British Journal of Mathematical and Statistical Psychology (2007), 60, 85-106

Detection of long-range dependence and estimation of fractal exponents through ARFIMA modelling

Kjerstin Torre, Didier Delignières, & Loïc Lemoine

EA 2991 Motor Efficiency and Deficiency
University Montpellier I, France

Abstract: The aim of this paper was to evaluate the performances of ARFIMA modelling for detecting long-range dependence and estimating fractal exponents. More specifically, our aim was to test the procedure proposed by Wagenmakers, Farrell and Ratcliff (2005), and to compare the results obtained with the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC). The present studies show that ARFIMA modelling is able to adequately detect long-range dependence in simulated fractal series. Conversely, this method tends to produce a non-negligible rate of false detections in pure ARMA series. Generally, ARFIMA modelling presents a bias favouring the detection of long-range dependence. AIC and BIC gave dissimilar results, due to the different weights attributed by the two criteria to accuracy and parsimony. Finally, ARFIMA modelling provides good estimates of fractal exponents, and could adequately complement classical methods, such as spectral analysis, detrended fluctuation analysis or rescaled range analysis.

Introduction

Some recent experiments, generally underlain by the framework of dynamical systems theory, tried to analyze the dynamical structure of time series of psychological or behavioural variables. A quite intriguing result in these experiments was the recurrent discovery of fractal properties in the studied series. Fractals were evidenced, for example, in self-esteem (Delignières, Fortes, & Ninot, 2004), in mood (Gottschalk, Bauer, & Whybrow, 1995), in serial reaction time (Gilden, 1997; van Orden, Holden, & Turvey, 2003), in finger tapping (Gilden, Thornton, & Mallon, 1995; Delignières, Lemoine, & Torre, 2004), in stride duration during walking (Hausdorff, Peng, Ladin, Wei, & Goldberger, 1995), or in relative phase in a bimanual coordination task (Schmidt, Beek, Treffner, & Turvey, 1991).

All these variables were previously conceived as highly stable over time, and fluctuations in successive measurements were considered as randomly distributed, and uncorrelated in time. As a consequence, a sample of repeated measures was assumed to be normally distributed around its mean value, and noise could be discarded by averaging. This methodological standpoint was implicitly adopted in most classical psychological research (for a deeper analysis, see Gilden, 2001; Slifkin & Newell, 1998). In other words, temporal ordering of data points was ignored and the possible correlation structure of fluctuations was clearly neglected.

Fractal analysis focuses, in contrast, on the time-evolutionary properties of data series and on their correlation structure. Fractal processes are characterized by a complex pattern of correlations appearing following multiple interpenetrated time scales. In such processes, the value at a particular time is related not just to immediately preceding values, but also to fluctuations in the remote past. Such series are said to present long-term memory, or long-range dependence. This property is typically revealed by a very slow decay over time of the auto-correlation function, which tends to follow a power law.

Detecting fractal properties in empirical time series could have important theoretical implications. Fractals are considered as the natural outcome of complex dynamical systems behaving at the frontier of chaos (Bak & Chen, 1991; Marks-Tarlow, 1999). Psychological variables should then be conceived as the macroscopic and dynamical products of complex systems composed of multiple interconnected elements. Moreover, psychological and behavioural

Send correspondence to:
Kjerstin Torre
Faculty of Sport Sciences – University Montpellier 1
EA 2991, "Motor Efficiency and Deficiency"
700 Avenue du Pic Saint Loup
34090 Montpellier – France
Email: ktorre@univ-montp1.fr

time series often present fractal characteristics close to a very special case of fractal process, called 1/*f* or *pink* noise. With regard to the power spectrum of such time series, '1/*f* noise' signifies that each frequency has power proportional to its period of oscillation. As such, power is distributed across the entire spectrum and not concentrated at a certain portion. Consequently, fluctuations at one time scale are only loosely correlated with those of another scale. This relative independence of the underlying processes acting at different time scales suggests that a localized perturbation will not necessarily alter the stability of the global system. In other words, 1/*f* noise renders the system more stable and more adaptive to internal and external perturbations (West & Shlesinger, 1989).

The fGn/fBm model

A deeper presentation of fractal processes could be necessary to ensure a better understanding of the following parts of this article. A good starting point for this presentation is Brownian motion, a well-known stochastic process that can be represented as the random movement of a single particle along a straight line. Mathematically, Brownian motion is the integration of a white Gaussian noise. As such, the most important property of Brownian motion is that its successive increments in position are uncorrelated: each displacement is independent of the former, in direction as well as in amplitude. Einstein (1905) showed that, on average, this kind of motion moves a particle from its origin by a distance that is proportional to the square root of the time.

Mandelbrot and van Ness (1968) defined a family of processes they called *fractional Brownian motions* (fBm). The main difference with ordinary Brownian motion is that in an fBm successive increments are correlated. A positive correlation signifies that an increasing trend in the past is likely to be followed by an increasing trend in the future. The series is said to be persistent. Conversely, a negative correlation signifies that an increasing trend in the past is likely to be followed by a decreasing trend. The series is then said to be *anti-persistent*.

Mathematically, an fBm is characterized by the following scaling law:

$$<\Delta x> \propto \Delta t^{H} \qquad (1)$$

which signifies that the expected displacement $<\Delta x>$ is a power function of the time interval (Δt) over which this displacement is observed. H represents the typical scaling exponent (or *Hurst* exponent) of the series and can be any real number in the range $0 < H < 1$. Ordinary Brownian motion corresponds to the special case $H = 0.5$ and constitutes the frontier between anti-persistent ($H < 0.5$) and persistent fBms ($H > 0.5$).

Fractional Gaussian noise (fGn) represents another family of fractal processes, defined as the series of successive increments in an fBm. Note that fGn and fBm are interconvertible: when an fGn is cumulatively summed, the resultant series constitutes an fBm. Each fBm is then related to a specific fGn, and both are characterized by the same H exponent. These two processes possess fundamentally different properties: fBm is non-stationary with time-dependent variance, while fGn is a stationary process with a constant expected mean value and constant variance over time.

Several methods have been proposed for quantifying the fractal properties of a given series. The most often used is the *Power Spectrum Density* (PSD) method, which exploits the property that in fractal processes, spectral power is a power function of the corresponding period:

$$S(f) \propto 1/f^{\beta} \qquad (2)$$

In this equation, f represents frequency and $S(f)$ the corresponding squared amplitude. This relationship is revealed by the obtaining of a linear regression in the double-logarithmic plot of the power spectrum. β can be easily estimated by calculating the negative slope ($-\beta$) of the line relating log (S(f)) to log f. As previously evoked in the introduction, this property led to the designation of '1/*f* noise' which corresponds to the special case $\beta = 1$. The definition is not so strict, nevertheless, and generally one considers a broader family of processes, called $1/f^{\beta}$ noises, β ranging from 0.5 to 1.5.

Another set of methods, such as *Detrended Fluctuation Analysis* (DFA) *or Rescaled Range Analysis* (R/S) exploits a corresponding relationship, in the time domain, which states that in a fractal process variance is a power function of the time over which it is computed:

$$\mathrm{Var}(x) = \Delta t^{2H} \qquad (3)$$

This equation derives directly from Eq. 1, H representing the Hurst exponent. Time-domain methods generally aim at estimating variability over a number of intervals of different lengths, and then to estimate H through the double logarithmic plot of variability against interval length. If the series under study is a fractal process, a linear regression is expected.

Long-range and short-range dependence

Long-range dependence, nevertheless, is not the only conceivable model of dependence in times series. More simply, one could conceive the current observation as being only influenced by the preceding one, or by a limited set of preceding

observations. For instance, this kind of short-term dependence could be conceptualized in terms of feed-back processes: during the repetitive performance of a given task, any information concerning the previous trial can be exploited for improving the current realization (Delcor, Cadopi, Delignières & Mesure, 2003; Spray & Newell, 1986). Another kind of short-term dependence was evidenced by Fortes, Delignières and Ninot (2004), which showed that the evolution of self-esteem over time could be adequately modelled by a short-term process, according to which any perturbation is partly corrected during the next assessment, in order to allow self-maintenance. The well-known timing model of Wing and Kristofferson (1973) also suggests a kind of short-term dependence in serial finger tapping: in this model, serial dependence is not determined by a feed-back process, but just by the contamination of contiguous inter-tap intervals by a common motor error term. Pressing and Jolley-Rogers (1997) proposed a modified version of the previous model for accounting for serial dependence in synchronization tapping. This model included an error correction process, each asynchrony between signal and tap being compensated at the following tap. This typical short-term dependence model was shown to adequately fit experimental data. Another well-known family of short-term dependence models gathers the sequential sampling models, developed in the domain of choice reaction time tasks (Laming, 1979; Ratcliff & Smith, 2004).

The most important point to note is that long-range dependence is not the only available hypothesis for accounting for serial dependence in psychological time series. Short-term dependence models offer an alternative framework, allowing a suitable fitting of empirical data, and supporting interesting and innovative models of psychological processes.

Statistical limitation of classical methods

As previously indicated, the identification of fractal properties with classical methods is only based on the visual inspection of the power spectrum or the diffusion plot, in bi-logarithmic coordinates. Generally authors are satisfied with obtaining an approximate linear fit to their data (see, for example, Yamada, 1995). Nevertheless, this visual evaluation remains qualitative, and there is no statistical test for judging if a regression line is appropriate or not (Wagenmakers, Farrell & Ratcliff, 2005).

Another problem is that short-term dependence processes can sometimes mimic the spectrum or the diffusion plot of a 1/*f* series (Farrell, Wagenmakers & Ratcliff, 2005; Rangarajan & Ding, 2000; Wagenmakers, Farrell & Ratcliff, 2004). Wagenmakers et al. (2004) proposed a number of examples of such ambiguous results obtained with short-range dependence processes. An auto-regressive process, for example, is supposed to present in the log-log spectral power a typical flattening in the low frequency region, reflecting the fact that there is no long-range dependence in the series. In a wide range of frequencies, nevertheless, the power spectrum also presents a 1/*f*-like linear trend: short-range dependence (i.e., a relation between the value at time *t* and time *t*-1) leads to the occurrence of low frequency components that have more power than high frequency components (Wagenmakers et al., 2004). The difference from a genuine 1/*f* spectrum often lies in two or three ambiguous points in the lowest frequencies.

Rangarajan and Ding (2000) called for the complementary use of different methods, in the frequency and time domains, in order to avoid false conclusions. They present a number of simulated examples in which spectral and rescaled range analyses gave dissimilar results, suggesting the limitation of the use of a unique method. Their approach, nevertheless, remains qualitative and the conclusions drawn could stay ambiguous.

Some authors proposed the application of a so-called *surrogate data test* (Theiler, Eubank, Longtin, Galdrikian, & Farmer, 1992; Hausdorff, Peng, Ladin, Wei, & Goldberger, 1995). This method consists in randomly shuffling data sets and estimating the fractal exponents of the obtained series. The expected mean scaling exponent for these surrogate data sets is about 0.5. An inferential test is then applied to compare the exponents obtained from the original series and the exponents obtained from the surrogate data sets. The aim of this method was to determine whether the detected fractal behaviour reflected reality or was due to chance or to the applied methods. Nevertheless, the null hypothesis that is tested in this procedure is the absence of correlation in the series. This null hypothesis is surely not the most relevant, as the absence of correlation (purely white noise) in psychological time series should be considered more as an exception than as the rule (Slifkin & Newell, 1998). According to Wagenmakers et al. (2004) the fundamental question is about the nature (short-term vs long-term) of dependencies in the series. Clearly, classical methods are unable to adequately answer this question. Spectral analysis, DFA or R/S analysis could be relevant for quantifying long-range dependence, when long-range dependence is a priori supposed to be present. But they seem per se unable to validate such a presence. Wagenmakers et al. (2004) and Farrell et al. (2005) recently proposed an inferential test for the presence of long-range dependence in time series, based on ARFIMA (auto-regressive fractionally integrated moving average) modelling.

Modelling short-range and long-range dependence

In short-range dependence processes, each value can be predicted by a limited set of preceding values. Box and Jenkins (1970) introduced a family of linear models, called ARIMA (for auto-regressive, integrated, moving average), intended to represent a variety of short-term relationships in time series.

ARIMA models are potentially composed of three components. The auto-regressive component suggests that the current observation y_t is determined by a weighted sum of the p previous observations, plus a random perturbation ε_t:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \quad (4)$$

In this equation ϕ_i represent the influence of the i^{th} previous value, and is assumed to progressively decay over time.

The moving-average component supposes that the current observation depends on the value of the random perturbations that affected the q preceding observations, plus its own specific perturbation:

$$y_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (5)$$

The integrated component of the model determines whether the observed values are modelled directly, or whether the differences between consecutive observations are modelled instead. The differencing parameter d indicates the number of differencing that should be applied to the series before modelling.

An ARIMA model is a combination of these three components, and can be designated by the respective orders of the three combined processes as (p,d,q). As an example, a (1, 1, 1) model should obey the following equation:

$$y_t - y_{t-1} = \phi_1(y_{t-1} - y_{t-2}) + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (6)$$

ARIMA models could be more conveniently expressed using the so-called *backshift operator*, defined as:

$$By_t = y_{t-1} \quad (7)$$

The generic ARIMA (p,d,q) model can then be rewritten as:

$$\phi(B)(1 - B)^d = \theta(B)\varepsilon_t \quad (8)$$

where $\phi(B)$ and $\theta(B)$ are, respectively, the auto-regressive and the moving average operators, represented as polynomials in the backshift operator:

$$\phi(B) = 1 - B\phi_1 - B^2\phi_2 - \ldots - B^p\phi_p \quad (9a)$$

and

$$\theta(B) = 1 - B\theta_1 - B^2\theta_2 - \ldots - B^q\theta_q \quad (9b)$$

Granger and Joyeux (1980) showed that it is possible to provide this model with long-range dependence properties by allowing the differencing parameter d to take on fractional values, thereby obtaining an ARFIMA model. ARFIMA models provide a very parsimonious account for long-range dependence, by the addition of a single parameter to classical ARMA models. Importantly, they allow the simultaneous modelling of short-term processes (by the combination of the p and q parameters), and long-range dependence through the d parameter, and as such the isolation of their respective effects. Finally, the ARFIMA parameters can be estimated using exact maximum likelihood, allowing the significance of the difference of d from 0 to be tested. As such, ARFIMA modelling can effectively be used for detecting the presence of long-range dependence in the series. Moreover, the estimation of d allows the quantification of the intensity of the long-range correlations within the series, as d is related to the spectral exponent β by the simple equation:

$$\beta = 2d \quad (10)$$

Note that d is bounded within the interval [-0.5; 0.5]; that is, ARFIMA modelling can only take stationary signals with spectral exponents β between –1 and +1 (fGns) into account. For fBm signals, it is possible to apply ARFIMA to the corresponding fGn (obtained by differentiation). One can then estimate the theoretical fractional parameter of the fBm series by adding 1 to the d value obtained from the fGn (Diebolt & Guiraud, 2005).

Wagenmakers et al. (2005) recently proposed a complete inferential procedure, based on ARFIMA modelling, for ascertaining the presence of long-range dependence in time series. Their method consists in fitting 18 models to the studied series. Nine of these models are ARMA (p,q) models, p and q varying systematically from 0 to 2. These ARMA models do not contain any long-range serial correlation. The other nine models are the corresponding ARFIMA (p,d,q) models, differing from the previous ARMA models by the inclusion of the fractional parameter d representing persistent serial correlations. One supposes that if the series contains long-range dependence, ARFIMA models should present a better fit than the transient ARMA models. Note that in a previous paper, Wagenmakers et al. (2004) proposed to only contrast an ARMA (1,1) and an ARFIMA (1,d,1) model. The present approach, based on the test of a wider range of ARFIMA models, constitutes an interesting improvement, as the correct specification of parameters p and q allows a better estimation of the

long-range estimator *d* (Taqqu & Teverovsky, 1998; Wagenmakers et al., 2005).

The determination of the best model is not so straightforward. The examination of the likelihood scores provided by the fitting procedure is not sufficient, as the capacity of models to account for the data is partly related to their number of free parameters. The selection of models has to be based on a trade-off between accuracy and parsimony: the best model is the one that gives a good account of the data with a minimum number of free parameters.

One of the most popular methods for combining accuracy and parsimony is the Akaike Information Criterion (Akaike, 1973), which can be computed according to the following equation:

$$\mathrm{AIC} = -2\log L + 2k \tag{11}$$

where L represents the maximum likelihood for the model under study, and k is the number of free parameters in the model, plus an additional parameter for the variance of the error series (i.e. for an ARMA(p,q), $k = p + q + 1$, and for an ARFIMA(p,d,q), $k = p + q + 2$). As can be seen, the first term rewards accuracy, and the second penalizes the lack of parsimony. The lower the AIC, the better the model is supposed to be.

An alternative criterion, the Bayes Information Criterion, is often used for model selection. BIC is defined as :

$$\mathrm{BIC} = -2\log L + k\log N \tag{12}$$

where N represents the number of observations in the series. BIC differs from AIC with the second term, which invokes a different penalty for parsimony: one can easily conclude from Eq. 11 and 12 that BIC should penalize complex models more severely than AIC, especially for long series.

The raw values of these criteria remain difficult to interpret and to compare between models. Wagenmakers and Farrell (2004) proposed a convenient transformation of the raw values into weights. Consider that the goal is to select the best model among m candidates. The first step is to compute the difference, for each model, between the criterion for this model and for the best model. That is, for the AIC criterion and the i^{th} model :

$$\Delta_i(AIC) = AIC_i - \min AIC \tag{13}$$

This difference in AIC can then be converted into an estimate of relative likelihood through the following transform :

$$L_i(AIC) \propto \exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\} \tag{14}$$

Finally, these relative likelihoods are transformed into weights by normalization (e.g. by division by the sum of the relative likelihood of all models) :

$$w_i(AIC) = \frac{L_i(AIC)}{\sum_{j=1}^{m} L_j(AIC)} \tag{15}$$

$w_i(AIC)$ can be conceived as the probability for the i^{th} model being the best model, given the data and the set of candidate models (Wagenmakers & Farrell, 2004). Similar calculations can be performed on BIC, leading to specific weights $w_i(BIC)$.

On the basis of these weights, two criteria could be proposed for detecting the presence of long-range dependence in the series : (1) the best model (i.e. the model with the highest weight) should be an ARFIMA (p, d, q), d being significantly different from 0, and (2) the sum of the weights of the ARFIMA models should be higher than the sum of the weights of the ARMA models. Note that the weights computed among a given set of models sum to one. The sum of ARFIMA models weights represents the overall probability of ARFIMA models to overcome their ARMA counterparts.

The aim of the present study was to test the capacity of ARFIMA modelling to detect the presence of long-range dependence in simulated fractional Gaussian noise series. We limited the analyses to stationary processes, which represent the generally expected behaviour of psychological variables. We tested this capacity over a wide range of H exponents, from 0.1 to 0.9. We also tested the effect of the length of the series, and the effect of the addition of white noise to the studied series. Secondly, we submitted simulated ARMA series to the same procedure, in order to analyze the occurrence of false detections of long-range dependence in such series.

Thirdly, we analyzed the performance of ARFIMA modelling in estimating the H exponent. In this final step, we tested again the effect of series length, and the effect of the addition of white noise. In all cases, we compared the respective performances of AIC and BIC. Finally, we applied the method to a set of empirical series, collected in an experiment on bi-manual coordination.

Detection of long-range dependence in simulated fGn series

We used the algorithm proposed by Davies and Harte (1987), for generating fGn series of known H exponent (for a detailed presentation, see Caccia, Percival, Cannon, Raymond, and Bassingthwaigthe, 1997). We generated 40 fGn series of 2048 data points for each of 9 values of H ranging from 0.1 to 0.9 by steps of 0.1. In order to test the effect of series

length, we applied the analysis on the entire series (2048 points), and then on the first 1024 points, the first 512 points, the first 256 points, and finally the first 128 points (i.e. series of 2^{11}, 2^{10}, 2^9, 2^8 and 2^7 points).

In a second step, we added to each original series a white noise series (fGn with $H = 0.5$). The added white noise series were different for each fGn series. We tested four noise/signal *SD* ratios: 0.00 (no added white noise), 0.33, 0.66, and 1.00 (equal variance for white noise and signal). The analysis was then applied to these contaminated signals. These tests were performed for a single series length (1024 points).

Models' fitting was conducted using the ARFIMA package (Doornik & Ooms, 1999; Ooms & Doornik, 1998) for the matrix computing language Ox (Doornik, 2001). We used, with some minor adaptations, the Ox code provided by Simon Farrell, available at the following web address: http://eis.bristol.ac.uk/~pssaf/ (for details, see Farrell et al., 2005).

Figure 1 reports the percentage of correct specifications (i.e. the best model was an ARFIMA model, with parameter *d* significantly different from zero), according respectively to AIC and BIC. Generally fGn series were recognized as long-range dependence processes, except for $H = 0.5$ (in this case, the series were simply white noises). Note, nevertheless, that in this case the best model was generally an ARFIMA model (and not as expected an ARMA (0,0) model), but the coefficient *d* was not significantly different from 0. Generally both methods selected ARFIMA models (for 100.0% of the series for BIC and 95.4% for AIC). The simplest models (0,*d*,0), (1,*d*,0) and (0,*d*,1) represented 99.1% of the selected models for BIC but only 60.7% for AIC.

Clearly BIC produced a better percentage of correct detections than AIC. For the longest series (1024 and 2048 points), the detection appeared perfect for BIC. The percentage of errors was higher for AIC, even for long series. The number of errors increased when series length decreased, especially for weakly persistent fGn ($H = 0.6$ or $H = 0.7$). Misspecifications with AIC were due in part to the selection of complex ARMA models, such as (2,2) or (1,1), and in part to the selection of complex ARFIMA models, such as (2,*d*,2) or (1,*d*,1), *d* being in this case not significantly different from zero.

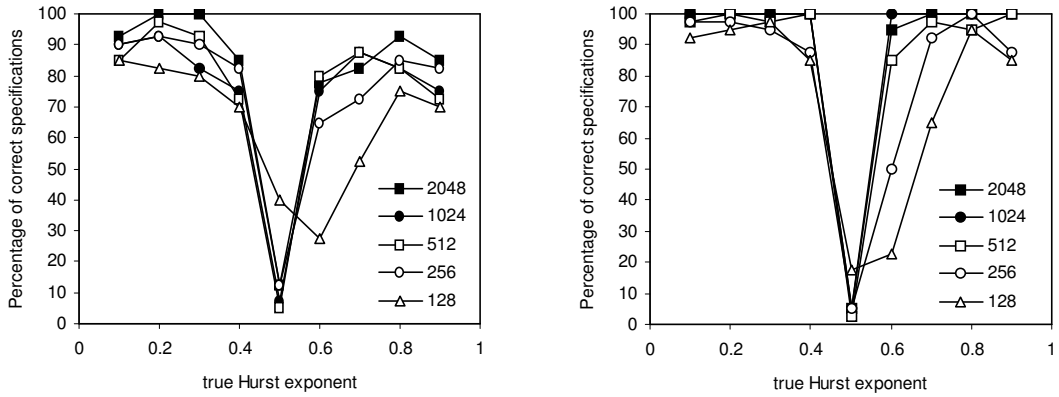

Figure 1: Percentage of correct specifications, according to AIC (left) and BIC (right). fGn series, from 2048 to 128 data points

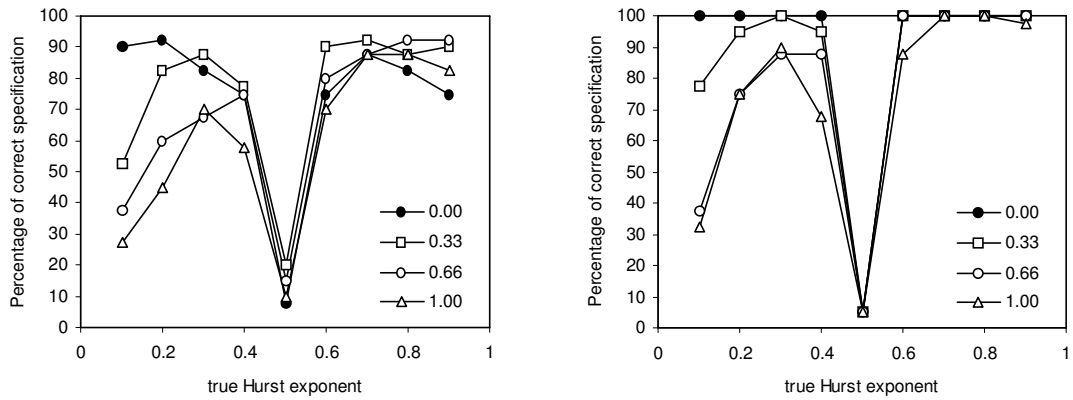

Figure 2: Percentage of correct specifications, according to AIC (left) and BIC (right). fGn series (1024 data points) with added noise.

Figure 2 indicates the percentage of correct specification for series of 1024 points contaminated by white noise. The effect of noise was clearly different for anti-persistent and persistent fGns: the addition of noise didn't affect the performances of BIC for persistent series, and even tended to enhance the performance of AIC. On the contrary, the addition of white noise generated a number of incorrect specifications for anti-persistent series, especially for the lowest H values (H = 0.1 and H = 0.2). Generally, BIC gave better results than AIC.

In these tests ARFIMA models represented 99.8% of the models selected by BIC, and 94.8% of the models selected by AIC. BIC selected the simplest model (0,*d*,0) for 96.6% of the series, and AIC only for 45.5%. Misspecifications were generally due to the selection of complex ARFIMA models, with parameter d not significantly different from zero. Note that the increase of misspecification rate in anti-persistent fGns with BIC was due to the appearance of non significant values for d in very simple models, such as (0,*d*,1).

Similar results were obtained considering the sums of ARFIMA weights (Figure 3 and 4). Generally these sums were higher for BIC than for AIC. As can be seen in Figure 3, the sums remained close to 1 for the longest series (2048, 1024, and 512 points), and tended to decrease for the shortest series, especially for weakly persistent noises (H = 0.6 and H = 0.7). The same phenomenon was observed for AIC, but the sums were generally lower. Note that in both cases a mean sum of ARFIMA weights of about 0.6 was obtained for series with H = 0.5.

Figure 3 also reports the variability of the obtained sums. The most interesting observation is the very low variability of BIC sums, for the longest series (2048 and 1024 points). The sums obtained were much more variable with AIC, even with long series.

Figure 4 reports the effect of the addition of white noise on the sums of ARFIMA weights. Here also, the effect of noise was clearly different for anti-persistent and persistent fGns: the addition of noise didn't affect the performances of BIC for persistent series, except for H = 0.6. The performances of AIC were degraded for H = 0.6, but enhanced for the highest H values. As previously, the addition of white noise induced a decrease of the sums of ARFIMA weights for anti-persistent series, especially for the lowest H values (H = 0.1 and H = 0.2). Generally, BIC gave better results than AIC.

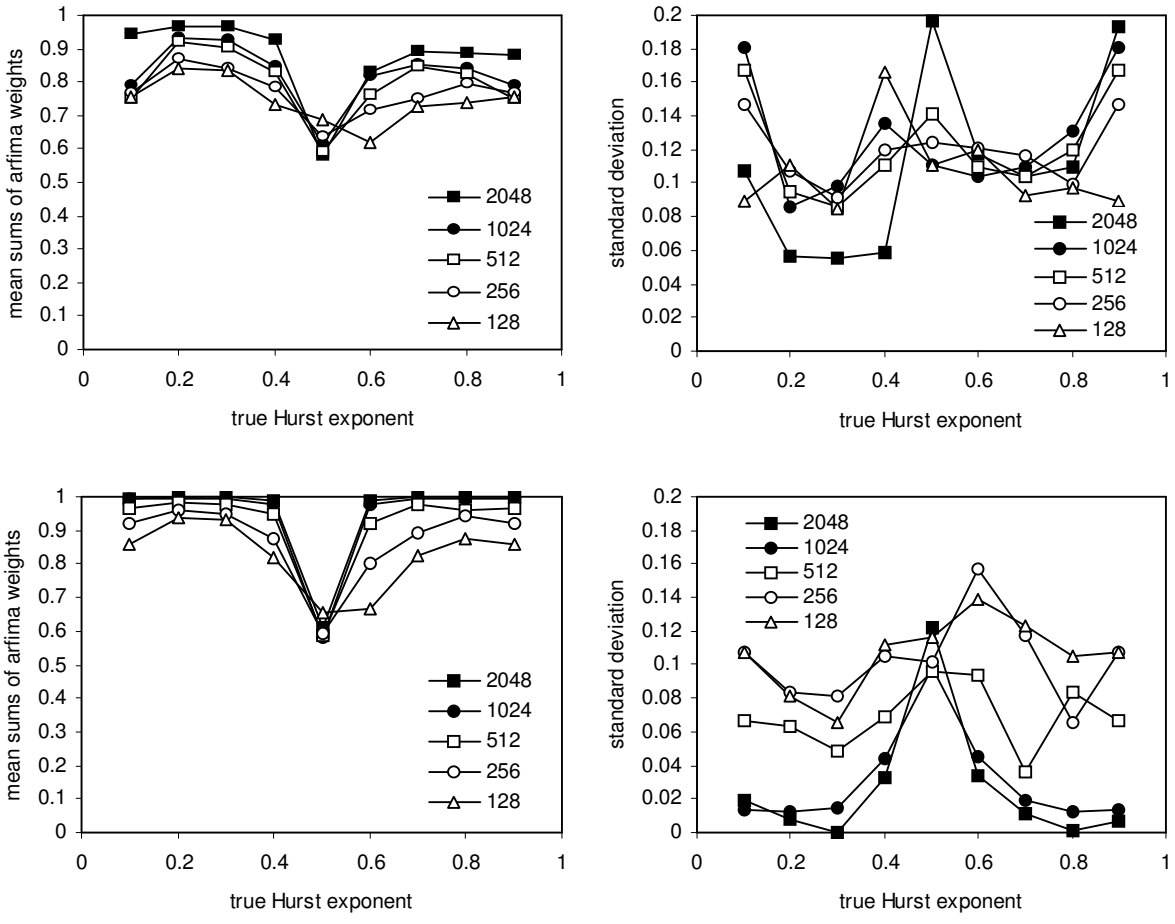

Figure 3: Mean (left panels) and standard deviation (right panels) of the sums of weights of ARFIMA models, according to AIC (top) and BIC (bottom). fGn series, from 2048 to 128 data points.

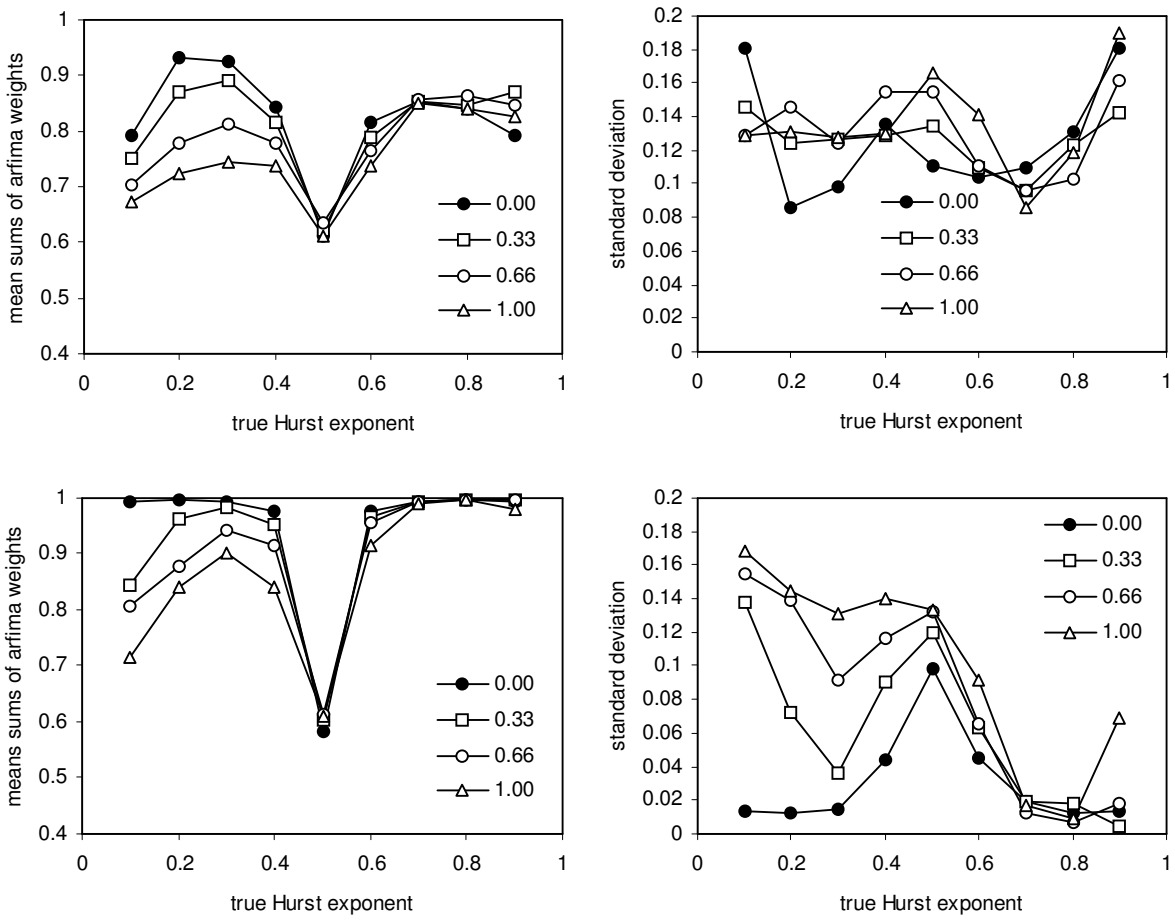

Figure 4: Mean (left panels) and standard deviation (right panels) of the sums of weights of ARFIMA models, according to AIC (top) and BIC (bottom). fGn series (1024 data points) with added noise.

Figure 4 also indicates the effect of white noise on the variability of the obtained sums. Variability remained weakly affected for BIC, concerning persistent series. Conversely, the addition of noise dramatically increased sums variability for anti-persistent series. Variability appeared higher with AIC, whatever the *H* value or the percentage of contamination by white noise.

False detections in simulated short-range dependence series

We simulated AR(1) and MA(1) series, according to Eq. (4) and Eq. (5), respectively, ϕ and θ both varying from 0.1 to 0.9 by steps of 0.1. Forty series were simulated in all cases. In order to test the effect of series length, we applied the analysis on the entire series (2048 points), and then on the first 1024 points, and the first 512 points. We report in Figure 5 the percentage of false detections of long-range dependence (i.e. the best model was an ARFIMA model, with coefficient *d* significantly different from 0).

Whatever series length, long-range dependence was erroneously detected for a number of series. The percentage of misspecification remained limited (around 10%) for BIC, when ϕ or θ exceeded 0.5. In this range of values, the use of AIC resulted in slightly higher levels of false detections. The percentage of misspecification dramatically increased for the lowest values of ϕ and θ. BIC seemed particularly affected, with percentages of false detections between 40 and 60% when ϕ or θ equalled 0.1.

The most frequently selected model was the ARFIMA (0,*d*,1) for MA(1) series (for 56.6% of the series with AIC, and 75.9% with BIC), and the ARFIMA (1,*d*,0) for AR(1) series (for 60.6% of the series with AIC, and 79.5% with BIC). The ARMA model that was actually underlying the series was rarely selected as the best model: the model (0,1) was selected by AIC for 1.0% of the MA(1) series, and by BIC for 2.5%, and the model (1,0) was selected by AIC for 1.3% of the AR(1) series, and by BIC for 2.0%.

Estimation of *H* exponent through ARFIMA modelling

In this part we used again the simulated fGn series, with *H* varying from 0.1 to 0.9. We fitted the nine ARFIMA models (*p*, *d*, *q*), *p* and *q* belonging to {0, 1, 2}, and we retained the coefficient *d* obtained from the best model. *d* was converted into *H* according to the following equation:

$$H = \frac{2d+1}{2} \qquad (16)$$

as $\beta = 2d$, and, for fGn, $H = (\beta + 1)/2$.

Figure 6 indicates the *H* estimations obtained, according to AIC and BIC, as a function of the true exponent. As can be seen, both methods gave rather good mean estimates, at least for the longest series (2048 and 1024 points), despite a slight underestimation of *H* for AIC all along the continuum, and for BIC for anti-persistent fGn ($H < 0.5$). The underestimation dramatically increased for AIC for the shortest series, and especially for $H > 0.3$. On the contrary, BIC seemed to function quite well with short series, except for the shortest ones (128 points), for which an underestimation tendency appeared for $H > 0.3$.

Figure 6 also indicates that both criteria differed in terms of estimation variability. BIC provided very consistent assessments with long series (1024 and 2048 points), except for highly anti-persistent fGns ($H = 0.1$). The variability in *H* estimation increased when series length decreased, especially for persistent fGn. Variability was clearly higher for AIC, even with long series.

The ARFIMA models (0,*d*,0), (1,*d*,0) and (0,*d*,1) represented more than 99% of the models selected by BIC, but in contrast, only 64% of the models selected by AIC. More complex models, as (1,*d*,1) or (2,*d*,2), were selected by AIC for 8.5% and 16% of the series, respectively. Note that these percentages were not affected by series length, suggesting that the degradation of the results with the decrease of series length, in terms of bias or variability, cannot be attributed to a poorer model selection.

Finally we compared the present results with those obtained by analysing the same series with some classical methods in fractal analysis. We applied three analyses: ${}^{low}PSD_{we}$, DFA, and R/S-detrended analysis, selected for the quality of their performances in *H* estimation (Delignières et al., 2005).

${}^{low}PSD_{we}$ is an improved version of spectral analysis, proposed by Fougère (1985) and modified by Eke et al. (2000). This method uses a combination

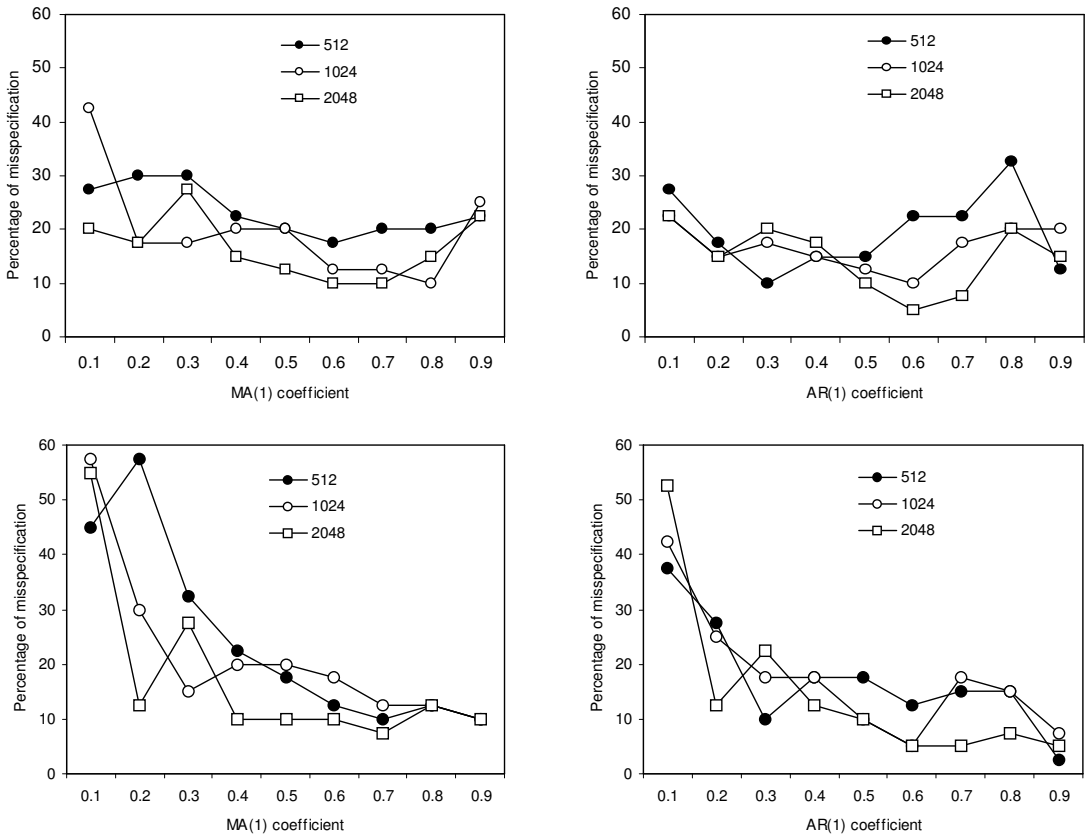

Figure 5: Percentage of misspecifications of MA(1) series (left) and AR(1) series (right). Top panel: AIC; bottom panel: BIC.

of preprocessing operations: Firstly the mean of the series is subtracted from each value, and then the series is tapered by the application of a parabolic window. Thirdly a bridge detrending is performed by subtracting from the data the line connecting the first and last point of the series. Finally the fitting of β excludes the high-frequency power estimates ($f > 1/8$ of maximal frequency). This method was proven by Eke et al. (2000) to provide more reliable estimates of the spectral index. β can be converted into H according to the following equations, which holds for fGn series (Eke et al., 2000):

$$H = \frac{\beta + 1}{2} \tag{17}$$

Detrended fluctuation analysis (DFA) is a well-known method, initially proposed by Peng et al. (1993). This method computes the standard deviation of an integrated, and locally detrended version of the original series, over intervals of increasing length. DFA allows estimation of an exponent α, ranging from 0 to 2. fGn are characterized by α exponents ranging from 0 to 1, and in this range α corresponds to H.

R/S-detrended analysis is an improved version of the classical Hurst method, proposed by Caccia, Percival, Cannon, Raymond, and Bassingthwaigthe. (1997). This method differs from the traditional algorithm by the application of a local detrending of the series of cumulative sums before the calculation of the local range. This method was proved to give more reliable estimates of the fractal exponent than the classical R/S analysis (Caccia et al., 1997; Eke et al., 2000).

Figure 7 allows comparison of the results obtained with ARFIMA modelling (using BIC) and these three methods. In terms of accuracy, ARFIMA modelling gave satisfactory results, as compared with the other methods. The best one was clearly DFA, which provided unbiased estimations, whatever the true underlying H and the length of the series. ${}^{low}PSD_{we}$ tended to underestimate H for anti-persistent fGns, even for the longest series. For shorter series, a progressive underestimation tendency appeared, for $H > 0$. R/S analysis was characterized by a known bias of overestimation for $H > 0.4$ (Caccia et al., 1997), and a slight underestimation for $H = 0.9$. ARFIMA modelling, albeit presenting its own biases, could support the comparison with the other methods.

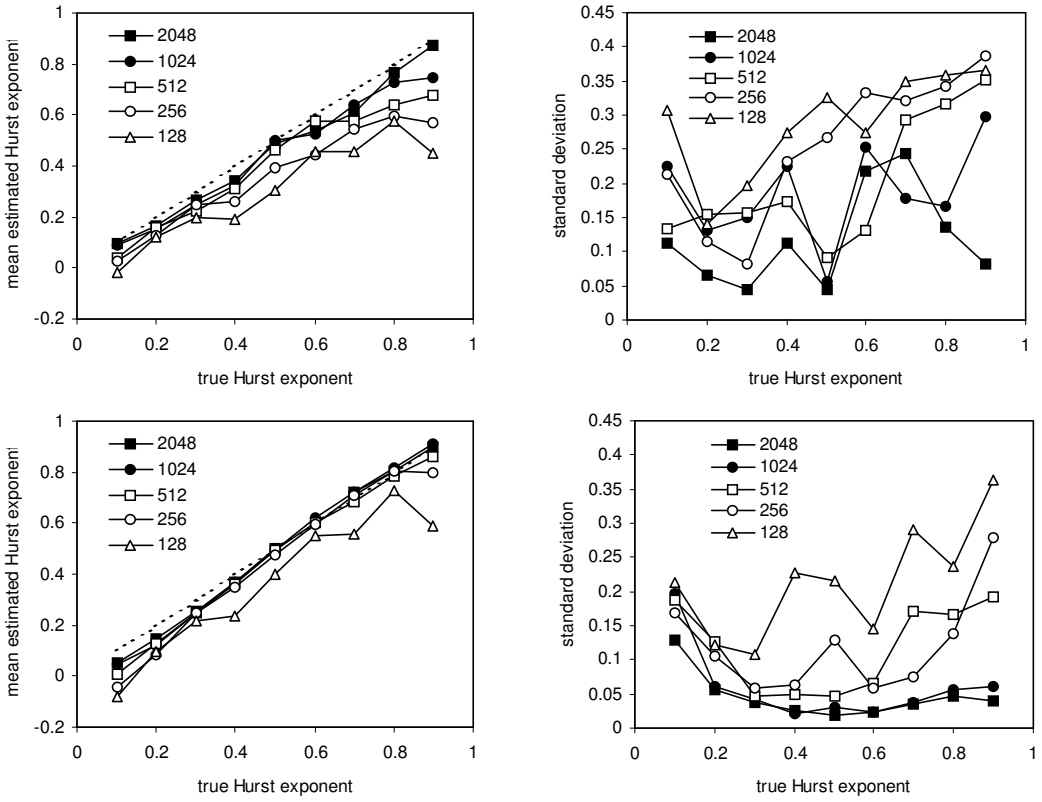

Figure 6: Estimation of the Hurst exponent by ARFIMA modelling, according to the AIC criterion (top panel) or to the BIC criterion (bottom panel). Simulated series of fGn with lengths of 2048, 1024, 512, 256 and 128 data points were used. Left panels report the estimated mean exponent against the true exponent. The dashed line indicates the theoretical equality between estimated and true exponents. Right panels report the variability (standard deviation) of estimation, against the value of the true exponent.

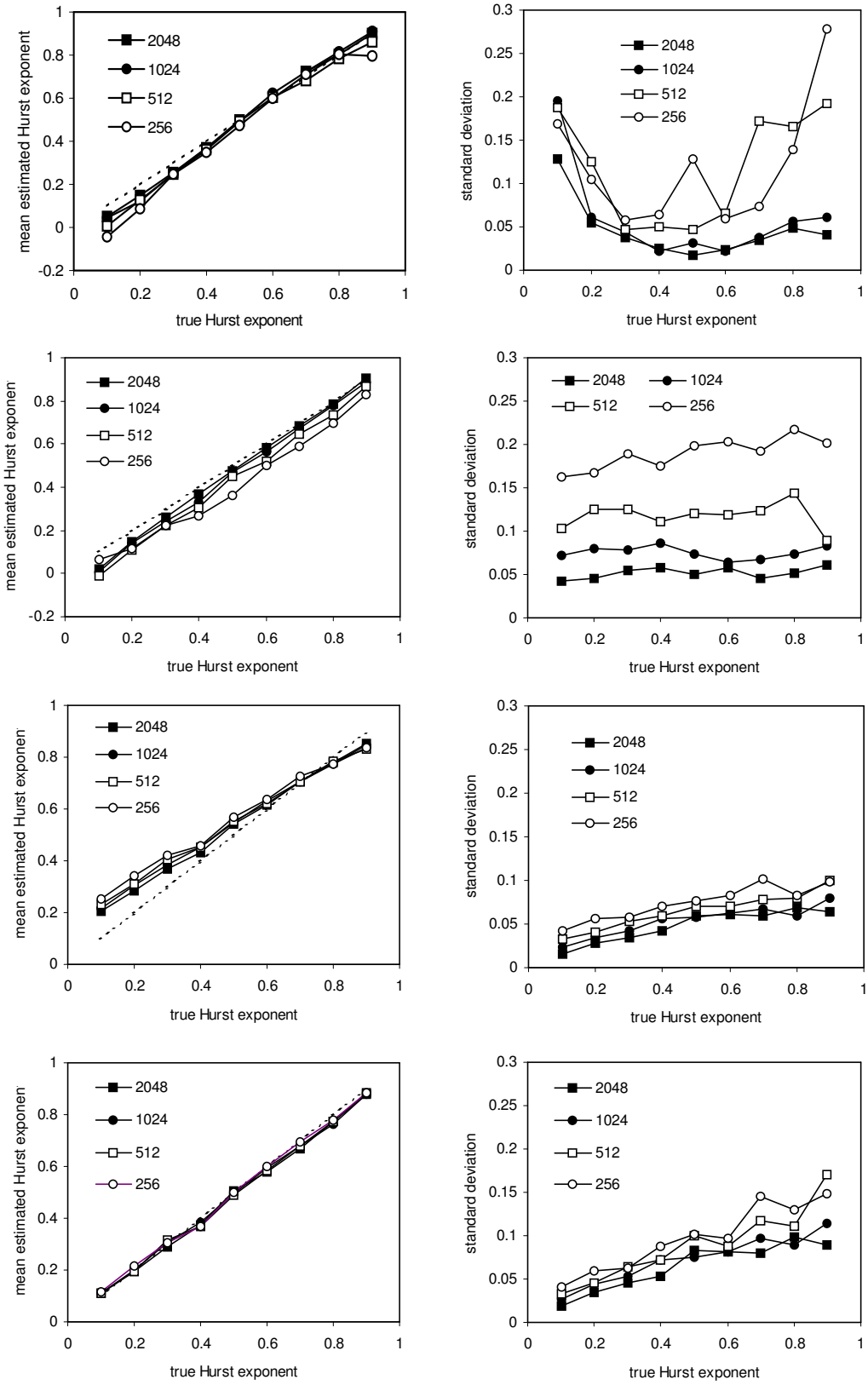

Figure 7: Estimation of Hurst exponent, according, from top to bottom, to ARFIMA modelling (BIC criterion), $^{low}PSD_{we}$, R/S analysis, and DFA. Simulated series of fGn with lengths of 2048, 1024, 512, and 256 data points were used. Left panels report the estimated mean exponent against the true exponent. The dashed line indicates the theoretical equality between estimated and true exponents. Right panels report the variability (standard deviation) of estimation, against the value of the true exponent.

In terms of variability, the performances of the four methods were very different. As can be seen, $^{low}PSD_{we}$ was very affected by series length, with a dramatic increase of variability with the shortest series (512 and 256 points). R/S analysis produced the lowest variability, and was moderately affected by the decrease of series length. Variability, nevertheless, tended to increase with increasing values of *H*. DFA presented a similar pattern of results with, nevertheless, a slightly higher level of variability, especially for persistent fGn. ARFIMA modelling was characterized by the highest levels of variability (1) for short series (512 and 256 points), and (2) for highly anti-persistent fGn ($H = 0.1$ and $H = 0.2$). Conversely, for the longest series (2048 and 1024 points), and for fGn with $H > 0.2$, ARFIMA modelling produced very consistent estimations, comparable in variability to those of R/S analysis.

An empirical example

Finally, we tried to apply ARFIMA modelling to series collected during a recent experiment focusing on the time-evolutionary properties of bi-manual coordination. Thirteen participants performed simultaneous oscillations of the forearms, according to two coordination patterns (in-phase and anti-phase), and following two frequencies (low vs high). The variable of interest was the relative phase between the two effectors, computed punctually at each initiation of the cycle of the dominant limb. Each participant performed these four conditions twice, allowing the collection of 104 series of 1024 points.

We applied ARFIMA modelling to these series, following the previously described procedure. AIC preferred an ARFIMA model for 72 series (70%), and BIC for 92 series (93%). The mean weight for these best ARFIMA models was respectively 0.32 for AIC, and 0.68 for BIC. The mean sum of weights for ARFIMA models was 0.74 for AIC and 0.93 for BIC. Only 4 series were simultaneously classified as ARMA processes by the two criteria (AIC and BIC). Considering the previous indications from our simulation studies, these results provided quite strong evidence for the presence of long-range dependence in these series of relative phase.

We then estimated *d* from the best model selected among the nine ARFIMA candidate models. According to AIC, we obtained a mean value of 0.263 ($SD = 0.167$), corresponding to an *H* value of 0.763, and according to BIC, a mean value of 0.258 ($SD = 0.098$), corresponding to an *H* value of 0.756. These mean *H* values were consistent with those obtained with the other methods ($^{low}PSD_{we}$: $H = 0.723$; DFA: $H = 0.786$, R/S analysis: $H = 0.785$). The samples of exponents obtained with ARFIMA modelling were significantly correlated with the exponents obtained from DFA (AIC: $r = 0.476$; BIC: $r = 0.494$), and from R/S analysis (AIC: $r = 0.462$; BIC: $r = 0.481$). With $^{low}PSD_{we}$ exponents, the correlation was significant for BIC ($r = 0.472$) but not for AIC ($r = 0.157$).

Discussion

A recurrent result, in this paper, was the better performances of BIC, as compared to AIC, in the detection of long-range dependence as well as in the estimation of fractal exponents. Clearly the penalty imposed by BIC on complex models led to a better estimation of the nature of the underlying processes. In most cases the best models for BIC were very simple: often the model didn't contain any short-term processes ((0,*d*,0) models), and in the others cases auto-regressive and moving average processes were rarely combined, and their orders never exceeded 1 ((1,*d*,0) or (0,*d*,1) models). In contrast, AIC generally preferred more complex models, using auto-regressive and moving average terms in combination and with higher orders (e.g. (2,*d*,1) model). As pointed out in our first study, a number of errors with AIC was due in part to the selection of complex ARMA models, such as (2,2) or (1,1), and in part to the selection of complex ARFIMA models, such as (2,*d*,2) or (1,*d*,1), *d* being in this case not significantly different from zero. This suggests that a combination of auto-regressive and moving average terms could mimic long-term dependence and capture an important part of the variance contained in the series. Several authors have demonstrated the possibility of mimicking fractal behaviour by the aggregation of a limited set of short-term processes (Granger, 1980; Hausdorff & Peng, 1996; Pressing, 1999). As such, one could characterise the potential problem posed by AIC in favouring complex models: the combination of auto-regressive and moving average terms seems able in some cases to completely hide the presence of long-term dependence in the series, or to lead to a non-relevant estimation of the parameter *d*.

Nevertheless, it is important to keep in mind that our simulation studies were performed with simulated series of pure fGn, supposed to contain only long-range processes. The ARFIMA model supposed to adequately fit the series should thus be very simple, and this could explain why BIC, in this particular case, gave better results. In contrast, empirical psychological series could be composed of a combination of short-range and long-range processes (Wagenmakers et al., 2004), and AIC could possibly perform better with such series. Additional studies are necessary to test this hypothesis. However at this point, our results should not be considered as definitely banishing the use of AIC, as the overall superiority of BIC could be related to the nature of our simulated series.

We limited our investigations in the present paper to the two classical model selection rules used

by Farrell et al. (2005). A number of alternative criteria have been proposed, such as minimum description length (MDL; Rissanen, 1978) or combined information criterion (CIC; Broersen, 2000). Additional investigations should be necessary for assessing the performance of these model selection criteria for the present purpose. Some improved versions of our classical criteria, such as the corrected AIC (Hurvich & Tsai, 1989), have also been proposed for the analysis of short data records. A 'short series', nevertheless, is defined by the fact that the number of observations is not much larger than the number of parameters in the candidate models. This was clearly not the case in the present study, and we didn't consider it necessary to use these corrected criteria.

The first study suggests that ARFIMA modelling performs quite well for detecting long-range dependence, at least when series have a sufficient length (1024 or 2048 points). These lengths correspond to those generally used in psychological experiments, even if on some occasions shorter series could be encountered. (e.g. Delignières et al., 2004; Kadota, Kudo & Ohtsuki, 2004). Whatever the method, nevertheless, its seems unreasonable to use series shorter than 512 points, and 1024 points seem to represent the best compromise between the requirements of the methods and the inherent limitations of psychological experiments (Delignières et al., 2006). The shortest series we studied here (256 and 128 points) should only be considered as formal examples, aiming at analyzing the behaviour of methods in extreme conditions (for a notable exception, see Madison, 2004).

The two criteria tested in this study (the percentage of ARFIMA models selection and the mean sum of ARFIMA models weights) gave similar results. Nevertheless, they should not be considered as absolutely redundant: for example, the selection of an ARMA as best model can be associated with a sum of weights in favour of ARFIMA models. These occasional discrepancies are not important when an extensive set of series is analyzed, as in the present studies. But when the analysis focuses on a single series, the sum of ARFIMA model weights should represent a better indication of the underlying presence of long-range dependence than the nature of the best model.

The performance of ARFIMA modelling in detecting long-range dependence appeared weakly affected by the presence of noise, at least for persistent fGns. This result is important, because noise is often present in the time series collected in experimental setting, and also because series collected in psychological experiments generally fall in this range of persistent fGns (Wagenmakers et al., 2004).

One can be surprised, in contrast, by the poor performance of ARFIMA modelling in the analysis of pure ARMA series. An important rate of false detections of long-range dependence was observed, especially when the auto-regressive or moving average coefficients were low. This result was particularly evident for BIC, with error rates exceeding 50% for the lowest p and q values. In most cases, the best model was not an ARMA but an ARFIMA model (although d was often not statistically different from 0, except for the lowest p and q values), and the sum of weights of ARFIMA models was always higher (around 0.65) than its ARMA counterpart. Note that when an ARFIMA model was selected with a significant d, this parameter was always close to zero (especially for series of 2048 points), and clearly outside the typical $1/f^{\beta}$ range which should begin at $d = 0.25$ (i.e. $\beta = 0.5$).

These results point out a tendency of the method to favour the selection of ARFIMA models. Surprisingly, Wagenmakers et al. (2004) showed an opposite pattern of results: in their simulation experiment ARMA models were mistakenly identified as ARFIMA models in only 7.5% of the simulated series, whereas ARFIMA models were mistakenly identified as ARMA models in 26.2% of the simulated series, suggesting a bias favouring the selection of ARMA models. The method they used, however, compared only two models, an ARMA (1,1) and an ARFIMA (1,d,1), according to AIC, and their simulated series were for a first set pure $1/f$ noises (contaminated by white noise), and for the second set ARMA (1,1) processes. The relatively high rate of misspecification for the first set could be explained by the fact that the most plausible model - ARFIMA (0,d,0) – was not tested by the authors. On the other hand, the low rate of misspecification for the second set could be related to the presence, in the two tested models, of the ARMA (1,1) that was actually used for generating the series.

Our present results, based on the test of a wider set of candidate models, suggest clearly a bias favouring ARFIMA models. Fractional integration seems to endow ARFIMA models with a greater flexibility than their ARMA counterparts, leading to important rates of spurious detections of long-range dependence. These observations suggest avoiding testing a unique series for the presence of long-range dependence but, rather, to systematically collect a set of series obtained in similar conditions. The rates of misspecification observed in our simulation studies could provide some useful guidelines in the interpretation of results. On the basis of our first simulation study (see Figure 1), one could propose rejecting the hypothesis of the presence of long-range dependence in the series if the percentage of identification of an ARFIMA model is inferior to

90% using BIC, and 70% using AIC. Similar standards could be proposed for the mean sums of ARMIFA models weights, which should approximately reach 0.9 with BIC and 0.7 with AIC for accepting the long-range dependence hypothesis. The set of empirical series we tested in our final study satisfied these two criteria. Wagenmakers et al. (2005) performed the same analyses on the original series collected by van Orden et al. (2003) in a reaction time task and in a word naming task. In the reaction time task, AIC and BIC selected an ARFIMA model for 7 and 4 series over 10, respectively (70 and 40%), and in the word naming task, for 14 and 13 series over 20, respectively (70 and 65%). These results were obviously less convincing than those reported in our final study, and offered more limited support for the presence of long-range dependence in reaction time series.

ARFIMA modelling also appears as a valuable tool for the estimation of fractal exponents. This method gives acceptable results, as compared with classical methods, with limited biases, and a low variability, at least with relatively long series (1024 and 2048 points). Here also, BIC seemed to provide better and less variable estimates than AIC. One can conclude that BIC selects the model comprising the most relevant p and q parameters, allowing a more accurate estimation of d (Taqqu & Teverovsky, 1998).

Obviously, the different methods didn't give exactly the same results, in terms of H exponents, for a given series. The final study showed that with real series, the correlations between the samples of exponents remains moderate, albeit significant. It is important to keep in mind that each method gives an estimation of the fractal exponent. All these methods are supposed to converge toward a common value, with increasing series length. Nevertheless, these methods exploit different properties of fractal series and are based on different algorithms: as such, and with relatively short series, differences in estimated exponents were not surprising. An interesting strategy is to average the estimates obtained from different methods, exploiting different algorithms, in order to obtain a more accurate estimation of the true H value (e.g. Schmidt et al., 1991). ARFIMA modelling could, in this respect, offer a valuable additional estimate.

In conclusion, the method proposed by Wagenmakers et al. (2005) and Farell et al. (2005) constitutes a quite appealing solution for statistically testing for the presence of long-range dependence in times series. A number of problems remain, nevertheless, due to the tendency of the method to favour the selection of ARFIMA models. Further methodological efforts are needed for improving the suitability of the method and allowing less ambiguous identification.

Secondly, ARFIMA modelling provides an interesting method for estimating fractal exponents. Often researchers limit their investigations to a limited set of well-known methods (spectral analysis, DFA, or R/S analysis). ARFIMA modelling constitutes a valuable alternative and a complementary solution for the problem of fractal estimation.

Finally, one could consider ARFIMA as a useful tool for deriving suitable models for the systems underlying the analyzed series. Nevertheless, it's important to keep in mind that a number of models are able to generate series possessing 1/*f* properties, such as, for example, the aggregation of multiple auto-regressive processes (Granger, 1980), the summation of moving-average processes on different time scales (Wing, Daffertshofer, & Pressing, 2004), stochastic delay differential equations (Chen, Ding, & Kelso, 1997), or self-organized critical models (Davidsen & Schuster, 2000). 1/*f* fluctuations represent a ubiquitous phenomenon, and one could doubt that a unique model could work for the diversity of biological or physical systems exhibiting such long-range dependencies (Wagenmakers et al., 2004). A relevant model should present theoretical and empirical plausibility, with regards to current theories and knowledge about the system under study. In this respect, ARFIMA models may not necessarily constitute the most appropriate candidates.

Authors'notes
We thank Sofiane Ramdani for his helpful contribution in the simulation of fGn series.

References

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In B.N. Petrov and F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory*. Budapest: Akademia Kiado.

Bak, P., & Chen, K. (1991). Self-organized criticality. *Scientific American, 264*, 46-53.

Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control*., San Francisco, CA : Holden-Day.

Broersen, P.M.T. (2000). Finite sample criteria for autoregressive order selection. *IEEE Transactions on Signal Processing*, 48, 3550-3558.

Caccia, D.C., Percival, D., Cannon, M.J., Raymond, G., & Bassingthwaigthe, J.B. (1997). Analyzing exact fractal time series: evaluating dispersional analysis and rescaled range methods. *Physica A, 246*, 609-632.

Chen, Y., Ding, M. & Kelso, J.A.S. (1997). Long memory processes ($1/f^{\alpha}$ type) in human coordination. *Physical Review Letters, 79*, 4501-4504

Davidsen, J., & Schuster, H.G.(2000). $1/f^{\alpha}$ noise from self-organized critical models with uniform driving. *Physical Review E, 62*, 6111-6115.

Davies, R.B., & Harte, D.S. (1987). Tests for Hurst effect. *Biometrika, 74*, 95-101

Delcor, L., Cadopi, M., Delignières, D., & Mesure, S. (2003). Dynamics of the memorization of a morphokinetic movement sequence. *Neuroscience Letters, 336,* 25-28.

Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics in Psychology and Life Science, 8,* 479-510.

Delignières, D., Lemoine, L., & Torre, K. (2004). Time intervals production in tapping and oscillatory motion. *Human Movement Science, 23*, 87-103.

Delignières, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006) Fractal analysis for short time series : A reassessement of classical methods. *Journal of Mathematical Psychology, 50,* 525-544

Diebolt, C., & Guiraud, V (2005). A note on long memory time series. *Quality & Quantity*, *39*, 827-836.

Doornik, J.A. (2001). *Ox: An object-oriented matrix language*. London; Timberlake Consultants Press.

Doornik, J.A., & Ooms, M. (1999). A package for estimating, forecasting, and simulating Arfima models: Arfima package 1.0 for Ox [On-line]. Available: http://www.doornik.com/download/arfima.pdf

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendieren Teilchen [*The motion of particles suspended in static liquids as claimed in the molecular kinetic theory of heat*]. *Annalen der Physik, 322*, 549-560.

Eke, A., Herman, P., Bassingthwaighte, J.B., Raymond, G.M., Percival, D.B., Cannon, M., Balla, I., & Ikrényi, C. (2000). Physiological time series: distinguishing fractal noises from motions. *Pflügers Archives, 439*, 403-415.

Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2005). ARFIMA time series modeling of serial correlations in human performance. Manuscript submitted for publication.

Fortes, M., Delignières, D., & Ninot, G. (2004). The dynamics of self-esteem and physical self: Between preservation and adaptation. *Quality and Quantity, 38,* 735-751.

Fougère, P.F. (1985). On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods : simulation studies and application to geographical data. *Journal of Geographical Research, 90 (A5)*, 4355-4366.

Gilden, D.L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science, 8*, 296-301.

Gilden, D.L. (2001). Cognitive emissions of 1/*f* noise. *Psychological Review, 108*, 33-56.

Gilden, D.L., Thornton, T., & Mallon, M.W. (1995). 1/f noise in human cognition. *Science, 267*, 1837-1839.

Gottschalk, A., Bauer, M.S., & Whybrow, P.C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry, 52*, 947-959.

Granger, C.W.J. (1980). Long memory relationships and aggregation of dynamic models. *Journal of Econometrics, 14*, 227-238.

Granger, C.W.J., & Joyeux, R. (1980). An introduction to long-memory models and fractional differencing. *Journal of time Series Analysis, 1*, 15-29.

Hausdorf, J.M., & Peng, C.K. (1996). Multiscaled randomness: a possible source of 1/f noise in biology. *Physical Review E, 54*, 2154-2157.

Hausdorff, J.M., Peng, C.K., Ladin, Z., Wei, J.Y., & Goldberger, A.R. (1995). Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology, 78*, 349-358.

Hurvich, C.M., & Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297-307.

Kadota, H., Kudo, K., & Ohtsuki, T. (2004). Time-series pattern changes related to movement rate in synchronized human tapping. *Neurosciences Letters, 370*, 97-101.

Laming, D.R.J. (1979). Auto-correlation of choice-reaction times. *Acta Psychologica, 43*, 381-412.

Madison, G. (2004). Fractal modeling of human isochronous serial interval production. *Biological Cybernetics, 90*, 105-112.

Mandelbrot, B.B., & van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review, 10*, 422-437.

Marks-Tarlow, T. (1999). The self as a dynamical system. *Nonlinear Dynamics, Psychology, and Life Sciences*, 3, 311-345.

Ooms, M., & Doornik, J.A. (1998). *Estimation, simulation and forecasting for fractional autoregressive integrated moving average models*. Discussion paper, Econometric Institute, Erasmus University Rotterdam, presented at the fourth annual meeting of the Society for Computational Economics, June 30, 1998, Cambridge, UK.

Peng, C.K., Mietus, J., Hausdorff, J.M., Havlin, S., Stanley, H.E., & Goldberger, AL. (1993). Long-range anti-correlations and non-Gaussian behavior of the heartbeat. *Physical Review Letters, 70*, 1343-1346.

Pressing, J. (1999). Sources of 1:f noise effects in human cognition and performance. *Paideusis, 2*, 42-59.

Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics, 76*, 339-347.

Rangarajan, G., & Ding, M. (2000). Integrated approach to the assessment of long range correlation in time series data. *Physical Review E, 61*, 4991-5001.

Ratcliff, R., & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333-367.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14,* 465-471.

Schmidt, R.C., Beek, P.J., Treffner, P.J., & Turvey, M.T. (1991). Dynamical substructure of coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 635-651.

Slifkin, A.B., & Newell, K.M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7*, 170-177.

Spray, J.A., & Newell, K.M. (1986). Time series analysis of motor learning: KR versus no-KR. *Human Movement Science, 5*, 59-74.

Taqqu, M.S., & Teverovsky, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance time series. In R. Adler, R. Feldman, and M.S. Taqqu (Eds.), *A practical guide to heavy tails: Statistical techniques and applications* (pp. 177-217). Boston: Birkhauser.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J.D. (1992). Testing for nonlinearity in time series: The method of surrogate data. *Physica D, 58,* 77-94.

Van Orden, G.C., Holden, J.C., & Turvey, M.T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331-350.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192-196.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review, 11*, 579–615.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General, 134*, 108-116.

West, B.J., & Shlesinger, M.F. (1989). On the ubiquity of 1/f noise. *International Journal of Modern Physics B, 3*, 795-819.

Wing, A., Daffertshofer, A., & Pressing, J. (2004). Multiple time scales in serial production of force: A tutorial on power spectral analysis of motor variability. *Human Movement Science, 23*, 569-590.

Wing, A.M., & Kristofferson, A.B. (1973). Responses delays and the timing of discrete motor responses. *Perception and Psychophysics, 14*, 5-12.

Yamada, N. (1995). Nature of variability in rhythmical movement. *Human Movement Science, 14*, 371-384.