Long-range time dependence in the cross-correlation function

B. A. Carreras and D. E. Newman
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-8070

B. Ph. van Milligen and C. Hidalgo
Asociación Euratom-Ciemat, 28040 Madrid, Spain

(Received 30 July 1998; accepted 2 November 1998)

Detection of a long-range time dependence in the radial cross-correlation function is normally difficult because of the oscillatory behavior of the cross-correlation tail, its low level of coherence, and noise contamination. This problem persists, even with large statistical samples. In this paper, a method for investigating long-range dependence in a single time series is extended to the calculation of the cross-correlation function. With this method and for time series with long-range time correlations, the accuracy of the determination of the cross-correlation function for long time lags is improved. The method is tested by applying it to fractional Gaussian noise and to the fluxes in a running sandpile model. This analysis technique can be applied to the detection of avalanche-type transport in magnetic confinement devices. © *1999 American Institute of Physics.* [S1070-664X(99)02402-7]

I. INTRODUCTION

Some of the phenomena observed in plasmas confined by magnetic fields suggest that a broad range of space and time scales play an essential role in the dynamics of the plasma. One of the possible explanations[1,2] is that plasma dynamics is governed by self-organized criticality (SOC).[3] Under such an assumption, a feature of the dynamics is the existence of transport events of all sizes that we usually denote as avalanches. Some plasma turbulence models have shown the possibility for such a transport mechanism.[4,5]

To identify the existence of avalanches in the experiments is not easy. A first attempt has been directed toward the identification of algebraic tails in the autocorrelation function of the plasma edge fluctuations.[6] Results from the analysis of fluctuation data from several experiments, including tokamaks, stellarators, and reversed-field pinch, showed the self-similar character of the electrostatic fluctuations with a self-similarity parameter, H, in the range 0.6 to 0.74. Such a character of the plasma edge fluctuations is consistent with transport by avalanches, but that is far from identifying it as the only possible mechanism.

Transport by avalanches is characterized by a broad range of radial scales. This range of scales should appear as an algebraic tail in the radial autocorrelation function. An identification of an algebraic tail in the radial correlation function would require a large number of simultaneous measurements of fluctuations at different radial positions. The number required makes such an approach unthinkable; therefore, we need to find a different approach. Because simultaneous measurements of fluctuations at a few radial positions are possible, it is interesting to investigate the correlation of the events responsible for the long-range time correlations in each radial position. The problem in doing so is the statistical requirements.

In the case of the autocorrelation function, it is not easy to directly detect the long time lag algebraic tail. The reasons are the oscillatory character of the tail, its low level of coherence, and noise contamination. The use of the rescaled range (R/S) analysis[7] or some of its variants[8] allows one to surmount these difficulties. To study the radial correlation of events associated with long time lags is even an order of magnitude harder. The direct calculation of the cross-correlation function does not provide the needed accuracy.

In this paper, we propose a generalization to the cross-correlation function of the techniques used in investigating the long time lag tail of the autocorrelation function.[9,10] With this method, we are able to improve the accuracy of the determination of the existence of algebraic tails for long time lags. The method is tested by applying it to fractional Gaussian noise (fGn)[11] and to the fluxes in a running sandpile model. The results show that the proposed method gives an effective determination of the long-range correlations in situations for which the direct calculation of the cross-correlation function cannot detect it.

The remainder of this paper is organized as follows. In Sec. II, the analysis technique for cross-correlation measurements is introduced. An application to the cross-correlation of fGn sequences is discussed in Sec. III. In Sec. IV, the basic algorithm of the running sandpile model is presented, along with a discussion of the radial cross-correlation of fluxes. In Sec. V, the analysis presented in Sec. II is applied to the sandpile fluxes. It shows the higher accuracy in the determination of the long-range tails of the cross-correlation function. Finally, the conclusions of this paper are presented in Sec. VI.

II. ANALYSIS APPROACH

Many techniques have been proposed to determine the long-range correlations in a time series. One of the first approaches was the R/S method, pioneered by Mandelbrot and Wallis.[7,12] This was followed by several others, among them the scaled windowed variance method[13–15] and the disper-

sional method.[8,16] Some of these methods can be extended to the case of cross-covariance of two time series.

First, we briefly discuss the basis for these analysis methods.[9] Given time series of length n, $X \equiv \{X_t : t = 1,2,...,n\}$, corresponding to a stationary process, all second-order properties of this time series are given by the autocovariance function $\gamma_\Delta = \mathrm{cov}(X_t, X_{t+\Delta})$, where Δ is the time lag. However, an alternative representation of the second-order properties of this series can be constructed through averaging the original time series over nonoverlapping blocks. That is, for each $m = 1,2,...,n$, we construct a new time series, $X^{(m)} \equiv \{X_u^{(m)} : u = 1,2,...,n/m\}$, with n/m elements by first separating the original one in n/m nonoverlapping blocks of m elements. Each element of the new series is obtained by averaging the m elements in each of the blocks. That is,

$$X_u^{(m)} = \frac{X_{um-m+1} + \cdots + X_{um}}{m}. \tag{1}$$

For each $X^{(m)}$ series, we can define its variance $V^{(m)}$, with $\sqrt{V^{(m)}}$ measuring the relative dispersion at each scale m. It has been observed that the ratio of the logarithm of $\sqrt{V^{(m)}/V^{(1)}}$ to the logarithm of the scale m is $1-D$, where D is the fractal dimension of the series. Here D is also related to the Hurst exponent,[7,17] $H = 2 - D$.

The variance of the subseries $X^{(m)}, V^{(m)}$, can also be directly related to the autocovariance function,

$$V^{(m)} = \frac{V^{(1)}}{m} + \frac{2}{m^2} \sum_{s=1}^{m} \sum_{\Delta=1}^{s-1} \gamma_\Delta . \tag{2}$$

This relation separates the short and long time lag contributions. For a random variable X, the first term on the right-hand side (rhs) of Eq. (2) dominates, and in the limit of large m, we obtain the Gaussian statistics result, that is, the standard deviation decreases as one over the square root of the number of samples. However, for processes with long-range time dependence, the second term may diverge in the large m limit. In this situation, X does not verify the conditions of the Central Limit Theorem, and $V^{(m)}$ does not scale as m^{-1}. Equation (2) can also be used to calculate the autocovariance function of the original series in terms of the variance of the successive averaged subseries. Inverting Eq. (2), we obtain

$$\gamma_\Delta = \tfrac{1}{2} \partial^2 (\Delta^2 V^{(\Delta)}). \tag{3}$$

Here, the operator ∂^2 is the second-order central derivative operator in finite differences; that is,

$$\partial^2 (f_i) = f_{i+1} + f_{i-1} - 2 f_i . \tag{4}$$

Equation (3) shows that the information on the variance for all time series, $X^{(m)}$, resulting from averaging the original sequence is equivalent to the information contained in the autocovariance function for the original series, X. Therefore, for an infinite series, $V^{(m)}$ provides an alternative equivalent description of the second-order properties of the original series. The advantage of using $V^{(m)}$ to determine the asymptotic behavior of the autocorrelation function instead of using γ_m directly is that the latter is a second derivative of the former. Therefore, small oscillations in $V^{(m)}$ are amplified in γ_m. For noisy signals, it is practically impossible to extract the scaling behavior from γ_m, but in many cases it is possible from $V^{(m)}$.

Because of this advantage, we would like to generalize Eq. (3) to the cross-covariance function for two time series. We will take these two series to be measurements at two separate radial positions. This is motivated by the applications we have in mind, and it is not a limitation of the technique that can be applied to any two series. Let us consider two time sequences of length n: the first one $X \equiv \{X_t : t = 1,2,...,n\}$ is measured at a radial position r_1; and the second one $Y \equiv \{Y_t : t = 1,2,...,n\}$ is measured at a radial position $r_2 = r_1 + \delta$. For each series, we can calculate the series of partial averages, that is, $X^{(m)} \equiv \{X_u^{(m)} : u = 1,2,...,n/m\}$ and $Y^{(m)} \equiv \{Y_u^{(m)} : u = 1,2,...,n/m\}$, where the elements of these series, $X_u^{(m)}$ and $Y_u^{(m)}$, are given by Eq. (1). We assume that each of the two sequences of partial averages has zero mean to avoid unessential complications in the following calculations.

For each pair of time sequences, the cross-covariance is defined in the following way:

$$\gamma^{(m)}(\Delta, \delta) = \frac{m}{n} \sum_{u=1}^{n/m} X_u^{(m)} Y_{u+\Delta}^{(m)} . \tag{5}$$

Here, $\gamma^{(1)}(\Delta, \delta)$ is the usual cross-covariance of the original sequence, X and Y. Now the problem is to relate $\gamma^{(m)}(0,\delta)$ to $\gamma^{(1)}(\Delta, \delta)$ in a way analogous to Eq. (3). To do so, we must calculate $\gamma^{(m)}(0,\delta)$ in an explicit way:

$$\begin{aligned} \gamma^{(m)}(0,\delta) &= \frac{m}{n} \sum_{u=1}^{n/m} X_u^{(m)} Y_u^{(m)} \\ &= \frac{1}{mn} \sum_{u=1}^{n/m} \left(\sum_{i=1}^{m} X_{um-m+i} \right) \left(\sum_{j=1}^{m} Y_{um-m+j} \right). \end{aligned} \tag{6}$$

Expanding and regrouping terms, we obtain

$$\begin{aligned} \gamma^{(m)}(0,\delta) &= \frac{1}{m} \gamma^{(1)}(0,\delta) + \frac{1}{m^2} \\ &\quad \times \sum_{i=1}^{m-1} \sum_{\Delta=1}^{m-1} [\gamma^{(1)}(\Delta,\delta) + \gamma^{(1)}(-\Delta,\delta)]. \end{aligned} \tag{7}$$

This relation is a generalization of Eq. (2). It relates the radial correlation of the partial sums series to the radial correlation at nonzero lags. This relation can be inverted in the same way as it was done for Eq. (2), and leads to

$$2\gamma_s(m,\delta) \equiv \gamma^{(1)}(m,\delta) + \gamma^{(1)}(-m,\delta) = \partial^2 [m^2 \gamma^{(m)}(0,\delta)], \tag{8}$$

where ∂^2 is the second-order finite-difference operator, Eq. (4), and $\gamma_s(m,\delta)$ is the symmetric component of the cross-covariance. Equation (8) tells us that calculating the cross-covariance of the partial averages at zero lag is equivalent to calculating the symmetric part of the cross-covariance for a finite time lag. For the autocorrelation, this method was ad-

vantageous, because by taking the sums we eliminate some of the high frequency that we are not interested in and dominate the correlations.

The scaling of $\gamma^{(m)}(0,\delta)$ with m provides further information, as it was the case of the variance of a single series. When $\gamma^{(m)}(0,\delta)$ falls off with m as a power, $\gamma^{(m)}(0,\delta) \approx \gamma^{(1)}(0,\delta) m^{-\beta}$, Eq. (8) gives

$$\gamma_s(m,\delta) \approx \tfrac{1}{2}(2-\beta)(1-\beta)\gamma^{(m)}(0,\delta). \tag{9}$$

Here, we assume that $2>\beta>0$. Equation (9) shows that the symmetric part of the cross-covariance, $\gamma_s(m,\delta)$, has the same power tail in m as $\gamma^{(m)}(0,\delta)$. Therefore, to determine this power tail is easier to do it for $\gamma^{(m)}(0,\delta)$. If we try to calculate it from the cross-covariance, we have to take two derivatives of $\gamma^{(m)}(0,\delta)$, which clearly amplifies the noise. The advantages of this method are similar to those when it is applied to the autocorrelation function. With this method, we have the level of correlation when a power tail exists, and it can be detected even for very low cross-correlation values. However, a caveat is needed. In the case of zero cross-correlation, Eq. (8) only gives zero equal zero. Using real data, we do not get zero. Therefore, a determination of a power decay exponent in this situation is misleading. Only when the correlation is nonzero does the determination of the exponents make sense.

When the cross-covariance is a power function of the time lag, we can introduce an effective cross-correlation function based on Eq. (9). Dividing both sides of Eq. (9) by the square root of the product of the variance of each series, we obtain the symmetric component of this cross-correlation,

$$\rho_s(m,\delta) = \frac{(2-\beta)(1-\beta)}{2} \frac{\gamma^{(m)}(0,\delta)}{\sqrt{V_1^{(1)} V_2^{(1)}}}, \tag{10}$$

where $V_1^{(1)}$ and $V_2^{(1)}$ are the variance for the two input series. We determine β for each m by calculating $-\ln[\gamma^{(m)}(0,\delta)/\gamma^{(1)}(0,\delta)]/\ln(m)$. For $\delta=0$, this is the prescription of the dispersional method, and $\beta=2D-2$. Having β, we can calculate $\rho_s(m,\delta)$ from Eq. (10). This expression is only valid for a simple power dependence, but it is a useful measure if there are not strong deviations from a power law. In the latter case, we consider β to be a weak function of m and determine its value by doing a power fit over short m intervals. In analogy to the autocorrelation function, we can define an H exponent for the cross-covariance by $H=1-\beta/2$.

In what follows, we compare $\rho_s(m,\delta)$ calculated using Eq. (10) with

$$\hat{\rho}_s(m,\delta) = \frac{\gamma^{(1)}(m,\delta) + \gamma^{(1)}(-m,\delta)}{2\sqrt{V_1^{(1)} V_2^{(1)}}}, \tag{11}$$

where $\gamma^{(1)}(m,\delta)$ and $\gamma^{(1)}(-m,\delta)$ are calculated using the direct definition of cross-correlation, Eq. (5). From this comparison, we can assess the advantages of using Eq. (10) instead of the direct determination of the cross-correlation.

From the previous discussion, we have identified three different methods for calculating the symmetric component of the cross-correlation function. We have summarized these methods in Fig. 1.

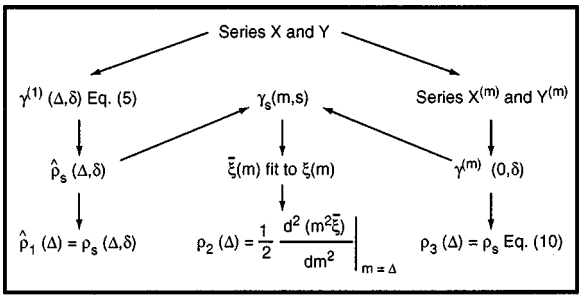

FIG. 1. Three different methods for calculating the symmetric component of the cross-correlation function of two time series.

(1) A direct determination of the covariance $\gamma^{(1)}(\Delta,\delta)$ using Eq. (5), followed by the calculation of $\hat{\rho}_s(\Delta,\delta)$ through Eq. (11). This is the usual method of calculating the cross-correlation function. We denote the symmetric component of the cross-correlation evaluated this way by

$$\rho_1(\Delta) \equiv \hat{\rho}_s(\Delta,\delta). \tag{12}$$

(2) Once we have evaluated $\gamma^{(m)}(0,\delta)$ as described above, we take $\xi(m) \equiv \gamma^{(m)}(0,\delta)$. Alternatively, if we have evaluated $\hat{\rho}_s(\Delta,\delta)$ by the previous method, we invert Eq. (8) by integrating twice; that is, we calculate

$$\xi(m) = \frac{2}{m^2} \int \int_0 d\Delta \hat{\rho}_s(\Delta,\delta). \tag{13}$$

For all cases we have considered, $\xi(m)$ is a smooth function of m with a clear algebraic tail. We can do a fit to this function using a simple parametrization, $\bar{\xi}(m)$. Then we can calculate analytically the second derivatives of $\bar{\xi}(m)$ and recover a smoother form for the cross-correlation function,

$$\rho_2(\Delta) \equiv \frac{1}{2} \frac{d^2(m^2\bar{\xi})}{dm^2}\bigg|_{m=\Delta}. \tag{14}$$

(3) We calculate the cross-covariance of the series $X_u^{(m)}$ and $Y_u^{(m)}$ as a function of time lag m. Using Eq. (10) and β determined by one of the methods previously discussed, we obtain a third form for the cross-correlation,

$$\rho_3(\Delta) \equiv \rho_s(\Delta,\delta). \tag{15}$$

In Sec. IV, we discuss an example of the second method applied to the fluxes of a running sandpile model. In this case, the cross-covariance of the partial sums, $\gamma^{(m)}(0,\delta)$, can be approximately described over a broad range of time lags by a function of the form

$$\gamma^{(m)}(0,\delta) \approx \frac{c_1}{(1+m/m_0)^{\beta}}, \tag{16}$$

where c_1 is a constant. We can use this analytical parametrization to test the approximations discussed above. We have calculated $\gamma^{(m)}(0,\delta)$ for $\beta=1.7$ and $m_0=700$. In Fig. 2, we have plotted $\gamma_s(m,\delta)$ calculated from Eq. (8) and its approximate form given by Eq. (9). We can see that the cross-covariance $\gamma_s(m,\delta)$ has a zero that distorts the algebraic tail for short time lags. However, the approximate form, Eq. (9), has a well-defined asymptotic tail, describes well the sym-

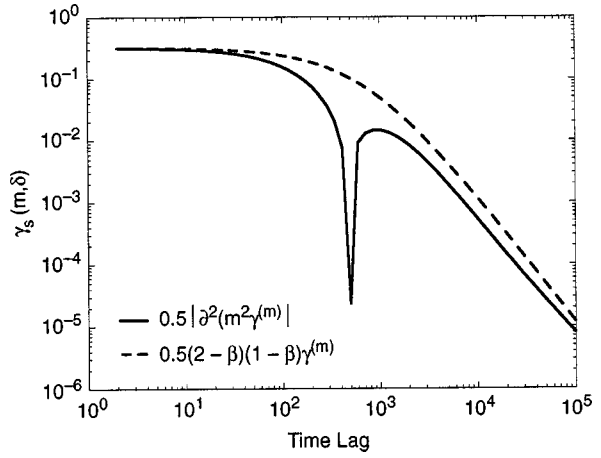

FIG. 2. Symmetric component of the cross-covariance $\gamma_s(m,\delta)$ calculated from a cross-covariance $\gamma^{(m)}(0,\delta)$ given by Eq. (16) and for $\beta=1.7$ and $m_0=700$. The figure shows $\gamma_s(m,\delta)$ calculated from Eq. (8) and its approximate form given by Eq. (9).

metric cross-covariance in the regions of power dependence, and gives its value within a factor of 2. The same can be said for the cross-correlation given by Eq. (10) because it only differs from the cross-covariance by a normalization constant.

In Eq. (5), the number of terms used to calculate $\gamma^{(m)}(0,\delta)$ decreases as n/m when m increases. We have found that unless $m>n/10$, the calculated $\gamma^{(m)}(0,\delta)$ is not reliable. In case of a noisy signal, it is also important to use an ensemble average of several subsamples. Both conditions set serious constraints on the required statistics for these studies.

III. ANALYSIS OF CROSS-COVARIANCE BETWEEN SERIES OF FRACTIONAL GAUSSIAN NOISE

After having developed a method of analysis for the cross-covariance of two time series, we would like to use some model-generated series to test the effectiveness of the analysis. To better understand these techniques, it is useful to start with fGn.[11] To do so, we first had to construct a fGn

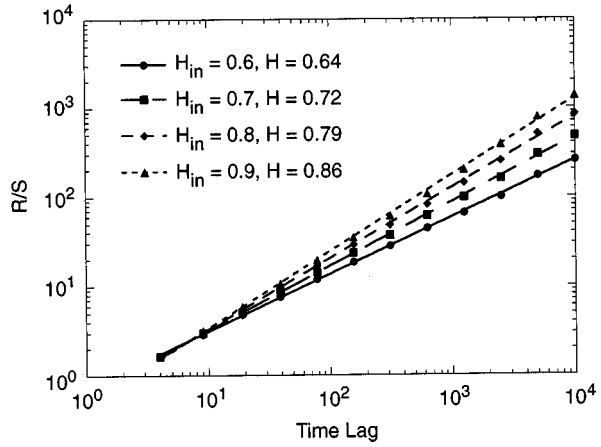

FIG. 3. The R/S analysis of four fGn time series. In this figure, $H_{\rm in}$ is the input value for generating the series and H is the result of the fit to the R/S values.

TABLE I.

$H_{\rm in}$	H_X	H_Y	$H_{\rm CC}$
0.9	0.83	0.85	0.99
0.7	0.66	0.69	0.96
0.5	0.52	0.52	0.82

time series. As a generator for these time series, we have used the method described in Ref. 18 and summarized in the Appendix. In generating the fGn sequences, we need as input a sequence of random numbers. Using this method, we have calculated several fGn sequences of 10 000 points. To test the reliability of the fGn generator, we have applied the R/S analysis to these time series; the results are shown in Fig. 3. In this figure, $H_{\rm in}$ is the input value for generating the series; the fit to the R/S values shows the resulting value of H. The results seem to be in agreement with the usual R/S slight bias toward the 0.7–0.8 range.

To calculate the cross-covariance, we have to generate pairs of sequences. We have considered two types of pairs.

(1) Two sequences with the same value of $H_{\rm in}$, generated with independent random number sequences. We expect these series to be totally uncorrelated. The results of the correlation analysis of these series are shown in Table I. In this table, H_X and H_Y are the H exponents of the series X and Y, respectively, and $H_{\rm CC}$ is the corresponding H for the cross-covariance.

(2) Two sequences with different values of $H_{\rm in}$, but generated with the same sequence of random numbers. This should create a correlation between the series. The results of the analysis of these series are shown in Table II. The notation in Table II is the same as for Table I.

For the time sequences of Table I, the $H_{\rm CC}$ is larger than both H_X and H_Y, and it is close to 1. However, for the sequences in Table II, the exponent of the cross-covariance is between the exponents of the two input series. The main difference between the cases in Table I and Table II is the level of correlation between the input pair of time sequences. In the case of decorrelated sequences, the cross-covariance tends to be independent of m, and its value decreases with increasing numbers of samples that we average. This is reflected in $H_{\rm CC}$ having no relation to the H exponent of the two input sequences. These results are illustrated in Fig. 4, where we have plotted $|\gamma^{(m)}(0,\delta)|$ for $H_{\rm in}=0.5$. We have plotted this function for a single sample and for the average over ten samples. In doing the average, the value of H only changes from 0.8 to 0.76. However, the covariance value decreases considerably, and the function becomes noisier. In the same figure, we also compare $\gamma^{(m)}(0,\delta)$ with $V^{(m)}$ for the input series, showing the clear change in the functional de-

TABLE II.

$H_{\rm in}$	H_X	H_Y	$H_{\rm CC}$
0.9–0.6	0.82	0.57	0.71
0.8–0.6	0.75	0.57	0.67
0.6–0.6	0.57	0.57	0.57

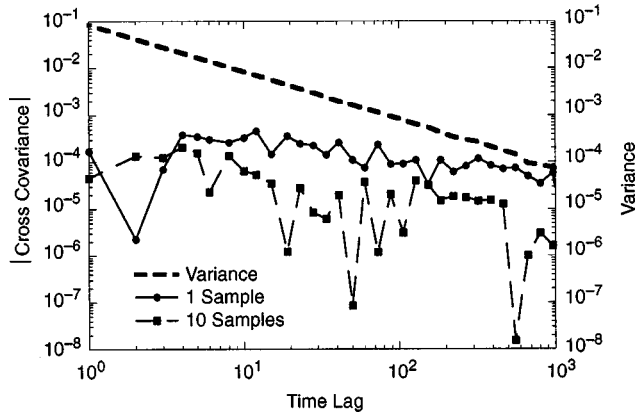

FIG. 4. Cross-covariance $|\gamma^{(m)}(0,\delta)|$ for two $H_{\text{in}}=0.5$ time sequences for a single time sample and averaged over 10 time samples. The covariance function is also compared to the variance of a single sequence.

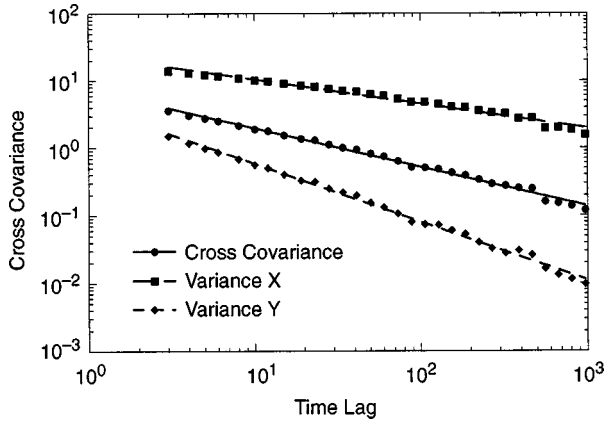

FIG. 6. Cross-covariance and the variances for two fGn correlated series.

pendence on m. If we look at both the autocorrelation and the cross-correlation of this series calculated using Eq. (10), we see that both are very low and oscillate (Fig. 5). The value decreases with the number of samples that we average. All of this indicates that no correlation exists.

When the time sequences are correlated (examples in Table II), $\gamma^{(m)}(0,\delta)$ has an m dependence close to the $V^{(m)}$ of the original input sequences. This is shown in Fig. 6. Calculating the autocorrelation for both series and their cross-correlation using Eq. (10), we have a well-defined behavior with m, and the three functions remain positive. This clearly indicates that both series are correlated, and we can obtain an estimate for the correlation level using the third method described in Sec. II; that is, applying Eq. (10) to $\gamma^{(m)}(0,\delta)$. It is interesting to compare the symmetric component of the cross-correlation function calculated directly by Eq. (5), $\rho_1 \equiv \hat{\rho}_s(m,\delta)$, with the one calculated by the third method proposed in this paper, that is, $\rho_3 \equiv \rho_s(m,\delta)$ from Eq. (10). Since the covariance in this case is a simple power law, the result of using the second method described in Sec. II is identical to using the third method. This comparison is shown in Fig. 7. For time lags below 50, the two calculations give very similar results. However, $\hat{\rho}_s(m,\delta)$ becomes very noisy for longer time lags, and it is impossible from the direct calculation to determine its functional dependence at long time lags. For $\rho_s(m,\delta)$, the oscillations and noise are minimal over the whole range of time lags considered.

IV. RADIAL CORRELATION OF FLUXES IN A SANDPILE MODEL

The running sandpile model has been suggested as a paradigm for SOC turbulent plasma transport in magnetic confinement devices.[1,2] The sandpile model has the instability gradients represented by the slope of the sandpile, while the turbulent transport is modeled by the local amount that falls (overturns) when the sandpile becomes locally unstable. The model system is driven by a random ''rain'' of sand grains on the pile. This model allows us to study the dynamics of the transport independent of both the local instability mechanism and the local transport mechanism. Because of the relative simplicity of the model, we are also able to do very long time calculations and collect reasonably large statistical samples.

A standard cellular automata algorithm[19] is used to study the dynamics of the driven sandpile. The domain is divided into L cells, which are evolved in steps. The number of sand

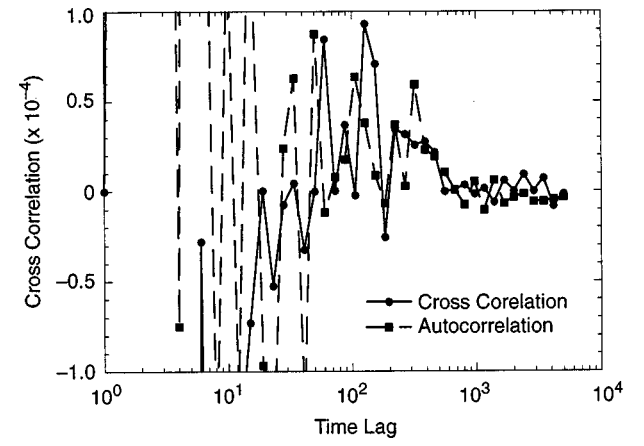

FIG. 5. Autocorrelation and the cross-correlation of the two $H_{\text{in}}=0.5$ time series used in Fig. 4.

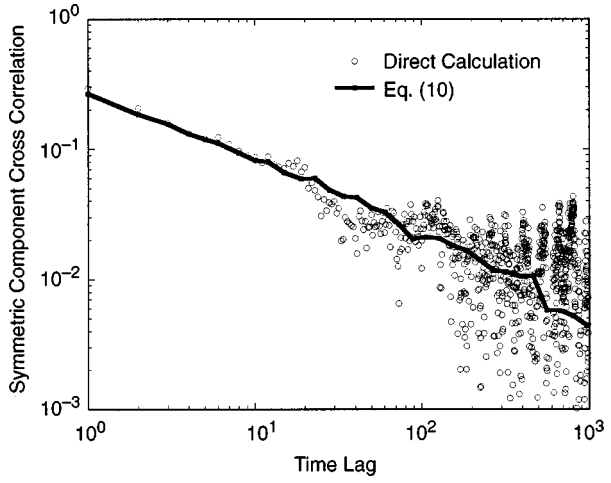

FIG. 7. A symmetric component of the cross-correlation function for two fGn series calculated by the direct method and using Eq. (10).

grains in a cell is h_n, called the height of cell n. We take as radial position the value n that identifies the cell. The local gradient is Z_n, the difference between h_n and h_{n+1}, and Z_{crit} is the critical gradient. The sandpile evolution is governed by the following simple set of rules.

(1) First, sand grains are added to the cells with a probability P_0. For each cell, a random number $0 \leqslant p \leqslant 1$ is drawn, if $p \geqslant 1 - P_0$; then

$$h_n = h_n + 1; \tag{17}$$

otherwise, the height h_n is not changed.

(2) Next, all the cells are checked for stability against a simple stability rule and either flagged as stable, $Z_n < Z_{\text{crit}}$, or not, $Z_n \geqslant Z_{\text{crit}}$.

(3) Finally, the cells are time advanced, with the unstable cells overturning and moving their excess ''grains'' to another cell. That is, if $Z_n \geqslant Z_{\text{crit}}$, then

$$\begin{aligned} h_n &= h_n - N_f, \\ h_{n+1} &= h_{n+1} + N_f. \end{aligned} \tag{18}$$

Here, N_f is the amount of ''sand'' that falls in an overturning event. In terms of the physical quantities that we associate with turbulent systems, each cell can be thought of as the location of a local turbulent fluctuation (eddy). Here Z_{crit} is the critical gradient at which fluctuations are unstable and grow, and N_f is the amount of ''gradient'' that is transported by a local fluctuation (local eddy-induced transport, for example). The average sandpile profile is equivalent to the mean temperature or density profile, while the total number of sand grains in the pile (the total mass) is the total energy/particle content of the device. At any given time, the local flux at a radial position is either zero, if this position is stable, or of N_f, if it is unstable. Time records of the flux at different radial positions are stored during the steady-state phase of the sandpile evolution. In the next section, we will apply the techniques described in Sec. II to analyze these time records, in particular, to the cross-correlation of fluxes in different radial positions.

An analysis of the sandpile fluxes has shown the existence of two characteristic time lag regions.[20] For time lags shorter than the time needed to cross the sandpile, $\Delta t < L$, the Hurst exponent is about 0.8. This indicates the existence of long-range correlations that are induced by avalanches. Similar values of H have been found in the analysis of plasma edge fluctuations.[6] For long time lags, longer than the time needed to cross the sandpile, $\Delta t \gg L$, the analysis of sandpile fluxes gives values of H below 0.5. This corresponds to long-range anticorrelations, which is an indication that once a large avalanche has happened, the probability of another large avalanche decreases. To detect if similar effects exist in plasma fluctuations requires very long time records, longer than the ones used up to now in those types of analysis.

In the sandpile model, each avalanche propagates both uphill and downhill.[21] This is a consequence of the mass conservation during the propagation; grains propagate downhill, while ''holes'' propagate uphill. This form of propagation translates to a particular structure of the cross-correlation of fluxes.[2] In Fig. 8, we show in schematic form three different types of avalanches crossing the radius $r = r_1$ and $r = r_2 = r_1 + \delta$. We measure time in relation to the r_1 position. Avalanches starting at $r < r_1$, for example, at point 1 in Fig. 8, cross r_2 with a time delay $\Delta = \delta$, while the ones starting at $r > r_2$, for example, at point 2 in the same figure, cross r_2 at $\Delta = -\delta$ before reaching r_1. Therefore, we expect the cross-correlation to have two maxima at $\Delta = -\delta$ and $\Delta = \delta$, respectively. The relative height of these maxima is related to the relative probability of avalanches starting at $r < r_1$, $P_0 r_1$, versus the probability of avalanches starting at $r < r_2$, $P_0(L - r_2)$. Also, avalanches start between r_1 and r_2, like at point 3 in Fig. 8. For these avalanches, the time lag lies in the interval $-\delta < \Delta < \delta$. These avalanches give the cross-correlation between the maxima and can result in filling up the minimum between them and giving just a flat region between the peaks.

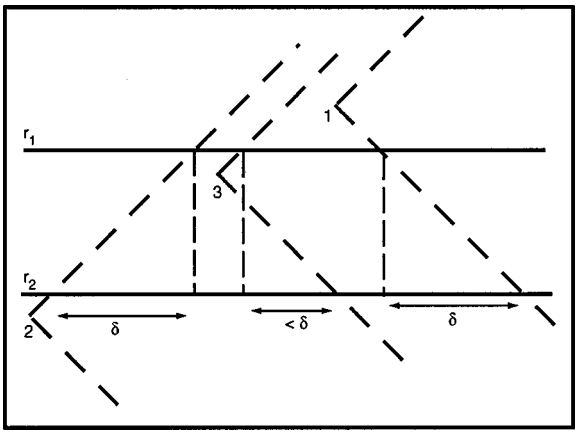

FIG. 8. Three types of avalanches depending on the position of the starting point with relation to the two radial positions where the flux is measured. The vertical axis indicates the radial position in the sandpile and the horizontal axis indicates the time.

The change of the antisymmetric component of the cross-correlation function with a radial position is illustrated in Fig. 9. In this figure, we have plotted the cross-correlation

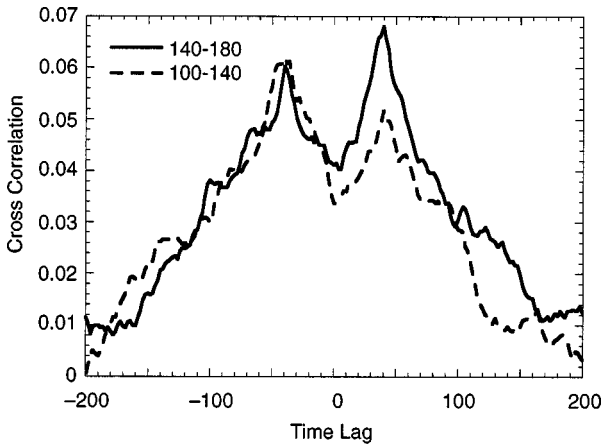

FIG. 9. The cross-correlation function of the fluxes of a sandpile with $\delta = 40$ in relation to the radius of $r = 140$. The antisymmetric component of the cross-correlation function depends on the relative positions of the surfaces.

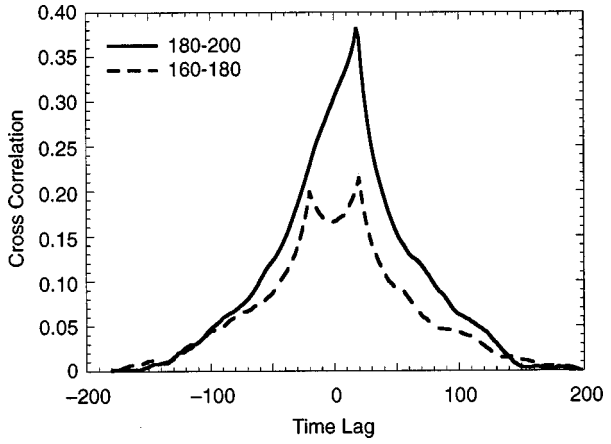

FIG. 10. The cross-correlation function of the fluxes of a sandpile with $\delta=20$ in relation to the radius of $r=180$. At the edge of the sandpile, there is only one possible propagation direction, outward. In this case, the asymmetry is extreme and the cross-correlation larger than toward the inside.

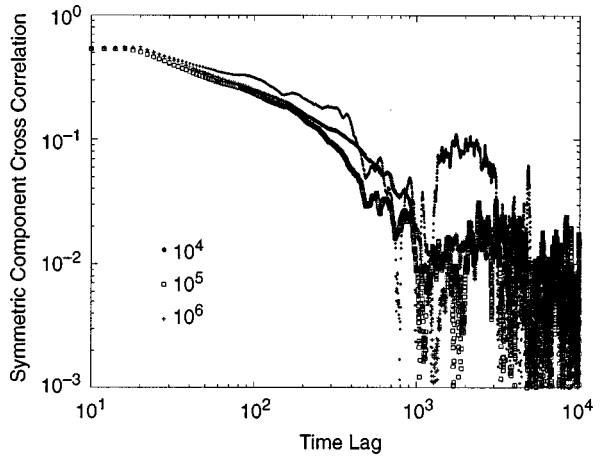

FIG. 12. The absolute value of the cross-correlation function for $\delta=20$ and for three record lengths: 10^4, 10^5, and 10^6 points. As the record length increases, the power dependence of the tail before the first zero becomes better defined.

function for a fixed radial separation, $\delta=40$, at two different radii r_2 on the sandpile. As discussed above, the cross-correlation function has a clear peak at $\Delta=\pm 40$. The antisymmetric component of the cross-correlation changes sign when we move from near the center, $r_1=140$ and $r_2=100$, to near the edge, $r_1=140$ and $r_2=180$, because near the edge there are not many upward propagating avalanches and downward propagation dominates. The avalanche propagation in opposite directions also causes an increased decorrelation. At the edge of the sandpile, there is only one possible propagation direction, outward. In this case, the asymmetry is extreme and the cross-correlation larger than toward the inside (Fig. 10). As δ increases, the correlation decreases, and the point of maximum correlation, $(\Delta)_{\max}$, moves toward high time lags in such a way that $(\Delta)_{\max}=\delta$. This is shown in Fig. 11, where we have plotted the cross-correlation function for a fixed reference position, $r_1=140$, and changing r_2. For the same δ, the symmetric part of the cross-correlation is practically the same, while the antisymmetric contribution varies with r_2, as was shown in more detail in Fig. 8.

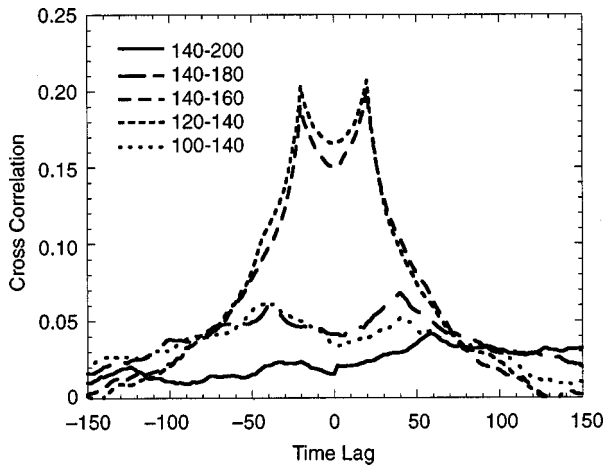

FIG. 11. The cross-correlation function of the fluxes of a sandpile as a function of δ for a fixed reference position, $r_1=140$, and changing r_2.

For time lags in the range of $10^2<\Delta<3\times10^3$, the symmetric component of the cross-correlation goes through zero and has a clear region of anticorrelation. For long time lags, $\Delta>10^3$, the function $\hat{\rho}_s(\Delta,\delta)$ is very noisy and oscillates. From this result, it is clear that we cannot use the direct determination of $\hat{\rho}_s(\Delta,\delta)$ to assess the functional dependence of the cross-correlation at long time lags. For a given set of parameters, the position of the first zero does not sensitively depend on the specific realization or the length of the time series for series lengths above 10^5 points. In Fig. 12, we have plotted the absolute value of the cross-correlation function for $\delta=20$ and for three record lengths: 10^4, 10^5, and 10^6 points. As the record length increases, the power dependence of the tail before the first zero becomes better defined. The position of the first zero of the cross-correlation function depends on the value of P_0. As P_0 increases, the zero moves to shorter time lags and the overall cross-correlation decreases.

The results shown in Figs. 9–12 are for a sandpile of length $L=400$, $Z_{\text{crit}}=10$, $N_f=3$, and the evolution is driven by a rain of grains with a probability varying between $p_0=0.0005$ and $p_0=0.0025$. The cross-correlation is calculated in the direct way using the $\gamma^{(1)}(\Delta,\delta)$ covariance given by Eq. (5).

V. ANALYSIS OF THE SANDPILE RADIAL CORRELATION OF FLUXES

In the previous section, we have discussed the properties of the $\gamma^{(1)}(\Delta,\delta)$ cross-covariance and cross-correlation $\hat{\rho}_s(\Delta,\delta)$ of the sandpile fluxes. Both functions have been calculated by the direct method. From the calculated $\hat{\rho}_s(\Delta,\delta)$, it is difficult to discern the functional form of a high time lag tail. In this section, we test the alternative determination of the long time lag tail of the cross-correlation by the other two methods discussed in Sec. II. Since $\hat{\rho}_s(\Delta,\delta)$ has been already calculated, following the second method, we can calculate $\xi(m)$ by Eq. (13). This function should be equivalent to $\gamma^{(m)}(0,\delta)/\sqrt{V_1^{(1)}V_2^{(1)}}$, calcu-

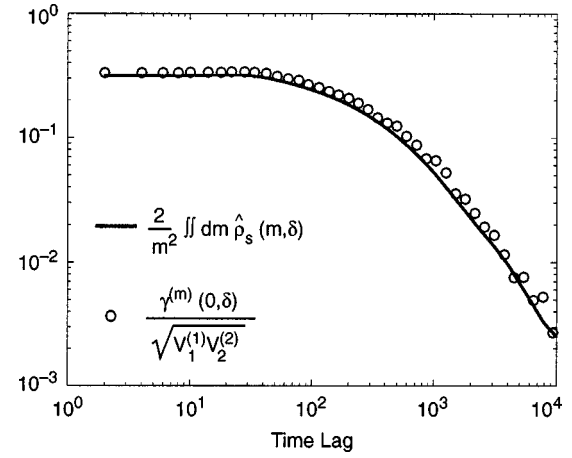

FIG. 13. A comparison between $\xi(m)$ for the sandpile flux calculated using Eq. (13) and $\gamma^{(m)}(0,\delta)/\sqrt{V_1^{(1)}V_2^{(1)}}$, calculated independently by the third method in Sec. II.

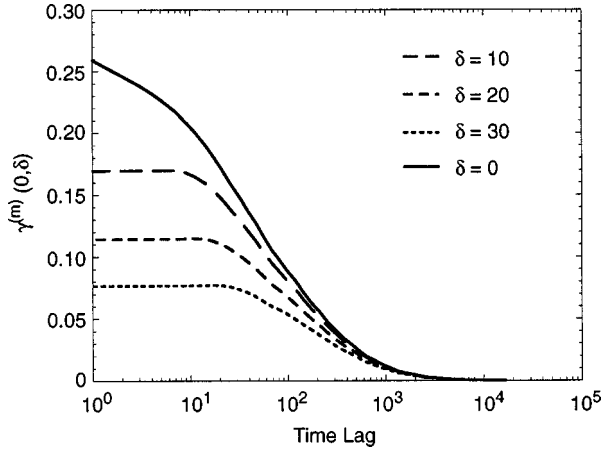

FIG. 15. A symmetric component of the cross-covariance function of the sandpile fluxes for different values of δ.

lated independently by the third method in Sec. II. We have plotted both functions in Fig. 13. The first interesting result is that the functions are smooth and positive. The result of both calculations agree well, and they show the existence of a power tail for long time lags. This result is particularly interesting given the apparently noisy character of the $\hat{\rho}_s(\Delta,\delta)$ function, as shown in Fig. 12. The next step in the second method is to fit by a simple analytical function, $\overline{\xi}(m)$. We used the parametrization given in Eq. (16), which provides a good fit for all the cases that we have considered. Once we have obtained the analytical form, $\overline{\xi}(m)$, we can carry out the final step in evaluating the symmetric component of the cross-correlation function by the three methods outlined in Fig. 1. The functions $|\rho_1|$, $|\rho_2|$, and $|\rho_3|$ are compared in Fig. 14 for the average of three time records of 10^5 points each. Both ρ_2 and ρ_3 show the clear algebraic tails at long time lags that it was not possible to identify in ρ_1. They also show similar correlation levels in the region of the tail. Additionally, the three functions have a zero of the correlation function at about a time lag of 400. This indicates anticorrelation for all lags above the zero crossing. For ρ_3, the zero is always somewhat higher than the other two methods, because in its derivation, we have assumed a power law behavior for $\gamma^{(m)}(0,\delta)$. This approximation fails near the zero. The change from correlation to autocorrelation can also be detected directly in $\gamma^{(m)}(0,\delta)$ (Fig. 13) when the exponent β becomes larger than 1.

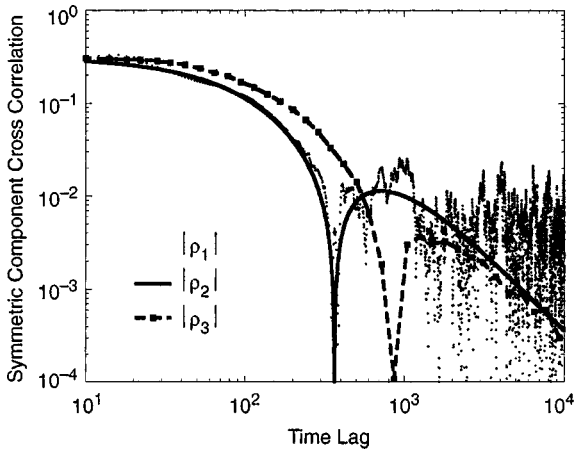

FIG. 14. A symmetric component of the cross-correlation function of sandpile fluxes calculated by the three methods outlined in Fig. 1. The functions $|\rho_1|$, $|\rho_2|$, and $|\rho_3|$ are the average of three time records of 10^5 points each.

To determine the algebraic tail of the cross-correlation function of the sandpile fluxes under different conditions, we can look at $\gamma^{(m)}(0,\delta)$. The results when δ changes are plotted in Fig. 15. At each radial position, we have used six time records of 10^5 points each. For the different time records, $\rho_3(m,\delta)$ has a very similar dependence on the time lag; it does not oscillate or change sign. This is a clear indication that a correlation exists. In Fig. 15, we can see how increasing the radial separation, δ, changes the magnitude of the cross-correlation. The cross-correlation decreases at short time lags because the short avalanches (in length and duration) cannot cross both of the radial surfaces considered. Therefore, as the separation increases, the peak at $\Delta=0$ decreases, and the point of change of $\gamma^{(m)}(0,\delta)$ moves to larger values of the time lag. This change, moving toward longer time lags; corresponds to the shift of the peak of the cross-correlation function, as shown in Fig. 10. The decrease in the peak of $\gamma^{(m)}(0,\delta)$ causes a flattening of this function at short time lags, but its tail for long time lags seems to remain the same. If the tail of $\gamma^{(m)}(0,\delta)$ remains the same, regardless of the value of δ, the flux events associated with it are correlated for relatively large radial separations. To better examine the structure of the tail for long time lags, we plotted in Fig. 16 some functions of Fig. 15 in a logarithmic scale. The agreement on the high time lag tails is clear, which is evidence for the radial correlation of the long avalanches. We can arrive at a similar conclusion by looking at the cross-power spectrum (Fig. 17). The high-frequency range, associated with small nonoverlapping avalanches, is strongly modified from the autopower spectrum. At high frequencies, the autopower spectrum falls off as $\omega^{-\alpha}$, with α in a range 2 to 4, while the cross-power spectrum in the same range of frequencies shows strong oscillations with peaks at frequencies of $1/\delta$ and higher harmonics.[2] However, the $1/f$ and

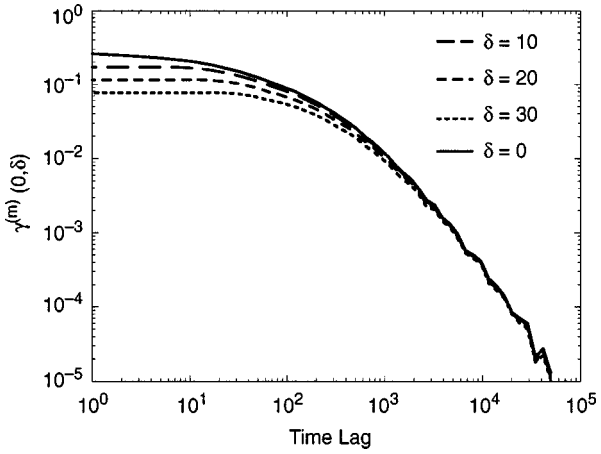

FIG. 16. A symmetric component of the cross-covariance together with the variance for the reference position for the same cases as Fig. 15. The plot is logarithmic scale to better examine the structure of the tail for long time lags.

low-frequency ranges, which are associated with large-scale and long correlation events, are hardly changed.

It is interesting to compare the level of cross-correlation as a function of δ for short time lags (decorrelation of the fluctuations and short avalanches) and for long time lags (correlation of large events). We can see in Fig. 18 how the correlation at short time lags ($m=10$ in the figure) decays systematically with δ, while at longer time lags ($m=967$ in the figure) it stays at a constant level.

The tail of the $\gamma^{(m)}(0,\delta)$ function at long time lags has a power dependence with exponent β varying between 1.3 and 1.7 for the different parameters considered. As previously indicated, the autocorrelation tails have the same type of power dependence. In the case of the autocorrelation, the exponent provides an interpretation for the type of correlation. This β range corresponds to a range in the Hurst exponent from 0.35 to 0.15, which indicates that the events in the tail are anticorrelated. This interpretation is consistent with the cross-correlation having a zero and being negative at the long time lags. Therefore, for the sandpile fluxes and time lags longer than 1000, the events are anticorrelated. This is

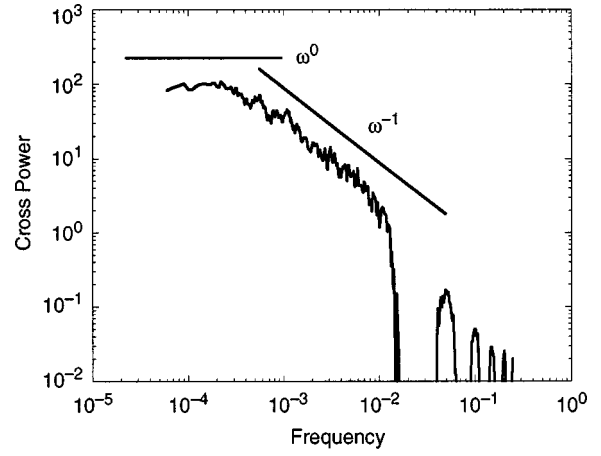

FIG. 17. The cross-power spectrum for sandpile fluxes with $\delta=20$.

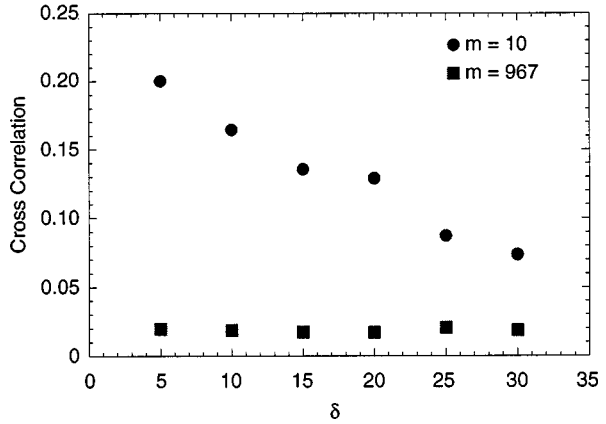

FIG. 18. The cross-correlation of sandpile fluxes as a function of δ for two values of the time lag, $m=10$ and $m=967$.

probably an indication that once a large avalanche has occurred, the probability of another large avalanche decreases.

The alternative methods for determining the long time tail can indeed give insight into the dynamics of the running sandpile. Here, we have limited the study to show the effectiveness of these methods. The implication of the analysis for the sandpile dynamics will be discussed elsewhere.

VI. CONCLUSIONS

The extension of the method to investigate long-range dependence in a single time series to the cross-correlation function has been shown to be effective. Two possible new methods have been suggested. With both methods, and for time series with long-range time correlations, we are able to improve the accuracy on the determination of the cross-correlation for long time lags. For both fractional Gaussian noise and fluxes in a running sandpile model, these new analysis techniques allow the detection of cross-correlation tails well beyond the values of the time lag for which the usual cross-correlation function is dominated by noise. Using the usual determination of the radial cross-correlation function, it is normally difficult to detect those long-range time dependences because of the oscillatory behavior of the cross-correlation tail, its low level of coherence, and the noise contamination.

This approach could be useful in identifying radial correlation of transport events in magnetically confined plasmas. Such types of events are expected if plasmas satisfy the SOC paradigm. In this case, transport is dominated by avalanches. Transport by avalanches is characterized by a broad range of radial scales. This range of scales should appear as an algebraic tail in the radial autocorrelation function. Such an identification could be done by applying the techniques described in this paper to simultaneous measurements of fluctuations at a few radial positions. At present, such work is in progress.

APPENDIX: FRACTIONAL GAUSSIAN NOISE

Let us consider a particle moving on a line and taking a step $\pm\xi$ every unit of time. If the step length is a random variable with Gaussian probability distribution function

$$p(\xi)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\xi^2}{2}\right), \tag{A1}$$

then the sequence of independent Gaussian steps is called Gaussian noise. The position of the particle after a time $t=\tau m$ is given by

$$B_{1/2}(t)=\sum_{i=1}^{m}\xi_i, \tag{A2}$$

with m being the number of steps and τ being the time taken per step. The function $B_{1/2}(t)$ is, therefore, the corresponding Brownian motion associated with the Gaussian noise. This Brownian motion has no time correlations. Mandelbrot and Van Ness[11] introduced a generalization of the Brownian motion, the fractional Brownian motion, that does have long-range time correlations. Given the value at $t=0$, the fractional Brownian motion with index H is defined by

$$B_H(t)-B_H(0)=\frac{1}{\Gamma(H+\frac{1}{2})}\int_{-\infty}^{t}K(t-t')dB(t'). \tag{A3}$$

This formulation shows that the value of the random function $B_H(t)$ at a time t depends on all previous increments $dB(t')$, with the kernel defined as

$$K(t-t')=\begin{cases}(t-t')^{H-1/2}, & 0\leqslant t'\leqslant t,\\ (t-t')^{H-1/2}-(-t')^{H-1/2}, & t'<0.\end{cases} \tag{A4}$$

Since the kernel is a power law function, the fractional Brownian motion is self-similar with self-similarity parameter H,

$$B_H(\lambda t)-B_H(0)=\lambda^H[B_H(t)-B_H(0)]. \tag{A5}$$

Differentiating the fractional Brownian motion, we obtain the fractional Gaussian noise. To simulate fractional Brownian motion, it is useful to express the integral as a summation on discrete time steps. A practical formula for this was given by Mandelbrot and Wallis,[22–24]

$$\begin{aligned}B_H(t)-B_H(t-1)&=\frac{n^{-H}}{\Gamma(H+\frac{1}{2})}\left(\sum_{i=1}^{n}i^{H-1/2}\xi_{1+n(M+t)-i}\right.\\&\left.+\sum_{i=1}^{n(M-1)}[(n+i)^{H-1/2}-i^{H-1/2}]\xi_{1+n(M-1+t)-i}\right).\end{aligned} \tag{A6}$$

Here, $\{\xi_i\}$ with $i=1,2,...,M,\ldots$, is a set of random Gaussian variables with unit variance and zero mean. Here, each integer time step has been divided in n intervals to approximate the integral and the Gaussian noise is averaged over M time steps. Therefore, the number of time steps taken over which the Gaussian noise is averaged should be increased with the series length. The number of intervals, n, controls the accuracy. A more detailed discussion of the fGn can be found in Ref. 18.

ACKNOWLEDGMENT

This work was supported by Oak Ridge National Laboratory, managed by Lockheed Martin Energy Research Corp., for the U.S. Department of Energy under Contract No. DE-AC05-96OR22464.

[1] P. H. Diamond and T. S. Hahm, Phys. Plasmas **2**, 3640 (1995).
[2] D. E. Newman, B. A. Carreras, P. H. Diamond, and T. S. Hahm, Phys. Plasmas **3**, 1858 (1996).
[3] P. Bak, C. Tang, and K. Weisenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[4] B. A. Carreras, D. Newman, V. E. Lynch, and P. H. Diamond, Phys. Plasmas **3**, 2903 (1996).
[5] X. Garbet and R. E. Waltz, Phys. Plasmas **5**, 2836 (1998).
[6] B. A. Carreras, B. van Milligen, M. A. Pedrosa *et al.*, Phys. Rev. Lett. **80**, 4438 (1998).
[7] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **4**, 909 (1969).
[8] J. B. Bassingthwaighte, L. S. Liebovitch, and B. J. West, *Fractal Physiology* (Oxford University Press, New York, 1994).
[9] D. R. Cox, in *Statistics: an Appraisal*, edited by H. A. David and H. T. David (The Iowa State University Press, 1984), p. 55.
[10] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman and Hall, New York, 1994).
[11] B. B. Mandelbrot and J. W. V. Ness, SIAM (Soc. Ind. Appl. Math.) Rev. **10**, 422 (1968).
[12] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 967 (1969).
[13] B. B. Mandelbrot, Phys. Scr. **32**, 257 (1985).
[14] J. G. Moreira, J. K. L. de Silva, and S. O. Kamphorst, J. Phys. A **27**, 8079 (1994).
[15] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond, and J. B. Bassingthwaighte, Physica A **241**, 606 (1997).
[16] J. B. Bassingthwaighte and G. M. Raymond, Ann. Biomed. Eng. **23**, 491 (1995).
[17] H. E. Hurst, Trans. Am. Soc. Civ. Eng. **116**, 770 (1951).
[18] J. Feder, *Fractals* (Plenum, New York, 1988).
[19] L. P. Kadanoff, S. R. Nagel, L. Wu, and S.-M. Zhou, Phys. Rev. A **39**, 6524 (1989).
[20] B. A. Carreras, B. van Milligen, M. A. Pedrosa *et al.*, Phys. Plasmas (to be published).
[21] T. Hwa and M. Kadar, Phys. Rev. A **45**, 7002 (1992).
[22] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 228 (1969).
[23] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 242 (1969).
[24] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 260 (1969).