Proceedings of the Sixth IAHR International Symposium on Stochastic Hydraulics, Taipei, 1992

A CONCEPTUALLY-BASED ARMA MODEL FOR MONTHLY STREAMFLOWS

P. CLAPS, Ph.D.
Dept. of Envir. Engrg. and Phys., Università della Basilicata, Potenza, 85100 Italy
and
F. ROSSI, Professor
Inst. of Civil Engineering, Università di Salerno, Fisciano, 84080 Italy

ABSTRACT: Based on physical considerations, the runoff process is considered as the output of a linear conceptual system whose structure depends on the hydrogeological characteristics of the watershed. At the monthly scale, runoff from Central-Southern Italy basins are considered as produced by two components coming from distinct aquifers, with long-term and short-term recession, and a component with under-month lag which includes the subsurface and the surface runoff. The system is hence considered as made up of two linear reservoirs in parallel plus a zero-lag linear channel. The input to the system has the meaning of effective precipitation. In this way, from volume balance equations a constant parameters ARMA(2,2) model is obtained. Model residual has pseudo-periodic structure, since it is proportional to the system input. The effectiveness of this model was verified either with respect to the validity of conceptual hypotheses and as regards statistical efficiency.

INTRODUCTION

The identification of stochastic models of streamflows is usually achieved through empirical procedures, by comparing the performances of different models with respect to the reproduction of the stochastic features of the process, such as mean, variance and autocorrelation structure (e.g. *Lawrance and Kottegoda*, 1977; *Salas and Obeysekera*, 1982; *Noakes et al.*, 1985; *Bowles et al.*, 1987). By following this approach, modeling of monthly runoff is commonly performed through AR or ARMA models (*Box and Jenkins*, 1970, *Salas et al.*, 1980). To allow reproduction of the periodicity displayed by mean, variance and sample autocorrelation of the runoff at monthly scale, models which require the estimation of one or more parameters per month are preferred, i.e. PAR or PARMA (see e.g. *Noakes et al.*, 1985; *Jimenez et al.*, 1989). Their use, yet, becomes particularly critical when dealing with inadequate data or ungauged stations, where it is difficult to assess model reliability because of the lack of significance of the goodness-of-fit tests.

An alternative approach to model selection is the one that makes use of a-priori physical information on the phenomenon (e.g. *Klemes*, 1978). This information is represented by the role that different factors, i.e. the climate and the hydrogeology, have on the stochastic features of the runoff process. This approach ensures model parsimony with regards to the number of parameters, and physical interpretability of parameters.

In this paper, the latter approach is pursued, in order to establish an objective criterion for model selection within the framework of linear models. According to this rationale, the characterization of the runoff process is based on the identification of conceptual structures responsible for regularities within the phenomenon. This allows tho formulate a linear conceptual model of the process, which is equivalent to a linear system. With regard to an input process, the system behaves as a linear filter just like a linear stochastic model acts towards the residual process.

CONCEPTUAL BASES OF THE MONTHLY RUNOFF STOCHASTIC MODEL

Time series of runoff relative to watersheds in Central-Southern Italy were considered where the climate is qualified by two distinct seasons within the year: autumn and winter are wet while spring and summer are dry. In these basins, snowmelt runoff can be often neglected. These watersheds are characterized by the geology of Apennines mountains which features the presence of several great fractured carbonate massifs, containing large aquifers at their base.
The observation on a semi-logarithmic scale of the series of daily runoff highlights that discharges decrease in different ways over time during spring and summer. Thus, runoff can be considered as made up of different components: (*a*) a long-term baseflow, deriving from aquifers with over-year recession, such as the ones contained in the fractured carbonate massifs; (*b*) a short term baseflow, coming from aquifers where recession occurs within a few months at the end of the wet season; (*c*) the direct runoff component, which has sub-monthly lag and includes the subsurface and the surface runoff. Direct runoff at monthly time scale is proportional to the effective precipitation, in that it represents the component of rapid response (less than one month) to the net rainfall input.

The two aquifers are considered as linear reservoirs, while the direct runoff component is regarded as the outlet of a linear channel with zero lag. The runoff is thus considered as the outlet of a linear system (Fig. 1). The conceptual input to the system is represented by the effective rainfall, which however is not considered a data in this problem.

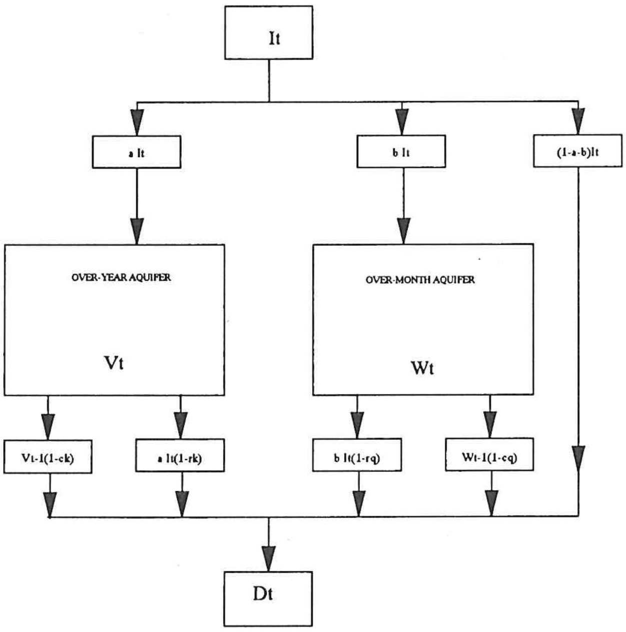

Fig.1. Conceptual model of monthly runoff

Net rainfall I_t is subdivided into: aI_t = recharge to the over-year recession aquifer; bI_t = recharge to the over-month recession aquifer; $(1- a - b)I_t$ = direct runoff. Parameters a and b, which are, in general, variable (they depend, for instance, on the degree of soil saturation), are considered to be constant in this analysis, in order to simplify the representation. The other two conceptual model parameters are the storage constants, k and q, of the two aquifers.

Stochastic Model of Monthly Runoff

Let us consider the volume D_t which outflows from a linear reservoir in a unit time interval t with reference to an inflow of volume R_t. If V_{t-1} indicates the volume stored in the reservoir at the beginning of the interval, D_t is given by the linear storage equation:

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

In this relation, k is the *storage coefficient* and c_k is the *recession coefficient;* c_k equals $e^{-1/k}$. The quantity r_k is called *recharge recession coefficient;* it has the same meaning of c_k but only applies to the recession of the recharge volume. r_k depends on the storage coefficient and on the form of the recharge function.

If the recharge function is an impulse input occurring at the time $T \in [0,1]$ it is $r_k = e^{-(1-T)/k}$ (*Moss and Bryson,* 1974). For uniform input, as considered in this work with reference to the monthly scale, it is $r_k = k(1 - e^{-1/k})$.

With regard to the linear system shown in Fig. 1, the sum of the groundwater and direct runoff components is given by:

$$D_t = V_{t-1}\,(1-c_k) + a\,I_t\,(1-r_k) + W_{t-1}\,(1-c_q) + b\,I_t\,(1-r_q) + (1-a-b)\,I_t \tag{2}$$

Volume balance equations of the two aquifers are:

$$V_t = V_{t-1}\,c_k + a\,I_t\,r_k \tag{3}$$

$$W_t = W_{t-1}\,c_q + b\,I_t\,r_q \tag{4}$$

By substituting in Eq. 2 the expressions of W_{t-1} and V_{t-1} obtained from Eqs. 3 and 4 and rearranging, we obtain one equation in D_t, D_{t-1}, D_{t-2}, I_t, I_{t-1}, I_{t-2}:

$$\begin{aligned} & D_t - [c_k + c_q]D_{t-1} + [c_k c_q]D_{t-2} = \\ & = (1 - a\,r_k - b\,r_q)I_t - [c_k + c_q - a\,r_k(1+c_q) - b\,r_q(1+c_k)]I_{t-1} \;+ \\ & - [a\,r_k c_q + b\,r_q c_k - c_k c_q]\,I_{t-2} \end{aligned} \tag{5}$$

This expression can be identified with an ARMA(2,2) stochastic model

$$d_t - \Phi_1\,d_{t-1} - \Phi_2\,d_{t-2} = \varepsilon_t - \Theta_1\,\varepsilon_{t-1} - \Theta_2\,\varepsilon_{t-2} \tag{6}$$

where d_t equals $D_t - E[D_t]$ while ε_t is the model residual. The structure of the residual is not that of a white noise, as is to be expected given its hydrological meaning.

The formal correspondence between the stochastic and conceptual representations is obtained through the relations:

$$\Phi_1 = c_k + c_q\,; \qquad \Phi_2 = -c_k c_q \tag{7}$$

$$\Theta_1 = \frac{c_k + c_q - a\,r_k(1+c_q) - b\,r_q(1+c_k)}{1 - a\,r_k - b\,r_q} \tag{8}$$

$$\Theta_2 = \frac{a\, r_k c_q + b\, r_q c_k - c_k c_q}{1 - a\, r_k - b\, r_q} \qquad (9)$$

and imposing proportionality between the residual ε_t and the zero mean net rainfall i_t:

$$\varepsilon_t = (1 - a\, r_k - b\, r_q) i_t \qquad (10)$$

Through the parameters of the stochastic model expressed in Eq. 6 it is possible to estimate the parameters and the input process of the conceptual model, by means of

$$c_k = \frac{\Phi_1 + \sqrt{(\Phi_1^2 + 4\Phi_2)}}{2} \quad ; \quad c_q = \frac{\Phi_1 - \sqrt{(\Phi_1^2 + 4\Phi_2)}}{2} \qquad (11)$$

$$a = \frac{M - N - M\, b\, r_q}{M\, r_k} \qquad (12)$$

$$b = \frac{-(\Theta_1 - \Theta_2)N + (\Phi_1 - \Phi_2)M + (1 + 2\, c_q)(N-M)}{2\, M\, (c_k - c_q) r_q} \qquad (13)$$

where: $N = (1 - \Phi_1 - \Phi_2)$ $\qquad M = (1 - \Theta_1 - \Theta_2)$

The residual ε_t actually represents an estimate of the conceptual variable i_t. This proportionality is affected by an error, which takes into account that the conceptual model is only an approximation of reality. Consequently, Eq. 6 should be reconsidered, substituting ε_t with

$$\varepsilon'_t = \varepsilon_t + \xi_t \qquad (14)$$

where the quantity ε'_t is the actual residual of the stochastic model, ε_t represents the component of ε'_t which has conceptual meaning and ξ_t is the error component, considered with zero mean and constant variance. ARMA representation thus becomes:

$$d_t - \Phi_1\, d_{t-1} - \Phi_2\, d_{t-2} = \varepsilon'_t - \Theta_1\, \varepsilon'_{t-1} - \Theta_2\, \varepsilon'_{t-2} \qquad (15)$$

In practical terms, until the characteristics of the error ξ are specified, reference is made to the net rainfall estimated with error:

$$I'_t = i'_t + \mu = \frac{\varepsilon'_t}{1 - a\, r_k - b\, r_q} + \mu \qquad (16)$$

where μ equals $E[D_t]$ since D_t and I_t are equivalent in average (*Claps*, 1990).

Monthly net rainfall process has complex links with climate, which induces periodicity, and with geomorphologic and soil characteristics. This allow to define the net rainfall as a *pseudo-periodic* process and, consequently, the monthly runoff stochastic model is an ARMA(2,2) model with *pseudo-periodic* residual (PPR-ARMA(2,2)).

Estimation of Parameters of the PPR-ARMA(2,2) Model

From the above discussion emerges that the model identified for monthly runoff is a constant parameters PPR-ARMA model, in which no deseasonalization procedure is applied. Moreover, in order to preserve the formal correspondence between the conceptual and stochastic representations of the process, no data transformation is made. The reproduction of series skewness is referred to the probabilistic analysis of the residual.

Least-Squares estimates of the parameters of this model have finite variance and are asymptotically normal [*Pierce* 1971]. However, for reason related to the constraints imposed to the stochastic parameters by their conceptual meaning (*Claps*, 1989; *Claps*, 1990) a non-standard estimation procedure was utilized, based on the use of information obtained from two different time scales of aggregation. In fact, only the over-month aquifer parameters are estimated on the monthly time scale, while the over-year recession aquifer parameters are estimated on the annual scale (*Claps*, 1989, *Claps*, 1990).

PROBABILISTIC MODEL OF THE RESIDUAL

The residual ε'_t is made up of a conceptual component ε_t, which is proportional to the net rainfall I_t through Eq. 10, and an error component ξ_t, which represents a stationary and continuous variable, with zero mean and constant variance $\sigma_\xi^2 = \sigma_0^2$. The two components may be considered as uncorrelated and the following hold:

$$E[\varepsilon'_t] = E[\varepsilon_t]; \qquad \sigma^2[\varepsilon'_t] = \sigma^2[\varepsilon_t] + \sigma_0^2 \tag{17}$$

For convenience's sake, reference is made to the net rainfall variable

$$I_t = \frac{\varepsilon_t}{(1 - a\, r_k - b\, r_q)} + \mu$$

Because of its very meaning, the variable I_t should assume only positive values and present finite probability at the value zero. A satisfactory probabilistic representation of I_t can be obtained by considering the sum of a poissonian number of events, with parameter ν, whose intensity is distributed exponentially with parameter λ. The correspondent probability density function has the expression (*Johnson and Kotz*, 1971):

$$P[x=0] = e^{-\nu} \qquad \text{(finite probability for x=0)} \tag{18}$$

$$f_X(x) = e^{-\lambda x-\nu} \sqrt{\nu\lambda/x}\; \mathcal{I}_1(2\sqrt{\lambda x \nu}) \qquad \text{for } x > 0 \tag{19}$$

where $\mathcal{I}_1(x)$ is the modified Bessel function of order 1. The function introduced in Eqs. 16 and 19 will be referred to as the *Bessel distribution*. The relations between the sample moment, and the distribution parameters are ($\beta = 1/\lambda$):

$$\mu_x = \nu\beta; \qquad \sigma_x^2 = 2\nu\beta^2 \tag{20}$$

The continuous part of the Bessel distribution can be approximated to a Normal distribution and, in particular, to a square root Normal distribution (*Johnson and Kotz*, 1971). By considering this approximation, if σ_y^2 is the (constant) variance of the variable $y=\sqrt{x}$, through first order analysis we get

$$\sigma_y^2 \sim \frac{1}{4} \sigma_x^2 \mu_x^{-1} = \frac{\beta}{2} \qquad (21)$$

In this way it is possible to establish a correspondence between the first two moments of I_t:

$$\sigma^2[I_t] = 2\beta\mu[I_t] \qquad (22)$$

From Eq. 16, the second of the relations in Eq. 17 and Eq. 22, we get

$$\sigma^2[I'_t] = \frac{\sigma_0^2}{c^2} + 2\beta\mu = m + 2\beta\mu \qquad (23)$$

In Eq. 23 c equals $(1 - a\, r_k - b\, r_q)$, m is proportional to the error variance and β has the meaning of a climatic parameter which is function of the average intensity of events. Both m and β are assumed to be constant, i.e. season-independent. By means of Eq. 23 σ_0^2 can be estimated through a linear regression.

The linearity hypothesis described in Eq. 23 can be verified by analyzing the net rainfall series deriving from the model application. This constitutes also an indirect check on the assumed hypothesis on the distribution of the residual ε'_t, considered the sum of a Normal and a Bessel random variable.

APPLICATION TO RIVERFLOW TIME SERIES IN CENTRAL-SOUTHERN ITALY

An application of the proposed procedure was carried out on three time series of monthly runoff relative to gauging stations on rivers in Central-Southern Italy (Table 1). Table 2 shows the estimates obtained for the stochastic and conceptual parameters.

Table 1. Characteristics of the stations and the time series considered

Station	Area (Km^2)	obs. (years)	Mean Annual Rainfall (mm)
1 Giovenco at Pescina	139	11	277
2 Nera at Torre Orsina	1445	25	606
3 Tiber at Rome	16545	50	448

Table 2. Estimates of the PPR-ARMA(2,2) model monthly runoff parameters.

Station	a	b	k (months)	q (months)	Φ_1	Φ_2	Θ_1	Θ_2
1	0.61	0.277	35.3	2.23	1.61	-.621	1.35	-.387
2	0.71	0.256	50.4	3.56	1.74	-.741	1.19	-.228
3	0.53	0.363	40.6	1.66	1.52	-.535	1.16	-.202

By using the net input estimated by means of Eq. 16, the reconstructed over-month and over-year groundwater runoff can be examined, partly so as to validate the conceptual parameters estimates. An example of the reconstruction of groundwater runoff is shown in Fig. 2, in which it is to note the goodness of the reconstruction of the minima of the observed series.

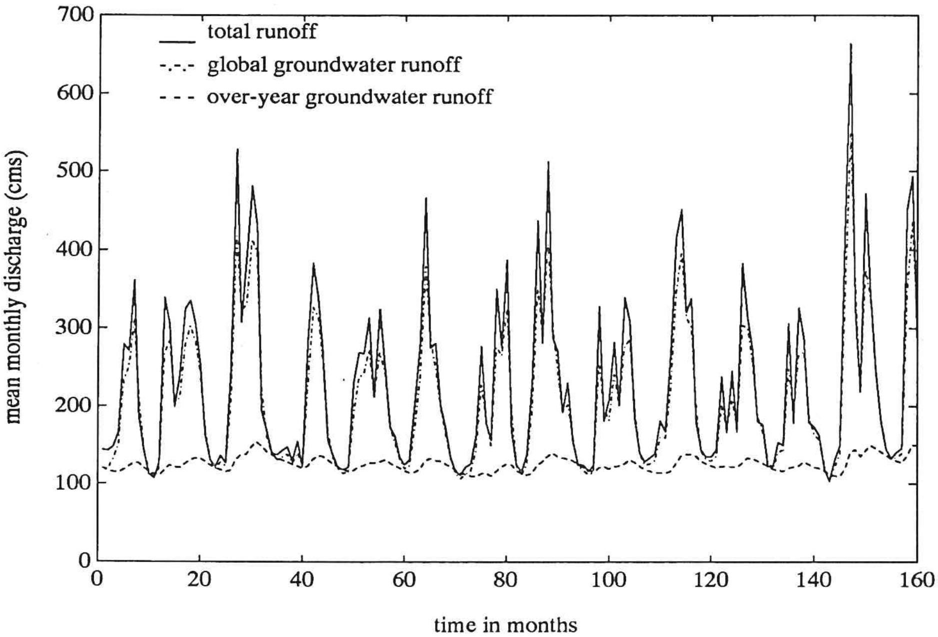

Fig. 2. Reconstruction of the groundwater streamflow: River Tiber at Rome, years 1921 - 1935.

The regression model of the estimated net rainfall variance with respect to the mean was then applied, according to Eq. 23, with results which can be considered as confirming the assumption of linear dependence, since the coefficients σ_0^2 / c^2 were found to vary from 470 to 3000 mm^2 with R^2 varying between 0.43 and 0.86. The hypothesis that the two characteristic parameters, σ_0^2 / c^2 and β , are constant from month to month allows to reduce the error associated with the sample estimate, even if this means paying a price in not taking into account parameter heterogeneity within the year. Considering the shortness of the single station data set, which is made up of 12 data, the regression analysis could also be made on regional basis. In this case the hypothesis of parameter homogeneity within a region could be checked.

In order to test model's statistical efficiency, a comparison was made with the PAR(1) model, which was fitted to the logarithm of the dimensionless runoff data, as suggested by *Noakes et al.* (1985). The comparison of the statistical performances of the two models was based on the R^2, without taking the probabilistic analysis of the residual into account.

The PPR-ARMA model shows higher R^2 than the PAR(1) (Table 3). Since the PPR-ARMA requires the estimation of 12 mean monthly values, of the error and the residual variance and of 4 stochastic parameters while the PAR(1) requires the estimation of 12 means, 12 variances and 12 autoregressive parameters, even better results would arise by considering the *adjusted* R^2, whose value decreases when the number of parameters to estimate increases.

Table 3. R^2 values for PAR(1) and PPR-ARMA(2,2) models.

Station	1	2	3
R^2 *PAR(1)*	.417	.682	.363
R^2 *PPR-ARMA(2,2)*	.551	.745	.496

CONCLUSIONS

The advantages of following a conceptually-based approach in runoff modelling are as follows: (*a*) the use of a-priori informations provides an objective criterion for model identification and tends to determine parsimony in the number of parameters; (*b*) model parameters are physically interpretable. In such a way, parameters can be validated even in situations of limited data; (*c*) it is made possible, in principle, to evaluate model parameters even in ungauged stations.

In this paper, with regard to the monthly runoff process, an ARMA(2,2) stochastic model with pseudo-periodic residual (PPR-ARMA) is identified by considering the runoff as the sum of different processes, each one with a characteristic time scale. A priori informations, concerning the characteristic time scale of each sub-process, are introduced in the statistical criteria used in the estimation. Parameters accounting for long-term persistence are actually estimated on the annual time scale while only parameters connected with short-term persistence are estimated on the monthly scale.

The connection established between model residual and net rainfall allows to estimate the latter process, which is modeled by means of a mixed probability distribution depending only on the monthly means and on a climatic parameter which can be considered constant within the year.

REFERENCES

Bowles,D.S.,W.R.James and N.T. Kottegoda (1987) "Initial model choice: An operational comparison of stochastic streamflow models for drought", *Water Resour. Management*,**1**,3-15.

Box, G.E. and G. Jenkins (1970) *Time Series Analysis, Forecasting and Control*, Revised Edition, Holden-Day, San Francisco, (Reprint 1976).

Claps, P. (1990) "Modelli stocastici dei deflussi dei corsi d'acqua" (in italian), *Ph.D. Dissertation*, Università degli Studi di Napoli, Italy, 279 pp.

Claps, P. (1989) "Conceptual Basis of Stochastic Models of Monthly Streamflows", poster paper, *NATO A.S.I. "Stochastic Hydrology and its use in Water Resources Systems"*, Peñiscola, Spain (in publ.)

Jimenez C., A.I. McLeod and K.W Hipel (1989) "Kalman filter estimation for periodic autoregressive-moving average models", *Stochastic Hydrol. and Hydraul.*, **3**, 227-240.

Johnson, N.L. and S. Kotz (1971), *Continuos Univariate Distributions*, John Wiley & Sons, New York.

Klemes,V. (1978) "Physically based stochastic hydrologic analysis", *Advan. in Hydroscience*,**11**,285-352

Lawrance, A.J. and N.T. Kottegoda (1977) "Stochastic modelling of riverflow time series", *J.R. Statist. Soc.,A*, **140**, 1-47.

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

Noakes, D.J., A.I. McLeod and K.W. Hipel (1985) "Forecasting monthly riverflow time series", *Int. J. Forecast.*, **1**, 179-190.

Pierce, D.A. (1971), Least squares estimation of the regression model with autoregressive-moving average errors, *Biometrika*, vol. 58, pp. 299-312.

Rao,A., R.L. Kashyap and L.Mao (1982) "Optimal choice of type and order of river flow time series models", *Water Resour. Res.*, **18**(4),1097-1109.

Salas, J.D., J.W. Delleur, V. Yevjevich and W.L. Lane (1980) *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado.

Salas, J.D. and J.T.B. Obeysekera (1982) "ARMA models identification of hydrologic time series", *Water Resour. Res.*, **18**(4), 1011-1021.