Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process

Hristos Tyralis and Demetris Koutsoyiannis

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

Abstract Various methods for estimating the self-similarity parameter (Hurst parameter, H) of a Hurst-Kolmogorov stochastic process (HKp) from a time series are available. Most of them rely on some asymptotic properties of processes with Hurst-Kolmogorov behaviour and only estimate the self-similarity parameter. Here we show that the estimation of the Hurst parameter affects the estimation of the standard deviation, a fact that was not given appropriate attention in the literature. We propose the Least Squares based on Variance estimator, and we investigate numerically its performance, which we compare to the Least Squares based on Standard Deviation estimator, as well as the maximum likelihood estimator after appropriate streamlining of the latter. These three estimators rely on the structure of the HKp and estimate simultaneously its Hurst parameter and standard deviation. In addition, we test the performance of the three methods for a range of sample sizes and H values, through a simulation study and we compare it with other estimators of the literature.

Key words Hurst phenomenon; Hurst-Kolmogorov behaviour; long term persistence; hydrological statistics; hydrological estimation; Hurst parameter estimators.

INTRODUCTION

Hurst (1951) discovered a behaviour of hydrological and other geophysical time series, which has become known with several names such as Hurst phenomenon, long-term persistence and long-range dependence, and has subsequently received extensive attention in the literature.

Earlier, Kolmogorov (1940), when studying turbulence, had proposed a mathematical model to describe this behaviour, which was further developed by Mandelbrot and van Ness (1968) and has been known as simple scaling stochastic model or fractional Gaussian noise (see Beran 1994; Embrechts and Maejima 2002; Palma 2007; Doukhan et al. 2003; Robinson 2003; and the references therein). Here the behaviour is referred to as the Hurst-Kolmogorov (HK) behaviour or HK (stochastic) dynamics and the stochastic model as the HK process (HKp).

A lot of studies on this kind of behaviour regarding actual data have been accomplished. To mention two of the most recent, Buette et al. (2006) studied the Irish daily wind speeds, using ARFIMA and GARMA models, whereas Zhang et al. (2009) studied the scaling properties of the hydrological series in the Yellow River basin. Of critical importance in analyzing hydrological and geophysical time series is the estimation of the strength of the HK behaviour. The parameter H, known as the Hurst or self-similarity parameter of the HKp arises naturally from the study of self-similar processes and expresses the strength of the HK behaviour. A number of estimators of H have been proposed. These are usually validated by an appeal to some aspect of self-similarity, or by an asymptotic analysis of the distributional properties of the estimator as the length of the time series converges to infinity.

Rea et al. (2009) present an extensive literature review dealing with the properties of these estimators. They also examine the properties of twelve estimators, i.e. the nine more classical estimators (aggregated variance, differencing the variance, absolute values of the aggregated series, Higuchi's method, residuals of regression, R/S method, periodogram method, modified periodogram method, Whittle estimator) discussed in Taqqu et al. (1995) plus the wavelet, GPH and Haslett-Raftery estimator. Weron (2002) discusses the properties of residuals of regression, R/S method and periodogram method. Grau-Carles (2005) also analyzes the behaviour of the residuals of regression, the R/S method and the GPH.

Besides new estimators are proposed, for example Guerrero and Smith (2005) presented a maximum likelihood based estimator, while Coeurjolly (2008) presented estimators based on convex combinations of sample quantiles of discrete variations of a sample path over a discrete grid of the interval [0, 1]. Some authors propose improvements of existing estimators. For example, Mielniczuk and Wojdyllo (2007) improve the R/S method. Other authors like Esposti et al. (2008) propose methodologies which use more than one methods simultaneously to estimate the H parameter.

Because the finite sample properties of these estimators can be quite different from their asymptotic properties, some authors have undertaken empirical comparisons of estimators of H. The nine classical estimators were discussed in some detail by Taqqu et al. (1995) who carried out an empirical study of these estimators for a single series length of 10 000 data points, 5 values of H, and 50 replications. All twelve above estimators were discussed in more detail by Rea et al. (2009) who carried out an empirical study of these estimators for series lengths between 100 and 10 000 data points in steps of 100, H values between 0.55 and 0.90 in steps of 0.05 and 1000 replications. Rea et al. (2009) also presented an extensive literature review about the same kind of empirical studies.

These studies did not include two methods. The maximum likelihood (ML) method discussed by McLeod and Hippel (1978) and McLeod et al. (2007), probably due to computational problems (Beran 1994, p. 109), and the method by Koutsoyiannis (2003), hereinafter referred to as the LSSD (Least Squares based on Standard Deviation) method, which was also articulated recently by Ehsanzadeh and Adamowski (2010). The ML method estimates the Hurst parameter based on the whole structure of the process, i.e. its joint distribution function. The LSSD method relies on the self-similarity property of the process. One common characteristic of the ML and LSSD methods is that they estimate simultaneously the Hurst parameter H and the standard deviation σ of the process. This is of great importance, because both parameters are essential for the construction of the model and, as we will show below (see also Koutsoyiannis 2003) their estimators generally are not independent to each other. In addition, the classical statistical estimator of σ encompasses strong bias if applied to a series with HK behaviour (Koutsoyiannis 2003; Koutsoyiannis and Montanari 2007). It is thus striking that some of the existing methods do not remedy or even pose this problem at all, and estimate H independently of σ and vice versa, e.g. assuming that σ can be estimated using its classical statistical estimator, which does not depend on H.

The focus of this paper is the simultaneous estimation of the parameters H and σ of the HKp. We use the ML and LSSD methods that have the capacity of simultaneous estimation, after appropriate streamlining of the former in a more practical form, and we propose a third method which is an improvement of the LSSD method (referred to as LSV method —Least Squares based on Variance) retaining the simultaneous parameter estimation attitude. We apply the three methods to evaluate their performance in a Monte Carlo simulation framework and we compare the results with those of the estimators presented in Taqqu et al. (1995) with the exception of the Whittle estimator, which we replaced by the local Whittle estimator presented in Robinson (1995).

METHODS

Definition of HKp

Let X_i denote a stochastic process with $i = 1, 2, \ldots$ denoting discrete time. We assume that there is a record of n observations which we write as a vector $\boldsymbol{x}_n = (x_1 \ldots x_n)^{\mathrm{T}}$ (where the superscript T is used to denote the transpose of a vector or matrix). We recall from statistical theory that each observation x_i represents a realization of a random variable X_i, so that $\boldsymbol{x}_n$ is a realization of a vector of identically distributed random variables $\boldsymbol{X}_n = (X_1 \ldots X_n)^{\mathrm{T}}$ (notice the upper- and lower-case symbols used for random variables and values thereof, respectively, and the arrangement of the sample members and observations from the latest to the earliest). It is assumed that the process is stationary, a property that does not hinder to exhibit multiple scale variability. Further, let its mean be denoted as μ := E[X_i], its autocovariance γ_j := Cov[X_i, X_{i+j}], its autocorrelation ρ_j := Corr[X_i, X_{i+j}] = γ_j / γ_0 (j = 0, ±1, ±2, …), and its standard deviation $\sigma := \sqrt{\gamma_0}$.

Let k be a positive integer that represents a timescale larger than 1, the original time scale of the process X_i. The mean aggregated stochastic process on that timescale is denoted as

$$X_i^{(k)} := (1/k) \sum_{l=(i-1)k+1}^{ik} X_l \quad (1)$$

The notation implies that a superscript (1) could be omitted, i.e. $X_i^{(1)} \equiv X_i$. The statistical characteristics of $X_i^{(k)}$ for any timescale k can be derived from those of X_i. For example, the mean is

$$\mathrm{E}[X_i^{(k)}] = \mu \quad (2)$$

whilst the variance and autocovariance (or autocorrelation) depend on the specific structure of γ_j (or ρ_j). Now we consider the following equation that defines the HKp:

$$(X_i^{(k)} - \mu) \stackrel{\mathrm{d}}{=} \left(\frac{k}{l}\right)^H (X_j^{(l)} - \mu), \; 0 < H < 1, \quad (3)$$

where the symbol $\stackrel{\mathrm{d}}{=}$ stands for equality in (finite dimensional joint) distribution and H is the Hurst parameter. We assume that equation (3) holds for any integer i and j (that is, the process is stationary) and any timescales k and l (≥ 1) and that X_i is Gaussian.

We also define the aggregated stochastic process for every time scale:

$$Z_i^{(k)} := \sum_{l=(i-1)k+1}^{ik} X_l = k\, X_i^{(k)} \quad (4)$$

For this process the following relationships hold:

$$\mathrm{E}[Z_i^{(k)}] = k\,\mu,\ \gamma_0^{(k)} = \mathrm{Var}[Z_i^{(k)}] = k^{2\cdot H}\,\gamma_0,\ \sigma^{(k)} = (\gamma_0^{(k)})^{1/2} \tag{5}$$

The autocorrelation function of either of $X_i^{(k)}$ and $Z_i^{(k)}$, for any aggregated timescale k, is independent of k, and given by

$$\rho_j^{(k)} = \rho_j = |j+1|^{2H}/2 + |j-1|^{2H}/2 - |j|^{2H} \tag{6}$$

Maximum likelihood estimator

In this section the method of maximum likelihood is employed for the estimation of the parameters of HKp, namely H, σ, μ. For a given record of n observations $\boldsymbol{x}_n = (x_1 \dots x_n)^{\mathrm{T}}$ the likelihood of $\boldsymbol{\theta} := (\mu, \sigma, H)$ takes the general form (McLeod and Hippel 1978):

$$p(\boldsymbol{\theta}|\boldsymbol{x}_n) = \frac{1}{(2\pi)^{n/2}}\ [\det(\sigma^2\,\boldsymbol{R})]^{-1/2}\ \exp[-1/(2\sigma^2)\ (\boldsymbol{x}_n - \mu\,\boldsymbol{e})^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,(\boldsymbol{x}_n - \mu\,\boldsymbol{e})] \tag{7}$$

where $\boldsymbol{e} = (1\ 1\ \dots\ 1)^{\mathrm{T}}$ is a column vector with n elements, $\boldsymbol{R}$ is the autocorrelation matrix, i.e., a n x n matrix with elements $r_{ij} = \rho_{|i-j|}$, and det( ) denotes the determinant of a matrix.

Then a maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{H})$, as shown in Appendix A, consists of the following relationships:

$$\hat{\mu} = \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\hat{\boldsymbol{R}}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\hat{\boldsymbol{R}}^{-1}\,\boldsymbol{e}},\quad \hat{\sigma} = \sqrt{\frac{(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e})^{\mathrm{T}}\,\hat{\boldsymbol{R}}^{-1}\,(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e})}{n}} \tag{8}$$

and $\hat{H}$ can be obtained from the maximization of the single-variable function $g_1(H)$ defined as:

$$g_1(H) := -\frac{n}{2}\ \ln\left[\left(\boldsymbol{x}_n - \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}\,\boldsymbol{e}\right)^{\mathrm{T}}\boldsymbol{R}^{-1}\left(\boldsymbol{x}_n - \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}\,\boldsymbol{e}\right)\right] - \frac{1}{2}\ln[\det(\boldsymbol{R})] \tag{9}$$

LSSD method

This method was proposed by Koutsoyiannis (2003). In his paper after a systematic Monte

Carlo study he found an estimator $\widetilde{S}$ of σ, approximately unbiased for known H and for normal distribution of X_i, where

$$\widetilde{S} := \sqrt{\frac{n - 1/2}{n - n^{2\mathrm{H} - 1}}}\, S = \sqrt{\frac{n - 1/2}{(n - 1)\,(n - n^{2\mathrm{H} - 1})}} \sqrt{\sum_{i = 1}^{n} (X_i - X_1^{(n)})^2} \qquad (10)$$

$$S^2 := \frac{1}{n - 1} \sum_{i = 1}^{n} (X_i - X_1^{(n)})^2 \qquad (11)$$

and $X_1^{(n)}$ from equation (1) equals the sample mean.

This algorithm is based on classical sample estimates $s^{(k)}$ of standard deviations $\sigma^{(k)}$ for timescales k ranging from 1 to a maximum value $k' = [n/10]$. This maximum value was chosen so that $s^{(k)}$ can be estimated from at least 10 data values.

Combining (5) and (10), assuming $\mathrm{E}[\widetilde{S}] = \sigma$ and using the self-similarity property of the process one obtains $\mathrm{E}[S^{(k)}] \approx \mathrm{c}_k(H)\, k^H\, \sigma$ with

$$\mathrm{c}_k(H) := \sqrt{\frac{n/k - (n/k)^{2H - 1}}{n/k - 1/2}} \qquad (12)$$

Then the algorithm minimizes a fitting error $e^2(\sigma, H)$:

$$e^2(\sigma, H) := \sum_{k = 1}^{k'} \frac{[\ln \mathrm{E}[S^{(k)}] - \ln s^{(k)}]^2}{k^p} + \frac{H^{q+1}}{q+1} = \sum_{k = 1}^{k'} \frac{[\ln\sigma + H \cdot \ln k + \ln \mathrm{c}_k(H) - \ln s^{(k)}]^2}{k^p} + \frac{H^{q+1}}{q+1} \qquad (13)$$

where a weight equal to $1/k^p$ is assigned to the partial error of each scale k. For $p = 0$ the weights are equal whereas for $p = 1, 2, \ldots,$ decreasing weights are assigned to increasing scales; this is reasonable because at larger scales the sample size is smaller and thus the uncertainty larger. Using Monte Carlo experiments it was found that, although differences in estimates caused by different values of p in the range 0 to 2 are not so important, $p = 2$ results in slightly more efficient estimates (i.e., with smaller variation) and thus is preferable. A penalty factor $H^{q+1}/(q+1)$ has been included in e^2 in (13) for a high q, say 50. The effect of this factor is that it excludes the value $\hat{H} = 1$ and forces $\hat{H}$ to slightly smaller values when it is close to 1. As a consequence this factor helps get rid of an infinite $\hat{\sigma}$ also forcing to smaller values for $\hat{H}$ close to 1 (see Appendix B).

An analytical procedure to locate the minimum is not possible. Therefore, minimization of $e^2(\sigma, H)$ is done numerically and several numerical procedures can be devised for this purpose. A detailed iterative procedure is given in Koutsoyiannis (2003).

LSV method

In the previous method an approximately unbiased estimator $\tilde{S}$ of σ was found after a systematic Monte Carlo simulation. However, if σ^2 is used instead of σ, we have the advantage that there exists a theoretically consistent expression, which determines $\mathrm{E}[S^2]$ as a function of σ and H. This is the basis to form a modified version of the LSSD method, the LSV method. From the general relationship (Beran 1994, p. 9)

$$\mathrm{E}[S^2] = \left(1 - \frac{\delta_n(\rho)}{n-1}\right) \sigma^2, \text{ where } \delta_n(\rho) := (1/n) \sum_{i \neq j} \rho(i, j) = 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho(k) \tag{14}$$

we easily obtain that for an HKp:

$$\mathrm{E}[S^2] = \frac{n - n^{2H-1}}{n-1} \sigma^2 \tag{15}$$

Due to the self-similarity property of the process the following relationship holds:

$$\mathrm{E}[S^{2(k)}] = \frac{(n/k) - (n/k)^{2H-1}}{(n/k) - 1} \gamma_0^{(k)} = \frac{(n/k) - (n/k)^{2H-1}}{(n/k) - 1} k^{2H} \sigma^2 = \mathrm{c}_k(H)\, k^{2H} \sigma^2 \tag{16}$$

where

$$c_k(H) := \frac{(n/k) - (n/k)^{2H-1}}{(n/k) - 1} \text{ and } S^{2(k)} = \frac{1}{n/k - 1} \sum_{i=1}^{n/k} (Z_i^{(k)} - k\,X_1^{(n)})^2. \tag{17}$$

Thus, the following error function should be minimized in order to obtain an estimation of H and σ:

$$e^2(\sigma, H) := \sum_{k=1}^{k'} \frac{[\mathrm{E}[S^{2(k)}] - s^{2(k)}]^2}{k^p} = \sum_{k=1}^{k'} \frac{[c_k(H)\, k^{2H} \sigma^2 - s^{2(k)}]^2}{k^p},\ k' = [n/10] \tag{18}$$

Taking partial derivatives, i.e.,

$$\frac{\partial e^2(\sigma, H)}{\partial \sigma^2} = 2\,[\sigma^2\, \alpha_{11}(H) - \alpha_{12}(H)] \tag{19}$$

where:

$$\alpha_{11}(H) := \sum_{k=1}^{k'} \frac{c_k^2(H)\, k^{4H}}{k^p},\ \ \alpha_{12}(H) := \sum_{k=1}^{k'} \frac{c_k(H)\, k^{2H}\, s^{2(k)}}{k^p} \tag{20}$$

and equating to zero we obtain an estimate of σ:

$$\hat{\sigma} = \sqrt{\alpha_{12}(\hat{H})/\alpha_{11}(\hat{H})} \tag{21}$$

An estimate of H can be obtained by minimizing the single-variable function:

$$g_2(H) := \sum_{k=1}^{k'} \frac{s^{4(k)}}{k^p} - \frac{\alpha_{12}^2(H)}{\alpha_{11}(H)},\ 0 < H < 1 \tag{22}$$

We prove in Appendix B that $e^2(\sigma, H)$ attains its minimum for $H \leq 1$. However, when $\hat{H}$ = 1, then from equations (21) and (31) we obtain that $\hat{\sigma} = \infty$. Accordingly, to avoid such behaviour (values of $\hat{\sigma}$ tending to infinity), a penalty factor $H^{q+1}/(q+1)$ for a high q is added again, as in method LSSD, to the error function.

So the function to be minimized becomes:

$$e^2(\sigma, H) := \sum_{k=1}^{k'} \frac{[c_k(H)\, k^{2H}\, \sigma^2 - s^{2(k)}]^2}{k^p} + \frac{H^{q+1}}{q+1} \tag{23}$$

An estimate of H can be obtained by the minimization of the single-variable function:

$$g_2(H) := \sum_{k=1}^{k'} \frac{s^{4(k)}}{k^p} - \frac{\alpha_{12}^2(H)}{\alpha_{11}(H)} + \frac{H^{q+1}}{q+1},\ 0 < H < 1 \tag{24}$$

and σ is again estimated from (21).

RESULTS

The three H and σ estimators, namely ML, LSSD and LSV are implemented in the popular computational software Matlab. We evaluated each estimator's performance in estimating H and σ for simulated HKp. HKp series were generated using the Matlab central file exchange function ffgn, written jointly by Yingchun Zhou (Jasmine) and Stilian Stoev (http://www.mathworks.com/matlabcentral/fx_files/19797/1/ffgn.m). This function generates "exact" paths of HKp by using circulant embedding. We ran 200 replications of simulated HKp series with eight different lengths and five different H values. The lengths were 64, 128, 256, 512, 1 024, 2 048, 4 096 and 8 192 data points. The H values were 0.60, 0.70, 0.80, 0.90 and 0.95. Without loss of generality, in all cases the true (population) value of σ was assumed 1.00.

For each series, H and σ were estimated by each of these three estimators. For each H value and series length we estimated from the simulated data the median, 75% and 95% confidence intervals and the square root of the mean square error (Taqqu et al. 1995)

$$\text{RMSE} := \sqrt{\frac{1}{K} \sum_{k=1}^{K} (H_k - H)^2} \tag{25}$$

where K = 200 the number of replications. The H or σ estimates were sorted into ascending order and the median obtained by averaging the 100th and 101st values. Similar calculations were done for the upper and lower values of the 75% and 95% confidence intervals.

Figures 1-5 depict some of the results in graphical form. (To present the results in tabular form would require a very large amount of space). In Figs. 1-3 the vertical axis ranges between -0.3 and 0.2 for ΔH (the estimated H minus the true H) to facilitate comparisons among the estimators' standard deviation of their estimates. Figure 4 shows the RMSE as a function of the series length. Again all vertical axes have the same range to facilitate comparisons. Figure 5 presents RMSE as a function of series length. Figures 1-5 also depict corresponding results for the σ estimators.

The results for the ML method are shown in Fig.1. The ML method is unbiased for H at all series lengths when true H = 0.6, but becomes biased and underestimated H when H increases, for low length of time series. This method is unbiased for σ at all series lengths when true H = 0.6 but becomes biased and underestimates σ when H increases. But even for values of H over 0.9, the method becomes unbiased when the time series length increases.

The results for the LSSD method are presented in Fig. 2. The LSSD method was unbiased for H and σ at all series lengths when true $H \leq 0.9$, but became biased and underestimated H and σ when true H = 0.95. We observed the same results for the LSV method (Fig. 3), but this method was slightly worse compared with the previous method. The 75% confidence intervals all contain the true values, except when true H = 0.95 and the LSSD or LSV method is used to estimate H or σ.

Figure 4 compares the RMSE of all three methods. We observe that when estimating H the ML method is best, followed by the LSV method, for all values of H. The same holds when estimating σ, except that the LSSD method behaves better than the LSV method.

Fig. 5 presents the variation of RMSE when H increases. We observe that when estimating H the RMSE increases when H increases for the LSSD and LSV methods but it remains stable for the maximum likelihood method. However, when estimating σ, the RMSE increases for increasing H in all methods.

Figure 6 presents the correlation between $\hat{H}$ and $\hat{\mu}$, for nominal $H = 0.8$. $\hat{H}$ does not seem to affect $\hat{\mu}$, in terms of bias and this holds for every time series length.

Figure 7 presents the correlation between $\hat{H}$ and $\hat{\sigma}$, for nominal $H = 0.6$ and 0.8. It seems that an increase of nominal H results to an increase of the correlation between $\hat{H}$ and $\hat{\sigma}$. We can see that a high $\hat{H}$ results in a high $\hat{\sigma}$, and a low $\hat{H}$ results in a low $\hat{\sigma}$.

A proof of the kind of dependence between the maximum likelihood estimates of the parameters could be given by the use of the Fisher Information Matrix $\boldsymbol{I}(\boldsymbol{\theta})$ with elements $\boldsymbol{I}_{ij}(\boldsymbol{\theta}) := -\,\mathrm{E}[\frac{\partial^2 \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial\theta_i\,\partial\theta_j}]$, where $\boldsymbol{\theta} := (\theta_1, \theta_2, \theta_3) \equiv (\mu, \sigma, H)$. We easily calculate $\boldsymbol{I}_{12}(\boldsymbol{\theta}) = \boldsymbol{I}_{13}(\boldsymbol{\theta}) = 0$ and $\boldsymbol{I}_{23}(\boldsymbol{\theta}) = (1/\sigma)\ \mathrm{Tr}(\boldsymbol{R}^{-1} \frac{\partial \boldsymbol{R}}{\partial H}) \neq 0$ (see Appendix C). Thus $\hat{\mu}$ and $\hat{H}$ are orthogonal and so are $\hat{\mu}$ and $\hat{\sigma}$, but not $\hat{\sigma}$ and $\hat{H}$.

Figure 8 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus q. An increase of q results to a decrease of bias when estimating H and an increase in the corresponding variance. The minimum bias when estimating σ, is achieved for values of q around 50 depending on the actual values of H, but there is also an increase in the corresponding variance when q increases as expected from equations (21) and (31). It should be noted that a change of q does not influence the estimates when H is low, because $H^{q+1} / (q+1)$ is negligible for values of H near 0.5.

Figure 9 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus p. There is a range of p between 5 and 6, where we achieve minimum bias when estimating H or σ, but the corresponding variance decreases when p increases. We also note the irregularity between the graphs, caused by the presence of q, which gives smaller standard deviation of estimator for a high $H = 0.95$ rather than smaller H (e.g. $H = 0.90$).

Figure 10 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus $m := n/k'$. We observe that up to a value of $m = 10 \approx 1024/100$ the results remain the same, while for values of m more than 10 there is a higher bias and lower variance.

Finally we can see from Tables 1 and 2 that these three methods perform better than the eight methods discussed in Taqqu et al. (1995) and the local Whittle estimator discussed in Robinson (1995).

CONCLUSIONS

It is clear from the simulations that the three estimators (ML, LSSD and LSV) are not equivalent, when compared to each other. Compared to other estimators of the literature, when estimating H, they seem to be more accurate and have a low error. This holds, because they have lower variance for large time series length and the other estimators rely on some asymptotic properties, whereas these estimators rely mostly on the structure of the HKp.

An additional advantage of these three estimators is that, in addition to H, they estimate σ which is essential for the model. As seen in Fig. 7, and also proved in section "Results", $\hat{H}$ and $\hat{\sigma}$ are correlated and thus their maximum likelihood estimators cannot be calculated separately. Cox and Reid (1987) outline a number of statistical consequences of orthogonality. They state that the maximum likelihood estimate of H or σ when μ is given varies only slowly with μ. But this is not the case when examining σ versus H. As a consequence a non simultaneous estimator of σ and H may be suboptimal in terms of robustness comparing to the ML, LSSD or LSV estimators which estimate H and σ simultaneously. From a more practical point of view, the importance of accounting for the dependence of the estimators, could be understood from the numerous publications that calculate the standard deviation by the classical statistical estimator while at the same time find an $H > 0.5$, and sometimes very close to 1. Apparently, such estimates of standard deviation are heavily biased and this is usually missed to note.

There are some problems with the choice of q or p in LSSD and LSV estimators. When choosing a large q we benefit from the fact that it decreases the variance of the σ estimator, but it causes an irregularity for high values of H, that cannot be controlled a priori. However we believe that the benefits from the presence of q are superior to the losses induced from its use, especially given that its presence does not affect the estimators for low values of H. For the choice of p the conflicting criteria of minimum bias and minimum variance of estimator should be considered. As a consequence, an a priori choice of p and q has a degree of subjectivity. In this study we chose $p = 6$ for LSV, $p = 2$ for LSSD and $q = 50$ for both methods, and the results were rather satisfactory. Additionally we chose $m = 10$, although Figure 10 allows to use lower m values. A choice of m below 10 does not influence the results.

Another strong point of these three estimators is that they are easy to understand, again because they rely on the structure of the HKp. They also enable some interesting theoretical analyses such as those presented here, namely the bracketing of H and the behaviour of the estimator for high values of H.

There is a problem with the implementation of the ML estimator, because it needs large computational times for large time series lengths (e.g. many thousands of data values). But in hydrology the available time series are usually short. Thus, we think that its use is preferable, when an estimation of the HKp parameters is required. When the time series length increases we can switch to the LSV or the LSSD method. Among the three estimators, the ML estimator is better when estimating H, followed by the LSV method. But when estimating σ the LSSD method is superior to the LSV method.

APPENDIX A: Proof of equations (8) and (9)

From equation (7) we obtain:

$$p(\boldsymbol{\theta}|\boldsymbol{x}_n) = \frac{1}{(2\pi)^{n/2}}\,\frac{1}{\sigma^n}\,[\det(\boldsymbol{R})]^{-1/2}\,\exp[-\frac{1}{2\,\sigma^2}\,(\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}\,(\mu - \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}})^2 +$$

$$+\frac{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}\,\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{x}_n - (\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e})^2}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}})] \tag{26}$$

Since $\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e} > 0$ ($\boldsymbol{R}$ is positive definite matrix) the maximum of $p(\boldsymbol{\theta}|\boldsymbol{x}_n)$ is achieved when

$$\hat{\mu} = \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}} \tag{27}$$

For that value of μ, taking the logarithm of the posterior density we obtain:

$$\ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)] = -\,(n/2)\,\ln(2\pi) - n\,\ln\sigma - (1/2)\,\ln[\det(\boldsymbol{R})] - \frac{1}{2\,\sigma^2}\,(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e})^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e}) \tag{28}$$

$$\frac{\partial \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\,(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e})^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,(\boldsymbol{x}_n - \hat{\mu}\,\boldsymbol{e}) \tag{29}$$

Thus, the logarithm of the maximum posterior density is maximized when $\partial\ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]/\partial\sigma = 0$. The solution of this equation proves equation (8) and gives the ML estimator of σ.

Substituting the values of μ and σ from equation (8), we obtain:

$$\ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)] = \frac{n}{2}\,\ln(\frac{n}{2\pi}) - \frac{n}{2} - \frac{n}{2}\,\ln[(\boldsymbol{x}_n - \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}\,\boldsymbol{e})^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,(\boldsymbol{x}_n - \frac{\boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{e}}\,\boldsymbol{e})] -$$

$$-\frac{1}{2}\cdot\ln[\det(\boldsymbol{R})] = \frac{n}{2}\,\ln(\frac{n}{2\pi}) - \frac{n}{2} + g_1(H) \tag{30}$$

which is a function of H through the matrix $\boldsymbol{R}$. So we maximize the above single-variable function, or equivalently the function $g_1(H)$, and find $\hat{H}$.

We may observe that it is not necessary to form the entire matrix $\boldsymbol{R}$ and invert it to compute $g_1(H)$ (It suffices to form a column $(\rho_0 \ \dots \ \rho_{n-1})^{\mathrm{T}}$). Since $\boldsymbol{R}$ is a positive definite Toeplitz matrix we can use the Levinson-Trench-Zohar algorithm as described in Musicus (1988). This algorithm can solve the problem of calculating $\boldsymbol{R}^{-1}\,\boldsymbol{e}$ and ln[det($\boldsymbol{R}$)] using only $O(n^2)$ operations and $O(n)$ storage. In contrast, standard methods such as Gaussian elimination or Choleski decomposition generally require $O(n^3)$ operations and $O(n^2)$ storage. This is of critical importance when the time series size is large and computer memory capacity restricts its ability to solve the problem.

APPENDIX B: Proof of the bracketing of the *H* in (0, 1] in the LSV solution

In order to examine the behaviour of $\hat{\sigma}$ and $g_2(H)$ from equations (21) and (22) we calculate the following limits:

$$\lim_{H\to 1} \frac{\alpha_{12}^2(H)}{\alpha_{11}(H)} = \left[\sum_{k=1}^{k'} \frac{\ln(n/k)\, k^2\, s^{2(k)}}{k^p} \right]^2 \Big/ \sum_{k=1}^{k'} \frac{\ln(n/k)\, k^2}{k^p} > 0 \text{ and } \lim_{H\to 1} \frac{\alpha_{12}(H)}{\alpha_{11}(H)} = \infty \qquad (31)$$

Therefore, there is a possibility that $g_2(H)$ could have a minimum for $H = 1$ and $\sigma = \infty$, when σ tends to infinity from this path: $\sigma = \sqrt{\alpha_{12}(H)/\alpha_{11}(H)}$.

Then $\lim_{H\to 1} g_2(H) = \sum_{k=1}^{k'} \frac{s^{4(k)}}{k^p} - \left(\sum_{k=1}^{k'} \frac{\ln(n/k)\, k^2\, s^{2(k)}}{k^p}\right)^2 \Big/ \left(\sum_{k=1}^{k'} \frac{\ln(n/k)\, k^2}{k^p}\right).$

Now we prove $e^2(\sigma, H)$ attains its minimum for $H \leq 1$. The proof is given bellow:

Suppose that $H_2 > 1$ and $\sigma_2 > 0$ (It's easy to prove that an estimated $\hat{\sigma} > 0$ always). Now for any $H_1 \in (0, 1)$ we can always find a $\sigma_1 > 0$, such that $c_k(H_1)\, k^{2H_1}\, \sigma_1^2 - s^{2(k)} < 0$ for every k. For these values of H_1 and σ_1: $|\, c_k(H_1)\, k^{2\cdot H_1}\, \sigma_1^2 - s^{2(k)} \,| < |\, c_k(H_2)\, k^{2\cdot H_2}\, \sigma_2^2 - s^{2(k)} \,|$ for every k. This proves that $e^2(\sigma_1, H_1) < e^2(\sigma_2, H_2)$. Thus, $e^2(\sigma, H)$ attains its minimum for $H \leq 1$.

APPENDIX C: Calculation of Fisher Information Matrix's elements

We can easily calculate the $\boldsymbol{I}_{12}(\boldsymbol{\theta})$, $\boldsymbol{I}_{13}(\boldsymbol{\theta})$ and $\boldsymbol{I}_{23}(\boldsymbol{\theta})$ elements of the Fisher Information Matrix (Robert 2007, p. 129):

$$\frac{\partial \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial \mu} = -\frac{1}{\sigma^2}(\boldsymbol{e}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{e}\, \mu - \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{e}) \tag{32}$$

$$\frac{\partial \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(\boldsymbol{x}_n - \mu\, \boldsymbol{e})^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{x}_n - \mu\, \boldsymbol{e}) \tag{33}$$

$$\frac{\partial^2 \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial \mu\, \partial \sigma} = \frac{2}{\sigma^3}(\boldsymbol{e}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{e}\, \mu - \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{e}) \tag{34}$$

$$\frac{\partial^2 \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial \mu\, \partial H} = \frac{1}{\sigma^2}\left(\mu\, \boldsymbol{e}^{\mathrm{T}} \boldsymbol{R}^{-1} \frac{\partial \boldsymbol{R}}{\partial H} \boldsymbol{R}^{-1} \boldsymbol{e} - \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}^{-1} \frac{\partial \boldsymbol{R}}{\partial H} \boldsymbol{R}^{-1} \boldsymbol{e}\right) \tag{35}$$

$$\frac{\partial^2 \ln[p(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial \sigma\, \partial H} = -\frac{1}{\sigma^3}(\boldsymbol{x}_n - \mu\, \boldsymbol{e})^{\mathrm{T}} \boldsymbol{R}^{-1} \frac{\partial \boldsymbol{R}}{\partial H} \boldsymbol{R}^{-1} (\boldsymbol{x}_n - \mu\, \boldsymbol{e}) \tag{36}$$

The expectations of the above expressions are easily calculated and give the corresponding elements of the Fisher Information Matrix $\boldsymbol{I}(\boldsymbol{\theta})$.

Acknowledgements: The authors wish to thank two anonymous reviewers for their constructive comments.

REFERENCES

Beran J (1994) Statistics for Long-Memory Processes, Volume 61 of Monographs on Statistics and Applied Probability. Chapman and Hall, New York

Bouette JC, Chassagneux JF, Sibai D, Terron R, Charpentier R (2006) Wind in Ireland: long memory or seasonal effect. Stoch Environ Res Risk Assess 20(3):141-151

Coeurjolly JF (2008) Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. Ann Statist 36(3):1404-1434

Cox DR, Reid N (1987) Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society, Series B. (Methological) 49(1):1-39

Doukhan P, Oppenheim G, Taqqu M (2003) Theory and Applications of Long-Range Dependence. Birkhauser

Ehsanzadeh E, Adamowski K (2010) Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. Hydrol Process 24:970–980

Embrechts P, Maejima M (2002) Self similar Processes. Princeton University Press

Esposti F, Ferrario M, Signorini MG (2008) A blind method for the estimation of the Hurst exponent in time series: Theory and application. Chaos 18(3). doi:10.1063/1.2976187

Grau-Carles P (2005) Tests of Long Memory: A Bootstrap Approach. Stoch Environ Res Risk Assess 25(1-2):103-113

Guerrero A, Smith L (2005) A maximum likelihood estimator for long-range persistence. Physica A 355(2-4):619-632

Hurst HE (1951) Long term storage capacities of reservoirs. Trans ASCE 116:776-808

Kolmogorov AE (1940) Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum. Dokl Akad Nauk URSS 26:115–118

Koutsoyiannis D (2003a) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol Sci J 48(1):3-24

Koutsoyiannis D (2003b) Internal report: http://www.itia.ntua.gr/getfile/537/2/2003HSJHurstSuppl.pdf

Koutsoyiannis D, Montanari A (2007) Statistical analysis of hydroclimatic time series: Uncertainty and insights. Water Resour Res 43(5). W05429, doi: 10.1029/2006WR005592

Mandelbrot BB, JW van Ness (1968) Fractional Brownian motion, fractional noises and applications. SIAM Rev 10:422–437

McLeod AI, Hippel K (1978) Preservation of the Rescaled Adjusted Range 1. A Reassesment of the Hurst Phenomenon. Water Resour Res 14(3):491-508

McLeod AI, Yu H, Krougly Z (2007) Algorithms for linear time series analysis: With R package. Journal of Statistical Software 23(5):1-26

Mielniczuk J, Wojdyllo P (2007) Estimation of Hurst exponent revisited. Computational Statistics & Data Analysis 51(9):4510-4525

Musicus B (1988) Levinson and Fast Cholesky Algorithms for Toeplitz and Almost Toeplitz Matrices. Research Laboratory of Electronics Massachusetts Institute of Technology. RLE Technical Report No. 538

Palma W (2007) Long-Memory Time Series Theory and Methods. Wiley-Interscience.

Rea W, Oxley L, Reale M, Brown J (2009) Estimators for long range dependence: an empirical study. Electronic Journal of Statistics. arXiv:0901.0762v1

Robert C (2007) The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer ,New York

Robinson PM (1995) Gaussian semiparametric estimation of time series with long-range dependence. Ann Statist 23:1630-1661

Robinson PM (2003) Time Series with Long Memory. Oxford University Press

Taqqu M, Teverovsky V, Willinger W (1995) Estimators for long-range dependence: an empirical study. Fractals 3(4):785-798

Weron R (2002) Estimating long-range dependence: finite sample properties and confidence intervals. Physica A: Statistical Mechanics and its Applications 312(1,2):285-299

Zhang Q, Xu CY, Yang T (2009) Scaling properties of the runoff variations in the arid and semi-arid regions of China: a case study of the Yellow River basin. Stoch Environ Res Risk Assess 23(8):1103-1111

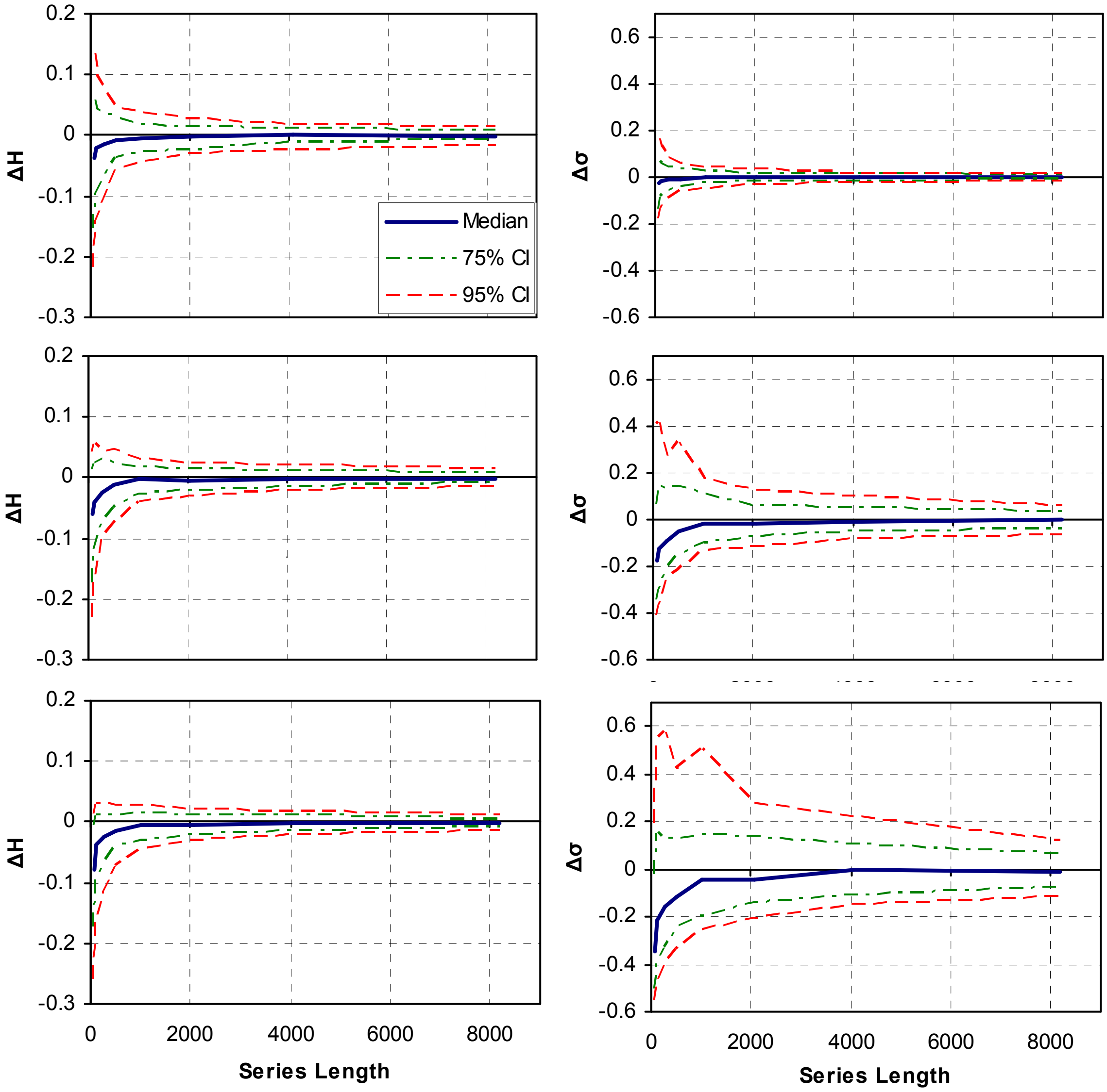

Fig. 1 Monte Carlo confidence intervals for the H and σ estimates with true H = 0.60, H = 0.90 and H = 0.95 (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the ML estimator.

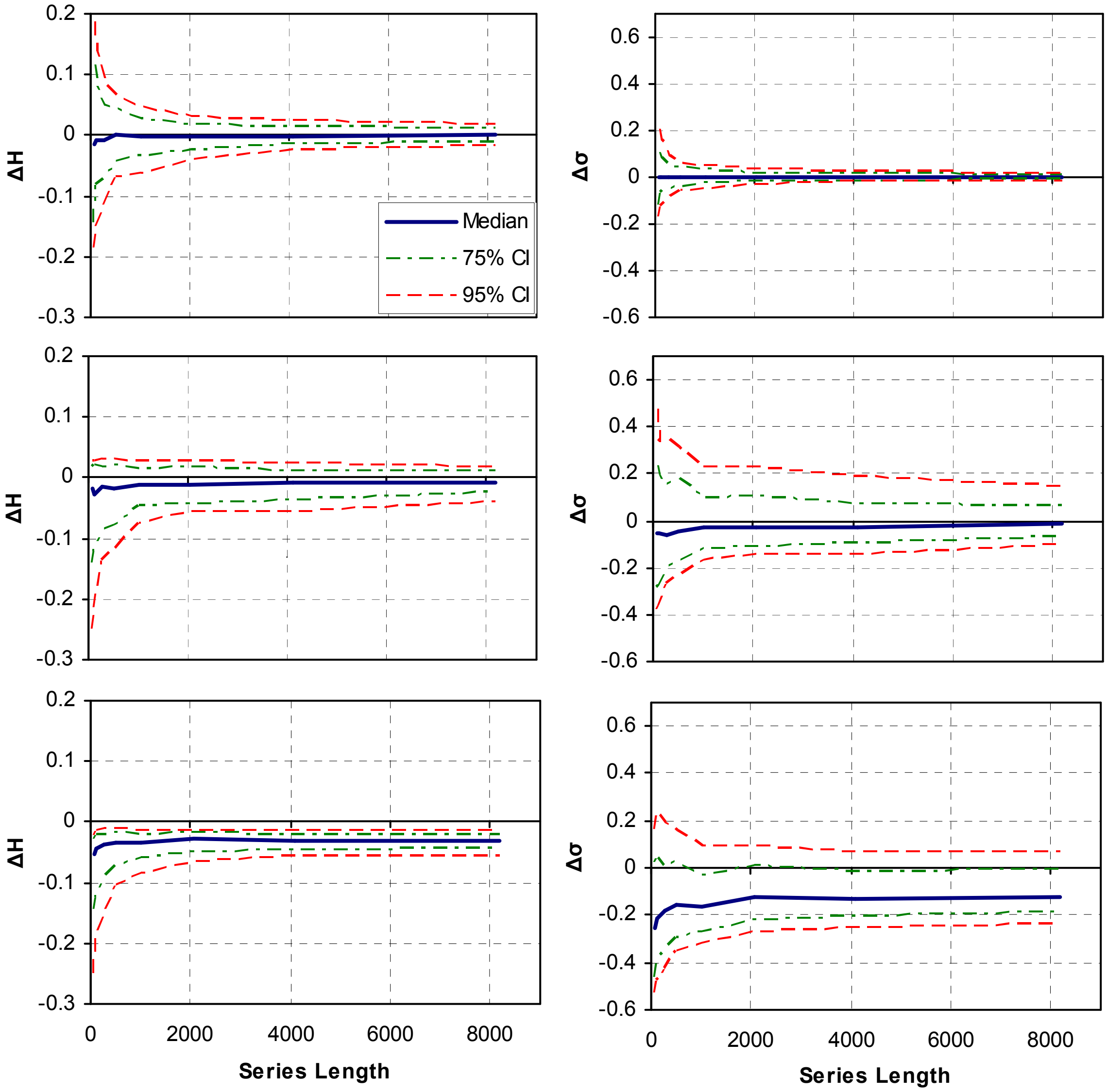

Fig. 2 Monte Carlo confidence intervals for the H and σ estimates with true H = 0.60, H = 0.90 and H = 0.95 (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the LSSD estimator.

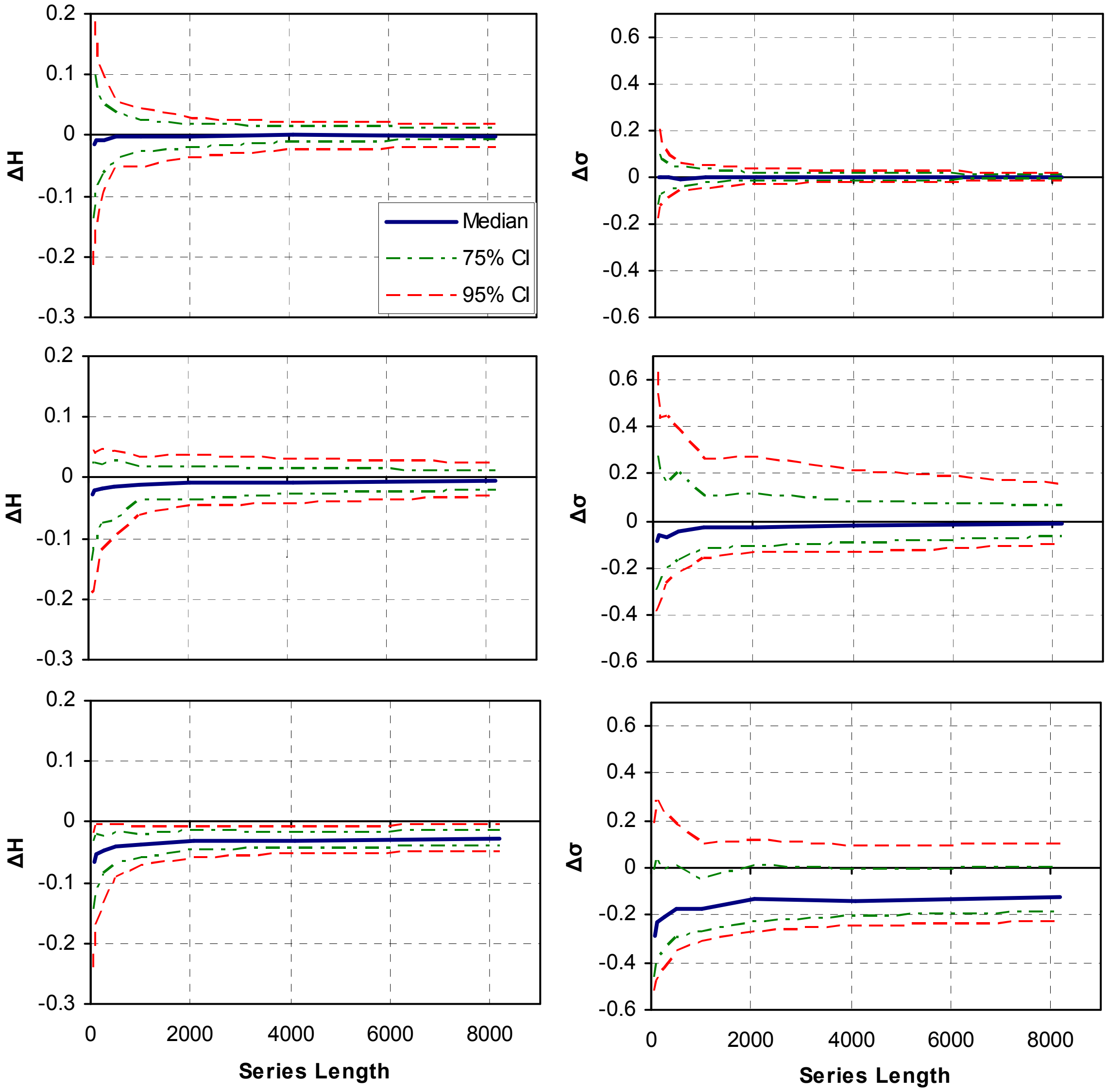

Fig. 3 Monte Carlo confidence intervals for the H and σ estimates with true H = 0.60, H = 0.90 and H = 0.95 (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the LSV estimator.

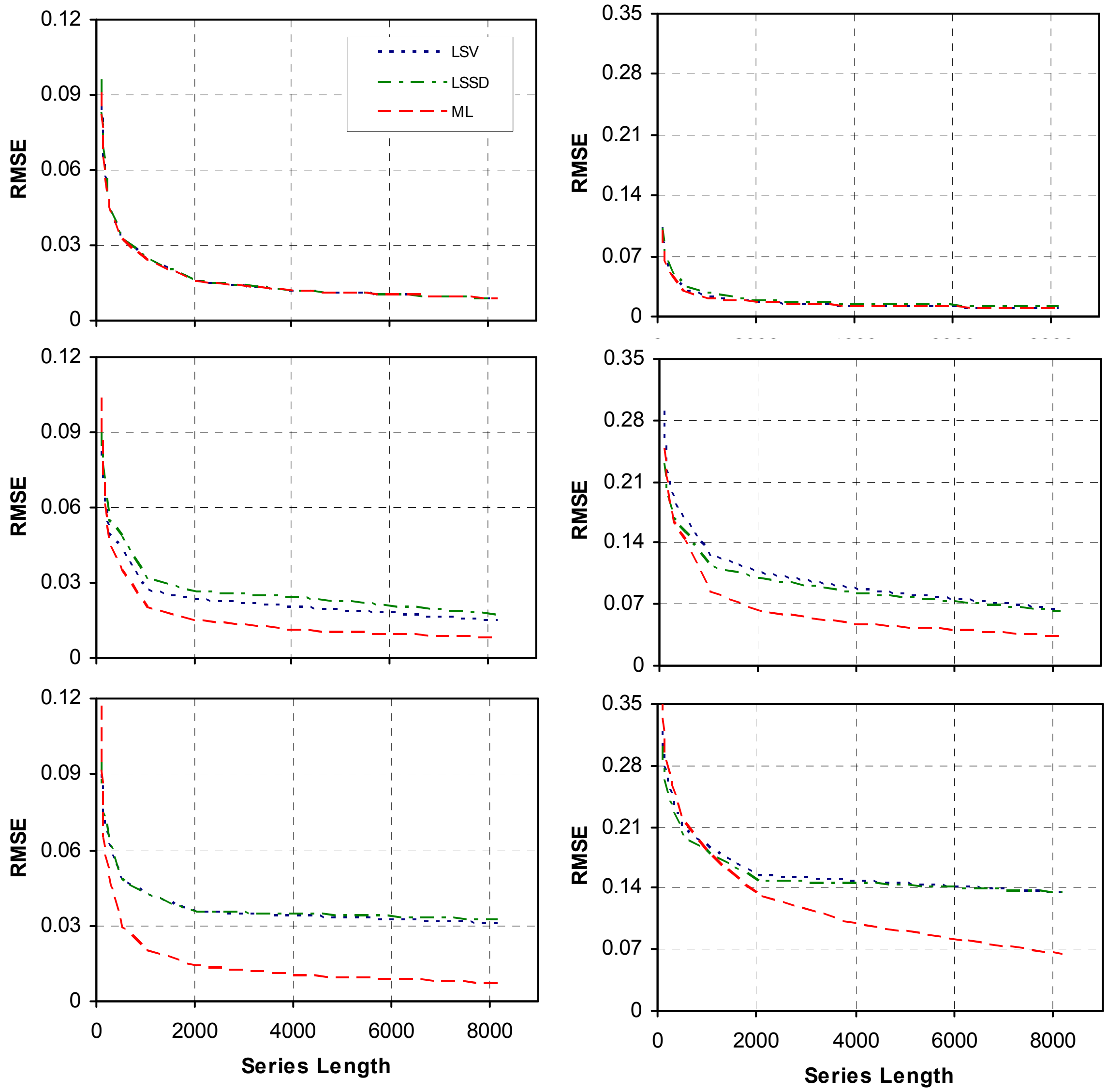

Fig. 4 Root mean square error (RMSE) (left of the estimated H and right of the estimated σ) as a function of series length for all three estimators, with $H = 0.60$, $H = 0.90$ and $H = 0.95$ (upper to lower panels) and $\sigma = 1$.

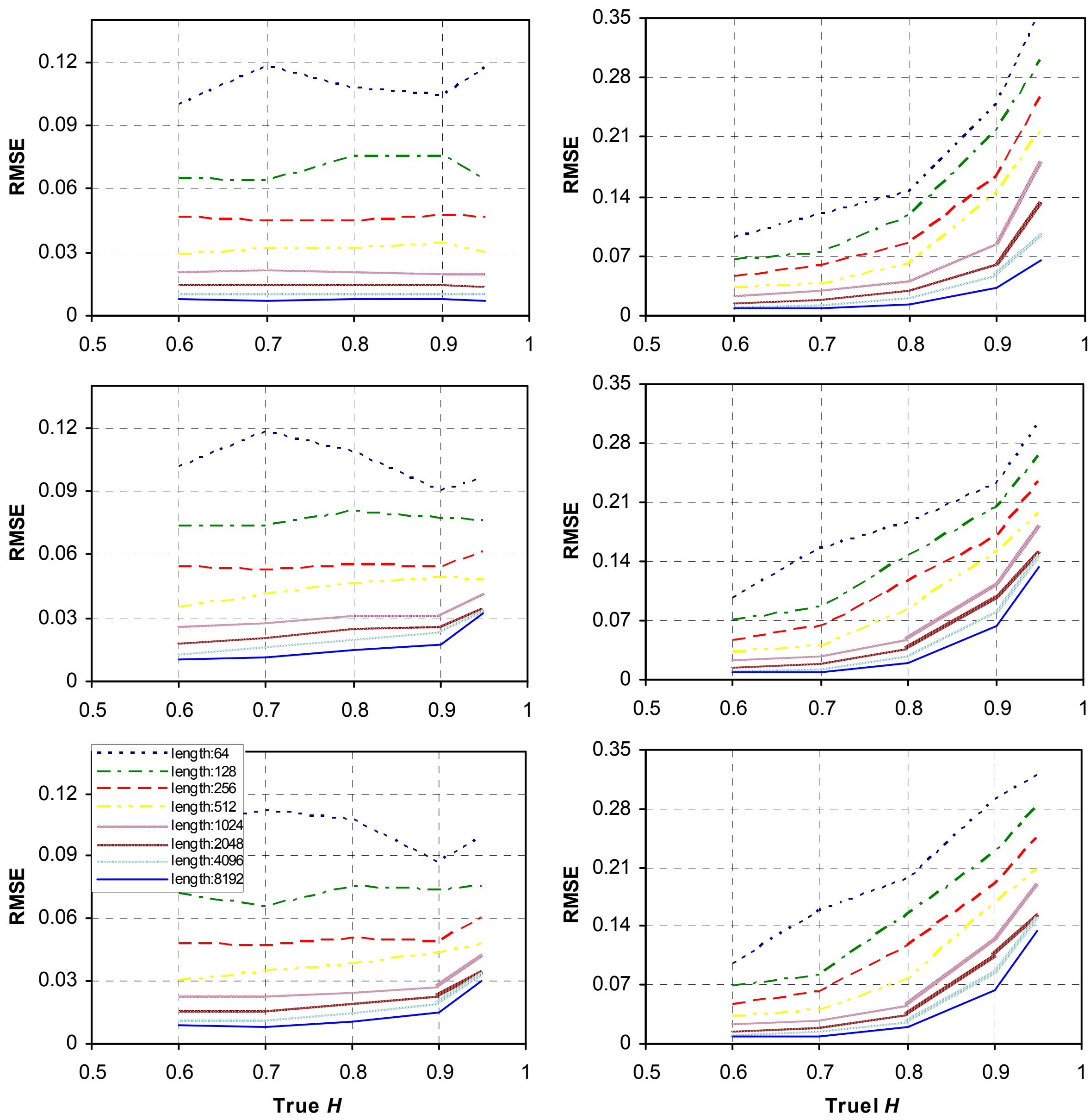

Fig. 5 Root mean square error (RMSE) of H (left) and σ (right) as a function of true H for all lengths. Upper to lower panels correspond to ML, LSSD and LSV methods.

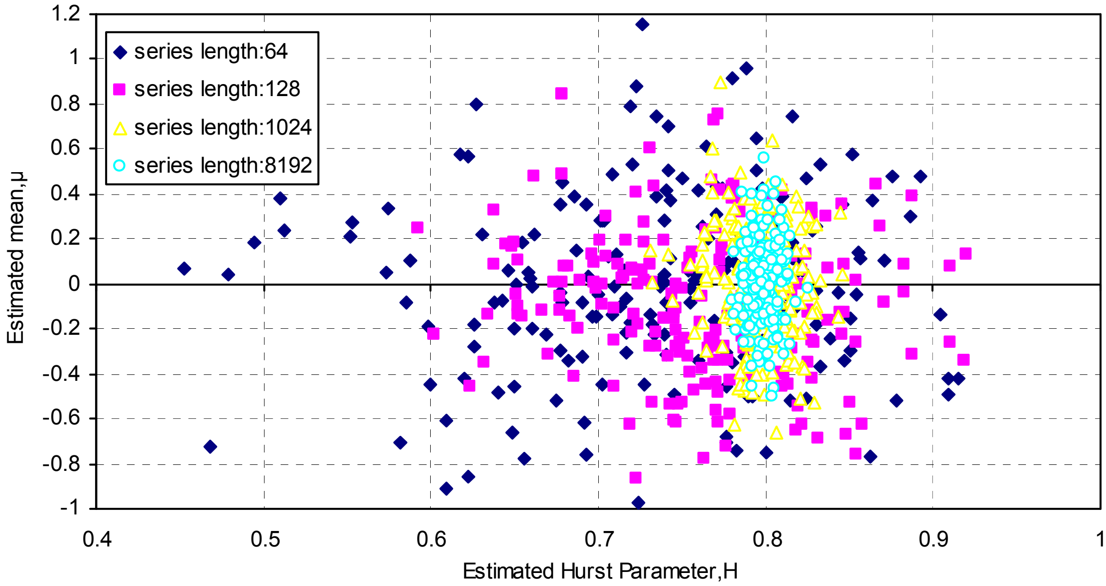

Fig. 6 Estimated Hurst parameter H versus estimated mean μ from the ML method from 200 ensembles of synthetic time series with various lengths for true $H = 0.8$.

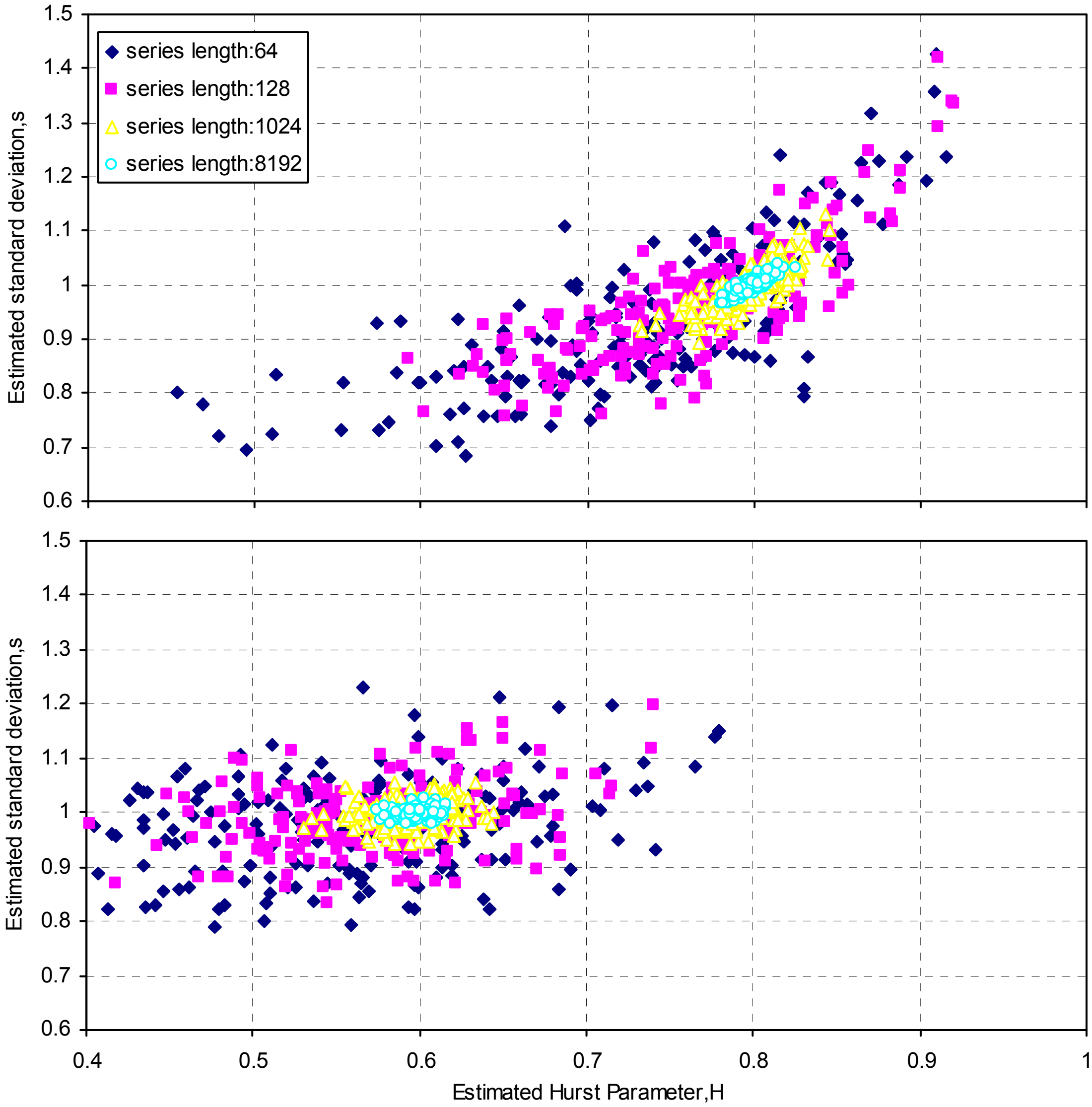

Fig. 7 Estimated Hurst parameter H versus estimated standard deviation σ from the ML method from 200 ensembles of synthetic time series with various lengths. The upper diagram corresponds to true $H = 0.8$ and the lower diagram coresponds to true $H = 0.6$.

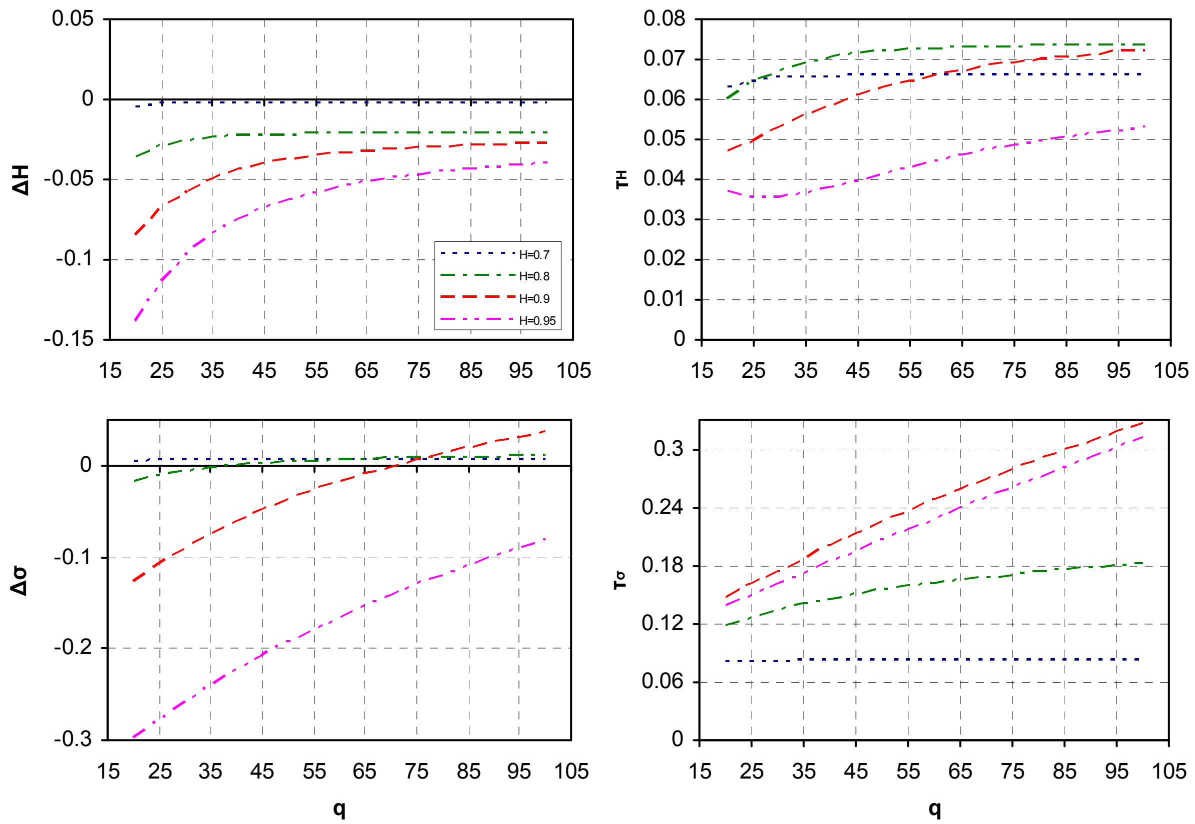

Fig. 8 Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 128 long (right) versus q, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations and $p = 6$ for the LSV estimator.

Definition of symbols used:

$$\tau_H := \sqrt{(1/(200\text{-}1)) \sum_{k=1}^{200} (\Delta H_k)^2},\ \tau_\sigma := \sqrt{(1/(200\text{-}1)) \sum_{k=1}^{200} (\Delta\sigma)^2}$$

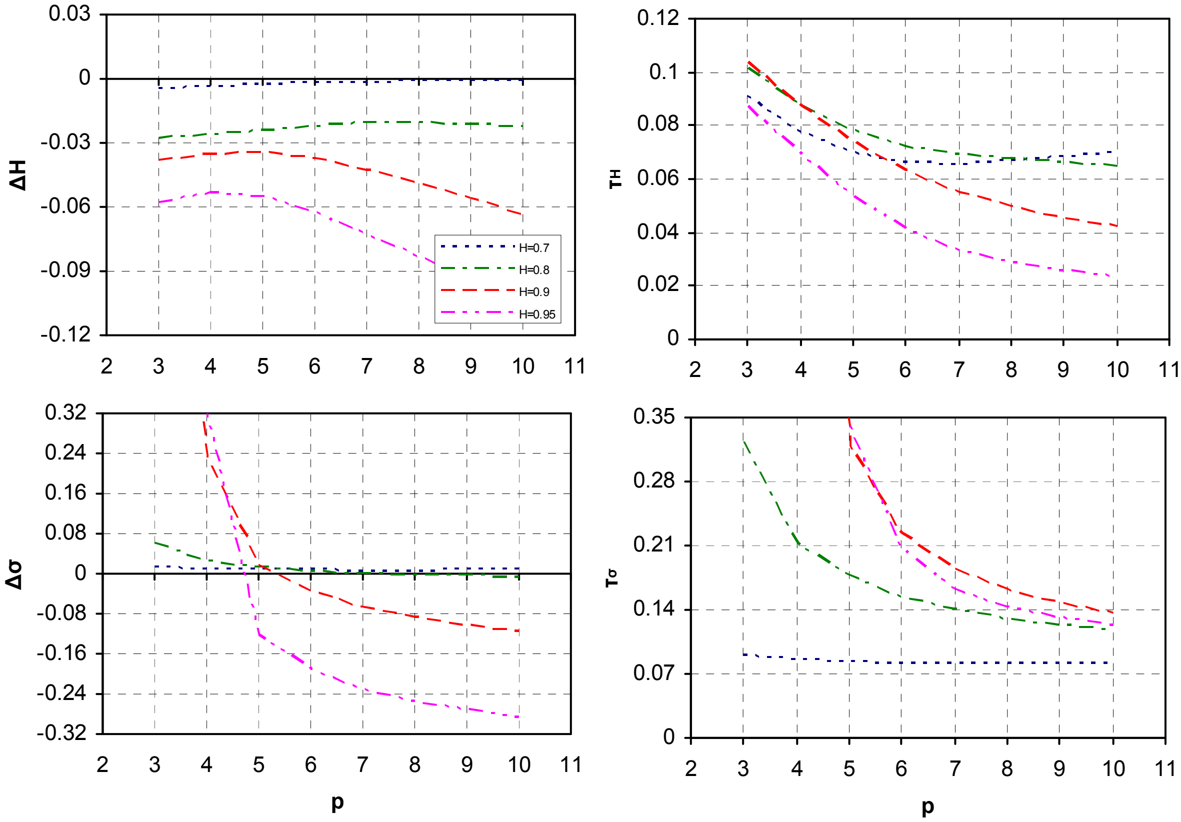

Fig. 9 Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 128 long (right) versus p, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations and $q = 50$ for the LSV estimator.

(See definition of symbols used in caption of Fig. 8).

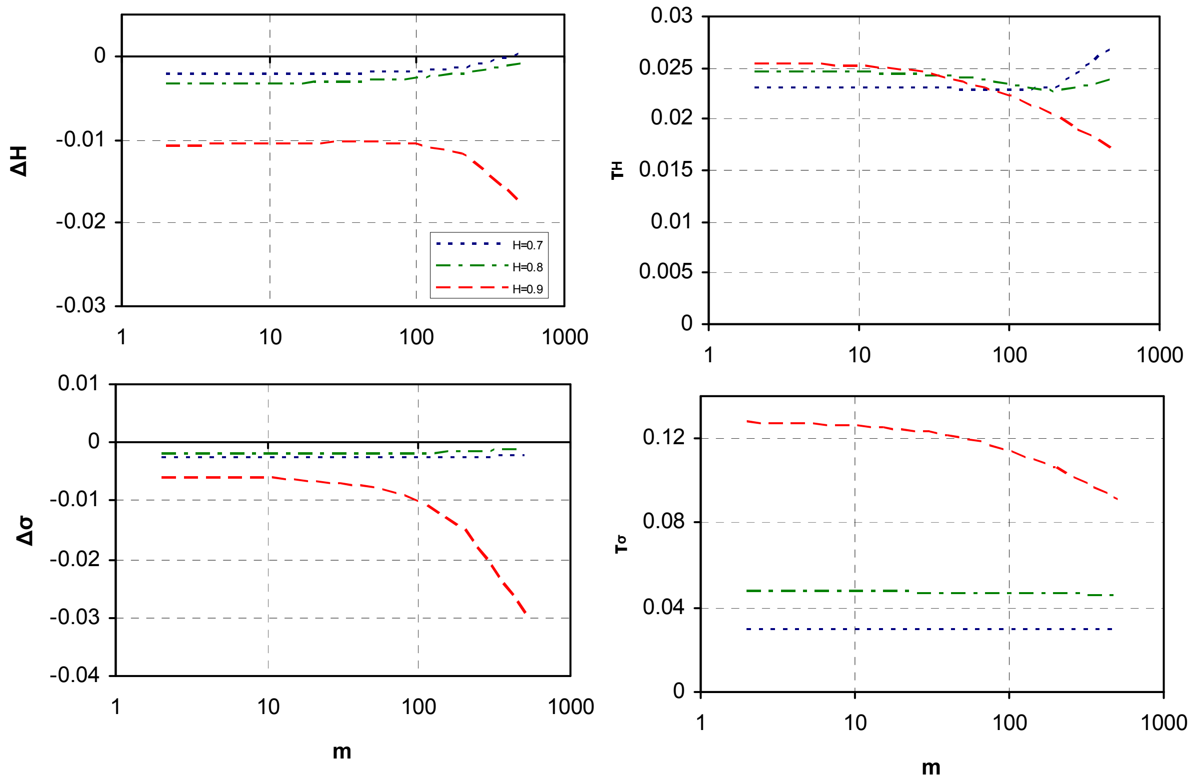

Fig. 10 Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 1024 long (right) versus m, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations, $p = 6$ and $q = 50$ for the LSV estimator. (See definition of symbols used in caption of Fig. 8).

Table 1 Estimation results for H using 200 independent realizations 8,192 long where τ is the standard deviation of the sample containing the estimated H's. H's were estimated using the Matlab central file exchange functions package "Hurst parameter estimate" written by Chu Chen (http://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate), except the local Whittle estimates, where the local Whittle estimator written by Shimotsu was used (http://qed.econ.queensu.ca/faculty/shimotsu/)

Estimation method		True H			
		0.6	0.7	0.8	0.9
Variance	$\hat{H}$	0.595	0.687	0.775	0.850
	τ	0.027	0.027	0.026	0.027
	RMSE	0.027	0.030	0.036	0.057
DiffVar	$\hat{H}$	0.567	0.667	0.771	0.864
	τ	0.073	0.068	0.067	0.061
	RMSE	0.080	0.076	0.073	0.070
Absolute	$\hat{H}$	0.594	0.686	0.775	0.849
	τ	0.028	0.027	0.028	0.029
	RMSE	0.029	0.031	0.038	0.059
Higuchi	$\hat{H}$	0.599	0.696	0.798	0.888
	τ	0.028	0.029	0.040	0.044
	RMSE	0.028	0.029	0.040	0.046
Var. of Residuals	$\hat{H}$	0.600	0.702	0.801	0.896
	τ	0.024	0.028	0.030	0.027
	RMSE	0.024	0.028	0.030	0.027
R/S	$\hat{H}$	0.619	0.706	0.784	0.854
	τ	0.031	0.032	0.031	0.032
	RMSE	0.036	0.033	0.035	0.055
Periodogram	$\hat{H}$	0.604	0.708	0.809	0.912
	τ	0.024	0.023	0.025	0.024
	RMSE	0.024	0.024	0.026	0.027
Modified Periodogram	$\hat{H}$	0.565	0.661	0.752	0.847
	τ	0.037	0.038	0.037	0.034
	RMSE	0.051	0.054	0.060	0.063
Local Whittle	$\hat{H}$	0.601	0.700	0.804	0.902
	τ	0.023	0.023	0.022	0.021
	RMSE	0.023	0.023	0.023	0.021

Note: *Variance*: a method based on aggregated variance; *DiffVar*: a method based on differencing the variance; *Absolute*: a method based on absolute values of the aggregated series; *Higuchi*: a method based on finding the fractal dimension; *Var. of Residuals*: a method based on residuals of regression, also known as Detrended Fluctuation Analysis (DFA); *R/S*: the original method by Hurst, based on the rescaled range statistic; *Periodogram*: a method based on the periodogram of the time series; *Modified Periodogram*: similar as the Periodogram method but with frequency axis divided into logarithmically equally spaced boxes and averaging the periodogram values inside the box (see details in Taqqu et al. 1995); *Local Whittle*: a semiparametric version of the Whittle estimator (see details in Robinson 1995).

Table 2 Estimation results for H using 200 independent realizations 8,192 long where τ is the standard deviation of the sample containing the estimated H's.

Estimation method		Nominal H			
		0.6	0.7	0.8	0.9
Maximum Likelihood	$\hat{H}$	0.599	0.700	0.799	0.899
	τ	0.008	0.007	0.008	0.007
	RMSE	0.008	0.007	0.008	0.007
Least Squares Standard Deviation	$\hat{H}$	0.599	0.699	0.799	0.892
	τ	0.011	0.011	0.015	0.015
	RMSE	0.011	0.012	0.015	0.017
Least Squares Variation	$\hat{H}$	0.599	0.700	0.800	0.895
	τ	0.009	0.008	0.011	0.014
	RMSE	0.009	0.008	0.011	0.015