OPTIMAL PARAMETER ESTIMATION OF CONCEPTUALLY-BASED STREAMFLOW MODELS BY TIME SERIES AGGREGATION

P. CLAPS[1] and F. MURRONE[2]
[1]Dept of Environm. Engineering and Physics, University of Basilicata
Via della Tecnica, 3 Potenza 85100 - Italy
[2]Dept Hydraul., Wat. Resour Manag. and Environm. Eng. Univ. of Naples "Federico II"
Via Claudio, 21 Napoli 80126 - Italy

In the framework of an integrated use, among different scales, of conceptually-based stochastic models of streamflows, some points related to efficient parameter estimation are discussed in this paper. Two classes of conceptual-stochastic models, ARMA and Shot-noise, are taken under consideration as equivalent to a conceptual system transforming the effective rainfall into runoff. Using these models, the possible benefits of data aggregation with regards to parameter estimates are investigated by means of a simulation study. The application made with reference to the ARMA(1,1) model shows advantageous effects of data aggregation, while the same benefits are not found for estimation of the conceptual parameters with corresponding Shot Noise model.

INTRODUCTION

Streamflow time series modelling if generally intended as the closest possible reproduction of the statistical features displayed by observed measures of the phenomenon under investigation. This is certainly what is needed in the majority of the practical cases for which time series are analyzed, for instance planning and management of water resources systems. Practical needs have lead, in the last decades, to a prevailing "operational" approach to time series modelling, in which little space has been left to the analysis of physical, observable aspects in riverflow series. On the other hand, a physically based approach to this problem addresses the reproduction as well as the interpretation of the features of the phenomenon.

One of the requirements of a correct reproduction of the phenomenon is that the model related to a given scale must be conceptually compatible with the models reffered to smaller or aggregated scales. Even out of a conceptual approach, the problem of determining stochastic models for aggregated data has received so far little attention. Among the few papers in this field Kavvas et al. (1977) Vecchia et al. (1983) and Obeysekera and Salas (1986) are worth mentioning.

With regard to the above requirement, using conceptually based models allows the basic advantages that the information related to a conceptual parameter can be transferred from a larger to a smaller scale, because its conceptual meaning does not depend on a particular time scale. Therefore, derivation of stochastic models from a general conceptual representation of the runoff process is a first step towards integration of models among different scales.

Claps and Rossi (1992) and Murrone et al. (1992) identified stochastic models of streamflow series over different aggregation scales starting from a conceptual interpretation of the runoff process. In this conceptual-stochastic framework there are conceptual parameters common to models related to different scales. The point of view characterizing this framework, which is summarized in the next section, is that the analysis of streamflow series should be extended beyond the scale at which data are collected, taking advantage of information available from models of the aggregated data.

The question arise if there is a particular time scale (and, consequently, a particular model) leading to an optional estimation of a given parameter. The choice of an optimal time scale is important because aggregation of data tends to reduce correlation effects due to runoff components with small lag time with respect to the effect produced by component with high lag time. At the same time aggregation reduces the number of data and, consequently, the quality of estimates.

In the above approach, Claps and Rossi (1992) and Murrone et al. (1992) considered a limited number of time scales, such as annual, monthly, and T-day (with T ranging from 1 to 7) and showed that conceptual parameters of models of monthly and T-day runoff are more efficiently estimated using different scales of aggregation.

An attempt to introduce a more systematic procedure in the selection of the optimal time scale for the estimation of each parameter is made in this paper. In this direction, simulation experiments are performed with regards to ARMA (Box and Jenkins, 1970) and Shot Noise (Bernier, 1970) stochastic models equivalent to a simple conceptual model of the runoff process.

CONCEPTUAL-STHOCASTIC MODELS AND TIME SCALES

The rationale of conceptualization

In the approach by Claps and Rossi (1992) and Murrone et al. (1992), formulation of a conceptual model for river runoff is founded on the "observation" of river flow series over different aggregation scales and on the knowledge of the main physical (climatic and geologic) features of basins.

Considering Central-Southern Italy watersheds, dominated by the hydrogeological features of Apennine mountains, distinct components can be recognized in the runoff: (1). the contribution provided by aquifers located within large carbonate massifs, that has over-year response time to the recharge (deep groundwater runoff); (2). a component, which is due to both overflow springs and aquifers within geological non-carbonate formations, which usually run dry by the end of the dry season (seasonal groundwater runoff); (3). the contribution by soil drainage, having a delay of several days with respect to precipitation (subsurface runoff); (4). the surface runoff, having lag-time that depends on the size of the watershed (for the rivers analyzed by Murrone et al. (1992), this lag ranges between a few hours to almost two days). In some cases, the deep groundwater component is lacking, reducing runoff components to three. The snowmelt runoff in the region considered is negligible. The above runoff components assume different importance with respect to the time scale of aggregation, leading to conceptual models of increasing complexity moving from the annual to the daily scale.

Bases for conceptual-stochastic model building proposed for the monthly scale (Claps and Rossi, 1992, Claps et al, 1993) and for the daily scale (Murrone et al., 1992) are essentially: (1) subsurface and. groundwater systems are considered as linear reservoir, with storage coefficients K_1, K_2, K_3, going from the smallest to the largest; (2). Runoff is the output of a conceptual system made up of the above reservoirs in parallel with a zero-lag linear channel reproducing the direct runoff component; (3) when a storage coefficient is small with respect to the time scale considered, the related groundwater component becomes part of direct runoff term, which is proportional to the system input; (4) the effective rainfall, i.e. total precipitation minus evapotranspiration, is the conceptual input to the system; this variable is not explicitly accounted into subsurface and groundwater systems at constant rates (recharge coefficients c_1, c_2, c_3, respectively) over time.

The main issues of model identification for annual, monthly and daily scales are summarized below.

Annual scale

Rossi and Silvagni (1980) first supported on conceptual basis the use of the ARMA(1,1) model for annual runoff series, based on the consideration that the correlation structure at that scale is determined by the deep groundwater runoff component. The use of this model for annual runoff modelling was proposed by O' Connell (1971) in virtue of its capacity of reproducing the long-term persistence displayed by annual runoff data. Salas and Smith (1981) showed how a conceptual system composed by a linear reservoir in parallel with a linear channel fed by a white noise input behaves as an ARMA(1,1) process.

Given an effective rainfall input I_t which infiltrates in the rate $c_3 I_t$ and whose part $(1-c_3)I_t$ goes in direct runoff, based on the hypothesis that the input is concentrated at the beginning of the interval [t–1, t]. the volume balance equations produce

$$D_t - e^{-1/K_3} D_{t-1} = (1 - c_3\, e^{-1/K_3})I_t - e^{-1/K_3}(1 - c_3)I_{t-1} \qquad (1)$$

where D_t is runoff in year *t*. This hypothesis can be removed considering different shapes of the within-year input function (Claps and Murrone, 1993). The hypothesis that I_t is a white noise process conducts to an ARMA(1,1) model

$$d_t - \Phi\, d_{t-1} = \varepsilon_t - \Theta\, \varepsilon_{t-1} \qquad (2)$$

in which d_t equals $D_t - E[D_t]$, Φ and Θ are the autoregressive and moving average coefficients, respectively, and ε_t is the zero-mean model residual. Conceptual and stochastic parameters in (2) and (3) are related by:

$$\Phi = e^{-1/K_3}; \qquad \Phi = \frac{\Phi(1-c_3)}{(1-c_3\Phi)}$$

$$K_3 = -1/\ln(\Phi); \qquad c_3 = \frac{\Phi-\Theta}{1-\Theta} \tag{3}$$

The expression of c_3 for uniform within-period distribution of input, is

$$c_3 = \frac{\Phi-\Theta}{(1-\Theta)K_3(1-e^{-1/K_3})} \tag{4}$$

The ARMA model residual is proportional to the effective rainfall by means of:

$$\varepsilon_t = (1-c_3\Phi)\{I_t - E[I_t]\} = (1-c_3\Phi)\{I_t - E[D_t]\}. \tag{5}$$

In absence of significant groundwater runoff Rossi and Silvagni (1980) showed that annual runoff in the hydrologic year is an independent process that follows a Box-Cox transformation of the Normal distribution. The notion of hydrologic year, which starts at the end of the dry season, is important because if a wet season and a dry season can be distinguished, the absence of significant runoff in the dry season determines absence of correlation in the hydrologic year runoff series.

Monthly scale

The assumptions recalled above on the role of the different components in streamflows lead to consideration that correlation effects in monthly runoff are due both to long-term persistence, due to the deep groundwater runoff, and to short-term persistence due to the seasonal groundwater runoff. The conceptual system identified by means of these considerations consists of two parallel linear reservoirs plus a zero-lag linear channel.
This latter account for the sub-monthly response components included into the direct runoff.

The share c_3I_t of the effective rainfall is the recharge of the over-year groundwater, with storage coefficient K_3, while c_2I_t is the recharge of the seasonal groundwater, with storage coefficient K_3. All c_j and K_j parameters are kept constant. Approximations determined by the latter assumption are compensated by parsimony in the number of parameters and by the greater significance given to the characteristics of the input I_t, considered as a periodic-independent process. Periodic variability of the recharge coefficients c_2 and c_3 is substantially due to the variability in soil moisture, which is a product of rainfall periodic variability.

Claps and Rossi (1992) and Claps et al. (1993) have shown that volume balance equations for the conceptual model under exam are equivalent to an ARMA(2,2) stochastic process with periodic-independent residual (PIR-ARMA), expressed as

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

with d_t and ε_t having mean zero. The formal correspondence between the stochastic and conceptual representations is obtained through the relations:

$$e^{-1/K_3} = \frac{\Phi_1 + \sqrt{\Phi_1^2 + 4\Phi_2}}{2} \tag{7}$$

$$e^{-1/K_3} = \frac{\Phi_1 - \sqrt{\Phi_1^2 + 4\Phi_2}}{2} \tag{8}$$

$$c_3 = \frac{M - N - Mc_2 r_2}{Mr_3} \tag{9}$$

$$c_3 = \frac{-(\Theta_1 - \Theta_2)N + (\Phi_1 - \Phi_2)M + (1 + 2e^{-1/K_2})(N - M)}{2M(e^{-1/K_3} - e^{-1/K_2})r_2} \tag{10}$$

where $N = (1 - \Phi_1 - \Phi_2)$, $M = (1 - \Theta_1 - \Theta_2)$, and $r_3 = k_3(1 - e^{-1/K_3})$ and $r_2 = k_2(1 - e^{-1/K_2})$. In addition, in the conceptual scheme proportionality is required between the residual ε_t and the zero-mean net rainfall i_t according to the relation

$$\varepsilon_t = (1 - a\, r_3 - b\, r_2) i_t \tag{11}$$

If the over-year groundwater component is negligible, for instance in practically impermeable basins, the conceptual system reduces to one reservoir in parallel with a linear channel, underlying a PIR-ARMA(1,1) stochastic process.

Probability distribution of monthly effective rainfall is assumed by Claps and Rossi (1992) as the sum of a Bessel distribution (Benjamin and Cornell, 1970, p. 310), arising from the sum of a Poissonian number of exponentially distributed events, and a Gaussian error term. A Box-Cox transformation of non zero data was also proposed by Claps (1992).

To preserve the formal correspondence between the conceptual and stochastic representations of the process, neither deseasonalization nor transformation procedures are applied on recorded data.

T-day scale: multiple Shot Noise model

The Shot Noise (Bernier, 1970) is a continuous-time stochastic process representing a phenomenon whose value, at a certain time, is determined additively by the effects of a random number of previous events. This process is determined by knowledge of: (1) the occurrence times of events, τ_i; (2) the input impulses intensity related to the events, Y_i; and (3) the response function of the system, $h(\cdot)$, describing the propagation in time of the effects of each impulse.

The hypotheses made for this kind of process are: (a). the $h(\cdot)$ function is continuous, infinitesimal for t tending to infinity and integrable; (b) intensities Y_i are random variables independent and identically distributed, with finite variance; and (c) event occurrence times τ_i are generated by homogeneous Poisson process. The process is stationary if its origin tends to –∞, meaning that the origin must be far enough by the time under consideration. Runoff D can be thus expressed, in continuous time, as

$$D(\tau) = \sum_{N(-\infty)}^{N(\tau)} Y_i h(\tau - \tau_i) \qquad (12)$$

where N(τ) is the realization of the Poisson process of occurrences.

In the conceptual framework considered (Murrone et al., 1992) the response function h(·) is a linear combination of responses of the conceptual elements. If the surface network is considered to behave as a linear reservoir, h(·) is expressed as

$$h(s) = c_0\, 1/K_0\, e^{-1/K_0 s} + c_1 1/K_1 e^{-1/K_1 s} + c_2 1/K_2\, e^{-1/K_2 s} + c_3 1/K_3 e^{-1/K_3 s} \qquad (13)$$

with $s = \tau - \tau_i$. The basin response is defined by 8 parameters: the four storage coefficients, K_j, and the four recharge coefficients, c_j, of which only 7 are to be estimated given the volume continuity condition, $\Sigma c_i = 1$. The c_j coefficients represent the share of runoff produced, in average, by each component. To limit the number of parameters and to take advantage by the linearity hypotheses, coefficients c_i and K_j are considered constant, i.e. the response function h(·) is kept constant.

The process (12) has infinite memory, which represents the current effect of previous inputs to the system. This effect can be evaluated at a fixed initial time $t_0 = 0$ by knowing the groundwater runoff quota at that time. At the beginning of the hydrological year (October 1 in our case), the seasonal and subsurface groundwater contributions are negligible relative to the deep groundwater runoff. Therefore the value D_0 of discharge at that time can be a good preliminary estimate of the groundwater runoff amount, thus expressing (12) as:

$$D(\tau) = D_0 e^{-\tau/K_3} + \sum_{N(0)}^{N(\tau)} Y_i h(\tau - \tau_i) \qquad (14)$$

The discretized form of the continuous process (14) is obtained by its integration over the interval [(t-1)T, tT], where t = 1, 2, … is the index describing the set of sampling instants and T is the sampling time interval. If the aggregation occurs on a T-day scale and integration is applied according to the linearity and stationary hypotheses, the following discretized formulation is obtained:

$$D(\tau) = K_3 e^{-t\tau/K_3}(e^{T/K_3} - 1)X_0 + \sum_{s=1}^{t} Y^*_{t-s+1} h_s \qquad (15)$$

where Y_t^* represents the sum of impulses occurred during the interval [(t-1)T, tT] and the integrated response is expressed as:

$$h_1 = \frac{1}{T}\sum_{i=0}^{3} c_i \left[K_i (e^{T/K_i} - 1) + T\right]$$
$$h_s = \sum_{i=0}^{3} c_i \frac{K_i}{T}\left[(e^{T/K_i} + e^{-T/K_i} - 2)\right] e^{-T(s-1)/K_i}, s > 1 \tag{16}$$

The function h_s represents the response of the system determined by a unit volume impulse of effective rainfall, uniformly distributed within the interval.

When a scale of aggregation T is chosen as considerably larger than the surface runoff lag-time, the surface runoff component can be considered as the output of a zero-lag linear channel, which has response function $c_0\ \delta(0)$, with $\delta(\cdot)$ as the Dirac delta function. This reduces to six the number of parameters to be estimated.

The structure of daily precipitation has been represented as uncorrelated, like in Poisson white noise models (Bernier, 1970) or characterized by Markovian arrival process (see *e.g.* Kron et al., 1990) or described with models based on the arrival of clusters of cells, such as the Neyman-Scott instantaneous pulse model (*e.g.* Cowpertwait and O'Connell, 1992). The distribution considered in the simulation for daily data is a Bessel distribution, corresponding to a Poisson white noise probabilistic model.

SIMULATION STUDY

Prerequisites to the simulation

For the reasons expounded in the introduction, the simulation study undertaken here aims preliminary to set a number of basic points in evaluating theoretically the effects of aggregation on parameter estimation. The problem here is not to identify the most correct model (as, for instance, in Jakeman and Hornenberg, 1993) but to understand if there are peculiar scales for estimation of parameters of a given model with pre-determined structure, as in Claps et al. 1993. Simple hypotheses in terms of input and system structure were adopted for the simulation, to grasp the basics of the positive or negative effects of aggregation in time.

A linear system was considered, which consisted of one linear reservoir, with storage coefficient K, and one linear channel, in parallel, with lag-zero. As shown with reference to annual runoff, this system, fed by stochastic input, is equivalent to an ARMA(1,1) model when the input is a continuous process. For input as a point process this system is equivalent to a single Shot Noise model (as compared to the multiple version arising from the presence of more than one reservoir).

For each set of “true” parameters **c** and **K** (written in bold) of the linear system, 20000 output data were generated. On the data obtained from the Gaussian input, parameters of the ARMA(1, 1) model were estimated and expressed, through (3), in terms of conceptual parameter estimates $\hat{c}$ and $\hat{K}$. Shot Noise model parameters were estimated on data generated from Bessel input. This first 10000 synthetic runoff data were not considered in the estimation, as warm-up length. (Salas et al., 1980, p.356) This length was set well beyond the suggested limits, to definitely eliminate possible "starting condition" effect.

The recharge coefficient **c**, indicating the amount of input entering the reservoir, ranged from 0.5 to 1. In the model of annual runoff **c** is less than 1 while the case **c** = 1corresponds to the model of a spring (see Claps and Murrone, 1993). The storage coefficient **K** was set in a range from 2 to 120 time units (t.u.). The 'time unit' is one unit of the time scale at which input and output data are generated and is also called the days because what is important is to indicate the value of the parameters in terms of a multiple of the scale of generation. Accordingly, time scales at different levels of aggregation are identified in number of time units.

In a preliminary set of simulations, the effect of the input standard deviation σ was recognized as null for Gaussian data and practically negligible for Bessel data. For this reason, only one level of input variability was considered for each distribution, namely $\sigma = 1/3$ for Gaussian input and $\sigma = 3$ for Bessel input. For both case the mean was set to 1.

To allow comparison of parameter estimates made on data obtained with the same "true" values but in different conditions, standard errors of parameters and the explained variance R^2 were used. R^2 is defined as $1-\sigma^2_\varepsilon/\sigma^2$, where σ^2_ε indicates the residual variance (taken as the variance of the surface runoff component in the Shot Noise model) and σ^2 indicates the variance of the synthetic runoff series.

Application

Main points to focus with the aid of simulations are: (1) In which manner the resolution of a linear reservoir depends on the relative mean (coefficient **c**) of its output with respect to total runoff? (2) Is there a prefential scale for the estimation of the storage coefficient? Results of parameter estimation simulated data, reported below, suggest a number of comments.

<u>ARMA(1,1) model</u>

The following comments arise from estimation of $\hat{c}$ and $\hat{K}$ through the ARMA(1,1) model:

1. Results of parameter estimation, reported in Table 1a to 1c (referred to **c** = 0.5, **c** = 0.8 and **c** = 1, respectively), show that aggregation reduces the variance of the fraction of input not entering the reservoir, producing higher values of the explained variance R^2. The obvious exception is the case **c**=1, in which there is no pure white noise component. For the case **c**=1, the model fitted to the data is still the ARMA(1,1), for it is the most general model of a single linear reservoir with generic within-period form of the input function (Claps and Murrone, 1993). This adds information in providing an estimate of **c**, obtained through (4).

2. A progressive increase in the standard error of the estimates also occurs with the aggregation, due uniquely to the decrease in the number of data. Table 2 shows that estimations made on the reference scale (1 t.u.) over limited samples produce standard

error greater than the corresponding standard errors for data aggregated on 7, 15 and 30 t.u.

3. For **c**<1, **K** is clearly underestimated. This tendency becomes more noticeable with increasing **K** and with decreasing **c**. Values found for R^2, which decreases in the same circumstances, reflect the poor estimate of **K**. A tendency toward a preferential scale for parameter estimation is not recognizable from the results in Table1a and 1b.

4. More understandable results are obtained by estimating $\hat{c}$ and $\hat{K}$ on scales aggregated in unit steps, from 1 to 15 t.u. In this regard, Figures 1-3, clearly shows that when **K** is much greater than the reference scale, aggregation produces better conditions for parameter estimation. The progressive increase in $\hat{c}$ and $\hat{K}$ up to a sill (Figure 1), give sufficient indications on these benefits. Therefore the preferential scale must be the one in correspondence of which the sill is reached (7 t.u. for this case), as a trade-off between the increase in R^2 and the decrease in the standard error of estimates.

Based on the results reported above, it seems that the preferential scale decreases with increasing **c** and with decreasing **K** (in general one should speak in terms of nondimentional preferential scale, i.e. divided by **K**). Figure 2, with **c**=0.5 and **K**=15 confirm this tendency showing a substantial constancy both in $\hat{c}$ and $\hat{K}$, that would indicate that the sill is reached at the reference scale. On the other hand, when **c**=1 the reference scale is the best one for estimation regardless of **K**, since quality of estimates degrades with aggregation (see the decrease of $\hat{c}$ and R^2 in Table 1c and the decrease of $\hat{c}$ in Figure 3).

TABLE 1a. Estimations from: ARMA(1,1) model, Gaussian input, **c** = 0.5 (scale in t.u.)

scale	K(t.u.)	c	Φ	Θ	R^2	K(t.u.)	c	Φ	Θ	R^2
		K=120					**K**=90			
1	10.49	0.227	0.909	0.884	0.002	17.01	0.322	0.943	0.917	0.004
7	104.5	0.490	0.935	0.877	0.029	89.53	0.508	0.925	0.853	0.037
15	96.18	0.508	0.856	0.727	0.063	81.30	0.524	0.832	0.677	0.078
30	83.31	0.528	0.698	0.457	0.100	72.00	0.548	0.659	0.383	0.117
60	84.12	0.538	0.490	0.171	0.111	73.20	0.553	0.441	0.102	0.117
		K=60					**K**=30			
1	29.08	0.461	0.966	0.938	0.008	22.74	0.509	0.957	0.914	0.018
7	67.72	0.522	0.902	0.805	0.051	34.42	0.513	0.816	0.657	0.074
15	61.82	0.537	0.785	0.588	0.098	36.08	0.537	0.660	0.393	0.117
30	57.12	0.565	0.592	0.272	0.135	36.30	0.559	0.438	0.092	0.128
60	57.82	0.557	0.354	0.012	0.111	36.67	0.523	0.195	-0.084	0.069
		K=15					**K**=7			
1	13.14	0.515	0.927	0.855	0.033	7.22	0.529	0.871	0.744	0.059
7	14.78	0.492	0.623	0.380	0.090	6.33	0.491	0.331	0.048	0.083
15	18.63	0.505	0.447	0.154	0.098	7.63	0.471	0.140	-0.083	0.046
30	25.99	0.483	0.315	0.041	0.078					

TABLE 1b. Estimations from: ARMA(1,1) model, Gaussian input, **c** = 0.8 (scale in t.u.)

scale	K(t.u.)	c	Φ	Θ	R^2	K(t.u.)	c	Φ	Θ	R^2
		K=120					**K**=90			
1	83.24	0.784	0.988	0.946	0.067	71.23	0.795	0.986	0.934	0.089
7	119.9	0.799	0.943	0.746	0.273	95.48	0.806	0.929	0.683	0.317
15	120.4	0.811	0.883	0.507	0.398	96.02	0.814	0.855	0.413	0.430
30	109.8	0.816	0.761	0.164	0.456	90.78	0.820	0.719	0.068	0.462
60	109.6	0.791	0.578	-0.079	0.386	87.66	0.781	0.504	-0.141	0.350
		K=60					**K**=30			
1	55.27	0.809	0.982	0.910	0.126	29.45	0.811	0.967	0.835	0.205
7	64.34	0.806	0.897	0.565	0.369	30.44	0.796	0.795	0.289	0.414
15	67.23	0.812	0.800	0.266	0.448	33.87	0.787	0.642	0.017	0.399
30	67.80	0.812	0.643	-0.039	0.436	40.41	0.747	0.476	-0.109	0.303
60	62.46	0.750	0.383	-0.192	0.273	36.44	0.654	0.193	-0.188	0.126
		K=15					**K**=7			
1	15.57	0.813	0.938	0.707	0.300	7.59	0.814	0.877	0.481	0.398
7	14.36	0.780	0.614	-0.008	0.384	6.59	0.743	0.345	-0.207	0.255
15	15.57	0.733	0.381	-0.168	0.255	5.51	0.762	0.066	-0.265	0.093
30	27.79	0.596	0.340	-0.039	0.137					

TABLE 1c. Estimations from: ARMA(1,1) model, Gaussian input, **c** = 1.0 (scale in t.u.)

scale	K(t.u.)	c	Φ	Θ	R^2	K(t.u.)	c	Φ	Θ	R^2
		K=120					**K**=90			
1	128.8	1.000	0.992	-0.981	0.996	98.27	1.000	0.990	-0.980	0.995
7	124.1	0.985	0.945	-0.301	0.937	94.22	0.980	0.928	-0.300	0.917
15	132.7	0.968	0.893	-0.260	0.863	99.47	0.957	0.860	-0.258	0.823
30	135.2	0.938	0.801	-0.252	0.752	104.7	0.920	0.751	-0.242	0.691
60	115.7	0.878	0.596	-0.285	0.546	89.46	0.848	0.511	-0.277	0.458
		K=60					**K**=30			
1	66.11	1.000	0.985	-0.980	0.993	32.74	1.000	0.970	-0.977	0.985
7	62.92	0.971	0.895	-0.299	0.878	30.81	0.943	0.797	-0.299	0.767
15	65.88	0.937	0.796	-0.255	0.747	32.13	0.879	0.627	-0.253	0.555
30	73.47	0.884	0.665	-0.221	0.581	41.82	0.781	0.488	-0.158	0.348
60	63.00	0.791	0.386	-0.253	0.326	36.34	0.663	0.192	-0.196	0.134
		K=15					**K**=7			
1	16.14	1.000	0.940	-0.969	0.969	7.46	1.000	0.875	-0.947	0.933
7	14.97	0.893	0.627	-0.303	0.585	6.82	0.812	0.358	-0.304	0.330
15	15.08	0.793	0.370	-0.266	0.309	4.98	0.858	0.049	-0.304	0.103
30	29.16	0.602	0.358	-0.030	0.144					

TABLE 2. ARMA(1,1) model: standard errors of estimates made on different scales compared to s.e. of estimates made on limited samples (case considered **K**=60, **c**=0.5).

	aggregated				limited sample			
n. of data	Φ	Θ	std. err. Φ	std. err. Θ	Φ	Θ	std. err. Φ	std. err. Θ
1448	0.902	0.805	0.0309	0.0429	0.6473	0.6061	0.1826	0.1906
666	0.785	0.588	0.0636	0.0838	0.4936	0.4596	0.345	0.3526
333	0.592	0.272	0.1180	0.1407	0.2146	0.1833	0.5067	0.5124

Shot noise model

The first consideration arising from the observation of Tables 3-4 and Figure 4 is that aggregation has quite different effects on the Shot Noise model estimates than for the ARMA model. With the Shot Noise model there are no evident benefits arising from aggregation, since best estimates are always obtained at the reference scale. The increasing bias of the estimated values of both parameters with aggregation does not leave much room for other considerations.

This outcome could be due to the alterations that aggregation induces in the impulse occurrence and intensity and reflects the different characters peculiar to this class of models. The positive aspect of this behavior is that even very large storage constants can be identified (with some bias) at minimum scale.

Another interesting aspect is the reduced increase of **c** reduces negative effect of aggregation on the estimate $\hat{c}$**.** This could be due to the reduced alteration of the white noise component, with aggregation, occurring when **c** increases.

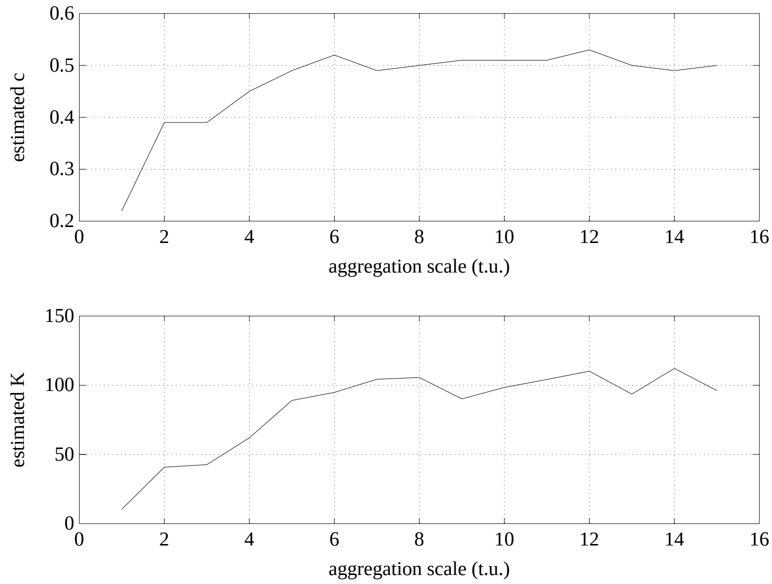

Figure 1. Arma(1,1) model: Parameter estimates on aggregated data (**c**=0.5, **K**=120 t.u.).

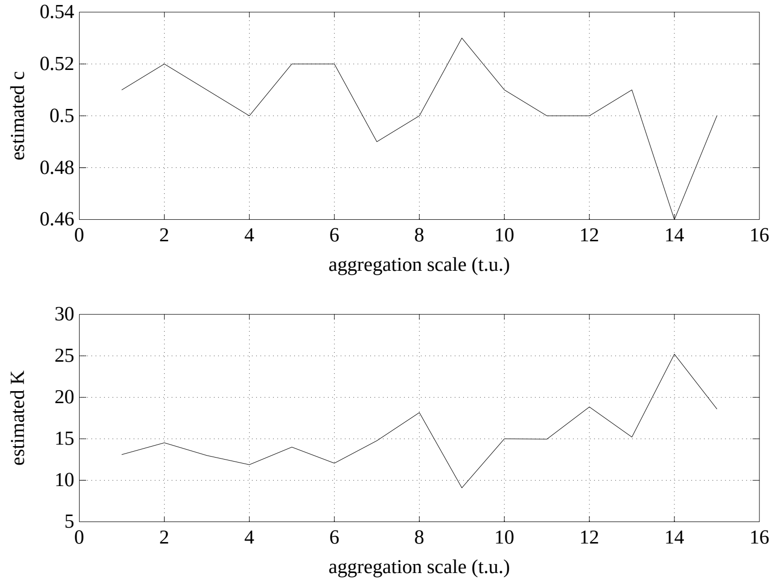

Figure 2. Arma(1,1) model: Parameter estimates on aggregated data (**c**=0.5, **K**=15 t.u.)

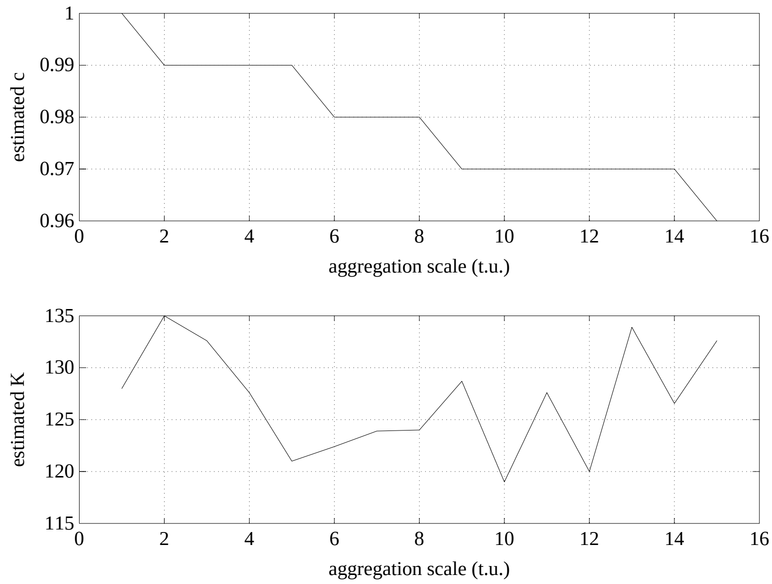

Figure 3. Arma(1,1) model: Parameter estimates on aggregated data (**c**=1, **K**=120 t.u.).

TABLE 3. Estimations from: shot noise model, Bessel input, conceptual model of one reservoir with a linear channel (**K**=60).

	C_1=0.5			C_1=0.8			C_1=1		
scale	c_1	K	R^2	c_1	K	R^2	c_1	K	R^2
1	0.532	48.52	0.026	0.813	55.19	0.092	0.993	66.39	0.933
7	0.704	49.54	0.210	0.872	77.68	0.386	0.969	110.56	0.796
15	0.756	82.36	0.242	0.896	112.24	0.474	0.953	172.23	0.707
30	0.837	214.76	0.263	0.912	261.86	0.419	0.942	229.29	0.613

TABLE 4. Estimations from: shot noise model, Bessel input, (**c**=0.5, **K**=120).

C_1=0.5			
scale	c_1	K	R^2
1	0.533	87.11	0.018
7	0.695	69.35	0.158
15	0.766	162.39	0.175
30	0.837	339.92	0.216

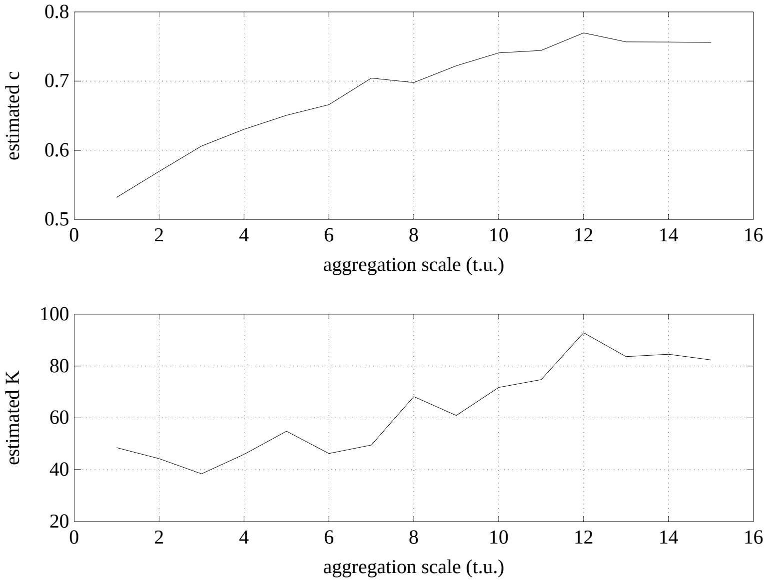

Figure 4. Shot noise model: Parameter estimates on aggregated data (**c**=0.5, **K**=60 t.u.).

FINAL REMARKS

In a conceptually-based stochastic framework for the analysis of the runoff data at different scales, a simulation study undertaken to assess possible effects of aggregation on parameter estimation. A simple linear conceptual system, made up of a linear reservoir and linear channel, was used to generate runoff data from Gaussian and Bessel input, and a conceptually based ARMA(1,1) model and a single Shot Noise model were respectively fitted to the data, providing estimates of the conceptual parameters.

Analysis of the results emerging by re-estimation of "true" parameters by means of these models showed that aggregation plays a significant role in achieving correct estimates for the ARMA(1,1) model. In particular, optimal aggregation scale is the one at which both estimates of the conceptual parameters attain a "sill" level, which is shown to correspond to the least biased value. On the other hand, aggregation does not produce the same effect on the Shot Noise model, for which the scale of generation was found to be the most significant for parameter estimation.

Although a more extensive work is needed to test the effect of aggregation on estimation of parameters of more complex systems, these results constitute an interesting starting point as a theoretical support to the use of integrated conceptually-based models.

REFERENCES

Bartolini, P. and J.D. Salas (1993) "Modelling of streamflow processes at different time scales", Water Resour. Res., **29**(8), 2573-2587.

Benjamin,J.R. and C.A. Cornell, Probability, Statistics and Decision for Civil Engineers, p. 310, Mc. Graw Hill Book Co., New York, 1970.

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

Bernier J. (1970) "Inventaire des Modeles des Processus Stochastiques applicables a la Description des Debits Journaliers des Rivieres", Rev. Int. Stat. Inst., 38, 49-61.

Claps, P. and F. Rossi (1992). "A conceptually-based ARMA model for monthly streamflows", in J.T. Kuo and G.F.Lin (Eds.) Stochastic Hydraulics '92, Proceedings of the Sixth IAHR International Symposium on Stochastic Hydraulics, Dept. of Civil Engrg., NTU, Taipei (Taiwan), 817-824.

Claps, P.(1992) Sulla validazione di un modello dei deflussi a base concettuale, Pro. XXII Conf. Hydraul. And Hdraul. Struct., Dept. Civil Eng., Univ. of Florence, D.91-D-102.

Claps, P. and F. Murrone (1993) "Univariate conceptual-stochastic models for spring runoff simulation", Proceedings of the XXIV Annual Pittsburgh Conference On Modelling And Simulation, May 10 - 11, 1993, Pittsburgh, USA, 491-494.

Claps P., F. Rossi and C. Vitale (1993) Conceptual-stochastic modelling of seasonal runoff using Autoregressive Moving Average models and different scales of aggregation, Water Resour. Res., **29**(8), 2545-2559.

Cowpertwait P.S.P. and P.E. O'Connell (1992) "A Neyman Scott shot noise model for the generation of daily streamflow time series", in J.P. O'Kane (Ed.) Advances in Theoretical Hydrology, Part A, chapter 6, Elsevier, The Netherlands.

Jakeman, A.J. and G.M. Hornberger (1993) "How much complexity is warranted in a rainfall-runoff model?", Water Resour. Res., **29**(8),2637-2649.Kavvas, M.L., L.J. Cote and J.W. Delleur (1977). "Time resolution of the hydrologic time-series models", Journal of Hydrology, **32**, 347-361.

Kron W, Plate E.J. and Ihringer J. (1990) "A Model for the generation of simultaneous daily discharges of two rivers at their point of confluence", Stochastic Hydrol. and Hydraul., (4), 255-276.

Obeysekera, J.T.B. and J.D. Salas (1986). "Modelling of aggregated hydrologic time series", Journal of Hydrology, **86**, 197-219.

O'Connell, P.E. (1971) A simple stochastic modelling of Hurst's law, Int. Symp. on Math. Models in Hydrology, Int. Ass. Hydrol. Sci. Warsaw.

Murrone, F., F. Rossi and P. Claps (1992). "A conceptually-based multiple shot noise model for daily streamflows", in J.T. Kuo and G.F.Lin (Eds.) Stochastic Hydraulics '92, Proc. of the Sixth IAHR Intl. Symp. on Stochastic Hydraulics, Dept. of Civil Engrg., NTU, Taipei (Taiwan), 857-864.

Rossi, F. and G. Silvagni (1980). Analysis of annual runoff series, Proc. Third IAHR Symp. on Stochastic Hydraulics, Tokio, A-18(1-12).
Salas, J.D. and R.A. Smith (1981) Physical basis of stochastic models of annual flows, Water Resour. Res., **17**(2), 428-430.
Salas J.D., Delleur J.W., Yevjevic V. and Lane W.L. (1980). Applied Modeling of Hydrologic Time Series. Water Resources Publications, Littleton, Colorado.
Vecchia, A.V., J.T.B.Obeysekera, J.D. Salas and D.C. Boes (1983) "Aggregation and estimation of low-order periodic ARMA models", Water Resour. Res., **19**(5),1297-1306.

Acnowledgements

This work was supported by funds granted by the Italian National Research Council – Group for Prevention from Hydrogeological Disasters, grant no. 91.02603.42.