ISMM

IASTED

Proceedings of the IASTED International Conference

MODELLING AND SIMULATION

Pittsburgh, PA, U.S.A.
May 10-12, 1993

Editor: M. H. Hamza

A Publication of
The International Association of Science
and Technology for Development - IASTED

ISBN: 0-88986-173-0
ISSN: 1021-8181

IASTED
ANAHEIM •CALGARY- ZURICH

UNIVARIATE CONCEPTUAL-STOCHASTIC MODELS FOR SPRING RUNOFF SIMULATION

Pierluigi Claps
Dept. Ingegneria e Fisica dell'Ambiente, Università della Basilicata, Via della Tecnica 3, 85100 Potenza- Italy

Federico Murrone
Dept. Idraulica, Gestione delle Risorse Idriche ed Ingegneria Ambientale, Università di Napoli "Federico II", 80126 Napoli - Italy

ABSTRACT

The spring runoff process is analyzed in this paper under a conceptual-stochastic framework, in order to identify stochastic models for simulation and forecasting purposes. The analysis is based on univariate stochastic models, in which maximum advantage is taken from the information contained in spring runoff series.

Considering a linear conceptual system transforming recharge into runoff it is shown that if the system is made up of a single linear reservoir the discretized spring runoff can be considered as following an ARMA(1,1) process (AutoRegressive and Moving Average process of order 1 and 1). This reduces to an AR(1) process for simplified form of the input. If two distinct components are recognized, for instance on the monthly scale, runoff can be considered as the output of two parallel linear reservoirs leading to an ARMA(2,2) stochastic process. On the monthly scale, models with periodic independent residual are proposed.

Data from two springs located in Southern Italy, which are part of the drinking water supply system of the city of Potenza, are analyzed with interesting results.

Key-words: Runoff, Spring, ARMA models, Recharge distribution

INTRODUCTION: SPRING RUNOFF MODELING ASPECTS

In water resources planning and management it is often required to evaluate potentiality of groundwater resources. In particular, springs can play an important role within drinking water supply systems. Management of water supply systems requires simulation models as support to procedures of optimization, and forecasting models as support to operation criteria in drought periods.

Stochastic models for spring runoff simulation have so far received little attention, probably due to the relative ease in a reasonable characterization of the conceptual scheme underlying the transformation of recharge into runoff. Usually, the aquifer is assumed as a single linear reservoir and modeling efforts are concentrated in the estimation of the recession coefficient, with techniques proposed for the evaluation of the storage coefficient from the base-flow recession curves in riverflow series (see, for instance, [6]; [2]; [9]). Evaluation of the recession coefficient can be sufficient for a short-term forecasting, whereas a more detailed analysis of the spring runoff process is needed for simulation purposes.

A correct reproduction of the stochastic characteristics of the process under study requires the characterization of the probabilistic structure of the recharge and a careful observation of the runoff series over different time scales, in order to recognize and to reproduce possible long-term persistence effects. Moreover, even in a linear conceptual framework is possible to build refined models by taking into account the within-period distribution of the recharge. In this paper these issues are addressed in view of the identification of accurate yet simple conceptually-based univariate models.

Univariate models discussed below allow estimation of the recharge as an hydrologic inverse problem. This can be of considerable help for simulation with respect to requirements of bivariate rainfall-runoff models, such as the determination of the hydrogeological basin area.

BASIS FOR THE CONCEPTUAL SCHEMATIZATION OF THE GROUNDWATER SYSTEM

A conceptual schematization of the physical characteristics of the phenomenon is the base for the stochastic model identification of the spring runoff process. It is, in fact, particularly convenient to characterize the spring runoff as the outlet of a system having rainfall as input, evapotranspiration and runoff as output and where the aquifer is the a linear element transforming the *effective rainfall* (*i.e.* rainfall–evapotranspiration) into runoff. An accurate description of the process under study can be attained by reproducing the random structure of precipitation, the nature of processes acting into the soil (infiltration, evapotranspiration, percolation) and the aquifer response to the hydrological input.

The hydrologic system transforming net rainfall into runoff should be, in general, considered as a nonlinear, rather than linear, reservoir, since it is intuitive that the characteristics of the response can vary according to the storage volume. On the other hand, patterns in disagreement with the response of the single linear reservoir can be also accounted for the presence of two distinct runoff components, that can be considered as the output of two parallel linear reservoirs with different storage constant.

Reproduction of nonlinearities in the reservoirs response causes the big disadvantage of not allowing the use of autocorrelation analysis, which is the most important tool in time series modeling. For this reason, reservoirs with linear response are commonly considered in applications. Moreover, it can be considered that nonlinearities tend to weaken as the time interval of aggregation increases and that a careful description of the input process can aid reducing global inaccuracies. For this reason, only the case of system made up of two reservoirs is examined here.

Runoff components with different responses to the recharge can be identified in presence of distinct hydrogeologic formations into the basin. Fig. 1 shows an example of change in slope (in semilog paper) of the recession occurring in the month of September, which is at the end of the dry season for the site examined. Since runoff from a linear reservoir not subject to recharge exhibits a linear recession in semilog paper, contributions from two different runoff components could be identified in this case.

Regarding the net rainfall structure, it is suggested that this variable retains the basic characteristics of the precipitation totals, such as

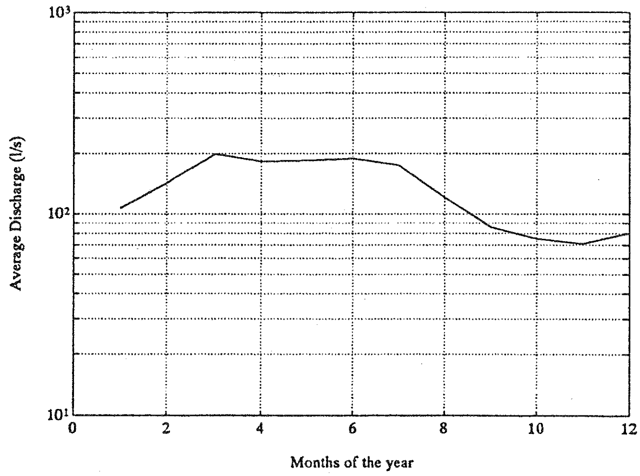

Fig. 1. Fossa Cupa spring. Average monthly discharges in year 1972.

independence and seasonality (on sub-annual scale). Therefore, net rainfall is considered as a stationary independent process on a annual basis and a periodic independent process on a monthly basis.

In order to preserve the formal correspondence between conceptual and stochastic representations of the process, no data transformation is made. Moreover, data are not deseasonalized, since this operation not only does not completely eliminate periodicity in the autocorrelation function but also causes the removal of characters having a sub-annual evolution, such as runoff components with sub-annual lag.

CONCEPTUAL-STOCHASTIC MODEL IDENTIFICATION

Discretized equations of the linear reservoir

The continuous-time response of a linear reservoir to an input $r(\tau)$ is (*e.g.*[4]):

$$Q(t) = Q_0\, e^{-t/k} + \int_0^t \frac{1}{k}\, e^{-(t-\tau)/k}\, r(\tau)\, d\tau \tag{1}$$

where Q is the discharge, k is the reservoir *storage coefficient* and $Q_0=Q(t=0)$. Integration of (1) over a unit-time interval gives

$$D_1 = \int_0^1 Q(t) = \int_0^1 Q_0 e^{-t/k}\, dt + \int_0^1 \frac{1}{k} e^{-t/k} \int_0^t e^{\tau/k}\, r(\tau)\, d\tau\, dt \tag{2}$$

where D_1 is the outflow volume. Given the initial volume $V_0=kQ_0$, by definition of linear reservoir, equation (2) becomes

$$D_1 = (1-e^{-1/k})V_0 + \int_0^1 \frac{1}{k} e^{-t/k} \int_0^t e^{\tau/k}\, r(\tau)\, d\tau\, dt \tag{3}$$

To express the above equation in discretized form, the input is often considered as an impulse occurring at the beginning of the interval. A generalized discretized form can be obtained by introducing the *recharge recession coefficient*

$$r_k = 1 - \int_0^1 \frac{1}{k}\, e^{-t/k} \int_0^t \frac{1}{R}\, e^{\tau/k}\, r(\tau)\, d\tau\, dt \tag{4}$$

where $R = \int_0^1 r(\tau)\, d\tau$ is the *recharge* volume within the interval.

Assuming, for the generic (t−1,t) unit-time interval, D_t as the runoff, V_{t-1} as the storage at the beginning of the interval, R_t as the recharge in the interval and $c_k=e^{-1/k}$, equation (3) can be expressed as

$$D_t = (1-c_k)\, V_{t-1} + (1-r_k)\, R_t \tag{5}$$

From (5) the meaning of coefficient r_k emerges clearly, as the equivalent of the *recession coefficient* c_k but responsible for the recession of the within-interval recharge volume. The form of the recharge function $r(\tau)$ has a direct influence on r_k. This is particularly important when a deterministic within-period variability of the input can be recognized, as in the case of series aggregated annually, where the input displays a seasonal variability. It is also worth noting that in (4) the storage coefficient k plays the role of a factor of reduction of the importance of the variability in $r(\tau)$.

In case of uniform recharge $r(\tau)=R/\Delta$ over a sub-interval of duration $\Delta\leq 1$ and initial time T_i, the expression of r_k is

$$r_k = \frac{k}{\Delta}\,(1-e^{-\Delta/k})\,[1-(1-e^{(1-T_i-\Delta)/k})] \tag{6}$$

that reduces to $r_k = e^{-(1-T)/k}$ (used in [8]) for $\Delta=0$. For $T_i=0$ and $\Delta=1$ relation (6) assumes the form

$$r_k = k\,(1-e^{-1/k}) \tag{7}$$

used in [1]. For $\Delta=0$ and $T=0$, which corresponds to impulse recharge at the beginning of the interval, (6) gives $r_k = e^{-1/k} = c_k$. Where there is no particular evidence of a within-period rainfall pattern, as in the case of monthly data, relation (7) for uniform $r(\tau)$ is chosen.

A seasonal pattern in hydrologic data can be estimated, with good accuracy, by means of Fourier series. The general expression of a Fourier series of a periodic function $f(t)$ is:

$$f(t) = A_0 + \sum_{n=1}^{N} A_n \cos\left(\frac{2n\pi}{T}t + \Phi_n\right) \tag{8}$$

where N is the number of harmonics and A_0 represents the mean of $f(t)$ over the period T. For fixed N, the above expression has $2N+1$ parameters, namely A_0, $\{A_1, A_2, \ldots A_N\}$ and $\{\Phi_1, \Phi_2, \ldots \Phi_N\}$, that can be estimated, for instance, by means of a least squares method.

Assuming that $f(t)$ is the series of non-dimensional monthly net rainfall means, $f(t)=r(t)/R$ and $A_0=1$. Substitution of (8) into (4), with the interval T set to 1 year, produces

$$r_k = 1-(ke^{-1/k}+1-k) - \sum_{n=1}^{N} \frac{A_n k e^{-1/k}\left[\cos\Phi_n - 2\pi nk\, \sin\Phi_n\right]}{4(\pi kn)^2+1} + \sum_{n=1}^{N} \frac{A_n\left[2\pi kn\, \cos(2\pi n+\Phi_n) - \sin(2\pi n+\Phi_n) + (4(\pi kn)^2+1)\sin\Phi_n\right]}{2\pi n(4(\pi kn)^2+1)} \tag{9}$$

Given the simple form of the monthly rainfall means curve, two harmonics are considered sufficient for an accurate fitting.

Stochastic model for a single linear reservoir

Consider a conceptual model, made up of a single linear reservoir. The outlet D_t of the reservoir in a unit-time interval (t−1,t) is expressed in (5) as a function of the initial storage V_{t-1} and of the recharge R_t. The mass-balance equation for the groundwater storage is

$$V_t = c_k V_{t-1} + r_k R_t \tag{10}$$

Substituting into (10) the expression of V_{t-1} resulting from (5), one obtains

$$V_t = \frac{c_k D_t}{1-c_k} - \frac{c_k(1-r_k)R_t}{1-c_k} + r_k R_t \tag{11}$$

Considering in (11) V_{t-1}, D_{t-1} and R_{t-1} instead of V_t, D_t and R_t, which does not affect generality, and resubstituting V_{t-1} as obtained from (5), gives

$$D_t - c_k D_{t-1} = (1-r_k)\, R_t - (c_k - r_k)\, R_{t-1} \tag{12}$$

If R_t is considered as an independent stochastic process, equation (12) represents an ARMA(1,1) (autoregressive and moving average of order 1 and 1) process [3]. Introducing the variables $d_t = D_t - E[D_t]$ and

$$\varepsilon_t = (1 - r_k)(R_t - E[R_t]) \tag{13}$$

both with zero mean, this model writes, in Box-Jenkins notation,

$$d_t - \Phi\, d_{t-1} = \varepsilon_t - \Theta\, \varepsilon_{t-1} \tag{14}$$

with meaning of the autoregressive and moving average parameters as

$$\Phi = c_k\,; \qquad \Theta = \frac{c_k - r_k}{1 - r_k} \tag{15}$$

Estimation of ARMA parameters allow the evaluation of both recession coefficient c_k and r_k while the ARMA model residual corresponds to the effective rainfall series through relation (13). On a monthly basis, given the seasonal character of the effective rainfall, the resulting model is an ARMA(1,1) with periodic independent residual (PIR-ARMA) [6].

If the estimate of Θ is statistically significant (*i.e.* standard error less than half the value of the parameter) estimate of r_k from Φ and Θ is reliable. Evaluation of r_k can be particularly useful if a riverflow series in a contiguous watershed must be analyzed, given that in that case a pre-determined r_k is required (see [5]). On the other hand, in case of scarcity of data, which affects especially the estimation of Θ, relations (6) or (9), together with a-priori information on the form of the within-period precipitation, can provide a more adequate estimation of r_k.

In the limit case of impulse recharge at the beginning of the interval, $r_k = c_k$ and (12) simplifies in

$$D_t - c_k\, D_{t-1} = (1 - c_k)\, R_t \tag{16}$$

which is equivalent to an AR(1) (autoregressive of order 1) process.

Stochastic model for two parallel linear reservoirs

With reference to [5], if aI_t is the input to the long-term response aquifer and $bR_t = (1-a)R_t$ is the input to the aquifer with short-term response, the corresponding expression of the spring runoff is

$$D_t = (1-c_k)V_{t-1} + (1-c_q)W_{t-1} + a(1-r_k)R_t + (1-a)(1-r_q)R_t \tag{17}$$

with c_q and W_t as parameters of the short-term aquifer. The volume balance equations for the groundwater storage are:

$$V_t = c_k\, V_{t-1} + a\, r_k\, R_t \tag{18}$$

$$W_t = c_q\, W_{t-1} + (1-a)\, r_q\, R_t \tag{19}$$

Putting in (17) expressions of W_{t-1} and V_{t-1} obtained from equations (18) and (19) and rearranging, gives one equation in D_t, D_{t-1}, D_{t-2}, R_t, R_{t-1}, R_{t-2}:

$$D_t - (c_k + c_q)D_{t-1} + (c_k c_q)D_{t-2} =$$

$$= [1 - a r_k - (1-a) r_q]R_t - [c_k + c_q - a\, r_k(1+c_q) - (1-a)\, r_q(1+c_k)]R_{t-1} + \tag{20}$$

$$- (a\, r_k c_q + (1-a)\, r_q c_k - c_k c_q)\, R_{t-2}$$

With reference to variables $\varepsilon_t = [1 - a r_k - (1-a) r_q](R_t - E[R_t])$ and d_t, with zero-mean, and considering on the monthly time scale R_t as a periodic independent stochastic process, the above representation is equivalent to a PIR-ARMA(2,2) process

$$d_t - \Phi_1\, d_{t-1} - \Phi_2\, d_{t-2} = \varepsilon_t - \Theta_1\, \varepsilon_{t-1} - \Theta_2\, \varepsilon_{t-2} \tag{21}$$

which becomes a PIR-ARMA(2,1) in the hypothesis of impulse input set at the beginning of the interval. Relations between conceptual and stochastic parameters for the ARMA(2,2) are reported in [5].

ISSUES ON MODEL APPLICATION

In [5] a probabilistic model of monthly net rainfall R_t was proposed, as the *Bessel* distribution. It results from the sum of a poissonian number of events, with parameter ν, with intensity distributed exponentially with parameter $1/\beta$. R_t assumes only positive values and presents finite probability at zero. Annual net rainfall can be considered as following a Box-Cox transformation of the Normal distribution, with index varying between 0 (lognormal distribution) and 0.5.

In the analysis of monthly riverflow time series it has been pointed out [6] the necessity of properly aggregating data when estimating parameters of conceptually-based stochastic models. Particularly, it was shown that the storage coefficient of aquifers with over-year recession is best estimated on the annual scale, while on the monthly scale only the parameter of the aquifer with over-month recession is estimated conveniently.

The presence of a runoff component with over-year lag can be recognized from the analysis of autocorrelation in annual data. If two distinct seasons exist, one wet and one dry, a correct evaluation of the autocorrelation in annual data is achieved by considering the hydrological year, that starts at the end of the dry season. In fact, at the end of the dry season possible short-term recession components practically vanish, so that correlation in runoff in the hydrological year can be ascribed only to an over-year recession component. If significant correlation is found on the annual scale, the possible presence of a faster, sub-annual, runoff component can be evaluated by visual inspection of monthly runoff in semilog paper and by application of the iterative estimation procedure proposed [6] for the PIR-ARMA(2,2) model.

Simulation of monthly data for a spring with one only over-year response component can be made by retaining in the PIR-ARMA(1,1) model of monthly runoff the value of k estimated on the annual scale. Coefficient r_k can be calculated through (7) and the structure of net rainfall can be obtained from the model residual (relation 13).

Data analysis

Monthly runoff data by two springs located in Southern Italy, near the city of Potenza, are analyzed. Main characteristics of series are shown on Tab. 1. The site under study presents a dry season during springtime and summertime, which is recognizable form the observation of Fig. 1.

Spring	Series length (Years)	Mean (l/s)	Std. Dev. (l/s)
Fossa Cupa	21	108.29	41.05
San Michele	21	105.89	49.76

Tab. 1. Characteristics of the time series analyzed.

The hypothesis of presence of correlation on annual data is first tested. Autocorrelation function of both series aggregated on the hydrologic year presents a $\rho(1)$ value within the confidence band and remaining $\rho(m)$ values which do not follow the exponential decay typical of the AR and ARMA models (see Tab. 2).

```
San Michele                               Fossa Cupa
                         lag   ac                                    lag   ac

.      |******         1  0.461      .      |******           1  0.451
.     *|      .        2 -0.090      .     *|      .          2 -0.096
.    **|      .        3 -0.142      .   ***|      .          3 -0.218
.      |*     .        4  0.096      .      |      .          4 -0.022
.      |*     .        5  0.052      .      |*     .          5  0.049
.   ***|      .        6 -0.205      .    **|      .          6 -0.153
.   ***|      .        7 -0.241      .      |      .          7 -0.001
.      |      .        8 -0.034      .      |*     .          8  0.082
.      |      .        9 -0.020      .      |      .          9  0.011
.     *|      .       10 -0.065      .    **|      .         10 -0.184
```

Tab. 2. Total autocorrelation function of average discharge in the hydrological year.

Based on this test and on the high standard error values of the estimated Φ and Θ parameters, observable in Tab. 3, correlation of annual data is assumed statistically negligible.

Given the results of the analysis of annual data, on the monthly scale a conceptual model with a single linear reservoir is considered. Estimation of the PIR-ARMA(1,1) model parameters on the two series produced the results shown in Tab.3. The effects of this schematization with respect to the recognition of two distinct recession curves on the observed data will be discussed hereafter.

Spring	Scale	Φ	std. err. Φ	Θ	std. err. Θ	R^2	k (months)
Fossa Cupa	1 Year	0.020	0.167	-0.953	0.028	0.490	3.07
	1 Month	0.740	0.028	-0.568	0.053	0.817	3.33
San Michele	1 Year	0.280	0.265	-0.425	0.276	0.351	9.42
	1 Month	0.786	0.026	-0.525	0.055	0.837	4.16

Tab. 3. Parameter estimates on annual and monthly basis.

As can be recognized from Tab. 3, standard errors of the estimates of Φ and Θ on annual data are comparable with parameter values, while standard errors of Φ and Θ for monthly data are very low with respect to parameter estimates. Given that the variance explained by the PIR-ARMA(1,1) model (R^2) applied to monthly data is quite high, one can conclude that even disregarding the presence of a possible component with slower response the stochastic description of both series is accurate. On the other hand, it cannot be excluded that the change in slope discussed earlier could be due to a systematic increase in recharge.

Finally, the comparison between observed and reconstructed series (Fig. 2 and 3) highlights small deviations in correspondence of runoff minima. These deviations cannot be attributed to the presence of a slower runoff component, which would have an opposite effect, and are more likely due to minor non-linearities in the groundwater system.

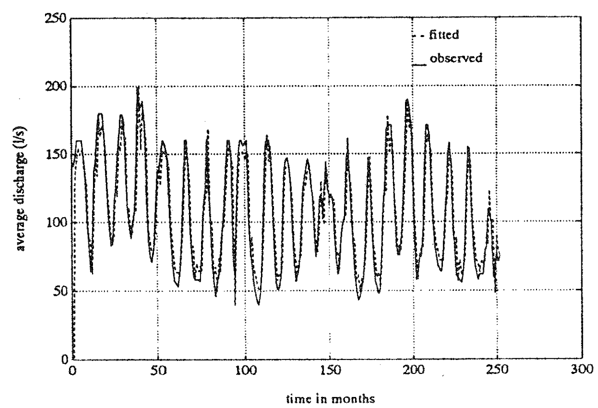

Fig. 2. Fossa Cupa. Data fitting with the PIR-ARMA(1,1) model.

As a verification of the efficiency of the estimates, r_k evaluated through the second of (15) was compared to values obtained from the hypothesis of uniform within-month net rainfall. From the second of (15) $r_k = 0.83$ for Fossa Cupa and $r_k = 0.860$ for San Michele were obtained, while from relation (7) $r_k = 0.86$ and $r_k = 0.89$ were found respectively. The difference between the two estimates is less than 4%.

CONCLUDING REMARKS

A conceptual-stochastic schematization is proposed in this paper for the spring runoff, considered as the output of a linear system feeded by a random input. It is shown that when two recession components are found in monthly data runoff follows an ARMA(2,2) stochastic process with periodic independent residual (PIR-ARMA). If runoff is the result of the response of a single aquifer to net rainfall, the appropriate model on the monthly scale is a PIR-ARMA(1,1).

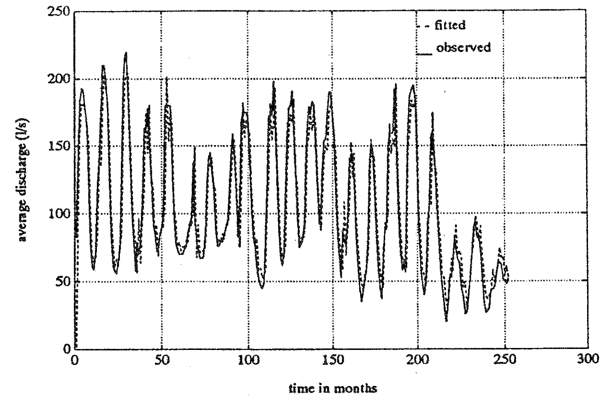

Fig. 3. San Michele. Data fitting with the PIR-ARMA(1,1) model.

A *recharge recession* coefficient, r_k is defined and expressions for different forms of the within-period recharge are obtained. These relations can be useful if the direct estimate of r_k through the ARMA parameters is not possible or unreliable, as in the case of data deficiency.

Estimation of model parameters is strongly affected by the scale of aggregation. This has been discussed in [6] for the case of monthly riverflows and analyzed in simulation experiments, that will be reported in a forthcoming paper. Consequently, data aggregated over the hydrologic year were first analyzed to detect the possible presence of a component with over-year lag, which is best estimated on the annual scale. Using information from the autocorrelation function and considering the variability of parameter estimates, this hypothesis was rejected for both series examined. Hence, the PIR-ARMA(1,1) model was fitted to monthly runoff with satisfying results in terms of explained variance and observed series reconstruction.

REFERENCES

[1] Andreu J. Modelos agregados y distribuidos, modelos unicelulares. Modelos Glover-Jenkins, in *Utilización conjunta de aguas superficiales y subterráneas,* ed. by M. Varela, S.G.O.P., Madrid, 1983 (in spanish).

[2] Bako, M.D. and D.N. Hunt. Derivation of baseflow recession constant using computer and numerical analysis, *Hydrol. Sci. Journal,* **33**(4), 357-367, 1988.

[3] Box, G.E. and G. Jenkins, *Time Series Analysis, Forecasting and Control,* Holden-Day, San Francisco, 1970. (2nd. Edition, 1976).

[4] Chow, V.T., Maidment, D.R. and Mays, L.W., *Handbook of Applied Hydrology,* p.208, McGraw-Hill, 1988.

[5] Claps, P. and F. Rossi, A conceptually-based ARMA model for monthly streamflows, in J.T. Kuo and G.F.Lin (Eds.) *Stochastic Hydraulics '92,* Proc. of the VI IAHR Intl. Symp. on Stochastic Hydraulics, Dept. of Civil Engrg., NTU, Taipei (Taiwan), 1992.

[6] Claps P., F. Rossi and C. Vitale, Conceptual-stochastic modeling of seasonal runoff using ARMA models and different scales of aggregation, to be published on *Water Resour. Res.,* 1993.

[7] Institute of Hydrology, *Low Flow Studies,* Research Report, Wallingford, Oxon, U.K., 1980.

[8] Moss, M.E. and M.C. Bryson, Autocorrelation structure of monthly streamflows, *Water Resour. Res.,* **10**(4), 737-744, 1974.

[9] Nathan, R.J. and T.A. McMahon. Evaluation of automated techniques for base flow and recession analyses, *Water Resour. Res.,* **26**(7), 1465-1473, 1990

Acknowledgments

This work was supported by funds granted by the Italian National Research Council - Group for Prevention from Hydrogeological Disasters, contracts no. 91.02633.42. and 91.02588.