

Detrended fluctuation analysis of significant wave height time series

L. Cabrera & G. Rodríguez

Departamento de Física, Universidad de Las Palmas de G. C., Spain

Abstract

Mean daily significant wave height time series are analysed by means of the detrended fluctuation analysis (DFA) method to examine the existence of long-range correlations, or long memory. The results indicate that the scaling behavior of significant wave height fluctuations is not constant over the considered time scales but there is a decrease of the scaling exponent with increasing time scale. Two crossover times have been identified, indicating three different scaling behaviors. Fluctuations associated to time scales between 10 and 100 days, approximately, show long-range correlation while fluctuations above and below this range present white and Brownian noise-like behavior, respectively.

1 Introduction

The knowledge of sea state conditions is critical not only for studying many oceanic and atmospheric processes but also for many offshore and nearshore operations. The evolution of sea state conditions is also of utmost importance in coastal erosion, several questions concerning safety, reliability and feasibility of offshore activities, as well as in maritime transport.

It has been highlighted by many researchers [1,2] that the design and construction of marine structures and the coastal and offshore development activities require a relatively accurate knowledge of wave conditions to be expected during those activities and the expected lifetime of the structures. However, acquiring accurate knowledge of wave conditions evolution proves to be a difficult task. Ocean wind waves are generated in the interface connecting the turbulent boundary layers of air and water. The coupling of these two different viscous fluids is strong and nonlinear. Waves are inherently forced by the wind, break intermittently and interact strongly with surrounding turbulence. Moreover, in



a given sea surface area, apart from an actively wind driven wave field, there may be one or more wave fields that have been generated elsewhere in the past and propagated toward this zone. These different wave fields interact through nonlinear interactions. As a result, in general, wind-generated waves are non-stationary, irregular and highly nonlinear, representing a process with a highly complex dynamics [3].

Many natural phenomena can be described as stochastic processes. A key feature of this type of process is the correlation time or scale, characterized by the autocorrelation function, involving either long range or short range correlation characteristics. The autocorrelation function decay for a long range correlated process lacks a characteristic time. In fact, the de-correlation time is infinite. On the other hand, short range correlation processes decay is described by an exponential function and therefore they exhibit finite time scales. Although, in recent years, several studies have shown that many complex natural systems, including hydro-meteorological processes, are characterized by long-range correlations [4, 5], a standard assumption of time series analysis is that observations separated by a large time span are roughly independent. However, for processes with long-range dependence, these observations are not independent.

Experimental data are often affected by non-stationarities, i.e. varying mean and standard deviation. These effects have to be well distinguished from the intrinsic fluctuations and correlations of the system in order to find the correct scaling behavior. Sometimes, type and reasons for underlying non-stationarities in collected data are not known [6]. Furthermore, it is difficult to identify and quantify long-range correlations using traditional tools like the autocorrelation function or power spectrum, partly because of the nonstationarity of the signal.

In this context, long-term time series of significant wave height (the parameter most frequently used to describe wave conditions) exhibit a number of features, namely random variability, serial correlation, seasonal periodicity, nonstationarity, and, possibly, long-term trends, evolving over different time scales. If there is long range dependence in significant wave height time series then forecasts in the long term, as well as the knowledge on the physics underlying the observed dynamical processes, could improve.

The key objective of this paper is to analyze significant wave height time series over a range of temporal scales to identify the possible existence of long-range correlations. For this, the detrended fluctuation analysis(DFA) technique is used, because it permits the detection of long-range correlations embedded in a seemingly nonstationary time series and also avoids the spurious results that are possibly caused by nonstationarity.

The rest of the paper is structured as follows. Section 2 describes some characteristics of the wave data used and introduces the detrended fluctuation methodology applied for detecting scaling properties in wave time series. Section 3 deals with a summary and discussion of results obtained. Conclusions are presented in Section 4.



2 Data and analysis methods

2.1 Significant wave height data

Wave conditions during a sea state can be properly characterised by the directional spectral density. However, the parameter most frequently used to describe a sea state is the significant wave height, defined as the average of the highest one-third of wave heights in a record or in terms of the spectral density function, as four times the square root of the zero order moment of the wave spectrum. These definitions are equivalent for seas with narrow band spectra.

Significant wave height data used in this paper have been provided by Puertos del Estado. Wave data were recorded during 12 years (from 1998 to 2009) using a SeaWatch directional waverider buoy placed at 382 m water depth, off Estaca de Bares Cape (Coruña), north-west of Spain. Wave data were recorded with 2 Hz sampling frequency during a period of 30 minutes, approximately, for each hour. These data were used to compute the wave spectrum and the significant wave height, as well as other spectral wave parameters. Estimated significant wave heights have been averaged to obtain a mean daily value.

One of the main obstacles in analysing long time series of significant wave height, as in almost any environmental application, is that usually datasets measured in nature are incomplete due to a number of reasons, such as interruption of measurements due to instrument failure or maintenance, accidental loss of data, or discard of erroneous measurements. A logical approach to alleviate this inconvenient is to use some of the various methods suggested for filling gaps in time series. Unfortunately, this type of approach is not free of problems. Chen et al. [6] observed that removing segments from a signal and stitching together the remaining parts does not affect the scaling behavior of positively correlated signals, even when up to 50% of the points are removed. Thus, stitching together segments of data obtained from discontinuous experimental recordings, or removing some noisy and unreliable parts from continuous recordings and stitching together the remaining parts has become a procedure commonly used in preparing data for scaling analysis and is the approach used in this study.

2.2 Detrended fluctuation analysis

The Detrended Fluctuation Analysis (DFA) technique was introduced by Peng et al. [7, 8] to investigate long-range power-law correlations along DNA sequences. The key advantages of the DFA over conventional methods are its ability to detect long-range correlations embedded in a seemingly non stationary time series, and also to avoid the spurious detection of apparent long-range correlations that are an artifact of nonstationarity. Thus, DFA has been established as an important method to reliably detect long-range correlations in data affected by trends.

Briefly, the detrended fluctuation analysis of a time series can be described as follows. Given a time series $\{x(i)\}$ with N values taken at equidistant intervals, a



new time series $\{y(k)\}$ is obtained by estimating its average value,

$$\bar{x} = \sum_{i=1}^N x(i) \quad (1)$$

and integrating the original time series,

$$y(k) = \sum_{i=1}^k (x(i) - \bar{x}) \quad (2)$$

Next, the integrated time series of length N is divided into $N_b = N/n$ nonoverlapping boxes, each containing n data points. Then, the local trend within each box is estimated by fitting an adequate function, $y_p(k)$, to the segment. The local trend in each box is commonly defined as the ordinate of a linear least-square fit of the data points in that box,

$$y_p(k) = ak + b \quad (3)$$

Thus, the possible influence of linear trends on scales larger than the segment are eliminated by subtracting $y_p(k)$ from the integrated time series in each box. However, it is interesting to remark that the use of other detrending functions may improve the accuracy of the DFA technique by removing higher order trends [9]. In this sense, the DFA method is conventionally denoted as DFA- q , with the index q denoting the order of the polynomial function used to characterise the local trend (DFA1, DFA2, ...). In the DFA- q approach, the influence of possible $(q - 1)$ th-order trends on scales larger than the segment size are removed by subtracting the best q th-order polynomial fit from the integrated time series, $y(k) - y_p(k)$, in each box.

The root-mean-square values of the integrated and detrended signal are calculated for each segment of length n to yield the detrended fluctuation function,

$$F_r(n) = \left[\frac{1}{n} \sum_{k=(r-1)n+1}^{rn} |y(k) - y_p(k)|^2 \right]^{1/2}, \quad r = 1, 2, \dots, N_b \quad (4)$$

Averaging $F_r(n)$ over the N_b intervals gives the mean value of the fluctuation function for a given subseries length n ,

$$\langle F(n) \rangle = \frac{1}{N_b} \sum_{r=1}^{N_b} F_r(n) \quad (5)$$

This computation is repeated over many time scales (box sizes) to provide a relationship between $\langle F(n) \rangle$, and the segment length n . Typically $\langle F(n) \rangle$ will increase with box size n .



A linear relation between $\log F(n)$ versus $\log n$ indicates the presence of scaling, which can be characterised by the slope α of the regression line fit

$$\langle F(n) \rangle \sim n^\alpha \quad (6)$$

A fluctuation plot is a graph of $\log F(n)$ versus $\log n$ and the slope of the fluctuation plot is the so called α scaling exponent which characterizes the correlation in a time series. A slope $\alpha = 0.5$ corresponds to a process with no long-term correlations (random uncorrelated process, i.e. white noise, or short-term memory) whereas $\alpha = 1.5$ corresponds to integrated random walk, i.e. Brownian noise. If $\alpha = 1$, temporal fluctuations are of flicker-noise type ($1/f$ noise), typical of systems in a self-organized critical state. The cases $\alpha > 1/2$ and $\alpha < 1/2$ should be physically distinguished. For $\alpha > 1/2$, there is persistence. In this case, if in the immediate past the signal has a positive increment, then on average an increase of the signal in the immediate future is expected. An exponent $\alpha < 1/2$ means antipersistence. In this case, an increasing value in the immediate past implies a decreasing signal in the immediate future, while a decreasing signal in the immediate past makes an increasing signal in the future probable (i.e. large and small values of the time series are more likely to alternate). Thus, data records with $\alpha < 1/2$ appear very noisy. They have a local noise of the same order of magnitude as the total excursion of the record.

Fluctuation analysis has been undertaken using custom-written IDL software. The computer program was tested by using synthetic signals with a known fractal dimension generated with the Weierstrass–Mandelbrot function [10].

3 Results and discussion

Significant wave height time series are obviously affected by seasonality and the concept of long-memory refers to non-periodic processes. Thus the seasonal variation of the sample mean, should be removed before analyzing the scaling behavior of the time series. Seasonality in each value of H_{s_i} has been removed by subtracting the mean of the corresponding day of the year, obtained by averaging over all years in the record. This approach has been also used by many authors (see e.g., [11]). Consequently, analysis of fluctuations is based on the departures of H_{s_i} from their mean daily value \overline{H}_{s_i} for each calendar date i ,

$$\Delta H_{s_i} = H_{s_i} - \overline{H}_{s_i} \quad (7)$$

The significant wave height time series is shown in Fig. 1a in which the seasonal behavior becomes evident. Time series of ΔH_s departures after removing the annual cycle is shown in Fig. 1b.

The integrated deviations function, or profile, of the mean daily significant wave height time series obtained by integration of the series of mean daily significant wave height departures from the corresponding calendar day mean is presented in Fig. 2. The complex behavior of H_s time series becomes apparent in this figure, since there seems to be structures over a variety of scales. Note that if during a



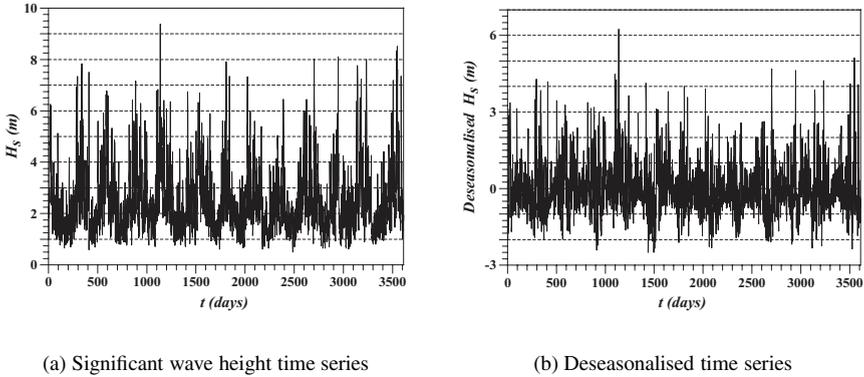


Figure 1: Time series of daily averaged significant wave heights measured during the period (1998–2009) off Estaca de Bares, Spain, in the Cantabric Sea, and Significant wave height time series deviations, ΔH_s , after removing the seasonal component

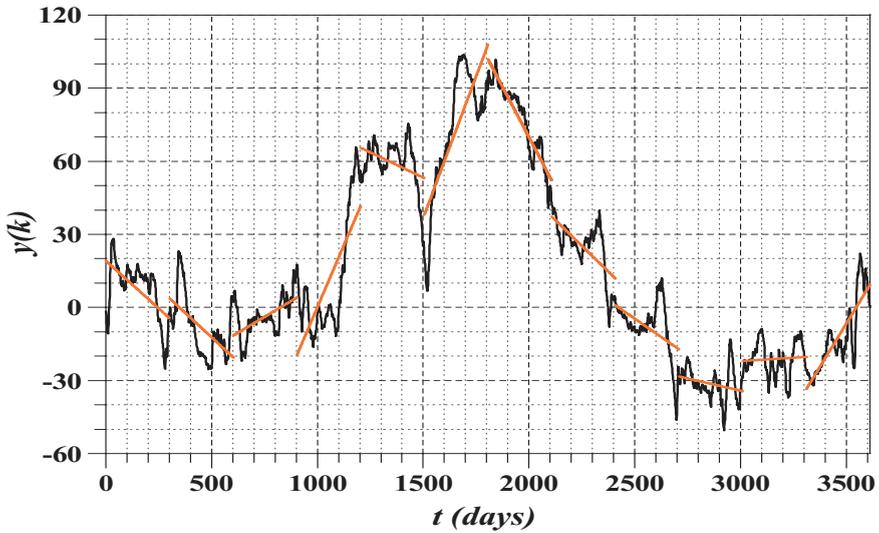


Figure 2: Mean-centred and integrated deviations of significant wave height. The straight lines are an example of the fitted local trend within windows of length $n = 301$ after partitioning the profile in $N_b = 12$ segments.

certain time period H_s is higher than the mean (stormy period), the slope of the integrated time series $y(n)$ for the specific period will be positive, while for periods with H_s lower than the mean (periods of calm), the slope will be negative.

The plot of $\log F(n)$ vs $\log n$ is shown in Fig. 3. Superposed straight lines are the best fit in the least-squares sense and indicate scaling with α exponent. To obtain statistically reliable estimates of the fluctuation function, the largest box size used is $n_{max} = N/4$, such as suggested in [12]. It can be noticed that with increasing the box size the fluctuation function increases as well. Nevertheless, it is interesting to observe that the plot of $\log F(n)$ vs $\log n$ does not fit a single line over the whole range of time scales. The plot, in fact, does not define a single long range correlation exponent but instead two or more correlation ranges.

Examination of Fig. 3 reveals the existence of two crossover and provide evidence of three distinct scaling exponents over an equal number of time scales. While the data resemble Brownian noise, $\alpha = 1.36$, dominance over short time scales, up to 10–11 days, persistent long-range correlations, $\alpha = 0.79$, are identified over time scales between 10 days and 115 days, approximately. This persistence crosses over to an almost independent random process scaling behavior, $\alpha = 0.47$, for larger time scales. This value of the DFA scaling exponent, lower than 0.5 may indicate a crossover from persistent to anti-persistent random behavior. However, this value is almost 0.5 and probably represents a random behavior for larger time scales.

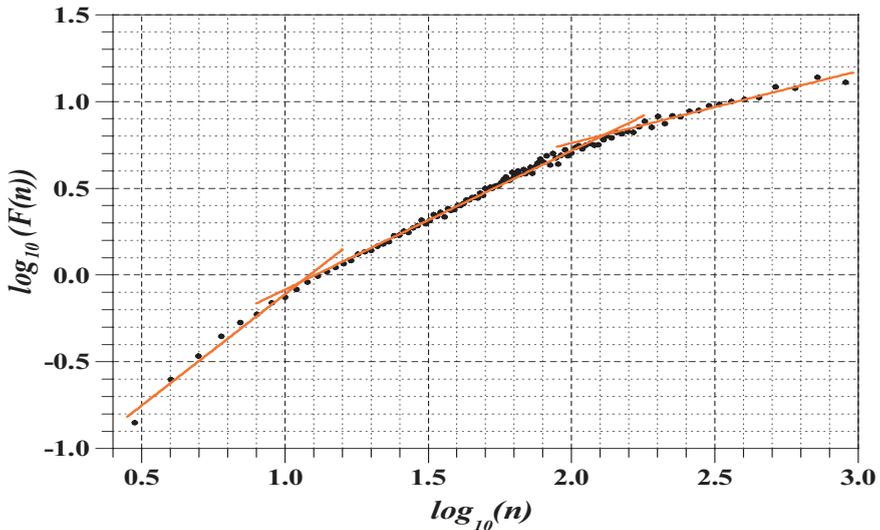


Figure 3: Logarithm of the fluctuation function $F(n)$ versus logarithm of time scale n for daily average of significant wave height measured during 12 years off Estaca de Bares, Spain, in the Cantabric Sea.

4 Conclusions

Detrended fluctuation analysis of mean daily significant wave height fluctuations estimated from wave data taken in the Cantabric Sea, off Estaca de Bares Cape, Coruña, in deep waters, during the period 1998–2008 is presented.

As is common in the fluctuation analysis of natural phenomena, the observed DFA plot is not a single straight line but presents three different scaling regions separated one from another by a crossover time.

It is shown that the associated time series display long-range dependence for time scales from 10–11 days up to time scales close to 100 days, characterized by a scaling exponent $\alpha = 0.79$. For smaller time scales fluctuation analysis reveals a Brownian noise-like behavior while for time scales larger fluctuations it tends to depict a fully random behavior.

The observed difference in the scaling behavior may be due to variabilities in the underlying physics governing the process. In particular, it may be related to changes in the external atmospheric and oceanic forcing mechanisms over various time scales but the origin of this diversity in scaling behavior should be investigated in depth.

Acknowledgements

The authors are grateful to Puertos del Estado (Spain) for kindly provide the wave data used in this study. This research has been undertaken within scientific project TEC2009-14217 financed by the Spanish Ministry of Science and Innovation.

References

- [1] Goda, Y., *Random Seas and Design of Maritime Structures* (3rd Edition), World Scientific Publishing, 2010.
- [2] Ochi, M.K., *Ocean Waves: The Stochastic Approach*, Cambridge University Press, 2005.
- [3] Komen, G.J., L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. Janssen, *Dynamics and Modelling of Ocean Waves*, Cambridge University Press, Cambridge, 1994.
- [4] Beran, J., *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.
- [5] Fox R. and Taqqu M.S., Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series, *The Annals of Statistics*, 14, 517–532.
- [6] Chen, Z., P.C. Ivanov, K. Hu, and H.E. Stanley, Effect of nonstationarities on detrended fluctuation analysis, *Physical Review E*, 65(4), 15 pp, 2002.
- [7] Peng, C.K., S.V. Buldyrev, S. Havlin, M. Simmons, H.E. Stanley and A.L. Goldberger, Mosaic organization of DNA nucleotides, *Physical Review E*, 49(2), 1685–1689, 1994.



- [8] Peng, C.K., Havlin, S., Stanley, H.E., and Goldberger, A.L., Quantification of scaling exponents and crossover phenomena in non stationary heartbeat time series, *Chaos*, 5(1), 82–87, 1995.
- [9] Bunde, A., S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter and K. Voigt Correlated and uncorrelated regions in heart-rate fluctuations during sleep, *Phys. Rev. Lett.* 85, 3736, 2000.
- [10] M.V. Berry; Z.V. Lewis, On the Weierstrass–Mandelbrot Fractal Function, *Proc. R. Soc., A* 370, 459–484, 1980.
- [11] Koscielny-Bunde, E., A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, and H.J. Schellnhuber, Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.*, 81(3), 729–732, 1998.
- [12] Koscielny-Bunde, E., J.W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin, Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies, *J. Hydrology*, 322, 120–137, 2006.

