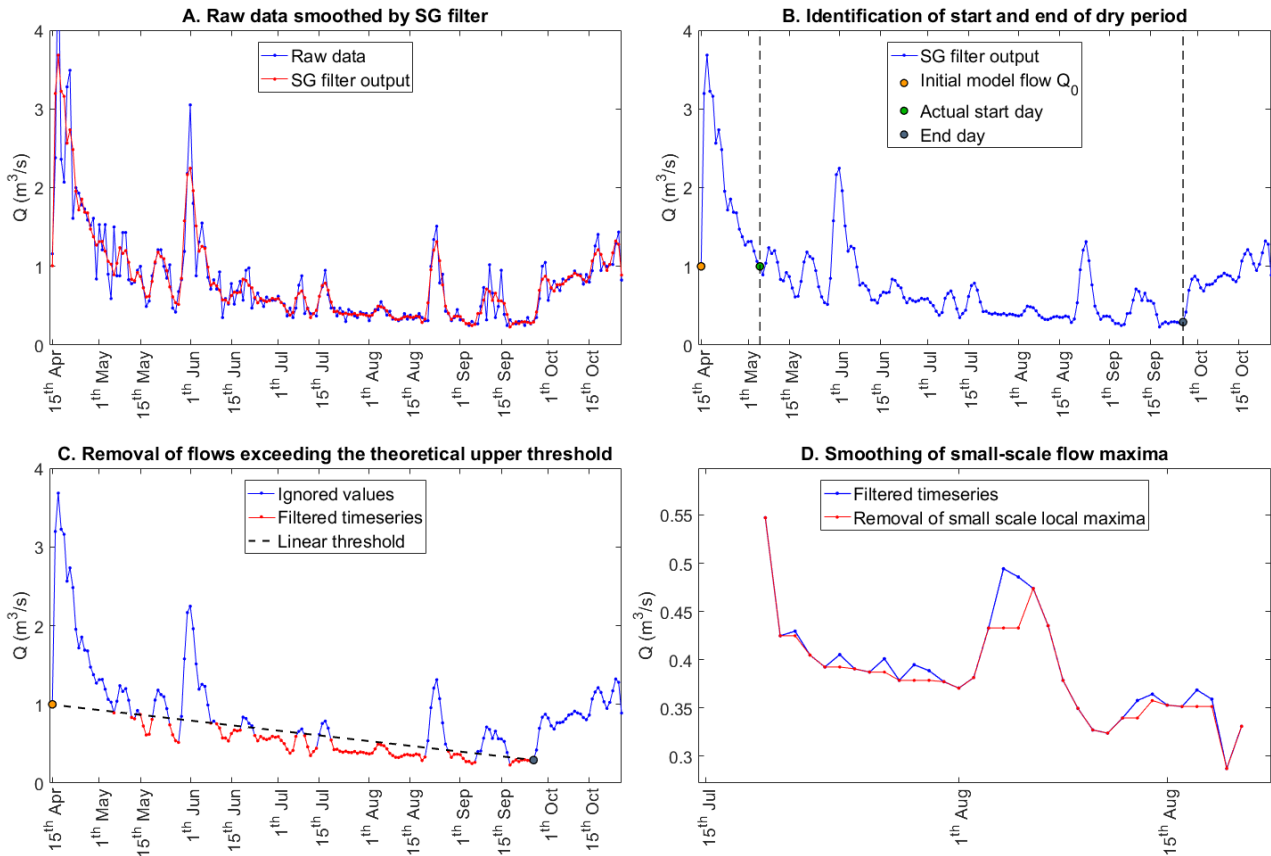

Figure 1. Example of extracting adjusted flow data from daily hydrographs; SG stands for Savitzky- Golay

2.3 Model calibration

For each year j, a one-dimensional calibration problem is solved, to identify the value of k_j that ensures the optimal fitting of eq. (1) to the adjusted data. Our objective function is a modified expression of efficiency (MEF), using as benchmark the expected flows over the reference period. This is of key importance, since the well-known Nash-Sutcliffe efficiency would compare the simulated values against the average flow, thus providing unrealistically high model performance. Apparently, the average is far from being representative of the river regime during the dry period, which exhibits a non-stationary behaviour due to the systematic flow decrease. This shortcoming has been widely discussed in the literature (e.g., Pushpalatha *et al.*, 2012).

Initially, we considered as benchmark not the overall mean flow over the reference time horizon but the mean flow value of each individual day, as well as the median, which is generally a better estimator than the mean, since the dry-period flows are highly skewed. However, due to noise effects, the two benchmarks (particularly the mean) exhibit random fluctuations, which are not desirable (Fig. 2). To remove such effects, we introduced three stepwise exponential functions that were fitted to the daily means and medians, as well as the lower envelope of medians. The latter is the most representative of the river regime during the dry period, and thus it was finally used as benchmark within calibrations.

2.4 Model formulation in forecast mode

The initial formulation of the low-flow simulation model (1), considering an annually varying recession coefficient, has limited practical usefulness, as we seek the development of a forecasting tool with parameters that are known in advance, before issuing a forecast. For employing eq. (1) in forecast mode, we need a formula for estimating the annual value of parameter k_j, based on easily retrievable information that is available before the start of the dry period (similar to q_{0j}, which is estimated on the basis of early April flows). In this respect, also accounting for the outcomes of preliminary calibrations, we tested the application of a linear relationship between k_j and q_{0j}, i.e. $k_j = a\, q_{0j} + b$, where a (s m^{-3} d^{-1}) and b (d^{-1}) are regional parameters, i.e. constant for each basin and independent of the year index j. Hence, the model in forecast mode is now written:

$$q_{jt} = q_{0j} \exp[-(a\, q_{0j} + b)\, t] \tag{3}$$

In order to obtain parameters a and b, we examine two procedures. First, we use a set of already calibrated values of k_j and fit a linear regression equation against the independent variable q_{0j}. By accepting the assumptions of the least squares approach, the slope and intercept of the regression model are optimal estimators of a and b. Alternatively, parameters a and b are considered as control variables of a global optimization problem, asking for fitting eq. (3) to the entire set of reference flow data (in contrast to eq. (1), where different values of k_j are obtained by calibrating against the low flows of each dry period).

For both procedures, the model performance was assessed in terms of modified efficiency through the calibration period, while its predictive capacity was assessed by calculating MEF across an independent data period (validation).

2.5 Uncertainty analysis

A renowned shortcoming of the classical calibration-validation paradigm, known as split-sample test, is the dependency of the model performance on the length and time window of the data sample. This may introduce significant uncertainty, not only to the model performance but also to the optimized parameter values.

In order to account for such uncertainties within the estimation of parameters a and b, we employed a Monte Carlo calibration-validation scheme. For both procedures, we calibrated a and b against 1000 randomly selected subsets (not necessarily continuous) over the whole period of adjusted low flows, and validated against the remaining subsets (a constant validation period of 10 years was systematically considered). For all results (parameter values and performance criteria) we estimated their average and standard deviation.

3. STUDY AREAS

The methodology was tested at four rivers of different scale (one large and three small) and hydrological regime (Risva, 2016). Their key characteristics as well as the corresponding daily average and median flows across the reference time horizon, are depicted in Fig. 2. Low flow regimes of the selected rivers are different. Specifically, Achelous, which produces the highest annual runoff in Greece (950 mm), retains significant discharge, while the departure of the mean daily flow from the median is practically negligible from May to August. In contrast, in Salso River there are large deviations between means and medians, since the dry period flows exhibit substantial asymmetry due to occasional yet quite severe flood events. Finally, in the two Cyprian rivers there are also systematic differences between means and medians, until the mid-summer; in particular, from early July, Xeros River retains a practically constant baseflow, while in Peristerona the flow is interrupted.
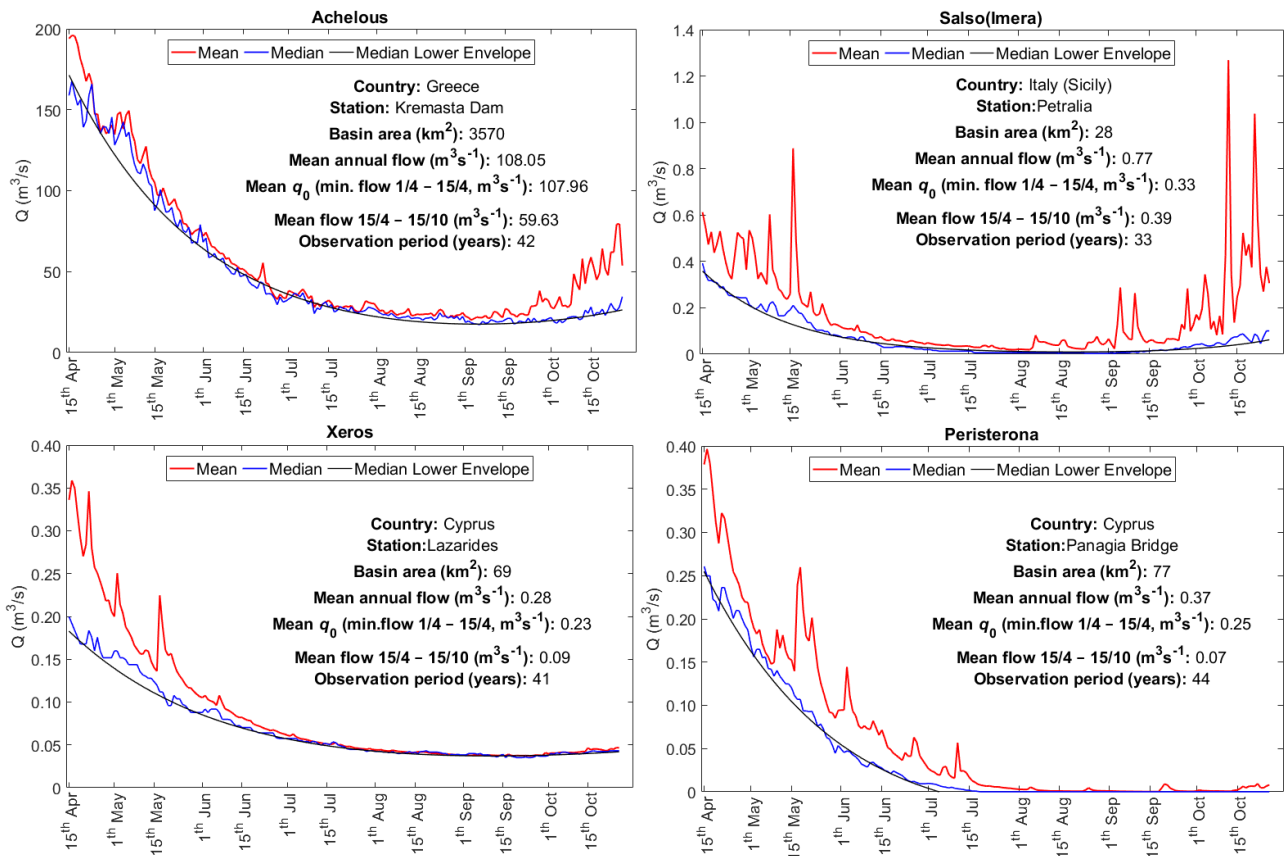
Figure 2. Synoptic information for case studies, plots of daily average and median flows from April 15th to October 15th and benchmark models fitted to the lower envelope of medians.

4. RESULTS AND DISCUSSION

Key results for all rivers are summarized in Table 1. The two parameter estimation approaches (regression-based and global optimization-based) are employed within a Monte Carlo context.

For the first approach, we demonstrate the optimized MEF considering eq. (1), with optimally fitted parameters k_j per calibration sample, and eq. (3), with a and b estimated through a posteriori regression between the optimized values of k_j and q_{0j}; in the last case, MEF refers to calibration and validation samples, and the full data sample (in forecast mode). Moreover, we show the optimized values of a and b, as well as the coefficient of determination (r^2) of the linear regression. Finally, using as typical starting flow the mean q_{0j} (average of minimum early April flows over the entire data period) we provide the corresponding recession rate, k. It is interesting to note that even a low correlation between k and q_0 may yield a satisfactory predictive capacity, as shown for Salso and, particularly, for Peristerona.

Regarding the second approach, we show MEF values for the calibration and validation samples, and the full data sample (forecast mode), the optimized parameters a and b, and the average recession rate, k.

By comparing MEF in forecast mode, it seems that both procedures ensure similar predictive capacity, which is reasonable, since the average recession rates are almost identical (except for Salso). However, in Achelous and Salso, the two approaches are not equivalently robust, as indicated from the variability of MEF. In such cases, we suggest selecting the parameterization ensuring the lowest uncertainty.

Characteristic examples of model fitting in forecast mode are given in Fig. 3. The model is absolutely consistent with the anticipated hydrological behavior of each river, which is well-represented by the average recession rate. As shown in Table 1, the lowest rate refers to Achelous, i.e. a large river with permanent flow, while the highest rate refers to Peristerona, which allows

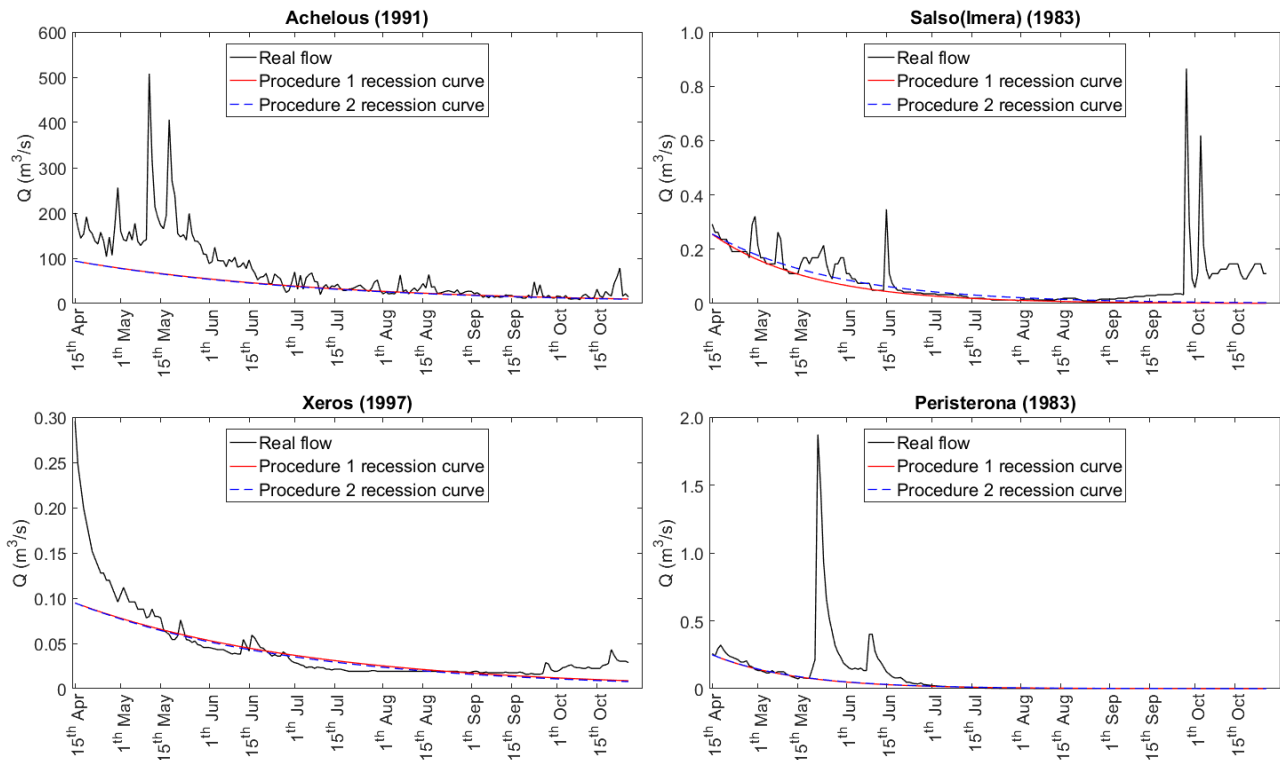reproducing the intermittent regime of this river.

Figure 3. Examples of forecasted low flows through the two parameter estimation procedures (1: regression, 2: global optimization) against the observed hydrographs over the reference period (15/4 to 15/10).

Table 1. Summary of model results (means out of 1000 simulations and standard deviations, in parentheses)

River name	Achelous	Salso (Imera)	Xeros	Peristerona
Parameter estimation through regression between k and q_0				
MEF, local calibration	0.643 (0.035)	0.790 (0.026)	0.889 (0.017)	0.831 (0.016)
Coefficient of determination, r^2	0.771 (0.028)	0.073 (0.029)	0.437 (0.031)	0.083 (0.019)
Parameter a (mm^{-1} d^{-1})	0.0034 (0.0001)	0.0052 (0.0011)	0.0204 (0.0012)	-0.0179 (0.0023)
Parameter b (d^{-1})	0.0037 (0.0002)	0.0244 (0.0014)	0.0098 (0.0003)	0.0400 (0.0009)
Recession rate k (d^{-1}) for mean q_0 (mm)	0.0126 (0.0003)	0.0297 (0.0018)	0.0157 (0.0005)	0.0350 (0.0011)
MEF, calibration	0.523 (0.044)	0.441 (0.070)	0.787 (0.032)	0.721 (0.021)
MEF, validation	0.496 (0.124)	0.414 (0.195)	0.752 (0.146)	0.713 (0.078)
MEF, forecast mode	0.523 (0.029)	0.437 (0.008)	0.788 (0.006)	0.721 (0.006)
Parameter estimation through global optimization				
Parameter a (mm^{-1} d^{-1})	0.0030 (0.0003)	0.0074 (0.0033)	0.0159 (0.0035)	-0.0170 (0.0070)
Parameter b (d^{-1})	0.0049 (0.0009)	0.0170 (0.0040)	0.0109 (0.0012)	0.0390 (0.0028)
Recession rate k (d^{-1}) for mean q_0 (mm)	0.0129 (0.0011)	0.0245 (0.0052)	0.0155 (0.0016)	0.0342 (0.0034)
MEF, calibration	0.529 (0.036)	0.483 (0.076)	0.800 (0.035)	0.725 (0.020)
MEF, validation	0.484 (0.143)	0.320 (0.196)	0.729 (0.173)	0.692 (0.113)
MEF, forecast mode	0.525 (0.006)	0.438 (0.084)	0.794 (0.006)	0.718 (0.004)

5. CONCLUSIONS

Low flow dynamics in Mediterranean rivers can be well approximated using the linear reservoir concept, which implies the determination of two parameters, i.e. the starting flow and the recession rate (respectively denoted as q_{0j} and k_j in eqs (1)-(4)). Our analyses indicated that the former can be easily defined as the lowest daily flow during early April, while the latter can be expressed as a linear function of q_0. In this respect, we can build a simple yet effective low flow forecasting model, by determining the slope and intercept of the above function, based on historical flow data.

Key innovations of the proposed framework are the extraction of the adjusted flow data from the total hydrograph, requiring a series of sequential transformations of the original data (smoothing,

peak removal, etc.), the selection of the lower envelope of daily medians as benchmark flows to be used within calibrations, and the stochastic implementation of calibrations, which allows quantifying the uncertainty of parameters and performance measures.

The case studies indicated that the proposed methodology is suitable for both large and small basins, producing both permanent and intermittent runoff. The strong advantage of the model is its parsimony, not only in terms of parameters but also in data requirements, since, after calibration, the sole input is the starting flow q_0. This makes the model easy to apply in an operational context, so as to provide reliable estimations of surface runoff during dry seasons.

REFERENCES

Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol.; 420-421: 171–182.

Risva, K., 2016. Model development for low flow estimations in river basins across the Mediterranean. Diploma thesis, School of Rural and Surveying Engineering, National Technical University of Athens.

Savitzky, A., Golay, M. J. E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analyt. Chem.; 36(8): 1627–39.

Smakhtin, V. U., 2001. Low flow hydrology: a review. J. Hydrol.; 240(3): 147–186.

World Meteorological Organization (WMO), 1974. International Glossary of Hydrology.