WATER RESOURCES ASSESSMENT UNDER WATER SCARCITY SCENARIOS

CONTINUOUS-TIME MODELLING OF HYDROLOGIC TIME SERIES:
SHOT NOISE MODELS

Paola ALLAMANO, Pierluigi CLAPS, Francesco LAIO

Editors

Goffredo La Loggia
Giuseppe T. Aronica
Giuseppe Ciraolo

centro studi IDRAULICA urbana

CONTINUOUS-TIME MODELLING OF HYDROLOGIC TIME SERIES: SHOT NOISE MODELS

Paola ALLAMANO*, Pierluigi CLAPS*, Francesco LAIO*

ABSTRACT

To simulate the runoff process at short-term aggregation scale (typically one day) the shot noise model has gained considerable reputation. In this paper, the model developed by Murrone et al. (1997) is taken as a notable example of physically-consistent framework to discuss the different aspects related to the building and estimation of a stochastic tool for time-series generation. In particular, applications of the model in various geographic, morphological and climatic conditions give the opportunity to present its features, also in relation to an alternative, bi-variate, model configuration.

After a general presentation of the model typology and of the literature background, the conceptually-based shot noise model framework is briefly resumed, with specific attention to the inverse estimation of the effective rainfall and to the assessment of the quality of runoff generation. Various issues related to model application in temperate, semi arid and alpine watersheds are then presented and discussed.

Even though additional efforts are required to adapt the conceptualization to alpine environments, the shot noise framework demonstrates to be a valid tool, also to accompany physically-based modelling or frequency analysis of extreme events, as it reproduces in a treatable way the basic mechanisms of runoff formation.

1 SHOT NOISE PROCESSES

1.1. PROPERTIES

A shot noise process is a random process in continuous time that considers an event in the present time to be the additive result of the effects of antecedent events.

* Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino

The process is completely defined by:

- the occurrences, i.e. the times at which events take place
- the pulses, i.e. the event intensities
- the response function, a continuous function that describes the temporal evolution and propagation of the pulse effect.

Defining X as the variable to describe, its value at time t is produced by:

$$X(t) = \sum_i Y_i \cdot h(t - \tau_i) \tag{1}$$

where Y_i is the input process and h(.) is the system response function that describes the time propagation of the impulse.
Process (1) is defined according to the following hypothesis:

- the magnitudes of the pulses are random variables, independent and identically distributed, having finite mean and variance
- the occurrences follow a Poisson process
- the response function h(.) is continuous, integrable and infinitesimal for $t - \tau_i \to \infty$.

The process is considered stationary when its origin tends at $-\infty$. However, under the third condition, the contribution to streamflow by very distant inputs in time can be disregarded and the response function can be truncated at a finite time origin:

$$X(t) = \sum_{t_0 < t_i < t} Y_i \cdot h(t - \tau_i). \tag{2}$$

In applying this continuous process to streamflow series, that have a discrete nature, it is necessary to represent the process in a discrete form, X_t, integrating equation (2) on the sampling interval T:

$$X_t = \int_{(t-1)T}^{tT} X(\tau) \cdot d\tau \tag{3}$$

where the index t describes all the sampling times and $X(\tau)$ has the form (2). Detailed representation if the integrated process form are available in Murrone et al. (1997).

As will be presented in the next section, literature shot noise models present different input processes, response functions and, over all, different identification and estimation phases.

2 LITERATURE REVIEW

In reproducing main features of daily flows the classical ARMA processes (Box and Jenkins, 1970; Salas et al., 1980) used for time series proved to be unsatisfactory. The reason are the steep gradients that produce marked intermittent nature of the process and the high skewness (Lawrence and Kottegoda, 1977; Battaglia, 1986). An evolution of the cited category of models is that of DARMA processes (Discrete ARMA) (Chang et al., 1987), in which the sequence of dry and rainy days is reproduced by a DARMA process combined with a probabilistic model of precipitation volumes (e.g. the exponential distribution). The obtained precipitation series is then transposed to a runoff series by a linear transfer function, that is completely defined only when the characteristics of precipitation are known. This is the reason why these models cannot be applied to the univariate processes.

Filtered Poisson processes were first investigated by Bernier (1970), who was followed by Weiss (1977). Weiss considered runoff to be the additive result of two different physical processes: the surface runoff and the aquifer contribution, each modelled as a linear system having its own input process and its own response function. The representation, in continuous time, is the filtered Poisson process:

$$X(t) = \sum_{N(-\infty)}^{N(t)} Y_i e^{-b(t-\tau_i)} \tag{4}$$

where N(t) is the counting function of the Poisson process and Y is an exponentially distributed random variable.

This runoff model has six parameters, three for each process considered. Parameter estimate is made by comparing first and second order model statistics, expressed in the daily discrete form (3), with the corresponding values of sample statistics. Series seasonality is taken into account by separately estimating the parameters for each month of the year. The model does not provide any information about the inputs structure, since the estimate of the relative input series is not carried out.

In the same year Treiber and Plate (1977), and later Kron et al. (1990), proposed a daily runoff model based on a shot noise process. The watershed is considered as a deterministic system, represented by an impulsive response function h(.), subject to a stochastic input, i.e. the pulses process, that gives the occurrence and intensity of the rainfall events. The estimation procedure determines the input series and the system response function, given the runoff observed series. The pulses, that are rigorously defined only under the hypothesis that h(.) monotonically decreases, occur in correspondence of positive flow variations.

A preliminary estimate of the response function is obtained applying the shot noise property:

$$K(s)=\frac{\int_0^{+\infty} h(t)\cdot h(t+s)\cdot dt}{\int_0^{+\infty} h^2(t)\cdot dt} \tag{5}$$

that relates h(.) and the process autocorrelation K(.). The least-squares estimate of the pulses intensities is obtained by imposing the non-negativity condition on the pulse value. Parameters of h(.) are estimated by introducing a dependency on the runoff value, in order to account for the presence of non-linearity in the daily runoff process. Final estimates come from the alternative optimization of the pulses and of h(.), by minimizing the residuals sum of squares. Precipitation seasonality is accounted for by estimating the stochastic parameters month by month. The autocorrelation structure of the pulses is reproduced by a two-parameters Markov model. The model has 67 parameters, of which 62 are referred to the effective precipitation process.

A bivariate model, based on the analysis of physical processes of the rainfall-runoff transformation, was introduced by Koch (1985). The impulsive response function (IUH) of the watershed is a combination of the responses of two linear storages, that represent the surface and the aquifer runoff. Precipitation is reconstructed according to a process with rectangular pulses of random length and intensity and Poissonian occurrence. A continuous representation following a filtered Poisson process is obtained, analogously to the Weiss model. The advantage of this approach is the possibility to assign a physical meaning to the stochastic parameters, but the problem of parameters estimates is not handled directly, making the model application impossible for operative purposes.

Battaglia (1986) proposed an original inference procedure referred to the standard shot noise models. He transposed expression (2), valid for a stationary process, to the discrete form:

$$X_t=\sum_{k=0}^{+\infty} h_k\cdot Z_{t-k}+E_t \tag{6}$$

where E_t is the minimum threshold representing the factors and components that influence additively the process but cannot be described by the shot noise mechanism. The following hypothesis were assumed:

- the h_k series is quadratically convergent, with h_0=1;
- Z_t is a stationary white noise with positive mean and finite variance;
- E_t is a stationary process with non negative mean and finite variance;
- Processes E_t and Z_t are uncorrelated.

A finite memory process is considered for the inference, with the summation term in (6) truncated at the finite value q, representing the maximum lag when the autocorrelation of the sample is found to be significatively positive. The occurrences are identified as the days when the following condition is verified:

$$x_t \geq x_{t-1} + L \tag{7}$$

being L the value of the threshold that optimizes the model fitting to the data and to the series of differences $Z_t = x_t - x_{t-1}$. The definitive estimate of parameters values of the model is obtained by iterating the following steps:

- given the initial series Z_t, the sum of the squares of the residuals, with respect to the unknown quantity h_k, is minimized;
- given the estimate of the weights of the response h_k, the minimizing procedure is applied to Z_t, for the instants in which the events occurred;
- the iteration procedure goes on until the value of the sum of the squares of the residuals can be considered stable.

The procedure suggested by Battaglia (1986) was retained in the model by Murrone et al. (1997) and represents, as will be explained in the following, a significant advance in the identification of a realistic input process.

3 SHOT NOISE MODELS OF DAILY FLOW TIME SERIES

The model by Murrone et al. (1997) is based on a conceptual watershed scheme, composed by three linear reservoirs in parallel, with storage constants k_i, plus a sub-daily-lag bypass channel (Figure 1). This configures a multiple shot noise model with the above conceptual elements:

- surface runoff having sub-daily lag time, with the lag depending on the watershed size
- interflow runoff with over-day storage
- contributions from seasonal aquifers, with over-month storage
- deep aquifer components with over-year storage

The effective rainfall represents the input process and is partitioned according to the recharge parameters c_i, that respect the continuity condition $\sum c_i = 1$. The basin response to an input represented by a unit Dirac pulse is a linear function and derives from the combination of the individual responses of the above components. In the most general form it can be written as:

$$h(t) = \frac{c_0}{k_0} e^{-\frac{t}{k_0}} + \frac{c_1}{k_1} e^{-\frac{t}{k_1}} + \frac{c_2}{k_2} e^{-\frac{t}{k_2}} + \frac{c_3}{k_3} e^{-\frac{t}{k_3}} \qquad \text{at } t \geq t_0 \tag{8}$$

in which it must be considered that:

$$h(t) = 0 \text{ at } t < t_0 \tag{9}$$

and that the term $\frac{c_0}{k_0} e^{-\frac{t}{k_0}}$, representing the instantaneous unit hydrograph (IUH) of the surface runoff, reduces to c_0 at aggregation times greater than the surface IUH base-width (t_b).

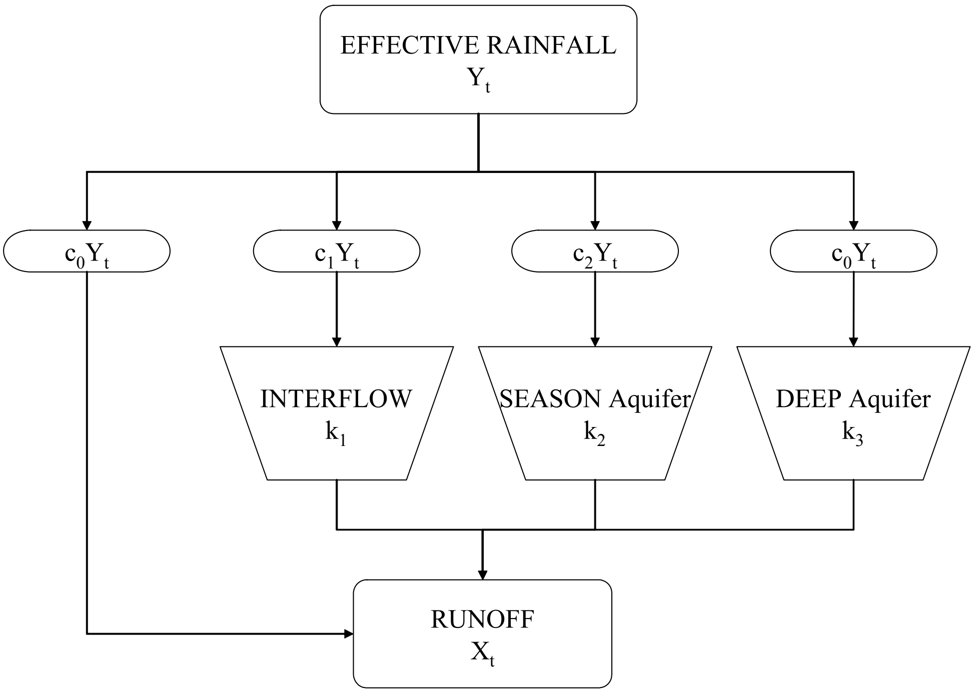

Figure 1 - Conceptual scheme proposed by Murrone et al. (1997) for the multiple shot noise model.

In fact, for $t_b << T$, being T the data aggregation time scale, the surface component recession period is less than the interval T, and the surface runoff process can be reasonably assimilated to the response of a linear channel with a zero-lag response. For a daily time scale, this condition is true for small and medium size basins. For large basins the above condition is valid only considering T > 1 day.

On the basis of the above assumptions, in the following eq. (8) will reduce to

$$h(t) = c_0 \delta(0) + \sum_{i=1}^{3} \frac{c_i}{k_i} e^{-\frac{t}{k_i}} \tag{10}$$

Under the hypothesis of linearity, streamflow X at time t is the additive result of the individual responses (Figure 2).

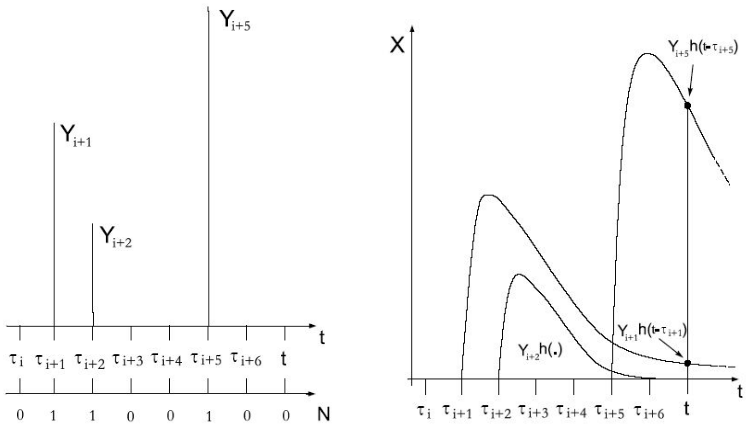

Figure 2 - The graph is an example of the weak influence of remote events (as Y_{i+2}) on the runoff formation at time t (from Giordano, 2004).

The number of events is given by the counting function N(t), that is 1 when the event occurs:

$$\mathrm{X}(\mathrm{t}) = \sum_{\mathrm{N}(-\infty)}^{\mathrm{N}(+\infty)} \mathrm{Y_i} \cdot \mathrm{h}(\mathrm{t} - \tau_\mathrm{i}). \tag{11}$$

The above relation does not have an operational connotation for the following reasons:

- it has infinite memory, since the cumulative sum of the terms formally expands from $-\infty$ to the time t;
- it describes a continuous process, while data is available at discrete times.

To overcome the problems due to the model's infinite memory it is necessary to know the runoff entity $\mathrm{X_0} = \mathrm{X}(\mathrm{t_0})$ in correspondence of the initial time t_0. Given $\mathrm{X_0}$, the equation (11) can be written as a piecewise function:

$$\mathrm{X}(\mathrm{t}) = \mathrm{X_0} + \sum_{\mathrm{N}(\mathrm{t_0})}^{\mathrm{N}(\mathrm{t})} \mathrm{Y_i} \cdot \mathrm{h}(\mathrm{t} - \tau_\mathrm{i}) \tag{12}$$

with the term:

$$X_0 = \sum_{N(-\infty)}^{N(t_0)} Y_i \cdot h(t - \tau_i) \tag{13}$$

representing the runoff generated by events occurred in the interval $(-\infty : t_0]$. The effects of the choice of X_0 are limited in time; therefore, its selection can be quite practical (see Giordano, 2004).

4 EFFECTIVE RAINFALL IDENTIFICATION

In building shot-noise models to represent daily streamflows, the choice of the probability distribution $F_Y(Y)$ of the random marks Y_i, and the definition of the shape of the response function are required. The random occurrence times $\{\tau_i\}$ are assumed to form a Poisson sequence, i.e. the number of occurrences N_t follows a Poisson distribution with rate λ.
The input amount $F_Y(Y)$ can be assumed to be exponential, with mean value α:

$$f_Y(y) = \frac{1}{\alpha} \cdot e^{-\frac{y}{\alpha}}. \tag{14}$$

The above corresponds to using a Poisson-Exponential (PE) model to simulate the effective rainfall input Y_t. Estimation of the PE model parameters was devised by Murrone et al. (1997) in four steps:

- The effective rainfall occurrences are identified in correspondence to the days when the streamflow increases; the magnitude of these events is initially taken equal to the amount of discharge increment, and the parameters of the PE model (α and λ) are estimated using the method of moments. This procedure is called "discharge increments pulses", or DIP, approach.
- Once the effective rainfall sequence is reconstructed, the parameters of the response function h(t) are found by minimizing the sum of squared distances between observed and reconstructed data
- A new pulse series is inversely estimated through deconvolution of the observed discharge time series and the estimated response function
- The second and third steps are repeated until convergence, that means stable Y(t). and h(t) estimates.

This approach allows one to clearly separate the estimation of the stochastic component (effective rainfall) from that of the deterministic one (response function).

Although quite efficient, the DIP approach presents some drawbacks and difficulties of application: first, the presence of measurements noise can produce small rises in discharge that should not be mistaken for effective rainfall events to avoid distortions in the modelling results. In fact the DIP procedure tends to produce pulse sequences with very large λ values (figure 3a) , sometimes greater than the average number of rainy days in a year.

Secondly, the basic hypothesis that the pulses are Poisson-distributed, and that they are mutually independent, are often not respected by the estimated sequences. In fact more complicate models have been proposed which account for the mutual dependence of peaks (e.g. Markov-chain models) and for the clustering of rainy days (Neymann-Scott models or similar). However, the increased complication of such models can hardly be properly supported, mainly due to the reduced length of the available time series. This make the simpler, yet possibly imperfect, Poisson independent model represents a good choice in many cases.

A method to derive a potentially more appropriate pulse sequence is the one based on the work by Claps and Laio (2003), in which the pulses can be identified by following a filtered peak over threshold (FPOT) procedure. The FPOT is summarized as follows:

- The peaks events are found in correspondence to all of the local maxima of the daily discharge time series.
- A sequence of filtered peaks (FP) is obtained by subtracting from each peak the discharge measured at the first minimum preceding the event. A similar approach to the selection of pulse intensities was used by Pegram (1980).
- A threshold filter is applied to the FP sequence to retain only the significant peaks.
- The appropriate threshold (s) that filters out noisy peaks is selected by testing the independence of the peaks in the sample (Kendall's τ test) and the distribution of occurrences (Cunnane test for the Poisson distribution). The threshold is gradually increased until the two test are jointly met. The convergence towards independence for large s values can be attributed to the increase of the distance between subsequent peaks. Analogously, when the threshold is increased the numbers of crossing of s in disjoint time intervals tend to become independent random variables, and this guarantees the asymptotic convergence towards Poissonianity (Claps & Laio, 2003).

The adoption of the FPOT approach allows one to avoid the deconvolution step in the peaks identification procedure, with substantial advantages in terms of simplicity and robustness of the procedure. Moreover, the method allows one to obtain a pulse sequence which automatically meets the independence and Poissonianity requirements. However, the number of selected peaks is reduced to 5-20 per year (figure 3b). This can result in a underestimation of the actual number of effective rainfall events. Since this result is obtained on a large number of runoff series, it can be a clue of the inadequacy of the Poisson independent model in the correct reproduction of the effective rainfall behaviour.

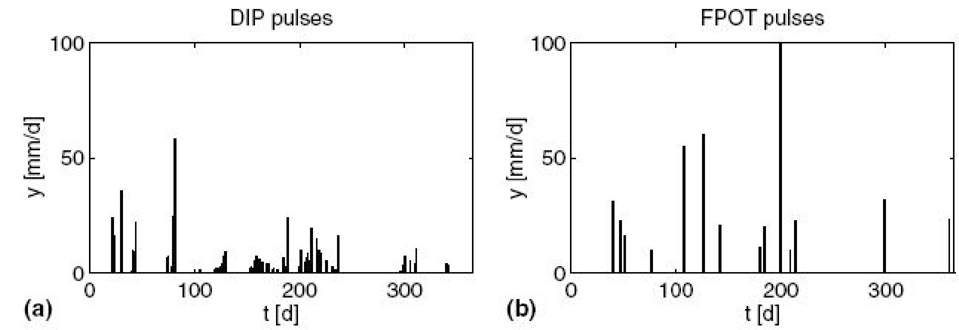

Figure 3 - Comparison of estimated effective rainfall sequences with the DIP (a) and FPOT (b) approaches. The example is relative to river Tanaro at Nucetto (from Claps et al. (2005)).

In order to estimate the response function coefficients, the determination of the full IUH base-width (t_c) is required, with t_c representing the time interval after which the input effects disappear. The procedure proves to be extremely conditioned by the value of the parameter t_c, and to define a unique base-time valid for the different basins is almost impossible. Therefore a control procedure is implemented by Claps et al. (2003a) to identify the optimal t_c for each basin.
The procedure evaluates the parameters for each linear reservoir using t_c = 50 days as first approximation. Then the impulse response h(t) is evaluated and the following condition is verified:

$$\sum_{i=1}^{t_c} h_i \geq 0.999 \tag{18}$$

If the condition is satisfied the current t_c is adopted; in the other case, the procedure is repeated with an increased t_c until convergence. No particular control is exerted on the response function parameters, except the condition (18). The number of identified reservoirs conditions the convergence of the procedure and, more in general, the convergence of the global procedure for estimation of parameters. Once the model structure is set, the values of estimated parameters can be validated with ancillary information and provide insights in the process features, as will be discussed in the applications.

5 ASSESSMENT OF MODEL RESULTS

The value of a stochastic model for data generation can be defined from its ability to correctly reproduce the statistical features of the observed records. Generally this ability is evaluated by comparing the moments of the observed and simulated series at different aggregation scales (e.g. Weiss, 1977) or, in some cases, the full probability density functions (e.g. Vandewiele and Dom, 1989) or, qualitatively, the flow duration curve.

Some confusion in the assessment procedures of model performances can derive from the use of both the reconstructed and the generated time series as possible terms of comparison with the real observations. The reconstructed time series is obtained from the convolution of the estimated input pulse sequence with the system response function. Reconstructed and observed series are easily compared using any measure of reciprocal distance between the two signals. However, the real value of the model must rather be judged by comparing observed and generated sequences: the time correspondence of the peaks of the two series is obviously lost in this case, and the distance between contemporary values becomes meaningless.

In this latter case, a better option is to consider the reciprocal distance between the observed and synthetic cumulative distribution functions, represented as flow duration curves. In order to build the flow duration curves, the observed discharge dataset of size n is sorted in ascending order, and an empirical frequency of occurrence $F_{(i)} = i/(n+1)$ is assigned to the i^{th} order statistic in the sample, $Y_{o(i)} = Y_o(F_{(i)})$. The same procedure is followed for the generated sample, whose size N is much larger than n, producing $F_{(j)} = j/(n+1)$ and $Y_{g(j)} = Y_g(F_{(j)})$. The flow duration curves are given by the $F_{(i)}$ and $F_{(j)}$ values plotted versus $Y_{o(i)}$ and $Y_{g(j)}$; the curves are not graphically represented in the following, because the resulting real and generated graphs are nearly indistinguishable.

To evaluate the distance between these curves, a value of generated discharge with the same frequency of occurrence as $Y_{o(i)}$, namely $Y_g(F_{(i)})$ need to be found. $Y_g(F_{(i)})$ is the empirical quantile corresponding to $F_{(i)}$, i.e. the j^{th} order statistic in the generated sample, with $j = (N+1)/(n+1)$ (j can be conveniently approximated to the closer integer because of the large sample size N).

The mean squared distance between the two flow duration curves (‘‘model error variance’’) is then evaluated as

$$s^2 = \frac{\sum_{i=1}^{n} \left[Y_o(F_{(i)}) - Y_g(F_{(i)})\right]^2}{n}. \tag{15}$$

In order to facilitate the comparison between different applications, the value in (15) is rescaled by the variance σ^2 of the observed discharges, and an index of performance similar to the coefficient of determination of linear regression models was proposed by Claps et al. (2005) as a measure of model adequacy:

$$I_1 = 1 - \frac{s^2}{\sigma^2} \tag{16}$$

The closer I_1 to its limit value 1, the more adequate is the model to represent the flow duration curve.

Other measures of model adequacy can be defined by considering other characteristics of the observed and generated sequences. For example, to test the correct reproduction of the annual maxima (AM) statistics (see Fig. 7e and f), a specific index can be defined as:

$$I_2 = 1 - \frac{s_{AM}^2}{\sigma_{AM}^2} \tag{17}$$

where $s_{AM}^2 = \frac{\sum_{j=1}^{k}\left[Y_o^{AM}(F_{(j)}) - Y_g^{AM}(F_{(j)})\right]}{k}$ is the mean squared distance between the empirical frequency curves of the k observed AM, Y_o^{AM}, and the corresponding curves for generated data, Y_g^{AM}, and σ_{AM}^2 is the variance of the observed AM sample.

6 MODEL PERFORMANCE IN DIFFERENT APPLICATION CONTEXTS

The Shot Noise procedure described here has not only the purpose of producing an efficient and parsimonious model formulation for time series simulation. In fact, this model can be also applied as a diagnostic tool, able to compare some characteristics of the long-term basin hydrologic response and some features of the inversely-estimated net rainfall. In general terms, one could expect that model building and application provides information on the main mechanisms underlying the long-term formation of the runoff. Presentation and discussion of some shot noise applications in the literature will support the above-mentioned purpose. Literature application of the Shot Noise model refer to watersheds taken from regions characterized by different climates and different sources of 'reservoir' effects. These effects were 'interpreted' according to the model conceptual structure.

In Murrone et al (1997) the multiple shot noise model was applied to 8 time series of daily flows, recorded in 7 watersheds located in the Apennine region of central-southern Italy. Two sub-series were considered for the Tiber river, since the streamflow record is interrupted. The basin under study are all characterized by the climate and the geology of Apennine mountains, in which the presence of large fractured carbonate massifs produces correlation also in the annual runoff, requiring the identification of a possible over-year groundwater component. Looking at values of the shot noise parameters estimated by Murrone et al. (Table 1) one can notice significant values of the parameter k_3 related to the deep groundwater component identified for six of the seven examined cases.

Table 1 - Estimated shot noise model parameters for the 'Appennine' time series

Name	Area (km^2)	c0	c1	c2	c3	k1 (d)	k2 (d)	k3 (d)
Alento @ Casalvelino	284	0.34	0.281	0.297	0.082	3.105	60.4	551.4
Calore Irpino @ Montanella	123	0.155	0.194	0.546	0.106	2.672	72.2	228.5
Tammaro @ Pago Veiano	555	0.272	0.236	0.492	-	2.912	35.6	-
Sacco @ Ceccano	922	0.319	0.242	0.261	0.178	2.706	53.9	507
Giovenco @ Pescina	139	0.106	0.112	0.191	0.59	3.274	56.3	1073
Tiber @ Rome	16545	0.098	0.158	0.225	0.52	4.076	43.7	1233
Tiber @ Rome	16545	0.101	0.136	0.243	0.52	5.552	40.2	1233
Nera @ Torre Orsina	1445	0.025	0.018	0.247	0.71	5.401	109.3	1533

In a semi-arid context, model identification resulted in the absence of the over-year component. Cannarozzo et al. (2003) applied the model to series from 11 watershed in Sicily, obtaining results reported in Table 2. The results show some variation in the k_2 and c_2 parameters while k_1 presents quite stable values in the interval of 1-3 days.

Based on the two series of results presented, one can conclude that in basins located in temperate regions, the only option in model identification and estimation is related to the presence of an over-year reservoir component. If this component is missing, the other sources of delay remain easily identifiable, as widely reported in the literature (see e.g. Jakeman and Hornberger, 1993). Therefore, shot noise application in temperate basins without over-year component presents no difficulties but can still provide interesting insights in net rainfall estimation, as it will referred later in this section.

Model application in non-temperate alpine basins revealed various interesting issues. First, identification of the over-year component was again registered, even in absence of the same hydro-geological features of the Apennine mountains. Claps et al (2003) considered 8 watershed located in Piemonte and Valle d'Aosta (North-Western Italy), presenting a wide spectrum of morpho-climatic characteristics. These basins present very variable average elevation (see Table 3) and some of them must be considered 'Alpine' basins, *i.e.* strongly influenced by the snow and ice contributions to runoff. Table 3 reports the model parameter estimates on these series, from which emerges that only the larger basin (Dora Baltea at Tavagnasco, also the second highest) presents a significant correlation at the over-year scale. This could be due to the presence of a strong runoff component coming from the melting of snow and glacier stocks. Beside this, it can be observed that other alpine watersheds, such as Ayasse and Chisone, show no deep component but rather high values of seasonal aquifer recharge coefficient c_2. This could be due to the seasonal-only effect of snow storage melting on the flow formation process.

Table 2 - Estimated shot noise model parameters for the time series in Sicily

Name	c0	c1	c2	c3	k1 (d)	k2 (d)	k3 (d)
Oreto @ Parco	0.198	0.253	0.549	-	1.14	70.59	-
Eleuterio @ Lupo	0.193	0.354	0.453	-	1.23	53.49	-
Salso @ Gagliano	0.23	0.402	0.368	-	1.56	24.45	-
Imera Mer. @ Capodarso	0.193	0.343	0.464	-	1.09	31.13	-
Imera Mer. @ Drasi	0.185	0.411	0.404	-	1.43	43.71	-
Imera Mer. @ Petralia	0.149	0.272	0.58	-	2.05	69.85	-
Belice @ Case Balate	0.196	0.346	0.458	-	1.42	54.2	-
Belice @ Finocchiara	0.314	0.351	0.336	-	2.79	81.56	-
Belice @ Ponte Belice	0.166	0.398	0.436	-	1.29	51.45	-
S. Leonardo @ Monumentale	0.357	0.256	0.387	-	1	53.71	-
Nocella @ Zucco	0.189	0.215	0.596	-	1.72	60.39	-

In alpine basins the presence of the snow-melting component represents a notable alteration of the net rainfall-runoff conceptual mechanism. Effective rainfall identified by the model are in fact not necessarily due to rainfall events, but can refer to snow melting events. For the purpose of time-series generation, an empirical use of the shot noise model raises the point of what is the most efficient estimation of effective rainfall, independently on its meaning. The application made by Claps et al. (2005) showed that more realistic runoff time series representation were obtained using the DIP type of input, as the one that better resembles the combination of rainfall and snowmelt event sequences.

If the rainfall-runoff conceptual mechanism is not sufficient for the process description in high alpine basins, better qualification of the runoff formation mechanisms are required. Also, runoff analysis in the 'transition' basins must be adequately addressed.

The Piemonte and Valle d'Aosta region presents a wide variety of climatic and morphological situations, offering cases for the analysis of different mountain basins. More detailed examination of the estimated effective rainfall series was presented in the work by Claps et al. (2003b) where the authors propose a comparison of direct and inverse effective rainfall estimates at the daily time scale. The comparison is carried out in terms of effective rainfall estimation by means of the IHACRES (Jakeman et al., 1990) and Shot Noise models, essentially to face model performances in transition and alpine basins. The Shot Noise and IHACRES models have a similar structure in terms of (linear) effective rainfall to runoff transformation, but the effective rainfall (ER) series is obtained by inverse estimation in the former model and directly from rainfall in the latter model.

Table 3 - Estimated shot noise model parameters for the Alpine series, with the DIP and the FPOT (values in italic) procedures used for effective rainfall identification

Name		Area (km^2)	Mean Elev.(m)	c0	c1	c2	c3	k1 (d)	k2 (d)	k3 (d)
Dora Baltea @ Tavagnasco	DIP	3313	2090	0.03	0.07	0.43	0.47	1.1	23.31	546
	FPOT			0.03	0.04	0.47	0.47	1.2	33.45	546
Ayasse @ Champorcher	DIP	42.2	2392	0.05	0.08	0.86	-	1.2	19.7	-
	FPOT			0	0.07	0.93	-	0.2	25.7	-
Borbera @ Baracche	DIP	202	880	0.1	0.2	0.71	-	2.4	53.6	-
	FPOT			0.08	0.19	0.72	-	3.3	76.3	-
Bormida @ Cassine	DIP	1483	493	0.12	0.48	0.39	-	2.2	55.9	-
	FPOT			0.12	0.38	0.49	-	2.5	89.2	-
Chisone @ S.Martino	DIP	580	1751	0.05	0.17	0.78	-	3.3	86.2	-
	FPOT			0.04	0.1	0.87	-	4.3	101.3	-
Orco @ Pont Canavese	DIP	617	1930	0.1	0.19	0.72	-	2.1	61.9	-
	FPOT			0.04	0.1	0.86	-	3.5	164.5	-
Scrivia @ Serravalle	DIP	605	695	0.12	0.34	0.54	-	2	43.7	-
	FPOT			0.12	0.28	0.6	-	2.7	57.1	-
Tanaro @ Nucetto	DIP	375	1227	0.1	0.25	0.65	-	2	135.3	-
	FPOT			0	0.18	0.82	-	0.4	104.7	-

In IHACRES, the rainfall-runoff transformation is obtained with two modules: a non linear loss module, that transforms precipitation to effective rainfall by considering the (direct -if available- or indirect) influence of temperature, and a linear module, based on the classical convolution of the effective rainfall by the unit hydrograph (UH) that produces the total streamflow. The non-linear loss module involves the calculation of an index of catchment storage *s(t)* for every time step *t*, based on a negative exponential weighting of precipitation and temperature:

$$s(t) = \frac{r(t)}{c} + \left(1 - \frac{1}{\tau_w[T(t)]}\right) \cdot z^{-1} \cdot s(t) \tag{18}$$

$$\tau_w[T(t)] = \tau_w \cdot e^{[0.062 \cdot f \cdot (20 - T(t))]} \tag{19}$$

In (18), $s(t)$ is the catchment storage index, $\tau_w[T(t)]$ is a variable controlling the rate at which the catchment wetness index $s(t)$ decays in the absence of rainfall, τ_w is the value of $\tau_w[T(t)]$ at T=20°C, c is a parameter chosen to constrain the volume of effective rainfall to equal runoff, f is a temperature modulation factor, z^{-1} is the backward shift operator.

The effective rainfall $ER(t)$ is computed as the product of total rainfall $r(t)$ and the storage index $s(t)$,

$$ER(t) = s(t) \cdot r(t) \tag{20}$$

and then convolved with the unit hydrograph of the two-reservoirs-in-parallel linear system,

$$h(t) = \frac{\nu_q}{\tau_q} \cdot e^{-\frac{t}{\tau_q}} + \frac{\nu_s}{\tau_s} \cdot e^{-\frac{t}{\tau_s}} \tag{21}$$

The above relation (21) is a function of the basin dynamic response characteristics (DRCs) (Littlewood *et al*, 2003) that depends on parameters ν_q and ν_s (relative volumetric throughputs for quick and slow flow), τ_q and τ_s (characteristic decay time constants for quick and slow UHs), and of the time step *t*.

For the comparison, six basins were examined: three rainfall-driven coastal watersheds in British Columbia (Canada) and three basins located in northern Italy, respectively a temperate, a transition and a pure alpine watershed (see Table 4). Authors observe that in temperate watersheds precipitation series show a strong correlation with observed runoff, while in alpine environments runoff derives also from snowmelt and precipitation values are often affected by significant errors with a consequent reduction of cross correlation coefficients.

The parameters to be set in the basin response functions are reported in Table 5 for each of the 6 basins. Note that for the Evançon at Champoluc the calibration procedure of the IHACRES method did not converge, so it was impossible to find the parameter values. The lack of correlation between rainfall and runoff in alpine basins is a clue of possible problems in the direct estimate of effective rainfall. In general terms, shot noise estimation of ER resulted more robust than the IHACRES ones, with slightly better perfpormance also in temperate basins. Performances were assessed by considering the time series of occurrences of direct and inverse ER estimates, starting from basins in temperate regions.

Table 4 - Watersheds main characteristics (area, mean elevation, average annual rainfall, average annual discharge) and cross correlation coefficients R and Q (see text for details).

	Area [km^2]	Mean Elev. [m]	r [mm/y]	y [mm/y]	R	Q
San Juan Riv. @ Port Renfrew	580	663	3452	2604	0.85	0.84
Kanaka Creek @ Webster Corners	48	460	1807	1818	0.75	0.78
Roberts Creek @ Roberts Creek	33	697	1383	993	0.6	0.78
Scrivia @ Serravalle	611	695	1389	827	0.67	0.71
Chisone @ S.Martino	580	1730	1058	694	0.45	0.29
Evançon @ Champoluc	102	2631	1084	977	0.15	*

Table 5 - IHACRES and Shot Noise parameters calibrated for three Canadian and three Italian basins

Basin	IHACRES		SHOT NOISE	
San Juan Riv. @ Port Renfrew	τ_q	0.76	k_1	1.86
	τ_s	53.15	k_2	70.63
	ν_q	0.47	c_1	0.45
	ν_s	0.53	c_2	0.41
			c_0	0.14
Kanaka Cr. @ Webster Corners	τ_q	1.69	k_1	1.28
	τ_s	89.20	k_2	61.63
	ν_q	0.67	c_1	0.38
	ν_s	0.33	c_2	0.44
			c_0	0.18
Roberts Cr. @ Roberts Creek	τ_q	3.47	k_1	1.62
	τ_s	114.13	k_2	61.48
	ν_q	0.54	c_1	0.35
	ν_s	0.46	c_2	0.50
			c_0	0.15
Scrivia @ Serravalle	τ_q	1.00	k_1	2.14
	τ_s	17.84	k_2	45.66
	ν_q	0.49	c_1	0.34
	ν_s	0.51	c_2	0.54
			c_0	0.12
Chisone @ S.Martino	τ_q	1.29	k_1	6.31
	τ_s	29.15	k_2	220.03
	ν_q	0.20	c_1	0.28
	ν_s	0.76	c_2	0.66
			c_0	0.06
Evançon @ Champoluc	τ_q	*	k_1	22.29
	τ_s	*	k_2	3507.80
	ν_q	*	c_1	0.64
	ν_s	*	c_2	0.30
			c_0	0.06

To this end, the time series of the occurrences are first reverted into binary time series, by attributing a value 1 to the days when an effective rainfall occurs, and a value 0 to the days when it does not. Standard statistical tools can then be applied to compare the two binary series. In this case, a modified form of the cross-correlation coefficient is used, the Goodman and Kruskal (1979) Yule's Q coefficient, which is well suited for applications to binary time series. The Yule's Q gives a measure of the proximity of the two binary series, based on the 2x2 contingency table below:

	0	1
0	*N1*	*N2*
1	*N3*	*N4*

In the table, N1 represents the frequency of occurrence, in the two series, of the (0,0) couple of values, N2 of the (0,1) couple, N3 of (1,0) and N4 of (1,1). Accordingly, the Yule's Q is written as:

$$Q = \frac{N1 \cdot N4 - N2 \cdot N3}{N1 \cdot N4 + N2 \cdot N3} \tag{22}$$

and varies between −1 and 1, with large values implying highly correlated binary series.

The Yule's Q values were computed in relation to the ER series estimated by IHACRES and Shot Noise. The results show that in basins with rainfall-driven streamflows, characterized by a high correlation coefficient between observed rainfall and runoff, it is possible to achieve a reasonable synchronicity between directly and inversely computed ER series, as proved by the high value of the Q coefficient. On the contrary, in alpine environments direct and inverse ER estimates are poorly correlated. For the Chisone basin the Yule's Q coefficient is very low and for the Evançon basin direct ER estimates become even unreliable.

The application demonstrated that the two models exhibit similar behaviour in temperate climates, in terms of values of conceptual parameters and characteristics of the estimated effective rainfall. As one moves from temperate to alpine basins the reliability of areal rainfall weakens and the role of snow in moderating runoff increases, so that the estimates of parameters, the identification of events and the magnitude of ER differ more and more. None of the two models has a specific module to deal with the effect of snow accumulation and melting. However, the features of the Shot Noise model (preservation of runoff volumes, objective evaluation of ER from runoff) produce a more reliable representation of the streamflow process, in particular for basins in a transition (from temperate to alpine) environment.

7 CONCLUSIONS

Univariate shot noise modelling of 'high resolution' time series has been outlined in different phases of model building: identification, estimation and validation. Possibilities to use the conceptually-based modelling for diagnosis of the rainfall-runoff process are discussed also in comparison with others bi-variate models, such as the IAHCRES.

Of particular importance is the possibility to give insight to the inversely estimated effective rainfall, by the viewpoint of its stochastic nature and of its conceptual meaning. As an intermittent process derived by a continuous one, estimation of effective rainfall requires additional attention and modelling efforts. To this end, occurrence analyses are presented, that help evaluating the consistency of this unobserved process to the observational evidence of measured rainfall in alpine and transition basins.

Even considering the need for an extension of the shot noise conceptualization to account for processes in Alpine basins, the good model performances in temperate to transition basins make the shot noise a good choice for general purpose time series modelling and simulation.

8 REFERENCES

Battaglia F., *Modelli stocastici per la rappresentazione e la previsione dei deflussi, Quaderno n.3,* Dip. di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", 1986 (in Italian).

Bernier J., *Inventaire des Modeles des Processus Stochastiques applicables à la Description des Débits Journaliers des Rivières,* Rev. Int. Stat. Inst., 38, 49-61 ,1970 (in French).

Box G.E. and Jankins G., *Time Series Analysis, Forecasting and Control. Revised Edition,* Holden-Day, San Francisco, 1970.

Cannarozzo M., Noto L., Viola F., *Effetti del campionemnto sulla stima dei parametri dell'IUH in un modello di deflussi giornalieri,* Proceedings of *"*La difesa idraulica del territorio", Trieste, 2003 (in Italian).

Chang T.J., Delleur J.W., Kavvas M.L., *Application of discrete autoregressive moving average models for estimation of daily runoff,* Journal of Hydrology, 91, 119-135, 1987.

Claps P., Laio F., *Can continuous streamflow data support flood frequency analysis?An alternative to the Partial Duration Series approach*, Water Resour Res 39(8), 1216, 2003.

Claps P., Giordano A., Laio F., *Simulazione stocastica dei deflussi giornalieri: metodi di determinazione delle stime inverse di pioggia efficace e relative proprietà statistiche*, Proceedings of "Giornata di studio: Metodi stiatistici e matematici per l'analisi delle serie idrologiche", Roma, 2003a (in Italian).

Claps P., Giordano A., Laio F., *Advances in shot noise modeling of daily streamflows*, Adv. in Water Resour., 28, 992-1000, 2005.

Claps P., Allamano P., Laio F., *Some considerations on univariate and multivariate streamflow simulation methods*, Proceedings of theNational IHP-UNESCO Seminar: 2005-2015 United Nations Decade for Action: "WATER FOR LIFE", Perugia, 2003b.

Giordano A., *An application of shot noise models to daily riverflow time series in North-Western Italy*, DITIC working paper 2-2004, Polytechnic of Turin, Turin, Italy, 2004. Available from: [http://www.idrologia.polito.it/~giordano] (in Italian).

Jakeman A.J., Littlewood I.G & Whitehead P.G., *Computation of the instantaneous unit-hydrograph and identifiable component flows with application to 2 small upland catchments*, *Journal of Hydrology*, 117, 275-300, 1990

Koch R.W., *A stochastic streamflow model based on physical principles,* Water Resources Research, 21(4), 545-553, 1985.

Kron W., Plate E.J., Ihringer J., *A model for the generation of simultaneous daily discharges of two rivers at their point of confluence,* Stochastic Hydrology and Hydraulics, (4), 255-276, 1990.

Lawrence and Kottegoda, *Stochastic modelling of riverflow time series,* J.R. Statist. Soc. A, 140, Part 1, 1-31, 1977.

Littlewood I.G., Down K., Parker J.R. & Post D.A., T*he PC version of IHACRES for catchment-scale rainfall-streamflow modelling. User Guide.* Institute of Hydrology, Wallingford, UK, 2003.

Murrone F., Rossi F., Claps P., *Conceptually-based shot noise modeling of streamflows at short time intervals*, Stoch Hydrol Hydraul, 11, 483-510, 1997.

Pegram G.G.S., *A continuous streamflow model*, J Hydrol., 47, 65-89, 1980.

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

Treiber B. and Plate E.J., *A stochastic model for the simulation of daily flows,* Hydrologic Sciences Bulletin, XXII, 1 3/1977, 175-192, 1977.

Vandewiele GL, Dom A. *A non-Gaussian multicomponent model for river flow*, Water Resour Res, 25(3), 397-404, 1989.

Weiss G., *Shot noise models for the generation of synthetic streamflow data,* Water Resour. Res., 13(1), 101-108, 1977.