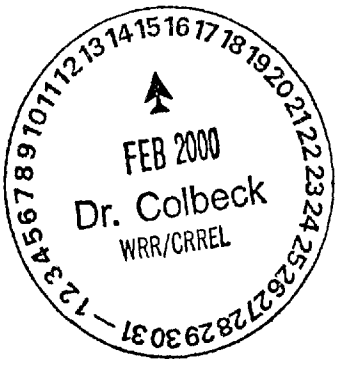
# A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering,
National Technical University, Athens
Heroon Polytechneiou 5, GR-157 80 Zographou, Greece
(dk@hydro.ntua.gr)




## Appendix (Supplement on microfiche)

### A1 Some examples for comparison of the generalized autocovariance structure to fractional Gaussian noise models

In this Appendix, we demonstrate the generalized autocovariance structure (GAS; Equation (7)) using synthetic and historical data records, also comparing it with the stochastic structure implied by the fractional Gaussian noise (FGN) model. We give five examples with record lengths between 44 and 100. The first two of them are synthetic samples generated by stochastic models. In this case the theoretical autocovariance function is known and, consequently, we can test the ability of model to capture this stochastic structure, and the appropriateness of the fitting method. The other three examples are annual streamflow and rainfall records of gauges at Greece and USA. In this case our purpose is to compare the appropriateness of each of the GAS and FGN models for fitting the data and investigate the model parameters.

For the GAS case, the most parameter parsimonious form was adopted for all examples, using the parameters $\gamma_0$, $\beta$ and $\kappa$ only. The fitting of $\beta$ and $\kappa$ was done by the least squares method on the empirical autocorrelation function (see section 2) for lags 0-20. The same method was used for the FGN case as well (symbolically, FGN/A). Fitting by means of the Husrt coefficient was also performed as an alternative for the FGN case (symbolically, FGN/H). The results are presented in graphical form in terms of autocorrelation functions and power spectra in Figure A1 through Figure A5.

**AR(1) example.** 100 data values were generated using a Gaussian AR(1) process with unit variance and lag-one autocorrelation coefficient equal to 0.5. In this case, the theoretical autocovariance function has the form (8). The fitted parameters for the GAS case are $\beta = 0.01$

(very close to the theoretical value 0) and $\kappa = 0.66$. The resulting autocorrelation function and power spectrum are almost identical to the theoretical ones. Apparently, the FGN model is not appropriate for this case since we know that the process is not long memory at all. Had we only the data record available, without knowing the theoretical autocorrelation, we possibly attempt to fit the FGN model. Then, applying the least squares method (FGN/A) we would find $\beta = 1.31$ ($H = 0.62$), which clearly underestimates the autocorrelation for small lags and overestimates it for large lags. Applying the Hurst coefficient method we would find $H = 0.44$ which would interpret as $H = 0.5$ (values smaller than 0.5 are not allowed by FGN) and we would assume that the process is white noise, which is not correct.

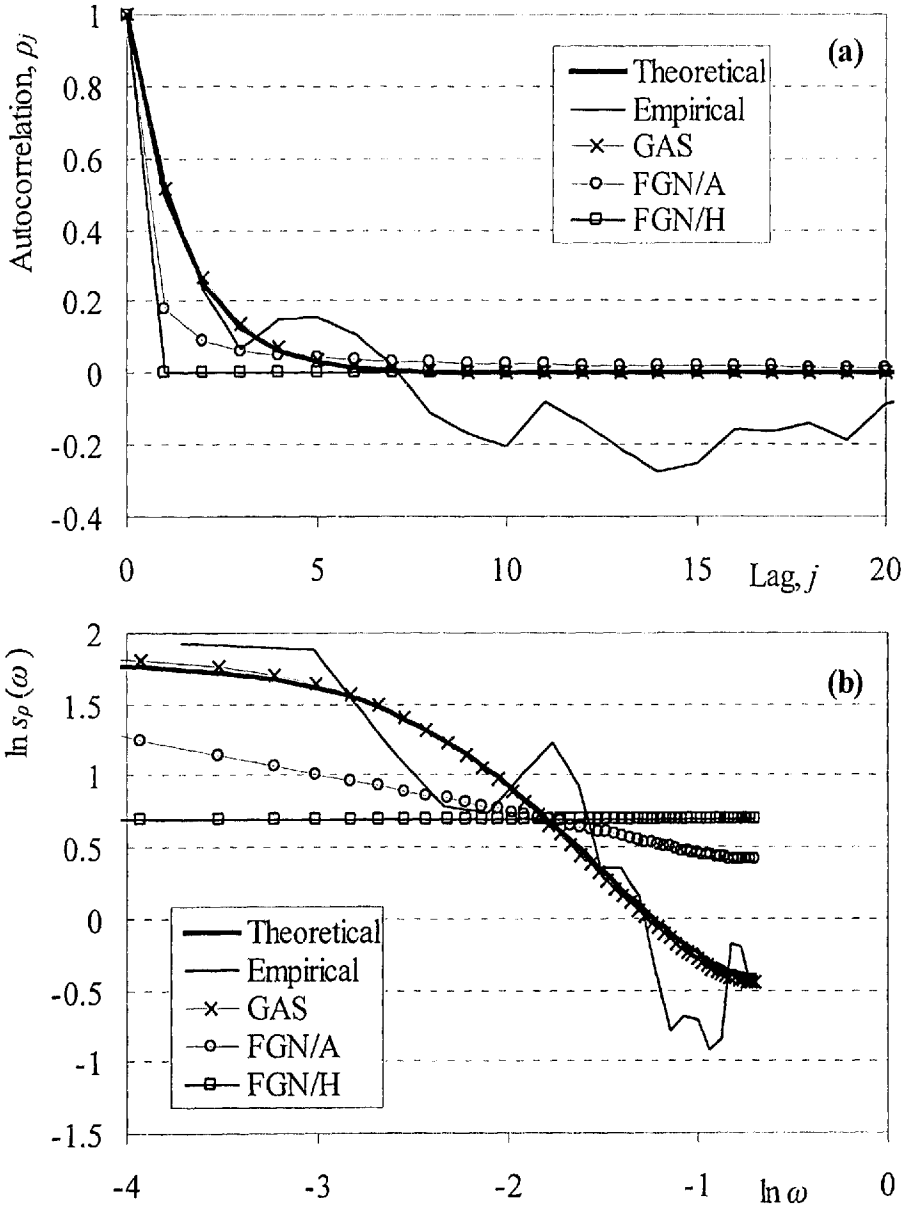



**Figure A1** Comparison of theoretical, empirical and fitted model autocorrelation functions (a) and power spectra (b) for a synthetic data set generated by an AR(1) process with 100 values.

**FGN example.** 100 data values were picked from the synthetic record generated in the application of section 6 (location 2). In this case, the theoretical autocovariance function has the form (5) with $H = 0.7$. The fitted parameters for the GAS case are $\beta = 1.48$ (>1, close to the theoretical value $\beta = 1 / [2(1 - H)] = 1.67$) and $\kappa = 3.28$. The resulting autocorrelation function and power spectrum are almost identical to the theoretical ones. The fitted parameter for the FGN/A case is $\beta = 1.67$ ($H = 0.70$), which is identical to the theoretical ones. However, applying the Hurst coefficient method we find $H = 0.98$, which is too high and results in autocorrelation function and power spectrum extremely departing from both the theoretical and empirical ones.

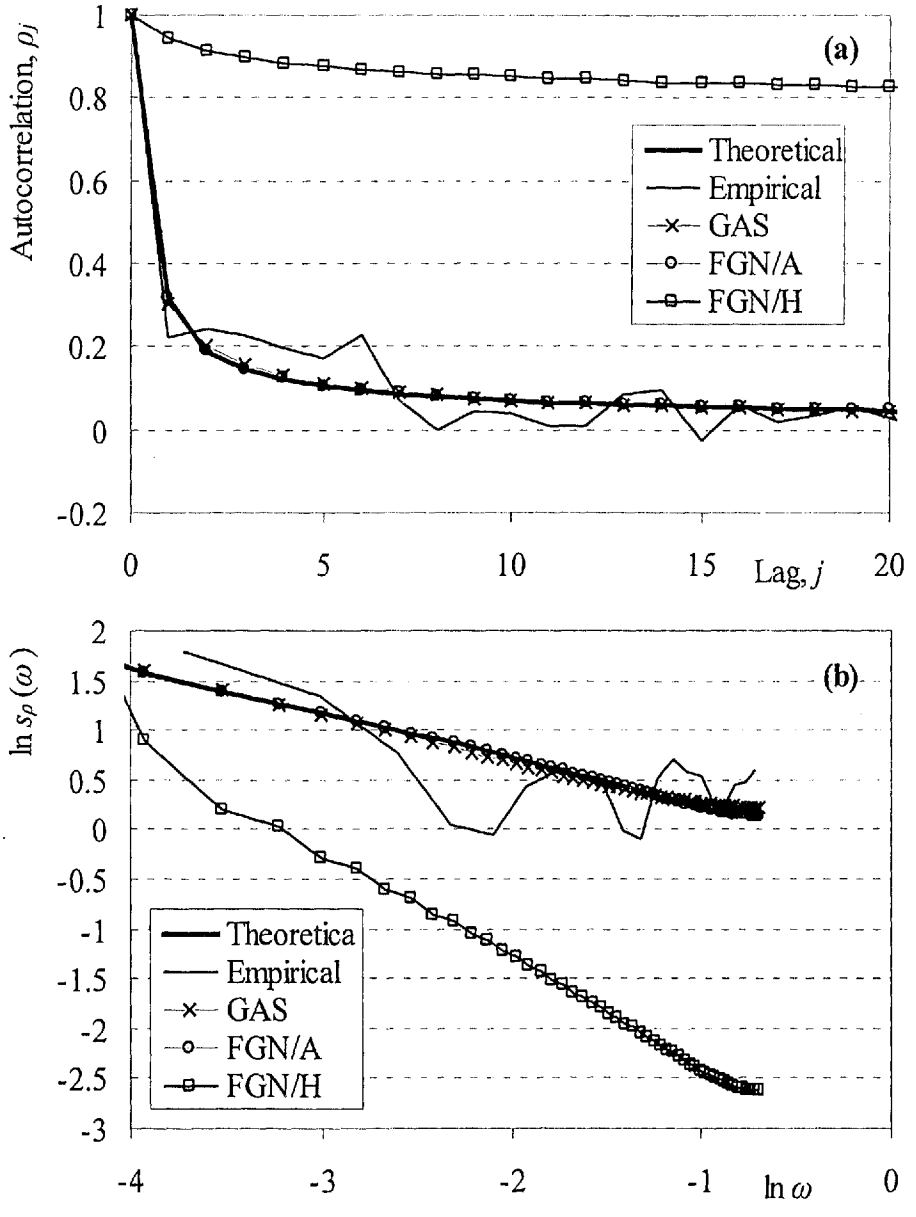



**Figure A2** Comparison of theoretical, empirical and fitted model autocorrelation functions (a) and power spectra (b) for a synthetic data set generated by a FGN process with 100 values.

**Kremasta streamflow example.** In this example we used a 44-year annual streamflow record of Acheloos River at Kremasta dam (Western Greece; overyear annual discharge 117.9 $m^3/s$). The estimated Hurst coefficient of the series is as high as 0.94, indicating a long memory. The fitted parameters for the GAS case are $\beta = 2.43$ and $\kappa = 12.6$, and, indeed, indicate a very long memory. Note that for empirical lag-one autocorrelation $\rho_1 = 0.22$, the characteristic parameter $\beta^*$ defined in section 2 is 1.54 and thus $\beta > \beta^*$. The fitted parameter for the FGN/A case is $\beta = 1.75$ ($H = 0.71 < 0.94$). Both GAS and FGN/A schemes agree well with the empirical autocorrelations and power spectra, the former outperforming the latter as better approaching the lag-one autocorrelation, which is important. The FGN/H scheme again results in autocorrelation function and power spectrum extremely departing from the empirical ones.

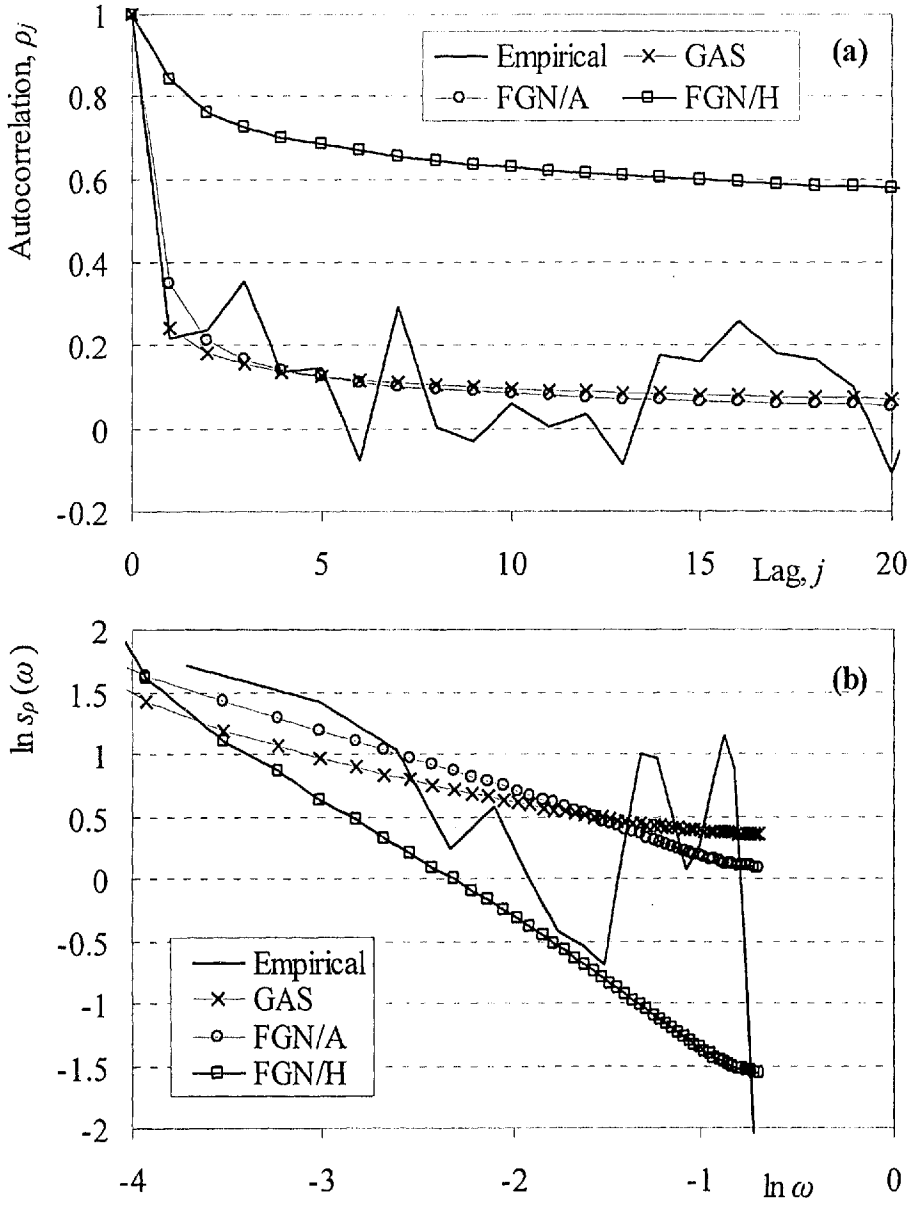



**Figure A3** Comparison of empirical and fitted model autocorrelation functions (a) and power spectra (b) for the 44-year annual streamflow record of Acheloos River at Kremasta (Western Greece).

**Coshocton runoff example.** In this example we used a 56-year annual runoff record at Coshocton, Ohio (for a catchment of 303 acres; overyear annual runoff 397.4 mm). The estimated Hurst coefficient of the series is as high as 0.89, indicating a long memory. However, the fitted parameters for the GAS case are $\beta = 0.16$ (< 1) and $\kappa = 0.90$, which do not correspond to very long memory). The fitted parameter for the FGN/A case is $\beta = 1.39$ ($H = 0.64 < 0.89$). Here the GAS scheme agrees well with the empirical autocorrelations and power spectra. The FGN/A scheme underestimates significantly the lag-one autocorrelation coefficient, whereas the FGN/H scheme again overestimates all autocorrelation coefficients.

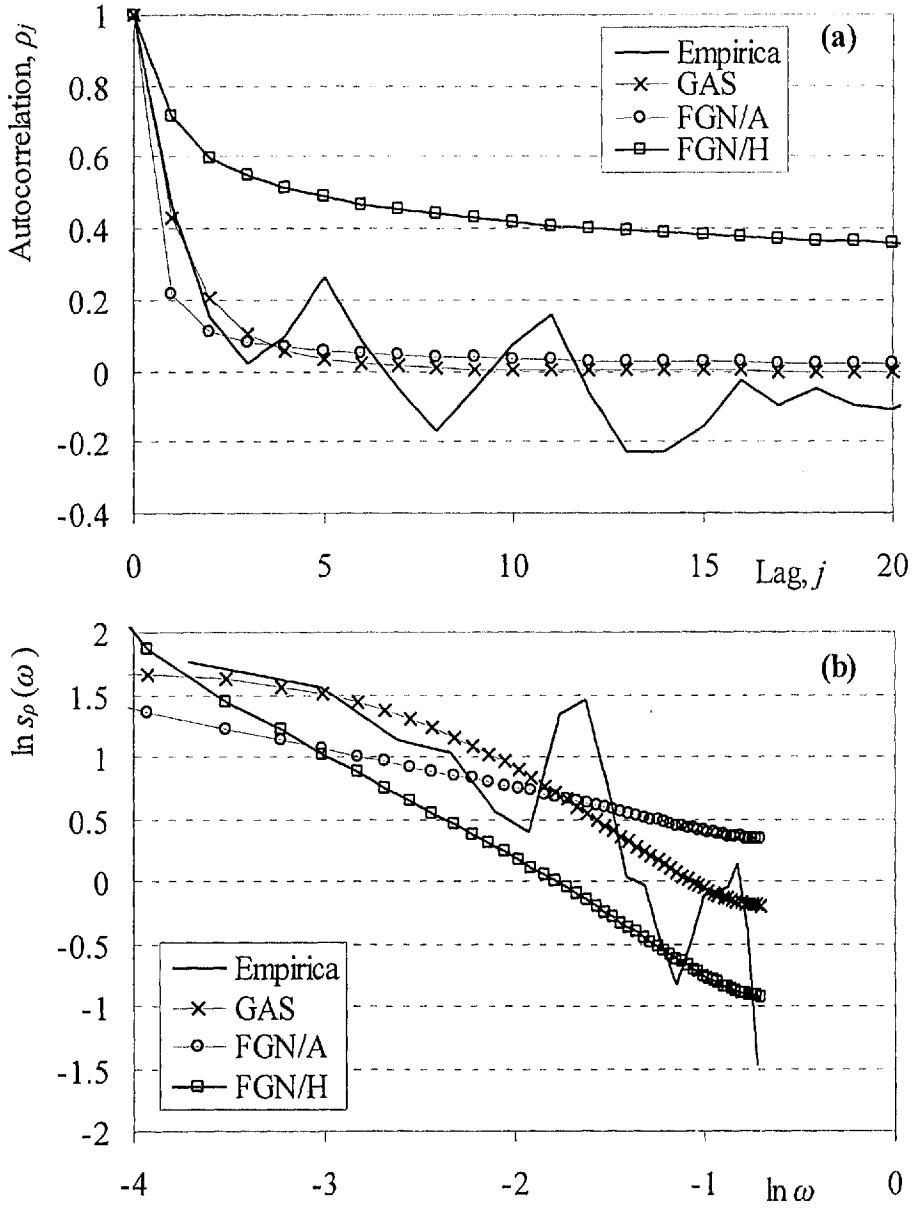



**Figure A4** Comparison of empirical and fitted model autocorrelation functions (a) and power spectra (b) for the 56-year annual streamflow record of Coshocton, Ohio.

**Aliartos rainfall example.** In this example we used an 86-year annual rainfall record at Aliartos, Eastern Greece (overyear annual rainfall 660.2 mm). The estimated Hurst coefficient of the series is as high as 0.93, indicating a long memory, and corresponding to lag-one autocorrelation coefficient equal to 0.82, although the empirical value of the latter is only 0.12. The fitted parameters for the GAS case are $\beta = 3.75$ (corresponding to $H = 0.87$, close to 0.93) and $\kappa = 300$. As in the Kremasta example, $\beta > \beta^* = 1.27$ (for $\rho_1 = 0.12$). Indeed, these parameters indicate long memory and simultaneously result in low autocorrelation for small lags (e.g. 0.15 for lag-one, which agrees well with the empirical value). Generally, the GAS scheme agrees well with the empirical autocorrelations and power spectra. The fitted parameter for the FGN/A case is $\beta = 1.53$ ($H = 0.67 < 0.93$). The FGN/A scheme overestimates the lag-one autocorrelation coefficient (0.27 versus 0.12), whereas the FGN/H scheme again overestimates all autocorrelation coefficients.

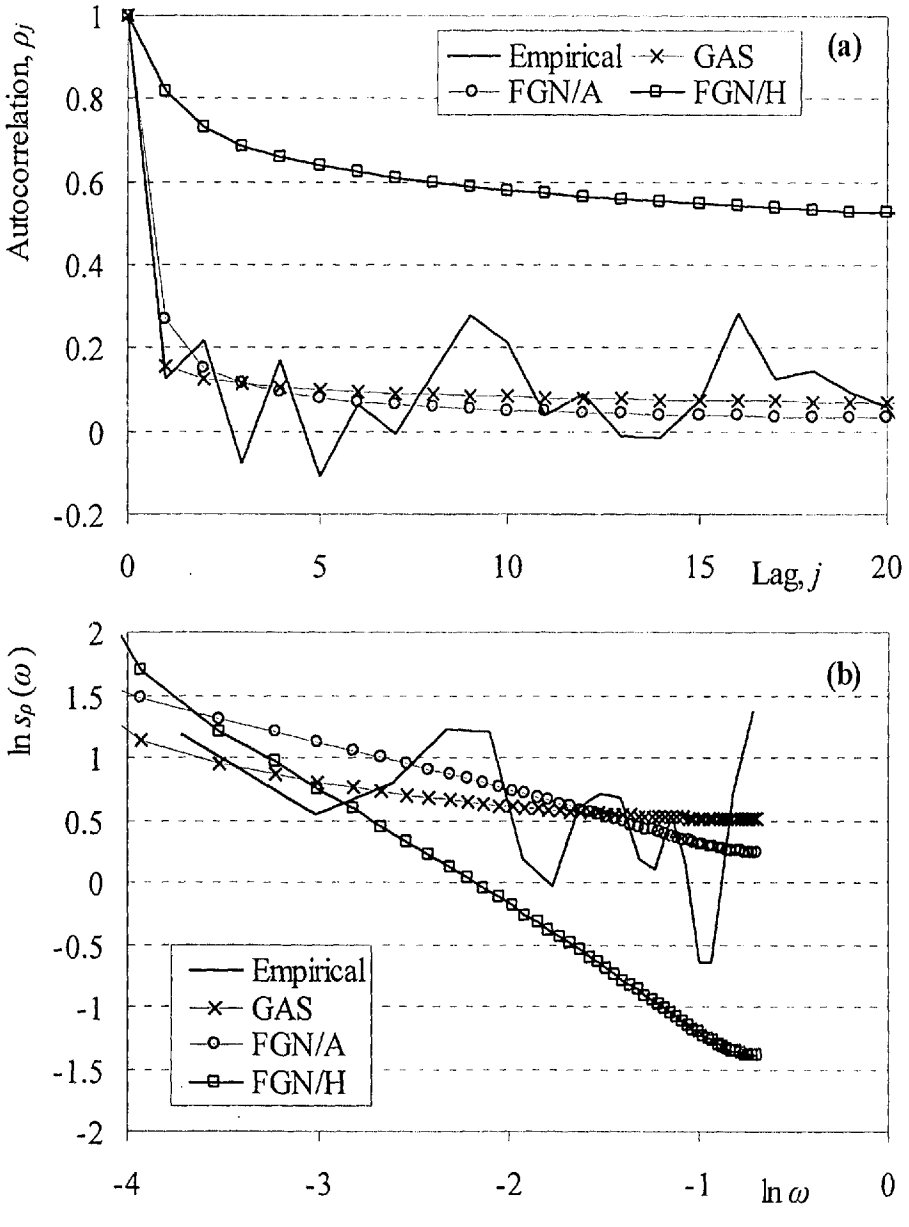



**Figure A5** Comparison of empirical and fitted model autocorrelation functions (a) and power spectra (b) for the 86-year annual rainfall record at Aliartos (Eastern Greece).

**Conclusion of Appendix 1.** From the examples with synthetic data, where the actual (theoretical) autocorrelation function and power spectrum are known, we may conclude that the GAS scheme is appropriate for both short and long memory processes, and the fitting method of least squares over the autocorrelation function results in reasonable fits, which are almost identical to the theoretical autocorrelation functions. The FGN scheme performs well if the underlying process is long memory and the scheme is fitted by the least squares method, but fails to resemble the actual process either if it is short memory or the fitting is done using the Hurst coefficient.

In all three examples with historic hydrologic data, a long memory structure emerges, as indicated by the high Hurst coefficients. However, the FGN scheme fitted in terms of the Hurst coefficient departs significantly from the empirical autocorrelation functions and power spectra. Better is its behavior if fitted by the least squares method. The GAS scheme fitted by the least squares method outperformed the FGN scheme in all cases. Interestingly, these three examples reveal that the two cases theoretically foreseen by the GAS scheme, but not by the ARMA of FGN schemes, may exist in reality. Thus, in the examples presented we have the cases: (a) $\beta < 1$ (Coshocton example) that indicates not too strong long-term persistence, and (b) $\beta > \beta^*$ (Kremasta and Aliartos examples) that indicates strong long-term persistence and simultaneously not too strong autocorrelation for small lags.

We must emphasize that the examples are presented here just to give some initial indications of the performance of the proposed generalized autocovariance function, also in comparison to that of the fractional Gaussian noise model. Before drawing final conclusions, more statistical research is needed about the model fitting method and more hydrological data sets must be investigated.

**Acknowledgments of Appendix 1.** The Coshocton runoff record is made available on the Internet by the Agricultural Research Service (ARS) research organization. The Kremasta streamflow record was compiled by data of the Public Power Corporation of Greece and the Aliartos rainfall record was compiled by data of the Hellenic National Meteorological Service raingauge and the earlier Kopais Organization raingauge operated at the same location. Both the Kremasta and Aliartos records were published in the reports of the project Evaluation and Management of the Water Resources of Sterea Hellas commissioned by the Greek Ministry of Environment, Regional Planning and Public Works to the National Technical University of Athens.

## A2 Closed solution of the internal parameter series

The complex form of the inverse finite Fourier transform of the series $\gamma_j$ is (in accordance to (13))

$$s_\gamma(\omega) = 2 \sum_{j=-\infty}^{\infty} \gamma_j \exp(2\, i\, \pi\, j\, \omega) \tag{A1}$$

where $i := \sqrt{-1}$. Using the BFMA model that incorporates as special cases both the BMA and the SMA models, and substituting $\gamma_j$ from (23) in (A1) we get

$$s_\gamma(\omega) = 2 \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_l\, a_{j+l} \exp(2\, i\, \pi\, j\, \omega) \tag{A2}$$

Interchanging the summations and setting $n = j + l$ we have

$$s_\gamma(\omega) = 2 \sum_{l=-\infty}^{\infty} a_l \sum_{j=-\infty}^{\infty} a_{j+l} \exp(2\, i\, \pi\, j\, \omega) = 2 \sum_{l=-\infty}^{\infty} a_l \sum_{n=-\infty}^{\infty} a_n \exp[2\, i\, \pi\, (n-l)\, \omega] \tag{A3}$$

or

$$s_\gamma(\omega) = 2 \sum_{l=-\infty}^{\infty} a_l \exp(-2\, i\, \pi\, l\, \omega) \sum_{n=-\infty}^{\infty} a_n \exp(2\, i\, \pi\, n\, \omega) \tag{A4}$$

which results in

$$s_\gamma(\omega) = 2\, s_a^*(\omega)\, s_a(\omega) = 2\, |s_a(\omega)|^2 \tag{A5}$$

where $s_a^*(\omega)$ is the complex conjugate of $s_a(\omega)$.

(A5) shows that (33) holds for any arrangement of the series of $a_j$ and consequently it holds for the BMA model as well. In case of the SMA model, since $a_j = a_{-j}$, the imaginary (sine) terms in its inverse finite Fourier transform vanish, so that $s_a(\omega)$ is a real function of $\omega$. Therefore, (A5) becomes

$$s_\gamma(\omega) = 2\, [s_a(\omega)]^2 \tag{A6}$$

which proves (31).

To show that there does not exist any other real valued transformation, different from DFT, that could result in an equation similar to (31) to enable a direct calculation of $a_j$ for the BMA model we use a counterexample. Specifically, we consider the simple autocovariance structure with all terms zero apart from the first two $\gamma_0$ and $\gamma_1$. This autocovariance is positive

definite if $2\,|\gamma_1| < \gamma_0$. From (19) we can verify that a solution for the series of $a_j$ is that with all terms zero apart from the first two $a_0$ and $a_1$, which are given by

$$a_0 = (1/2)\,(\sqrt{\gamma_0 + 2\,\gamma_1} + \sqrt{\gamma_0 - 2\,\gamma_1}),\quad a_1 = (1/2)\,(\sqrt{\gamma_0 + 2\,\gamma_1} - \sqrt{\gamma_0 - 2\,\gamma_1}) \tag{A7}$$

and satisfy

$$a_0^2 + a_1^2 = \gamma_0,\quad a_0\,a_1 = \gamma_1 \tag{A8}$$

Generally, we are seeking for some transformations $p_\gamma(\omega)$ and $p_a(\omega)$ of $\gamma_j$ and $a_j$, respectively, in the general form of (A1) but real valued, i.e.,

$$p_\gamma(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j\, g_j(\omega),\quad p_a(\omega) = \sum_{j=-\infty}^{\infty} a_j\, f_j(\omega) \tag{A9}$$

where $g_j(\omega)$ and $f_j(\omega)$ are sequences of orthogonal real functions of the real variable $\omega$, so that a relation similar to (A6) holds, i.e.,

$$p_\gamma(\omega) = [p_a(\omega)]^2 \tag{A10}$$

Note that such a relation is justified by dimensional analysis considerations, as well. The condition for orthogonality of the sequences $g_j(\omega)$ and $f_j(\omega)$ is needed because otherwise it will be not possible to invert the transformation, so that to derive $\gamma_j$ from $g_j(\omega)$ or $a_j$ from $f_j(\omega)$. The factors 2 in the right-hand sides of (A9) and (A10) are neglected for simplicity.

In our simple counterexample with two nonzero terms, the combination of (A9), (A10) and (A8) results in

$$(a_0^2 + a_1^2)\, g_0(\omega) + a_0\,a_1\, g_1(\omega) = [a_0\, f_0(\omega) + a_1\, f_1(\omega)]^2 \tag{A11}$$

or

$$(a_0^2 + a_1^2)\, g_0(\omega) + a_0\,a_1\, g_1(\omega) = a_0^2 f_0^2(\omega) + a_1^2 f_1^2(\omega) + 2\, a_0\, a_1\, f_0(\omega)\, f_1(\omega) \tag{A12}$$

From the condition that (A12) must hold for any couple of $a_0$ and $a_1$ we find that

$$f_0(\omega) = \pm f_1(\omega) = \pm\sqrt{g_0(\omega)}\,,\quad g_1(\omega) = \pm 2\, g_0(\omega) \tag{A13}$$

which violates the orthogonality assumption for both $g_j(\omega)$ and $f_j(\omega)$. Therefore, no real valued transformation with the desired properties exists.

## A3 Proof of the proposition of section 5

Firstly, we will prove that (39) preserves means and autocovariances. Taking average values in both sides of (39) we find that $E[X_i] = E[\widetilde{X}_i]$ (because by definition of $\mathbf{Z}$, $E[\mathbf{Z}] = E[\widetilde{\mathbf{Z}}]$), which proves preservation of means. Subtracting means from both sides of (39) we get

$$(X_i - E[X_i]) = (\widetilde{X}_i - E[\widetilde{X}_i]) + \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \{(\mathbf{Z} - E[\mathbf{Z}]) - (\widetilde{\mathbf{Z}} - E[\widetilde{\mathbf{Z}}])\}, \quad i = 1, 2, \ldots \tag{A14}$$

Writing (A14) for $X_j$ and then multiplying it with (A14) and taking expected values we get

$$\begin{aligned} \mathrm{Cov}[X_i, X_j] = {} & \mathrm{Cov}[\widetilde{X}_i, \widetilde{X}_j] - \mathrm{Cov}[\widetilde{X}_i, \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}] - \mathrm{Cov}[\widetilde{X}_j, \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}] \\ & + \mathrm{Cov}[\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}, \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}] + \mathrm{Cov}[\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \mathbf{Z}, \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \mathbf{Z}] \end{aligned} \tag{A15}$$

where we have omitted covariance terms among $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$ or $\widetilde{X}_i$, because $\mathbf{Z}$ is independent of $\widetilde{X}_i$ and consequently of $\widetilde{\mathbf{Z}}$. Observing that $\mathbf{h}$ is a symmetric matrix and $\mathrm{Cov}[\widetilde{\mathbf{Z}}, \widetilde{\mathbf{Z}}] = \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]$ by definition, we can write (A15) as

$$\begin{aligned} \mathrm{Cov}[X_i, X_j] = {} & \mathrm{Cov}[\widetilde{X}_i, \widetilde{X}_j] - \mathrm{Cov}[\widetilde{X}_i, \widetilde{\mathbf{Z}}] \mathbf{h}^{-1} \boldsymbol{\eta}_j - \mathrm{Cov}[\widetilde{X}_j, \widetilde{\mathbf{Z}}] \mathbf{h}^{-1} \boldsymbol{\eta}_i \\ & + 2\, \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \mathrm{Cov}[\widetilde{\mathbf{Z}}, \widetilde{\mathbf{Z}}] \mathbf{h}^{-1} \boldsymbol{\eta}_j \end{aligned} \tag{A16}$$

and using the definitions of $\boldsymbol{\eta}_i^T := \mathrm{Cov}[\widetilde{X}_i, \widetilde{\mathbf{Z}}]$ and $\mathbf{h} := \mathrm{Cov}[\widetilde{\mathbf{Z}}, \widetilde{\mathbf{Z}}]$,

$$\mathrm{Cov}[X_i, X_j] = \mathrm{Cov}[\widetilde{X}_i, \widetilde{X}_j] - \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \boldsymbol{\eta}_j - \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \boldsymbol{\eta}_i + 2\, \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \mathbf{h}\, \mathbf{h}^{-1} \boldsymbol{\eta}_j \tag{A17}$$

We note that $\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \boldsymbol{\eta}_j$ is scalar, so that $\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \boldsymbol{\eta}_j = (\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \boldsymbol{\eta}_j)^T = \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \boldsymbol{\eta}_i$. Besides, the last term of (A17) equals $2\, \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \boldsymbol{\eta}_j$. Thus (A17) is reduced to

$$\mathrm{Cov}[X_i, X_j] = \mathrm{Cov}[\widetilde{X}_i, \widetilde{X}_j] \tag{A18}$$

which proves our claim about preservation of covariances.

Secondly, we will prove (40). If we get covariances as above but conditionally on $\mathbf{Z} = \mathbf{z}$, the last term $\mathrm{Cov}[\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \mathbf{Z}, \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \mathbf{Z} \mid \mathbf{Z} = \mathbf{z}]$ of (A15) will now be zero. The other terms are not affected by the condition because of independence from $\mathbf{Z}$. Thus, setting $i = j$ and writing (A15) for $\mathbf{Z} = \mathbf{z}$, we get

$$\mathrm{Var}[X_i \mid \mathbf{Z} = \mathbf{z}] = \mathrm{Var}[\widetilde{X}_i] - 2\, \mathrm{Cov}[\widetilde{X}_i, \boldsymbol{\eta}_i^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}] + \mathrm{Cov}[\boldsymbol{\eta}_i^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}, \boldsymbol{\eta}_j^T \mathbf{h}^{-1} \widetilde{\mathbf{Z}}] \tag{A19}$$

which in a similar manner as previously takes the form

$$\mathrm{Var}[X_i \mid \mathbf{Z}=\mathbf{z}] = \mathrm{Var}[\widetilde{X}_i] - 2\,\boldsymbol{\eta}_i^T \mathbf{h}^{-1}\boldsymbol{\eta}_i + \boldsymbol{\eta}_i^T \mathbf{h}^{-1}\,\mathbf{h}\,\mathbf{h}^{-1}\boldsymbol{\eta}_i \tag{A20}$$

thus resulting in (40).

Next, we will show that $\mathrm{Var}[X_i \mid \mathbf{Z} = \mathbf{z}]$ coincides with the least mean square prediction error of $X_i$ from $\mathbf{Z}$. To this aim, we consider the linear prediction model

$$X_i = \boldsymbol{\kappa}^T \mathbf{Z} + U \tag{A21}$$

where $\boldsymbol{\kappa}$ is a vector of parameters and $U$ is a random variable whose deviation from mean represents the prediction error. We seek for the vector $\boldsymbol{\kappa}$ that minimizes $\mathrm{Var}[U]$. Taking expected values in both sides of (A21) and subtracting from (A21) we get

$$(U - E[U]) = (X_i - E[X_i]) - \boldsymbol{\kappa}^T (\mathbf{Z} - E[\mathbf{Z}]) \tag{A22}$$

so that

$$\mathrm{Var}[U] = \mathrm{Var}[X_i] - 2\,\mathrm{Cov}[X_i, \boldsymbol{\kappa}^T \mathbf{Z}] + \mathrm{Var}[\boldsymbol{\kappa}^T \mathbf{Z}] \tag{A23}$$

or equivalently,

$$\mathrm{Var}[U] = \gamma_0 - 2\,\mathrm{Cov}[X_i, \mathbf{Z}]\,\boldsymbol{\kappa} + \boldsymbol{\kappa}^T \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\,\boldsymbol{\kappa} \tag{A24}$$

Since by definition $\mathrm{Cov}[X_i, \mathbf{Z}] = \boldsymbol{\eta}_i^T$ and $\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}] = \mathbf{h}$,

$$\mathrm{Var}[U] = \gamma_0 - 2\,\boldsymbol{\eta}_i^T \boldsymbol{\kappa} + \boldsymbol{\kappa}^T \mathbf{h}\,\boldsymbol{\kappa} \tag{A25}$$

To find $\boldsymbol{\kappa}$ that minimizes $\mathrm{Var}[U]$ we take the derivative of the right-hand side of (A25) with respect to $\boldsymbol{\kappa}$ and equate it to 0. This results in

$$-2\,\boldsymbol{\eta}_i^T + 2\,\boldsymbol{\kappa}^T \mathbf{h} = \mathbf{0} \tag{A26}$$

or

$$\boldsymbol{\kappa} = \mathbf{h}^{-1}\,\boldsymbol{\eta}_i \tag{A27}$$

Substituting this result in (A25) we get

$$\mathrm{Var}[U] = \gamma_0 - \boldsymbol{\eta}_i^T \mathbf{h}^{-1}\,\boldsymbol{\eta}_i \tag{A28}$$

Thus, $\mathrm{Var}[U]$ is identical to $\mathrm{Var}[X_i \mid \mathbf{Z} = \mathbf{z}]$ given by (40).

Finally, we consider the case of application of (39) for one of the known $X_0, \ldots, X_{-k}$ in its left-hand side (i.e., for $-k \le i \le 0$). Apparently, in this case $\boldsymbol{\eta}_i$ will be equal to the $i$th column of $\mathbf{h}$. Since $\mathbf{h}\,\mathbf{h}^{-1} = \mathbf{I}$, $\boldsymbol{\eta}_i^T \mathbf{h}^{-1}$ will be equal to $i$th row of the identity matrix, i.e., a row vector with all elements zero apart from the $i$th element which will be one. Therefore, (39) becomes $X_i = \widetilde{X}_i + (X_i - \widetilde{X}_i) = X_i$, as it should. This proves that (39) remains consistent even when applied to the known present and past variables.

## A4 Proof of equation (57)

Let $\mathbf{d} := \mathbf{b}\,\mathbf{b}^T - \mathbf{h}$ so that $f(\mathbf{b}) := \|\mathbf{d}\|^2$. The $(k, l)$th element of $\mathbf{d}$ is

$$d_{kl} = \sum_{r=1}^{n} b_{kr}\, b_{lr} - c_{kl} \tag{A29}$$

so that

$$\frac{\partial d_{kl}}{\partial b_{ij}} = \sum_{r=1}^{n} b_{kr} \frac{\partial b_{lr}}{\partial b_{ij}} + \sum_{r=1}^{n} \frac{\partial b_{kr}}{\partial b_{ij}} b_{lr} \tag{A30}$$

Because $\mathbf{b}$, is symmetric, $\partial b_{lr} / \partial b_{ij}$ equals 1 if the element $b_{lr}$ coincides with $b_{ij}$ or its symmetric $b_{ji}$; otherwise equals zero. Symbolically,

$$\frac{\partial b_{lr}}{\partial b_{ij}} = \delta_{li}\,\delta_{rj} + \delta_{lj}\,\delta_{ri} - \delta_{lr}\,\delta_{ij}\,\delta_{li} \tag{A31}$$

where

$$\delta_{ij} := \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{A32}$$

Therefore,

$$\sum_{r=1}^{n} b_{kr} \frac{\partial b_{lr}}{\partial b_{ij}} = \sum_{r=1}^{n} b_{kr}\,\delta_{li}\,\delta_{rj} + \sum_{r=1}^{n} b_{kr}\,\delta_{lj}\,\delta_{ri} - \sum_{r=1}^{n} b_{kr}\,\delta_{lr}\,\delta_{ij}\,\delta_{li} \tag{A33}$$

or

$$\sum_{r=1}^{n} b_{kr} \frac{\partial b_{lr}}{\partial b_{ij}} = b_{kj}\,\delta_{li} + b_{ki}\,\delta_{lj} - b_{kl}\,\delta_{ij}\,\delta_{li} \tag{A34}$$

Likewise,

$$\sum_{r=1}^{n} b_{lr} \frac{\partial b_{kr}}{\partial b_{ij}} = b_{lj}\,\delta_{ki} + b_{li}\,\delta_{kj} - b_{lk}\,\delta_{ij}\,\delta_{ki} \tag{A35}$$

so that

$$\frac{\partial d_{kl}}{\partial b_{ij}} = (b_{kj}\,\delta_{li} + b_{ki}\,\delta_{lj} - b_{kl}\,\delta_{ij}\,\delta_{li}) + (b_{lj}\,\delta_{ki} + b_{li}\,\delta_{kj} - b_{lk}\,\delta_{ij}\,\delta_{ki}) \tag{A36}$$

The partial derivative of $\|\mathbf{d}\|^2$ with respect to $b_{ij}$ will be

$$\frac{\partial \|\mathbf{d}\|^2}{\partial b_{ij}} = \sum_{k=1}^{n} \sum_{l=1}^{n} 2\, d_{kl} \frac{\partial d_{kl}}{\partial b_{ij}}$$

$$= 2 \sum_{k=1}^{n} \sum_{l=1}^{n} d_{kl} \left(b_{kj}\, \delta_{li} + b_{ki}\, \delta_{lj} - b_{kl}\, \delta_{ij}\, \delta_{li} + b_{lj}\, \delta_{ki} + b_{li}\, \delta_{kj} - b_{lk}\, \delta_{ij}\, \delta_{ki}\right) \qquad \text{(A37)}$$

or

$$\frac{\partial \|\mathbf{d}\|^2}{\partial b_{ij}} = 2 \sum_{k=1}^{n} \left(d_{ki}\, b_{kj} + d_{kj}\, b_{ki} - d_{ki}\, b_{ki}\, \delta_{ij}\right) + 2 \sum_{l=1}^{n} \left(d_{il}\, b_{lj} + d_{jl}\, b_{li} - d_{il}\, b_{li}\, \delta_{ij}\right) \qquad \text{(A38)}$$

and, because both **d** and **b** are symmetric,

$$\frac{\partial \|\mathbf{d}\|^2}{\partial b_{ij}} = 4 \sum_{k=1}^{n} d_{ki}\, b_{kj} + 4 \sum_{k=1}^{n} d_{kj}\, b_{ki} - 4 \sum_{k=1}^{n} d_{ki}\, b_{ki}\, \delta_{ij} \qquad \text{(A39)}$$

We observe that each of the first and the second sums in the right-hand side of (A39) is the $(i, j)$th and $(j, i)$th element of the matrix $\mathbf{d}\,\mathbf{b} = \mathbf{e}$, which are equal due to symmetry. The sum in the third term equals the $(i, i)$th diagonal element of **e** if $i = j$; otherwise it is zero. This proves (57).