Nordic Hydroloy, 1983, 93-112

Kalman Filter Empirical Fitting on Monthly Rainfall-Runoff Responses

Maria Mimikou

Tech. Univ. of Athens and Public Power Corp.,
Athens, Greece

The majority of the methods for fitting rainfall-runoff Kalman filters use the statistical characteristics of the input (rainfall) and the output (runoff) series according to the systems analysis and control theory methodology. The problem arising with these techniques is essentially related with their inability to describe variations of the basin runoff response when the relevant information concerning the causes of such variations is not contained into the input to the filter. The presented empirical method of fitting solves this problem by taking into account additional input information and improving the knowledge about the hydrological system, but without changing the simple two-variable structure of the filter. This information consists of the percentage of snowfall into the monthly basin rainfall, the magnitude of basin rainfall at month t in comparison to the rainfall at month t-1 (increasing, constant, or decreasing) and the amount of basin rainfall associated with basin's moisture and runoff characteristics. The Kalman filter empirically fitted on monthly rainfall-runoff responses of the Aliakmon and Acheloos river basins in Greece is shown to be more accurate and adequate than the automatically fitted filter by using conventional statistical methods. The filter has been used in hydroelectric energy studies of the area, where accurate runoff estimation from rainfall was necessary.

Introduction

Rainfall-runoff models according to their structure are basically classified into two categories: the moisture accounting models and the systems approach type of models. Generally speaking, the first type of models are mathematically representing some of the physical processes which make up the dynamics of the basin response; the second type are relating the input-output processes, by using the systems analysis and control theory methodology and without representing, at least explicitly, the physical character of the process. One of the major innovations and contributions in modern control theory is the use of the recursive or Kalman filter, which under certain assumptions has been widely and successfully used for obtaining estimates either of hydrologic variables or of parameters of hydrologic models. It is a systems approach type of model and since its original development (Kalman 1960) much work has been done by the hydrologist either refining the technique or extending its applications (Hino 1973; Rodriguez-Iturbe et al. 1978).

The basic problem of the rainfall-runoff application of the filter is concentrated mainly on its parameters estimation. The majority of the proposed methods for the parameters estimation use the historical statistical characteristics of the input (rainfall) and the output (runoff) time series. These are regressional methods, mean error optimizing methods, self-adapting algorithms, methods based on the solution of input-output covariance matrices, etc. The problem arising with these techniques is essentially related with their inability to describe variations of the basin runoff response, when the relevant information concerning the causes of such variations is not contained into the input to the filter. On the other hand, it is true that the more we know about a hydrological system and the better we understand it, the easier it will be to interpret and control the parameters of the filter, improving thus their estimation. This means that there is a need of improving the input information to the filter in a way that important characteristics of the basin runoff response, associated with the additional inputs, are incorporated into the filter response mechanism.

This is exactly the basic idea behind the empirical method of filter fitting presented here. The method is empirically modifying the statistical filtering technique used by Schwartz and Shaw (1975) for filtering signals in noise. The improvement of the filter input information with some physical rainfall-runoff and basin characteristics, which experience has shown to affect significantly dynamic characteristics of the basin runoff response, the latter's incorporation into the filter response mechanism without changing its simple two-variable structure, the interpretation of the filter parameters and the control of their time variability, are the main objectives of the study. These objectives are successfully accomplished and are constituting the main advantages of the empirical method over other automatic "black-box" methods of fitting. Is is shown that the former improves significantly the filter's effiency for representing the physical process and its dynamics. It is found that for monthly rainfall-runoff Kalman filters the needed additional input information for a reliable representation of the basin runoff response and of its dynamics consists of the percentage of snowfall into monthly basin rainfall values, the magnitude of basin rainfall at month t in comparison to the magnitude of rainfall at month t-1 (increasing, constant, or decreasing) and the amount of monthly basin rainfall associated with basin's moisture and runoff characteristics. It must be noticed that the above mentioned basin rainfall consists of rainfall and snow water equivalent values combined. The filter is applied to the Aliakmon and Acheloos river basins located in Northwestern and Western Greece for estimating monthly runoffs at the Siatista and the Avlaki measuring stations correspondingly. The accurate estimation of monthly runoffs from rainfals at these two sites for time periods without flow measurements was necessary for extending the runoff records and studying the energy features of two dams programmed to be built at the sites. For comparison purposes the filter is fitted not only with the empirical method, but also with some conventional statistical methods.

Descriptions of the Filter and of the Systems of Application

The Model

The hydrological basin taken as a "black-box" system with rainfall as the input and runoff as the output is shown in Fig. 1. The mathematical representation of this process on a monthly basis by a recursive or Kalman filter is the following

$$q_{p,t} = a_i\, q_{p,t-1} + b_i P_{p,t} \quad , \qquad |a_i| < 1 \qquad (1)$$

where $q_{p,t}$ (p=1, 2, ..., n, years, t=1, 2, ..., 12 months) is the monthly runoff in m^3/sec, $P_{p,t}$ is the monthly rainfall mean over the basin with $E(km^2)$ surface area, containing not only the rainfall but also the snow water-equivalent on the basin, used here with the same units of m^3/sec ($P_{p,t}$ (m^3/sec) = $P_{p,t}$(mm) $\times$ $E(km^2)$ $\times$ $10^3/T \times 86,400$, with T the duration of month t in days) and a_i, b_i are the parameters of the filter which may be either constants or time variables.

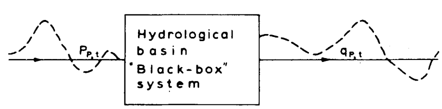

Fig. 1. The hydrological basin operating like a "black-box" system.

The Hydrological Systems of Application

The two hydrological systems, where the filter of Eq. (1) is applied for monthly rainfall-runoff modeling, are the Aliakmon river basin at Siatista, shown in Fig. 2 and the Acheloos river basin at Avlaki, shown in Fig. 3.

The Aliakmon river basin at Siatista with a surface area of 2,744 km^2 is located in Northwestern Greece and is equipped with 18 rainfall gauging stations and one staff gauge recorder with flow measuring station at Siatista. The available rainfall and discharge data are covering a common period of 16 years, which starts in 1962 and lasts up to 1978. The mean monthly rainfall values have been estimated

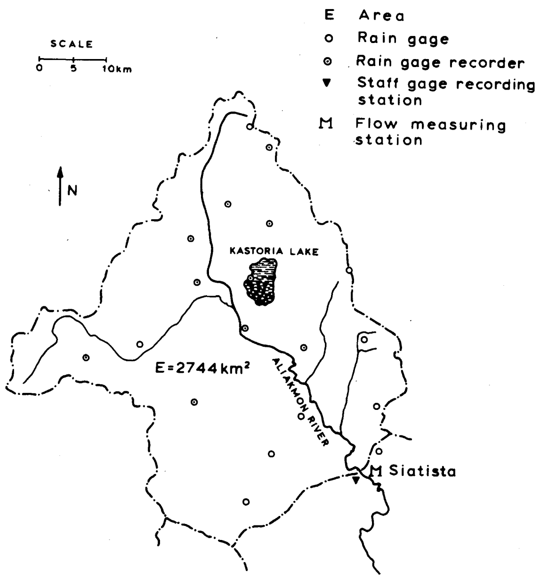

Fig. 2. Aliakmon river basin at the Siatista measuring station.

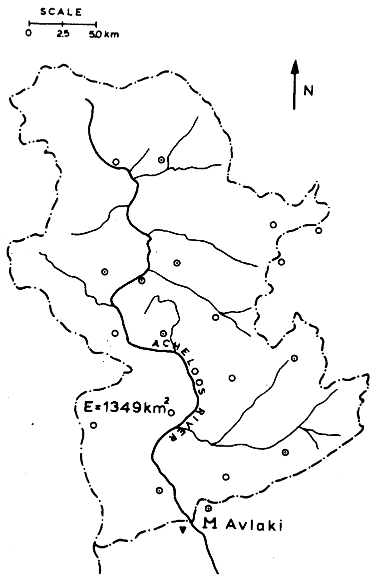

Fig. 3. Acheloos river basin at the Avlaki measuring station.

by using the Thiessen polygon method and afterwards they have been corrected because of the increase of rainfall with the altitude, taking into account the difference between stations' and basin's mean elevation. The 12-year period data of 1966 through 1978 are used for the calibration of the filter, whereas the 4-year period data of 1962 through 1966 are used for the verification of the adequacy of the filter.

The Acheloos river basin at Avlaki with a surface area of 1,349 km^2 is located in Western Greece and is equipped with 20 rainfall gauging stations and one staff gauge recorder with flow measuring station at Avlaki. The data of the basin are covering a 13-year period from 1965 through 1978 from which the 9-year period data of 1969 through 1978 are used for the calibration of the filter and the rest 4-year period data of 1965 through 1969 are used for verification. The Thiessen polygon method, with the previously mentioned corrections of its results, has been also used here to obtain the rainfall values. The estimated monthly basin rainfall values contain also the monthly snow water-equivalent amount; the contribution of the latter into the basin rainfall values in percentages per month has been also estimated for both basins, in order to be used as additional input information to the filter. The available rainfall information covers a 30-year period from 1950 to 1980, which is much longer than the period covered by runoff data for both studied basins. The filter was exactly used in extending the runoff records at Siatista and Avlaki for purposes of hydroelectric energy studies.

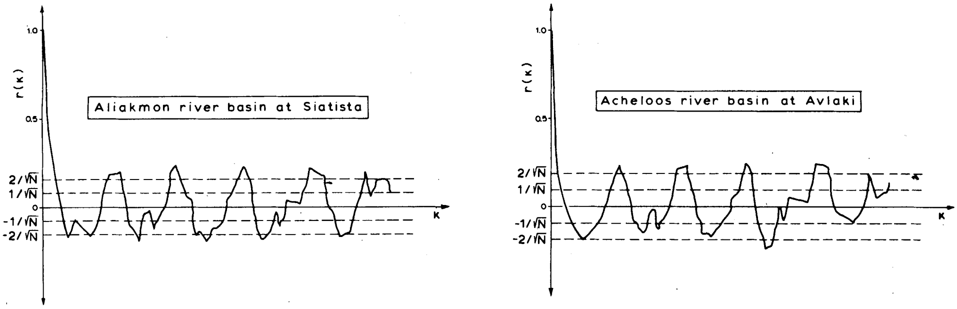

Fig. 4. Autocorrelograms of the input (rainfall) series.

Empirical Fitting on the Filter

Input-Output Series Identification

It is known that monthly hydrological series exhibit yearly and within the-year periodicities. The periodic cycle is very significant for monthly runoff series but is more or less insignificant for the majority of monthly rainfall series. This can be deduced also from the autocorrelograms of the monthly rainfall and runoff series of Aliakmon and Acheloos river basins shown in Figs. 4, 5a and 6a. The autocorrelograms of the rainfall series for both basins are given in Fig. 4, where one can see that they exhibit a very weak periodicity, since the successive every 12 lags peaks of the correlograms do not significantly exceed the magnitude of two typical errors equal to: $\pm 2/\sqrt{N}$, with $N=144$ for Aliakmon river and $N=108$ for Acheloos river, the number of the calibration period data. Therefore, the input series can be reasonably assumed to be stationary.

The autocorrelograms of the Aliakmon river basin runoff series shown in Fig. 5a and of the Acheloos river basin runoff series shown in Fig. 6a, with their significant peaks every 12 lags, are implying the existence of significant yearly periodicities. Therefore, the output series are not stationary and need to be stationarized before their use in the filter. The stationarization of the runoff series has been accomplished here by simply subtracting the monthly means, as follows

$$Q_{p,t} = q_{p,t} - \bar{q}_t \tag{2}$$

with

$$\bar{q}_t = \frac{1}{n} \sum_{p=1}^{n} q_{p,t} \tag{2a}$$

The autocorrelograms of the stationarized runoff series are shown in Figs. 5b and 6b, where one can see that no significant periodicity is left, since the correlograms after the first few lags are fluctuating inside the two typical error limits.

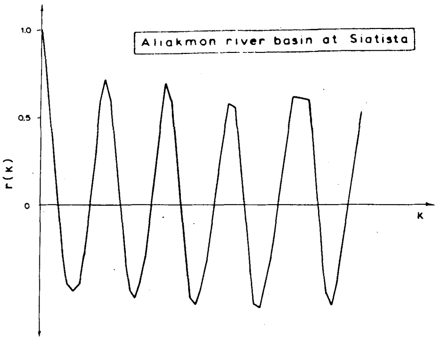

Fig. 5a.

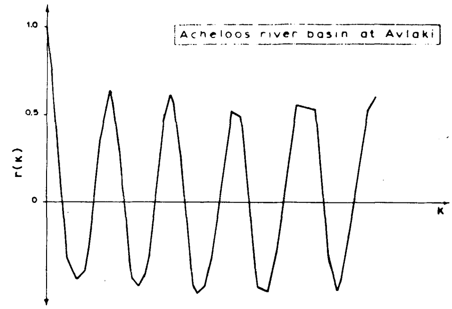

Fig. 6a.

Autocorrelogram of the non-stationary output (runoff) series.

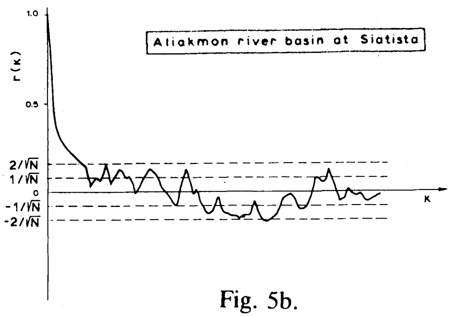

Fig. 5b.

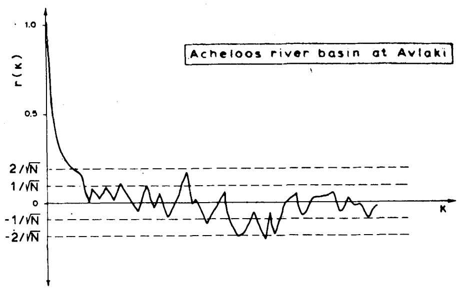

Fig. 6b.

Autocorrelogram of the stationarized output (runoff) series.

Examining then the shape of the autorcorrelograms of the stationarized output series for the first lags, in Figs. 5b and 6b one can easily see that for both basins they show an approximately exponential decay, implying that the output series can be represented by a first order recursive process driven by a zero mean white noise $w_{p,t}$, in the scheme shown in Fig. 7 and given by

$$Q_{p,t} = \alpha Q_{p,t-1} + w_{p,t}\,, \qquad |\alpha| \leq 1 \tag{3}$$

with α the first order autocorrelation coefficient of the runoff series. The variance σ_w^2 of the white noise series is given (Box and Jenkins 1970) by

$$\sigma_w^2 = \sigma_Q^2(1-\alpha^2) \tag{4}$$

with σ_Q^2 the variance of the runoff series.

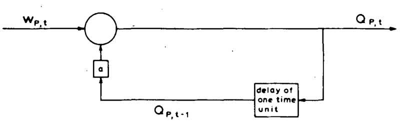

Fig. 7. The first-order output recursive process.

Estimation of the Filter Parameters under Certain Conditions

The parameters of a filter, with input and output series with the previously identified characteristics, have been estimated by Schwartz and Shaw (1975, pp. 331-337). The procedure of estimation, which is basically a signal detection procedure in the presence of noise, even though it uses the statistical characteristics of the stationary input, output and noises series, it is not restricted only to outputs represented by first order autoregressive models, but it can be also used for outputs with different stochastic models; in this case one has simply to replace Eq. (4) with another one appropriate expression giving the variance of the noise $w_{p,t}$. In order to obtain the final expressions for the parameters Schwartz and Shaw (1975, p. 334) are also using another characteristic input-output relationship, with the following expression.

$$P_{p,t} = Q_{p,t} + \eta_{p,t} \tag{5}$$

The relation between the output series and the input series (taken here with zero mean), appears to be obscured by an additive zero-mean independent noise $\eta_{p,t}$ called "observation noise", with variance given by

$$\sigma_\eta^2 = \sigma_P^2 - \sigma_Q^2 \tag{6}$$

where σ_P^2 is the variance of the stationary input series. Finally, the two expressions for the parameters given by Schwartz and Shaw (1975, pp. 335-336), are

$$b_i = \frac{A+\alpha^2 b_{i-1}}{1+A+\alpha^2 b_{i-1}} \tag{7}$$

$$a_i = \alpha(1-b_i) \tag{8}$$

where α is the first order autocorrelation coefficient of the output series and A is the ratio of the variance of the white noise part of the output series given in Eq. (4) to the variance of the »observation noise« given in Eq. (6), namely

$$A = \frac{\sigma_w^2}{\sigma_\eta^2} = \frac{\sigma_Q^2(1-\alpha^2)}{\sigma_P^2-\sigma_Q^2} \tag{9}$$

Eq. (7) is recursive with very limited variability, since it rapidly converges to a steady value. Therefore, the parameters given by this statistical filtering technique are practically constants. Examining now the validity of Eqs. (5) and (7) in the physical framework of a monthly rainfall-runoff process, one can see that Eq. (5) is valid, when nothing else besides noise $\eta_{p,t}$ interferes between rainfall and runoff. This situation can be practically approximated in months with heavy rainfall, producing undelayed runoff, when all components of the basin response reach their steady state and the runoff yield is maximized. Another observation is that, for b_i less than its steady limiting value, Eq. (7) is an increasing function on the (b_i, b_{i-1}) plane, and, since b_i is physically interrelated to rainfall, this implies certain limitations on the monthly rainfall values, which are given in the following paragraph. Therefore, Eq. (7) and also Eq. (8), whose derivation is based on Eq. (5), can be accepted only under certain conditions. For time periods with different conditions empirical modifications of these equations are made, using additional input information, which experience has shown to be related to dynamic characteristics of the basin runoff response. Through these modifications, the latter are incorporated into the filter response mechanisms improving thus its efficiency for representing the physical phenomenon and its dynamics.

Construction of the Filter Response Mechanism

The additional input information related to dynamic characteristics of the basin runoff response, its influence on the filter parameters and its usage for the construction of the filter response mechanism, obtained through a thorough searching of the monthly rainfall-runoff relations of the historical records of both basins under consideration, are discussed in the following. The interpretation of the filter parameters is that of a coefficient weighting the monthly rainfall contribution to the same month's runoff for b_i and of a memory weighting coefficient to the monthly runoff process for a_i.

Rainfall Amount and Magnitude

The information of the amount of rainfall at month t associated with some basin's moisture and runoff characteristics, like its seasonal water needs for saturation and for direct surface runoff production, is related to the dynamics of the monthly basin runof response, since it concerns the control of the basin's monthly rainfall-runoff balance. The information concerning the magnitude of the rainfall at month t in comparison to the rainfall at month t-1 (increasing, constant or decreasing) is also related to dynamic characteristics of the basin runoff response, since it is featuring the temporal variations of the basin's moisture condition. All above mentioned additional to the rainfall input information to the filter is taken into account, and hence the related dynamic characteristics of the basin runoff response are incorporated into the filter response mechanism, by properly modifying the mechanism of time variations of the rainfall weighting parameter b_i of the filter.

As it was previously explained, in order to use Eq. (7) one has to ensure first that rainfall is heavy, producing maximum undelayed runoff and that the rate of b_i, and thus of its previously defined conceptual analogue, is increasing. The former condition is met with monthly rainfall above an "upper critical" value P_s with none or "insignificant" snowfall contribution, whereas it was found that the latter is met, when rainfall is increasing from month to month or when it remains constant. Therefore, for months with increasing or constant ($P_{p,t} \gtrless P_{p,t-1}$) and heavy ($P_{p,t} \geqslant P_s$) rainfall without "significant" snowfall, b_i can be estimated from Eq. (7), which, by using the transformation $L = 1/(1 + A + \alpha^2 b_{i-1})$, is written as

$$b_i = 1 - L \tag{10}$$

For increasing or constant monthly rainfall, which is less than P_s but greater than or equal to a "lower critical" rainfall value P_o, under which no significant surface runoff is produced, the increasing rate of b_i is found to be retarded, namely, the slope of Eq. (10) on the (b_i, b_{i-1}) plane given by

$$\frac{db_i}{db_{i-1}} = \alpha^2 L^2 \tag{11}$$

must be reduced. It has been generally assumed here that all decreases or increases of the expressions of Eqs. (7) and (8) for b_i and a_i are proportional or inversely proportional to a power of the absolute value of the first order autocorrelation coefficient of the filter output process, namely of the parameter α, $|\alpha| \leqslant 1$.

It can be assumed then generally, that the slope given in Eq. (11) must be reduced $|a|^k$ times, with $k \geqslant 0$. Then, integrating this reduced slope one can obtain the expressions for b_i as

$$b_i = |\alpha|^k (1-L) \tag{12}$$

For decreasing rainfall, $P_{p,t} < P_{p,t-1}$, the coefficient of the $P_{p,t}$ contribution to runoff at month t is less than its value at the previous month. Thus, b_i decreases and the slope of Eq. (11) changes direction. The following modification of Eq. (11) was found to fit well the recession limb of b_i

$$\frac{db_i}{db_{i-1}} = 1 - \nu(\alpha^2 L^2)^{\mu} \tag{13}$$

The "weights" ν and μ are also related to the amount of monthly rainfall. If $P_{p,t} > P_s$, b_i is estimated to decrease slowly, according to a rate given by Eq. (13) with $\nu=1$ and $\mu=1$. Therefore, the expression for b_i becomes approximately

$$b_i = b_{i-1}(1-\alpha^2 L^2) \tag{14}$$

If $P_o \leqslant P_{p,t} < P_s$, b_i is estimated to decrease rapidly, according to a rate given by Eq. (13) with $\nu=1$, $\mu=½$; the expression for b_i becomes then approximately

$$b_i = b_{i-1}(1-\alpha L) \tag{15}$$

For rainfall values less than P_o, b_i is assumed to remain constant. Notice that the parameter b_i, given by Eq. (10) and by all modified expressions of it, in an always positive number less than one: $0 < b_i < 1$.

Snowfall

The other information found to be related to dynamic characteristics of the basin runoff response and to the memory parameter a_i of the filter, is the percentage of snowfall into basin rainfall. The existence of "significant" snowfall during the previous consecutive months (t-1, ...) and the absence of snowfall during month t, is affecting the basin runoff response for this month, which changes in order to account for the delayed runoff phenomenon due to snow melting. It must be noticed here that for the region of Greece under consideration the usual delayed time for melting of the major bulk of snow is one month after the last month of the period with "significant" snowfall.

The delayed runoff phenomenon is taken into account in the filter's operation by the augmentation of the contribution of the previous discharge $Q_{p,t-1}$ for the estimation of $q_{p,t}$ by Eq. (18). This means that if the parameter α is positive and $Q_{p,t-1}$ is negative, i.e.: $q_{p,t-1} < \bar{q}_{t-1}$, then the memory parameter a_i given by Eq. (8) must be decreasing, whereas if $Q_{p,t-1}$ is positive, i.e.: $q_{p,t-1} > \bar{q}_{t-1}$ then a_i is increasing. In the former case it has been assumed that the parameter a_i, given by Eq. (8), decreases proportionally to the k^{th} power of α, namely

$$a_i = a^{k+1}(1-b_i) \quad (16)$$

In the latter case it has been assumed that the increase of a_i is inversely proportional to α^k, namely

$$a_i = \frac{1}{a^{k-1}}(1-b_i) \quad (17)$$

If the paramter α is negative, which means that the parameter a_i given by Eq. (8) is also negative, then the modifications on Eq. (8) are made according to $|\alpha|^k$, and for $q_{p,t-1} < \bar{q}_{t-1}$, a_i is absolutely increasing and given by Eq. (17), whereas for $q_{p,t-1} > \bar{q}_{t-1}$, a_i is absolutely decreasing and given by Eq. (16). For $k \gtrless 2$, Eq. (17) is likely to give a value for $|a_i|$ greater than 1. This may be avoided by substituting $|a_i|$ with its maximum limiting value 0.999...

In the case of increasing or constant $P_{p,t}$ above the value P_s, the existence of "significant" snowfall at month t affects also the expression for b_i of the same month, which decreases proportionally to α^k. Namely, the parameter b_i is not any more given by Eq. (10), but by Eq. (12).

For the rest of the possible cases (i.e. none or "insignificant" snowfall, etc.), the memory parameter a_i is given by Eq. (8). The filter response mechanism with positive α, which is the case for both hydrological systems under consideration, is summarized in Table 1.

Control of Filter Parameters

In the above described procedure there are two key rainfall values, the P_s ("upper critical") and P_o ("lower critical"); a characteristic distinction is also used between "insignificant" and "significant" snowfall. The key values are estimated and the characteristic terms about snowfall are quantified according to the hydrologic experience obtained in dealing with the two specific basins of application.

The "upper critical" monthly rainfall value P_s, above which rainfall may be characterized as heavy with the previously described effects, is estimated to be equal to 25 % of the mean yearly basin's rainfall value $\bar{P}_t$, for the Aliakmon river basin and equal to 15 % of $\bar{P}_t$ for the Acheloos river basin, during the dry periods of the regions, which last from June to November for both basins. For the wet periods, which for both basins again last from December to May, the P_s values are equal to 15 % of $\bar{P}_t$ for the Aliakmon river basin and equal to 10 % of $\bar{P}_t$ for the Acheloos river basin. The "lower critical" value P_o is estimated to be approximately equal to the one twelfth of P_s for each of the periods of the year. The specific values estimated then for the Aliakmon river basin are: P_s (dry period) = 250 mm/mo, P_s (wet period) = 150 mm/mo and P_o (dry period) = 21 mm/mo, P_o (wet period) = 13 mm/mo, whereas for the Acheloos river basin the values are: P_s (dry period) = 300 mm/mo, P_s (wet period) = 200 mm/mo, andP_o (dry period) = 25 mm/mo, P_o (wet period) = 17 mm/mo.

The characterization "significant" snowfall is estimated to correspond to a situation where snow water – equivalent covers more than 5 % of the total basin rainfall of the Aliakmon basin and more than 10 % of the Acheloos basin during the snowy periods (consecutive months). Below this percentage, snowfall is found to behave like rainfall with direct melting and is characterized as "insignificant".

The rainfall and snowfall characteristic values are very important, since they actually control the time variability of the filter parameters. They depend on basin's climatic conditions, location, vegetation etc.

Empirical Fitting of the Filter

The filter, implemented in a computer code, is applied to both basins, with input the stationary monthly rainfall series and output the stationarized runoff series, as follows

$$\hat{q}_{p,t} = \bar{q}_t + a_i Q_{p,t-1} + b_i P_{p,t}, \quad (\text{for } t = 1, \quad Q_{p,t-1} = Q_{p-1,12}) \tag{18}$$

Using the input-output statistical characteristics the parameters, α and A are estimated as follows: for the Aliakmon river basin at Siatista as, α=0.44 and A=0.14 and for the Acheloos river basin at Avlaki as, α=0.52 and A=0.20. Then, starting with an arbitrary value b_1, i.e.: b_1=0.02, and following the empirical mechanism of variations of the filter parameters given in Table 1, the time series a_i and b_i, i=1...N, ($N = n \times 12$) have been estimated, by testing several times the "best" value for the exponent k, ($k \geqslant 0$). Two are the criteria for "best". First, the minimization of the mean squared error of the filter given by

Table 1 – Mechanism of variations of the filter parameters

Monthly rainfall $P_{p,t}$(mm) at month t	Formulae for b_i at month t	Snowfall during previous consecutive months (t-1,...)	Discharge at month t-1	Formulae for a_i at month t, with none or insignificant snowfall	Formulae for a_i at month t, with significant showfall
Increasing or constant					
$O \leqslant P_{p,t} < P_o$	$b_i = b_{i-1}$	None or insignificant		$a_i = \alpha(1-b_i)$	$a_i = \alpha(1-b_i)$
$P_o \leqslant P_{p,t} < P_S$	$b_i = \alpha^k(1-L)^{*1}$				
$P_{p,t} \geqslant P_S$	$b_i = 1-L$ or $b_i = \alpha^k (1-L)^{*2}$		$q_{p,t-1}<\bar{q}_{t-1}$	$a_i = \alpha^{k+1}(1-b_i)$	
		Significant			$a_i = \alpha(1-b_i)$
Decreasing					
$P_{p,t} \geqslant P_S$	$b_i = b_{i-1}(1-\alpha^2 L^2)$		$q_{p,t-1}>\bar{q}_{t-1}$	$a_i = \frac{1}{\alpha^{k-1}} (1-b_i)^{*3}$	
$P_S > P_{p,t} \geqslant P_o$	$b_i = b_{i-1}(1-\alpha L)$				
$P_o > P_{p,t} \geqslant O$	$b_i = b_{i-1}$				

(*1) $L = 1/(1 + A + \alpha^2 b_{i-1})$

(*2) In the presence of significant snowfall at month t.

(*3) If a_i exceeds 1, for $k \geqslant 2$, then it is set equal to 0.999...

Table 2 – Historical and simulated statistical characteristics, percentage errors and Portmanteau test. Empirical method.

Basin	Series	Mean (m^3/sec)	Variance (m^3/sec)^2	Skewness	Kurtosis	1st order autoc. coeff.	φ %	Portmanteau test $\bar{Q}$
Aliakmon r.	Historical	23.00	500.00	1.30	4.00	0.75	–	–
at Siatista	Simulated	25.00	476.20	1.27	3.78	0.76	3.90	31.10
Acheloos r.	Historical	50.30	2020.40	1.42	5.10	0.70	–	–
at Avlaki	Simulated	54.50	1858.80	1.36	4.72	0.71	4.30	28.50

$$\overline{\varepsilon^2} = \frac{1}{N} \sum_p \sum_t (q_{p,t} - \hat{q}_{p,t})^2, \quad p = 1 \ldots n, \; t = 1 \ldots 12, \tag{19}$$

with $N = n \times 12$ the number of data and $\hat{q}_{p,t}$ the estimated output value. The adequacy of the filter, checked by the Portmanteau lack of fit test on its residuals is the second criterion. The sum of the squares of the first ψ autocorrelation coefficients of the residual series $\hat{\xi}_i = q_{p,t} - \hat{q}_{p,t}$, of the model of Eq. (18), i.e. $r_y(\hat{\xi})$, multiplied by N, form $\tilde{Q}$ (Box and Jenkins 1970), given by

$$\tilde{Q} = N \sum_{y=1}^{\psi} r_y^2(\hat{\xi}) \tag{20}$$

$\tilde{Q}$ is supposed to be distributed as a χ^2 $(\psi$-$m)$ variable with $m=2$ the number of variables in the r.h.s. of Eq. (18). Then, if $\tilde{Q}$ is less than the χ^2 value, at a certain significance level with $(\psi$-$m)$ degrees of freedom, one can accept the model as adequate. Here, taking $\psi=25$, the χ^2 value at the 5 % significance level is 35.2 and at the 10 % level it is 32.0.

The value of the exponent k which minimizes the error and gives the most adequate model is: $k=1$, for both basins. It is worthwhile to mention here, that the empirically constructed response mechanism of the filter is the same for both basins, even though they have completely different hydraulicity; the Aliakmon river basin at Siatista has a mean specific runoff yield equal to: 0.011 $m^3/sec/km^2$ and the Acheloos river basin at Avlaki: 0.043 $m^3/sec/km^2$, namely, almost four times the previous one. It is also believed that this empirical fitting of the filter with the same mechanism for the time variation of the parameters can be applied to other basins too, at least of the same region. This belief is now being verified with an ongoing research on another two basins of the region.

The percentage error of the fitted filter is given by

$$\varphi = \frac{\overline{\varepsilon^2}}{\overline{q^2}} \times 100\% \tag{21}$$

where $\overline{q^2}$ is the mean of the squared input series; the smaller φ gives the more accurate model. The statistical characteristics (mean, variance, skewness, kurtosis and first order autocorrelation coefficients) of the simulated output series for the calibration period along with the historical statistical characteristics, the percentage errors and the Portmanteau statistics $\tilde{Q}$ on the residuals, are given in Table 2. The null hypothesis (Benjamin and Cornell 1970, pp. 416-417) that the historical statistical characteristics are preserved is accepted at the 10 % significance level, the percentage error is very small and the $\tilde{Q}$ value is lower than the χ^2 value at the 10 % significance level for both basins. These really satisfying accuracy and adequacy of the empirically fitted filter are to be further verified by comparison to the results of automatically filled filters, using three different "black-box" methods of fitting.

Automatic Black-Box Fitting of the Filter

The "black-box" methods used to fit the filter of Eq. (18) are: a multiple regressional method, a mean squared error optimizing method and a statistical filtering technique. The parameters a_i and b_i are to be taken either as constants or as time variables, with seasonal (monthly, $i=1,2,...,12$), but not yearly variation.

Multiple Regressional Method with either Constant or Variable Parameters

Assuming that $Q_{p,t-1}$ and $P_{p,t}$ are the independent and $Q_{p,t}$ the dependent variable of a three-variable linear regression, of the form

$$Q_{p,t} = C_i + a_i Q_{p,t-1} + b_i P_{p,t} \tag{22}$$

one can obtain the parameters a_i and b_i, the constant of regression C_i and the correlation coefficient r by using the least squares method and the formulas given in statistical books (Yevjevich 1982, pp. 264-268).

The estimated constant parameters of the regression are: $b_i = 0.25$, $a_i = 0.55$ and $c_i = -16.2$, with a correlation coefficient equal to $r = 0.76$ for Aliakmon river basin at Siatista, whereas: $b_i = 0.31$, $a_i = 0.60$ and $c_i = -20.3$, with a correlation coefficient equal to: $r = 0.78$, for Acheloos river basin at Avlaki.

For the case of variable parameters, twelve monthly regressions have been tried, but the majority of the monthly multiple correlation coefficients have been found to be insignificant at the 5 % significance level (Yevjevich 1972 p. 272) for both basins. This is why the multiple regressional method with variable parameters is not given herein.

Mean Squared Error Optimizing Method with either Constant or Variable Parameters

The method is based on the minimization of the error given by Eq. (19) and the trial and error selection of the parameters of the filter in Eq. (18). After several trials in the computer and for the case of constant parameters, the following values have been estimated; for Aliakmon river basin at Siatista: $a_i = 0.59$ and $b_i = 0.05$, whereas for Acheloos river basin at Avlaki: $a_i = 0.63$ and $b_i = 0.10$.

In the case of variable parameters, twelve monthly optimizations have been tried and the best results are given in Table 3.

Statistical Filtering Technique

The technique is based on the Schwartz and Shaw (1975) procedure of filtering previously described. It uses Eq. (7) and (8) for the estimation of the parameters which, as explained elsewhere in this paper, are practically constants. Using the α and A values already estimated for both basins, the limiting constant values for the parameters of the filter in Eq. (18) become: $b_i = 0.15$, $a_i = 0.37$ for the Aliakmon river basin at Siatista and, $b_i = 0.20$, $a_i = 0.42$ for the Acheloos river basin at Avlaki.

Results of the Automatic "Black-Box" Fitting of the Filter

The simulated by the filters statistical characteristics of the output series with the historical ones, the percentages errors and the Portmanteau tests $\tilde{Q}$ for all automatically fitted filters are given in Table 4.

The null hypothesis that the historical characteristics are preserved is accepted at the 10 % significance level with all methods, except the filtering technique. The $\tilde{Q}$-statistics shows an inadequacy of the filter fitted by the multiple regressional method and the mean squared error optimizing method with constant parameters at the 5 % and the 10 % levels for Aliakmon and Acheloos river basins, correspondingly. It shows also the inadequacy of the filter fitted by the statistical filtering technique, since its $\tilde{Q}$-statistics does not pass at the 5 % level for both basins. The $\tilde{Q}$ value for the mean squared optimizing method with variable parameters is the best of all previous ones, since it is lower than the limiting value 32 at the 10 % level for both basins. The percentage error varies from method to method and the smallest one belongs to the mean squared error optimizing method with variable parameters, whereas the biggest one to the statistical filtering technique, for both basins.

Comparison between Empirical and "Black-Box" Methods – Verification

Comparing Tabes 2 and 4 one can easily see that, even though the null hypothesis that the historical statistical characteristics are preserved is accepted at the 10 % level for all methods except the filtering technique, the simulated by the empirically fitted filter characteristics are the closest to the historical ones for both basins. Besides, the empirically fitted filter has the smallest percentage errors of all and the Portmanteau test values $\tilde{Q}$ on its residuals are the lowest of all and below the limiting value 32.0 at the 10 % significance level, for both basins. Therefore, the empirically fitted filter is shown to be more accurate and adequate in representing monthly rainfall-runoff responses than the automatically fitted filters. The superiority though of the empirical method over the other methods and especially over the optimizing method with variable parameters, which is found to be the second best method, is better shown in the following paragraph, where the filter is used to simulate data outside the calibration period. This is due mainly to the fact that the empirical method can change dynamically and continously the parameters according to real time information, which is improved in comparison to the information used by the other method. Additionally, the criterion of the minimum monthly squared errors, according to which the parameters of the optimizing method have been estimated, is not any more verified outside the calibration period. On the other hand, the two basic parameters α and A of the empirical method do not change significantly when extending the filter input and output stationary series.

Table 3 – Variable filter parameters by the mean squared error optimizing method.

Basin		O	N	D	J	F	M	A	M	J	J	A	S
Aliakmon r.	a_i	0.320	0.555	0.750	0.930	0.496	0.545	0.625	0.624	0.350	0.413	0.497	0.940
at Siatista	b_i	0.032	0.030	0.048	0.075	0.110	0.146	0.088	0.030	0.020	0.009	0.011	0.008
Acheloos r.	a_i	0.380	0.580	0.690	0.923	0.580	0.599	0.668	0.524	0.433	0.469	0.579	0.960
at Avlaki	b_i	0.094	0.084	0.094	0.110	0.181	0.203	0.156	0.084	0.077	0.034	0.036	0.017

Table 5 – Historical and simulated statistical characteristics, percentage errors and Portmanteau test. Verification period.

Basin	Method	Mean (m^3/sec)	Variance (m^3/sec)^2	Skewness	Kurtosis	1st order autoc. coeff.	φ %	Portmanteau test $\tilde{Q}$
	Historical series	30.60	865.00	1.77	7.00	0.75	–	–
Aliakmon r. at Siatista	Mean sq. error optim. Variable param.	24.00	571.60	1.17	4.55	0.84	12.90	33.50
	Empirical	27.70	727.30	1.49	5.60	0.78	6.40	19.20
	Historical series	63.10	2450.00	1.81	6.10	0.69	–	–
Acheloos r. at Avlaki	Mean sq. error optim. Variable param.	52.00	1572.50	1.18	3.90	0.78	13.20	34.30
	Empirical	57.10	1986.10	1.47	4.83	0.73	6.80	20.50

Table 4 – Historical and simulated statistical characteristics, percentage errors and Portmanteau test. "Black-box" methods.

Basin	Method	Mean (m^3/sec)	Variance (m^3/sec)^2	Skewness	Kurtosis	1st order autoc. coeff.	φ %	Portmanteau test $\tilde{Q}$
Aliakmon r. at Siatista	Historical series	23.00	500.00	1.30	4.00	0.75	–	–
	Multiple regr. Const. param.	25.10	424.00	1.52	4.78	0.77	6.10	40.00
	Mean sq. error optim. Const. param.	25.90	444.20	1.46	4.71	0.77	5.10	35.90
	Stat. filtering techn. Const. param.	33.50	325.50	0.85	2.56	0.84	13.50	56.20
	Mean sq. error optim. Variable param.	25.30	465.40	1.35	4.35	0.76	4.30	31.80
Acheloos r. at Avlaki	Historical series	50.30	2020.40	1.42	5.10	0.70	–	–
	Multiple regr. Const. param.	54.70	1660.00	1.70	6.24	0.72	6.50	34.60
	Mean sq. error optim. Const. param.	57.10	1717.40	1.63	6.15	0.72	5.80	33.30
	Stat. filtering techn. Const. param.	66.10	1273.80	0.91	3.21	0.80	13.80	43.30
	Mean sq. error optim. Variable param.	54.90	1818.40	1.51	5.67	0.71	4.70	29.10

Verification

The empirically fitted filter is to be verified for another four years of data outside the calibration period; namely for the data of the periods 1962-1966 and 1965-1969, for Aliakmon and Acheloos river basins, correspondingly. In order to prove further its superiority over the other methods, the best of these methods, namely the optimizing method with variable paramters, is also used for fitting the filter.

During the verification of the filter no historical runoff data are used and instead of using $Q_{p,t-1}$ in Eq. (18), the estimated value $\hat{Q}_{p,t-1}$ is used.

The simulated runoff statistical characteristics with the historical ones, of the verification period, the percentage errors and the Portmanteau tests $\tilde{Q}$ on the residuals for both methods and basins are given in Table 5.

The null hypothesis that the historical statistical characteristics are preserved is accepted at the 10 % level for the empirical method but it is not accepted at this level for the mean squared error optimizing method with variable parameters, for both basins. The percentage error of the empirically fitted filter is much smaller than the percentage error of the filter fitted by the other method and the $\tilde{Q}$-statistics on the residuals shows the adequacy of the empirically fitted filter and the inadequacy of the filter fitted by the optimizing method at the 10 % level, for both basins. Additionally, the superiority of the empirical method for representing the dynamics of the monthly rainfall-runoff responses can be seen from Fig. 8, where the historical and simulated hydrographs of the verification period are given. The empirically fitted filter fits much better than the other one fitted by the optimizing method on the historical pattern, especially in the rising limbs, the peaks and the Spring months (existence of pronounced delayed runoff phenomenon) of the hydrographs for both basins.

Conclusions

Conclusions of this research are the following:

1) The main advantages of the presented empirical method of fitting over an automatic "black-box" fitting of a monthly rainfall-runoff Kalman filter are:

a) Important dynamic characteristics of the basin runoff response, related to additional to the rainfall input information, are incorporated into the response mechanism of the filter, improving thus its efficiency for representing the physical process, without changing its simple two-variable structure.

b) The filter parameters are interpreted and their time variation is controlled.

2) For monthly rainfall-runoff Kalman filters the needed additional to the rainfall input information for reliable representation of the basin runoff response and of its dynamics consists of:

a) The percentage of snowfall into monthly basin rainfall values.

b) The magnitude of basin rainfall at month t in comparison to the magnitude of

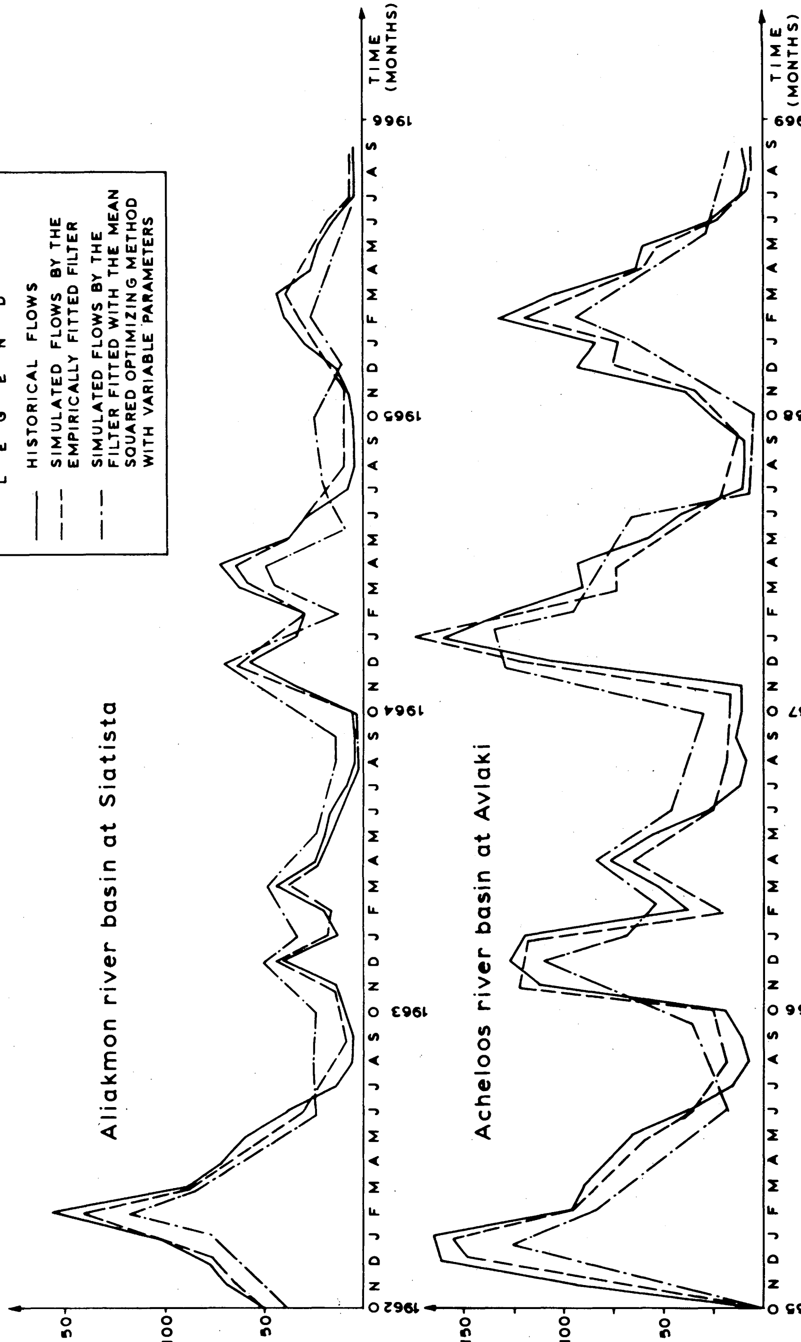

Fig. 8. Visual comparison between empirical and "black-box" method of fitting. Verification period.

rainfall at month t-1 (increasing, constant or decreasing), and

c) The amount of monthly basin rainfall associated with a basin's moisture and runoff characteristics

3) The empirically fitted filter, applied to two Greek basins, is shown to be more accurate and adequate to represent monthly rainfall-runoff responses and their dynamics, preserving simultaneously the runoff historical statistical characteristics, than three other filters automatically fitted by a multiple regressional method, a mean squared error optimizing method and a statistical filtering technique, with either constant or variable parameters. Its superiority is more pronounced when it is used outside the calibration period. For this reason it is very useful for accurate extension or filling-in of monthly runoff records from rainfalls for the purposes of hydrologic design.

4) The empirical mechanism for the time variations of the filter parameters (filter response mechanism) is found to be the same for both basins of application, even though they exhibit different hydraulicity.

Acknowledgements

The author wishes to thank the Environmental Research Center of the Greek Public Power Corporation for providing the monthly rainfall and runoff data.

References

Benjamin, T. R., and Cornell, A. C. (1970) *Probability, Statistics and Decision for Civil Engineers.* McGraw-Hill Co., New York, pp. 416-417.

Box, G. P., and Jenkins, G. M. (1970) *Time Series Analysis-Forecasting and Control.* Holden-Day Co., San Francisco, pp. 50-293.

Hino, M. (1973) On Line Prediction of Hydrologic Systems. Proc. XVth IAHR Conf., Istanbul, pp. 121-129.

Kalman, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems. *J. Basic Engi., Trans. Am. Soc. Mech. Eng., 82/2,* 35-45.

Rodriguez-Iturbe, I. B., Valdes, J. B., and Velasquez, J. M. (1978) Application of Kalman Filters in Rainfall-Runoff Studies. In: Chao-Lin Chiu (Editor), *Applications of Kalman Filter to Hydrology, Hydraulics and Water Resources.* Stochastic Hydraulics Program, University of Pittsburgh, Penn., pp. 223-253.

Schwartz, M., and Shaw, L. (1975) *Signal Processing, Discrete Spectral Analysis, Detection and Estimation.* McGraw-Hill Co., New York, pp. 331-337.

Yevjevich, V. (1972) Probability and Statistics in Hydrology. Water Resources Publ., Fort Collins, Colo., pp. 264-272.

Received: 3 December, 1982

Address: 6 Themistokleous Street,
Cholargos, Athens,
Greece.