NATIONAL TECHNICAL UNIVERSITY OF ATHENS
DEPT. OF NAVAL ARCHITECTURE AND MARINE ENGINEERING

Nonstationary stochastic modelling of time series with applications to environmental data

by

CH.N. STEFANAKOS

Submitted in partial fulfillment of the requirements for the degree of Doctor of Engineering at the Dept. of Naval Architecture and Marine Engineering of National Technical University of Athens

Advisory Committee

G.A. Athanassoulis, Professor NTUA, Thesis Supervisor
Ch.N. Psaraftis, Professor NTUA
E.G. Kounias, Professor Univ. of Athens

Examination Committee

G.A. Athanassoulis, Professor NTUA, Thesis Supervisor
E.G. Kounias, Professor Univ. of Athens
Th.A. Loukakis, Professor NTUA
S.A. Mavrakos, Professor NTUA
V.J. Papazoglou, Professor NTUA
Ch.N. Psaraftis, Professor NTUA
J. Spiliotis, Assistant Professor NTUA

ATHENS, March 1999

Synopsis

For any many-year long time series of an environmental parameter, and especially for time series of significant wave height, the prevailing qualitative features are the following: (i) a fluctuating character of the sequence of observations that prevents any exact reproducibility of the time series and calls for stochastic modelling, (ii) a statistical dependence between observations separated by a relative small time period, (iii) a pronounced seasonal variability within the annual cycle, (iv) a yearly statistical periodicity induced by the meteorological annual cycle, and, possibly, (v) an overyear trend, which is difficult to identify.

The main goal of this work is to elaborate an appropriate modelling of long-term time series of environmental parameters taking into account the specific features described above, and apply this modelling to solving concrete problems arising in ocean and coastal engineering. The proposed modelling consists in using nonstationary non-Gaussian stochastic processes with periodically (yearly) varying statistical characteristics. In order to assess the model and its components, an extensive analysis of existing time series has been performed. Since most part of measured time series of environmental parameters are incomplete, due to a number of reasons (as, e.g., interruption of measurements caused by instrument failure or by extreme natural phenomena etc.), the modelling is extended to cover the case of time series with missing values. In addition, a simulation procedure of nonstationary time series has also been developed, which is a very useful tool for various applications. The first- and second-order probability structure is studied and appropriately modelled using versatile bivariate structures. More precisely, a member of the Fréchet class of distributions (multivariate distributions with given marginals) is used for the representation of the second-order probability density function of the stationary component of the times series. Then, the corresponding second-order density of the nonstationary series is derived by applying a time-dependent transformation. Finally, the nonstationary modelling is combined with an enhanced method of extreme-value predictions, in order to calculate return periods of various levels from nonstationary time series of environmental parameters. The whole procedure is tested against more traditional methods, and it is found in accordance with them when the population of extreme values is sufficiently large. However, the present method exhibits a remarkable stability concerning the length of the examined time series, in contradiction to the traditional methods.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The analysis, modelling, and simulation of many-yearlong time series of environmental parameters is a common practice in various environmental engineering disciplines, especially in hydrology (Salas et al., 1980; Salas, 1993), solar radiation, and others. Ocean engineers also continuously deal with time series data (e.g., wave, wind, and current parameters, sea levels, etc.), but traditionally, they focus attention on specific statistical events, mainly extreme or cumulative quantities, and draw inferences by using simplified (random variable) models, not taking into account the correlation structure and the nonstationarity of the data (Ochi, 1982; Goda, 1988; Labeyrie, 1991; Castillo and Sarabia, 1992). Only in the last decade have there appeared some works dealing with time series analysis of sea state parameters, aimed mainly at developing wave climate simulators (Borgman and Scheffner, 1991; Medina et al., 1991; Scheffner and Borgman, 1992; Bettencourt, 1993; Frutuoso et al., 1993), and at the analysis and the modelling of *in situ* measurements (Guedes Soares and Ferreira, 1995; Guedes Soares et al., 1995). These works are briefly discussed herein.

In recent years the accumulation of measured (*in situ* and satellite) data and the production of a large amount of hindcast data by means of spectral wave models helped to understand the statistical structure of long-term time series of sea state parameters and revealed a hierarchy of timescales from half an hour up to a century. For example, the basic features of many-yearlong time series of significant wave height can be summarized as follows: they are non-Gaussian and nonstationary time series that also exhibit a year-to-year statistical variability and longer-term climatic variability (see Table 2.1). A more detailed description of the statistical structure of significant wave height is given in Chapter 2.

Various works have appeared dealing with the modelling of sea state parameter time series. Medina et al. (1991) studied the bivariate time series of significant wave height and period based on 6-hour spectral data from measurements off the Oregon coast. They assumed that significant wave height H_S and zero upcrossing period T_Z follow a lognormal distribution and

that their time series are transformed into normalized stationary Gaussian time series by means of the transformations (seasonal standardizations)

$$x(i) = \frac{h_S(i) - A_h(i)}{B_h(i)}, \tag{1.1}$$

$$y(i) = \frac{t_Z(i) - A_t(i)}{B_t(i)}, \quad i = 1, 2, \ldots, \tag{1.2}$$

where h_S=logH_S, t_Z=logT_Z, i is the discrete time, $A_{h,t}(i)$ and $B_{h,t}(i)$ are annually periodic parameters, accounting for the seasonal variability. These parameters are estimated on a month-by-month basis. The seasonally standardized time series $\{(x(i), y(i)),\ i = 1, 2, \ldots\}$ is given the structure of a first-order autoregressive model with constant coefficients, since no significant seasonality was found in the correlation coefficients. Bettencourt (1993) and Frutuoso et al. (1993) applied Medina's model (Medina et al., 1991) to 3-hour spectral data obtained by buoy measurements off the west coast of Portugal. They found that significant wave height H_S, spectral peak period f_p, and spectral moments m_{-1} and m_2 follow the lognormal probability law quite well and constructed Medina's type models for the pairs (H_S, m_1), (H_S, m_2), and (H_S, f_p). In checking these models, Bettencourt and Frutuoso et al. have found and reported various inadequacies that call for further developments in analysis, modeling, and estimation procedures. It should be noted, however, that the main goal of these works was to construct simple and easy-to-use wave climate simulators and not to study in detail the structure of time series of sea state parameters.

Borgman and colleagues (Borgman and Scheffner, 1991; Scheffner and Borgman, 1992) developed a different approach, based on normal-scores transformation and the assumption of piecewise (month-by-month) stationarity of data series. They studied the trivariate time series $\{(H_S(i),\ T_p(i),\ \Theta(i)),\ i = 1, 2, \ldots\}$, where T_p is the spectral peak period and Θ is the main direction of the sea state, using 20-year hindcast data obtained through the Wave Information Study (WIS) of the Coastal Engineering Research Center (CERC). After a normal-scores transformation of the data to a normal sequence, Borgman and Scheffner calculated the mean monthly correlation structure for each one of the 12 months of the year. Then, they used this normal, piecewise stationary model to simulate wave data, preserving the primary statistical properties of the finite data set.

Walton and Borgman (1990) presented another very interesting and systematic analysis and simulation procedure for univariate non-Gaussian and seasonally nonstationary time series in the context of water level modeling of the Great Lakes. This procedure can be applied without any changes to H_S time series, since all the basic characteristics of both (H_S and water level) time series are similar. This method utilizes a low-pass-filtering technique to obtain the time series of seasonal trends (for the mean and the standard deviation). The time series obtained after seasonal standardization (see equs. (1.1)–(1.2)) is assumed stationary, and it is transformed

to a Gaussian one by means of the normal-scores transformation. This is done by using the empirical distribution function, supplemented with appropriate analytical representations for its lower and upper tails. The reduced series is simulated in the frequency domain via its spectrum, and inverse procedures are used to simulate the initial series.

Another interesting means for modeling and simulating sea state parameter time series seems to be the low-order periodic autoregressive (PAR) models, which have been introduced since the early 1960s (Hannan, 1955; Thomas and Fiering, 1962) and used extensively in hydrology (Bartolini and Salas, 1993). However, no application of these models to wave data is known to the present authors.

Further, a problem that it is inherent to the measured time series is that they are frequently incomplete, i.e., a non-negligible number of observations is missing. Observations may be missing for a number of reasons such as, e.g., the interruption of measurements due to instrument failure, the effects of extreme natural phenomena such as hurricanes, the accidental loss of data files in the computer system, or the discard of a portion of measurements that have been taken in an erroneous way. Missing observations is also a common situation in a number of other applications, such as economics (Gómez and Maravall, 1994; Nieto and Martínez, 1996), electrical engineering (Giannakis and Zhou, 1994; Mirsaidi et al., 1997), hydrology (Shih and Cheng, 1989; Salas, 1993). Since the pioneering works of Jones (1962) and Parzen (1963), a lot of work has been made on modelling and analysis of time series with missing values, significant part of which is contained in the proceedings of a specialized conference edited by Parzen (1984). A survey of the more recent statistical literature can be found in Stefanakos and Athanassoulis (1997).

In ocean engineering, only few works exist for the modelling of missing values in time series of wave data. Hidalgo and colleagues (Hidalgo, 1995; Hidalgo et al., 1995) have developed a method for filling in missing values of time series of significant wave height. The method consists of three parts: (i) for small gaps a direct interpolation is used, (ii) for longer gaps predictions from high-order ARMA models are used, and (iii) for very long gaps additional information is used from neighbouring sites. According to the authors, the first two parts work satisfactorily well, whereas the third one work as long as the data from the neighbouring sites exhibit similar patterns with the studied site. Moreover, in order to stationarize the original series the authors take the first difference of the corresponding logarithms. However, differencing a time series with missing values will widen the gaps and/or introduce additional gaps into the series (Ljung, 1989). Moreover, overdifferencing may lead to parameter identifiability problems and noninvertibility of the series (Luceño, 1997).

In the present work, a flexible nonstationary stochastic model is developed for, and successfully applied to environmental data. The model is appropriately extended also to treat time series with missing values, and it is used for the missing-value completion of measured time series of wave data. According to this model, a many-year long time series of significant wave height

is considered to be a periodically correlated (cyclostationary) process, i.e., as a nonstationary stochastic process with yearly periodic mean value and standard deviation. It is noted that, the model can also incorporate additional features such as an additive overyear trend. Further, the first- and second-order probability structure is modelled by means of the so-called Plackett model, a versatile non-Gaussian tool of constructing bivariate probability models, which can handle any acquired information from the univariate case. Then, the nonstationary modelling is applied along with an enhanced extreme-value method to predict return periods from nonstationary time series of significant wave height.

In Chapter 2, and before the analysis and the modelling of many-year long time series of wave parameters, a presentation is given of the various long-term time-scales involved. The corresponding wave characteristics exhibited in each one of them are also discussed. Especially, concerning the possible existence of an overyear trend, a survey of the recent literature is made. In Chapter 3, a brief summary of cyclostationary processes is given and their main properties are stated. This kind of processes is used in the present work for modelling environmental time series. Based on this modelling, an analysis procedure is set up in Chapter 4, involving trend estimation, seasonal standardization, and spectral analysis of the residual time series. The procedure is applied to time series of hindcasted significant wave height in order to assess the wave climate at specific sea locations. Concerning the possible existence of a long-term overyear trend, which is a very important feature for the climate change, our results give a positive answer, in accordance with the most part of the ones in the existing literature.

In Chapter 5, a general procedure for the completion of missing values is described. The key idea is to perform the completion after having removed not only any deterministic character of the series, but also the correlation structure of the stationary series, and the uncorrelated residual series being only present. In this direction, the various deterministic components (additive trend and periodic or seasonal characteristics) of the series are first identified. Then, the series is appropriately detrended and seasonally standardized. The resulting time series, which is considered stationary and is modelled as a low-order ARMA process, is also incomplete and, thus, specific techniques are required for parameter estimation. The basic point here is the use of a consistent estimator for obtaining the autocovariance function (Parzen, 1963), which in turn is utilised to obtain the spectrum. The latter is essential for the estimation of ARMA parameters, which is performed by least-square fitting (Spanos, 1983). After finishing with the analysis of the incomplete series, an i.i.d. sequence with the same probability stucture as the uncorrelated residuals (existing values) is produced by means of numerical simulation, and is used to complete the latter. This completed version of uncorrelated residuals, in conjunction with the estimated ARMA parameters, is utilised to construct a complete version of the stationary series. Using, finally, the estimated deterministic components, a new nonstationary time series without missing values is constructed. The proposed method fulfills the four criteria proposed by Beveridge (1992, p. 3480). The procedure for the missing value completion is applied to

two synthetic (simulated) time series. The initial time series become incomplete using missing-value patterns from measured time series. Comparisons are performed between results from the initial (incomplete) and the reconstructed one, and presented results are very satisfactory. Then, the same procedure is applied to two measured time series of significant wave height.

In Chapter 6, after a brief survey of the Fréchet class of distributions (distributions with given marginals), a member of this class, namely the Plackett model, is presented, and a new procedure is given for the maximum likelihood estimation of its parameters. The (non-Gaussian) Plackett model is used for the representation of the second-order probability structure of the stationary component of the series, and, then, the corresponding structure of the nonstationary series is obtained by applying a time-dependent transformation. Results based on the Plackett model are compared with corresponding results based on the bivariate lognormal, and its performance seems to be better regarding the description of the empirical population, and its increased flexibility to describe the shape of the distribution. The probability modelling of the second-order structure is mainly exploited in Chapter 7, where it is used for extreme-value predictions in conjunction with the MENU method developed by Athanassoulis et al. (1995b). See also Soukissian (1995). This method, which provides a new definition of the return-period concept based on the Mean Number of Upcrossings of a level, is used for the calculation of return periods from nonstationary time series of significant wave height. Although in previous works (Athanassoulis et al., 1995b; Soukissian, 1995), an *a priori* assumption was made for the joint probability density function of the process and its time-derivative, in the present work this joint probability density is calculated exactly by means of the modelling presented in Chapter 6. The results of our method are compared with the ones of the Gumbel approach and of the POT method, and they are found to be in accordance with them when the population of extreme values is sufficiently large. However, our method exhibits a remarkable stability concerning the length of the examined time series, in contradiction to Gumbel approach and to POT method.

– *Directions of Further Applications*

The proposed model can be used for various applications in the context of probabilistic wave climate analysis and prediction, ocean engineering, offshore structures and ships' design, as, for example

1. Long-term stochastic simulation of wave climate. Such kind of results have been produced in the present work;

2. Construction of complete versions of measured time series of wave data, which are needed as input in numerical models concerning large-scale flow computations (Scheffner and Borgman, 1992);

3. Enhanced calculation of sea state duration;
4. Extreme-value predictions for marine structure responses; and
5. Input to coastal morphodynamics models. Especially, models used for the sediment transport, beach corrosion etc.

Let it be noted that the presented modelling can also be applied to time series of many other ocean and/or atmospheric parameters, such as, e.g., wave period, wave and wind direction, wind speed, sea temperature, air temperature, humidity, sea level, etc., which also exhibit seasonally varying randomness.

– *Main Original Contributions and Publications*

Finally, let us enumerate the main original contributions of the present work:

(i) The cyclostationary modelling proposed in Section 3.2, and the corresponding analysis procedure for a) complete time series in Sections 4.1–4.5, and b) incomplete time series in Sections 5.1–5.4. Although one may find simplified versions of this modelling in earlier works (Hurd, 1969; Parzen and Pagano, 1979; Salas, 1993), it is the first time that this is applied and assessed in its present integrated form,

(ii) The missing-value completion procedure summarized in Section 5.5.

(iii) The use of non-Gaussian Plackett model for the representation of the second-order probability structure of the nonstationary series, presented in Section 6.7,

(iv) The exploitation of the data-induced second-order probability structure to the implementation of the MENU method, in order to calculate return periods at various levels from nonstationary time series of significant wave height, presented in Section 7.4 and validated in Section 7.5.

In the course of this Thesis, and for the better implementation of the various topics, some totally new procedures/tools have been developed:

a) A nonstationary simulation procedure, presented and implemented in Section 5.6, and used in a slightly modified version in Section 7.5,

b) A method for the maximum likelihood estimation of the association parameter ψ of Plackett model, presented in Sections 6.5–6.6,

c) An integrated software package for: a) the analysis of nonstationary (complete and incomplete) time series of wave parameters, b) the analysis and modelling of their probability structure, and c) the return period prediction for various level values.

It should be finally noted, that some results of the presented work have already been published. There are two publications related to point (i) above (Athanassoulis and Stefanakos, 1995; Athanassoulis and Stefanakos, 1998), one related to point (ii) (Stefanakos and Athanassoulis, 1997), and one, which exploits further the result of point (i), is expected to be presented this year (Athanassoulis et al., 1999). Parts of Chapters 6 and 7 (points (iii) and (iv)) are under preparation to be submitted in appropriate scientific journals.

Chapter 2

Statistical structure of long-term wave data

Before proceeding to the analysis and modelling of long-term time series of wave parameters, it is expedient to carefully look into the statistical structure of wave data time series in order to realize their fundamental qualitative and quantitative characteristics. For this, a carefull definition of the various time and space scales involved should be first given.

According to climate modellers, see e.g. Berger (1991), the distinction of the various scales is imperative in order to identify the various wave characteristics in different scales, understand their impact, and appropriately incorporate them to climate models. Otherwise, it is very likely that a (wave) climate parameter either is modelled in an erroneous way, or is totally neglected in the modelling. It is noteworthy to say that the amount of information, that can be extracted from a time series of data, depends on three features (Burroughs, 1992): (i) the sampling interval, (ii) the length of the series, and (iii) the accuracy of the observations. As we will see in the next section, the first two refer to appropriate time scales. In the sequel, two examples are given to illustrate the proprer definition of time scales.

Let us first consider the investigation of the so-called climatic variability in the time series of wave data, i.e. a variability very slowly varying with a period, say, of fifty years. If we are working with a time series of one (or few, say five) year(s), there is no possibility to observe such kind of variability. If the time series become longer, say twenty-year long, a linear (increasing or decreasing) trend can be observed. If, further, a time series of one hundend (or more) years is analysed, a cyclic (periodic) trend is revealed.

Consider, now, the wave parameter "significant wave height H_S at a given site in the sea". If a single wave record with length of, say, six hours is available, then we can compute its power spectrum. Significant wave height is defined as a function of the zero-order moment of the

spectrum and, thus, is a single value. If, now, a bunch of wave records with total length of one month are taken into account, significant wave height is a sequence of values the statistical charcteristics of which remain constants, and it is fully characterised by its probability density function $f(H_S)$. If, finally, data with length of one year is used, the statistical characteristics of significant wave height are no more constants also, but they vary with the time, and it is, now, characterised by a family of probability density functions varying in time $f(H_S(\tau))$. See also Figure 2.1, where these three distinct behaviours of H_S are depicted. On the second column of this figure, a long-term time series of H_S is depicted in three different zooms. The three time windows, indicating the analysed data, are of length 1, 124 (one month) and 1460 (one year) six-hourly intervals, respectively. On the third column, the model proposed for H_S is shown.

2.1 Long-term time and space scales

Consider the sea-surface elevation $\eta(\mathbf{x}, t)$ as a function of space $\mathbf{x}$=$(x_1, x_2)\in\mathcal{D}_x$ and time $t\in\mathcal{I}_t$, where $\mathcal{D}_x$ denotes a space domain (sea area) and $\mathcal{I}_t$ a time interval. This function can be modelled as a random field, which, if we confine ourselves in an area of few (say ten) kilometers and few (say three) hours, can be considered as a homogeneous and stationary random field. By analyzing the function $\eta(\mathbf{x}, t)$, we obtain the general directional spectrum $S(\mathbf{k})$=$S(k_1, k_2)$, which fully describes (in probabilistic terms) the sea surface in the space domain $\mathcal{D}_x$ and in the time interval $\mathcal{I}_t$. Clearly, the spectrum $S(\mathbf{k})$ depends on $\mathcal{D}_x$ and $\mathcal{I}_t$, and thus, we better write $S(\mathbf{k};\mathcal{D}_x^{(m)},\mathcal{I}_t^{(n)})$. The indices $^{(m)}$ and $^{(n)}$ indicate that $S(\mathbf{k})$ varies as $\mathcal{D}_x$ and $\mathcal{I}_t$ change. To formulate this mathematically, it is convenient to define a new space variable $\boldsymbol{\chi}$ and a new time variable τ, representative of $\mathcal{D}_x$ and $\mathcal{I}_t$, respectively, with "infinitesimals" $\Delta\boldsymbol{\chi}$=O(10km) and $\Delta\tau$=O(3h). Then, the set of spectra $S(\mathbf{k};\mathcal{D}_x^{(m)},\mathcal{I}_t^{(n)})$, m=1,2, ... , n=1,2, ... , can be considered as the discrete family of $S(\mathbf{k})$ for a grid ($\boldsymbol{\chi}_m$, τ_n) in the new space-time variables. Further, the function $S(\mathbf{k};\mathcal{D}_x^{(m)},\mathcal{I}_t^{(n)})$ can be extended as a continuous function of $\boldsymbol{\chi}$ and τ, e.g. using some interpolation technique. Thus, the directional specrtum of sea waves over an extended sea area $\mathcal{D}_\chi$ and for a long time period $\mathcal{I}_\tau$ is written as $S(\mathbf{k};\boldsymbol{\chi}, \tau)$. Usually, instead of the full directional spectrum we confine ourselves to a number of spectral parameters (e.g. the significant wave height H_S, the zero-upcrossing period T_Z, and the mean direction $\overline{\theta}$), which describes the main features of the spectrum. In this case, a vector-valued function $\vec{\Lambda}(\boldsymbol{\chi}, \tau)$=$(\Lambda_1(\boldsymbol{\chi}, \tau), \Lambda_2(\boldsymbol{\chi}, \tau), \dots)$, where $\Lambda_1(\boldsymbol{\chi}, \tau)$=$H_S$, $\Lambda_2(\boldsymbol{\chi}, \tau)$=$T_Z$ etc., takes the place of the spectrum $S(\mathbf{k};\boldsymbol{\chi}, \tau)$. Two examples of special interest are: (i) $\vec{\Lambda}(\tau)$=$(H_S(\tau), T_Z(\tau), \overline{\theta}(\tau))$, which corresponds to in situ measurements at a fixed site in the sea, and (ii) $\vec{\Lambda}(\boldsymbol{\chi}(\tau))$= $(H_S(\boldsymbol{\chi}(\tau)), T_{alt}(\boldsymbol{\chi}(\tau)))$, which corresponds to satellite (altimeter) measurements. The $\boldsymbol{\chi}$-space and τ-time variability of the spectrum (or spectral parameters) is our main concern in this section.

In accordance with the above described conceptual framework, two space-time scales come

into play. The first one, with sampling intervals $\Delta\mathbf{x}$=O(5m) and Δt=O(1sec), is called the fast space-time scale. It is in this scale that the individual waves of the sea surface can be recorded. The second one, with sampling intervals $\Delta\boldsymbol{\chi}$=O(10km) and $\Delta\tau$=O(3h), is called the slow space-time scale. In the latter scale, the fluctuations (individual waves) of the sea surface are not "visible" and only the statistical characteristics of the sea waves (obtained by averaging over the population measured in the fast scale), modelled as functions of ($\boldsymbol{\chi}$,τ), are examined. The introduction of the slow time-scale has been made for the first time by Borgman (1973), and thoroughly studied by Labeyrie (1991) and Athanassoulis et al. (1992).

A specific sea state (at a given site) is characterized by the fact that the directional spectrum remains (almost) unaltered. The time period over which this event persists is called duration of the sea state. Various approaches have been used for calculating the duration of sea states; see, e.g., Lavielle et al. (1987), Labeyrie (1990), Athanassoulis et al. (1992), Athanassoulis and Soukissian (1993a). Numerical results giving first quantitative estimates and helping in understanding the qualitative features of sea-state duration statistics can be found in Athanassoulis and Soukissian (1993a) and Tournadre (1993). The duration of a sea state can be considered as the last time subscale in the fast time scale, and the first one in the slow time scale.

Other variabilities are also revealed if "slower" (either in space or in time) scales are considered. See, e.g., Table 2.1, where various long-term time scales are shown. The first one, as already mentioned, is the time period characterising the duration of an individual sea-state, and it can be viewed as a "transition" subscale from the fast-time (short-term) scale to the slow-time (long-term) scale. The second long-term subscale is the time period after which changes in weather conditions are not dependent on previous ones (i.e. it is the correlation time of a many-year long time series of wave parameters). The third one is very important, since wind waves (as any other climatic phenomenon) studied in the long-term scale possess a statistical periodic stucture with basic period of one year. Finally, the fourth and the fifth long-term time scales are these for which most applications need predictions. The basic difference between them is that for the former, there exist data for the corresponding predictions, whereas for the latter, up to now, the availability of sufficient amount of data is questionable. Thus, any kind of predictions in the latter scale is not lying on a sound basis, and should be reconsidered. See also next sections, where the long-term wave-climate variability is discussed.

For a description of the various statistical characteristics exhibited in the above time scales, and a review of the various findings in the existing literature, see Section 2.3.

2.2 The wave climate at a given site

A specific sea state is a phenomenon that lasts for some hours (e.g., 3-6 hours). When the time period of interest is much longer, many successive sea states occur and a population of spectra $S^{(n)}(\mathbf{k})$ (or $S^{(n)}(f,\theta)$), or spectral parameters $\vec{\Lambda}^{(n)}$=($\Lambda_1^{(n)}$, $\Lambda_2^{(n)}$, ...) come into play. The statistical characteristics of any one of these populations, associated with a given sea area, characterize the wave climate of it.

We can distinguish between two different types of data for the spectral population $S^{(n)}(f,\theta)$, n=1,2, ... :

(i) a time series of spectra $S^{(n)}(f,\theta)$=$S(f,\theta;\boldsymbol{\chi}_0,\tau_n)$, where $\boldsymbol{\chi}_0$ is a fixed location in the sea, or

(ii) a time series of spectra $S^{(n)}(f,\theta)$=$S(f,\theta;\boldsymbol{\chi}(\tau_n))$, where $\boldsymbol{\chi}(\tau_n)$ is the ground track of a satellite.

Alternatively, if we restrict ourselves only to some spectral parameters, we obtain time series of the form

(i′)

$$\{\vec{\Lambda}(\tau_n),\ \ n=1,2,\dots\}, \tag{2.1}$$

where $\vec{\Lambda}(\tau_n)$=$\big(H_S(\tau_n),T_Z(\tau_n),\theta(\tau_n)\big)$, or

(ii′)

$$\{\vec{\Lambda}(\boldsymbol{\chi}(\tau_n)),\ \ n=1,2,\dots\}, \tag{2.2}$$

where $\vec{\Lambda}(\boldsymbol{\chi}(\tau_n))$= $\big(H_S(\boldsymbol{\chi}(\tau_n)),T_{alt}(\boldsymbol{\chi}(\tau_n))\big)$.

There are two different approaches to analyze data of type (i) or (ii). The first one assumes that the time series of data is a realization of a stochastic process, and, then, it proceeds to an appropriate modelling and statistical treatment. The second one ignores the correlation structure of the time series of the data, and considers every signle value of it as a realization of a random variable. Some times a separation of the sample values by season is made before the statistical treatment. In this work, the second, more traditional, approach will not be considered.

As we will see in the next section, long-term wave data present themselves as time series of data, exhibiting random variability, serial correlation and seasonal periodicity. Thus, random-process

modelling ideally fits to wave climate description (Athanassoulis et al., 1992; Athanassoulis and Stefanakos, 1995).

Let us denote by $X(\tau)$ the time series of the wave parameters (or any appropriate transformation of it) at a given site $\boldsymbol{\chi}_0$, e.g., $X(\tau)$=$\ln(H_S(\boldsymbol{\chi}_0, \tau) + c)$, where c is appropriately selected. Then, $X(\tau)$ can be decomposed as follows:

$$X(\tau) = \overline{X}_{tr}(\tau) + m(\tau) + s(\tau)\ W(\tau), \tag{2.3}$$

where $\overline{X}_{tr}(\tau)$, $m(\tau)$, $s(\tau)$ are assumed to be deterministic functions, and $W(\tau)$ is considered as a realization of a stochastic process. It turns out that $W(\tau)$ is stationary (Athanassoulis and Stefanakos, 1995; Athanassoulis and Stefanakos, 1998), a fact that greatly facilitates the statistical characterization of the data $X(\tau)$. See also Figure 2.2. Further details of the analysis will be given in Chapter 4.

The function $\overline{X}_{tr}(\tau)$ is modelled as a linear function (over a period of one or two decades) and represents the possible overyear trend of the time series (see Sections 4.1 and 4.6.1). See also next section, where a survey of the existing literature on the long-term trends of wave heights is given. The functions $m(\tau)$ and $s(\tau)$ are considered deterministic and periodic with period one year, and they stand for the seasonal mean value and standard deviation, respectively. As we will see in Sections 4.2 and 4.6.2, for modeling purposes, low-order Fourier series representations of seasonal mean value and standard deviation, denoted by $\mu(\tau)$ and $\sigma(\tau)$, respectively, are used; see also Athanassoulis and Stefanakos (1995). Thus, time series $X(\tau)$ is now decomposed as

$$X(\tau) = \overline{X}_{tr}(\tau) + \mu(\tau) + \sigma(\tau)\ W(\tau). \tag{2.4}$$

If, further, $W(\tau)$ is Gaussian, then it can be given the structure of an ARMA process; see, e.g., Athanassoulis and Stefanakos (1997; 1998). See also Figure 2.3 and Section 5.4.

The above time series modelling can also be used for the integration of wave data from different sources. It will be possible, for instance, seasonal characteristics $\mu(\tau)$ and $\sigma(\tau)$ to be estimated by means of satellite data and $W(\tau)$ by in situ measurements (or model data). This direction is currently under study (Athanassoulis et al., 1999).

Let us now turn to the statistical characteristics exhibited by a many-year long time series of wave parameters and modeled by (2.4).

2.3 Statistical characteristics of wave parameters in long-term scales

For any many-year long time series of a spectral wave parameter, and especially for time series of significant wave height $H_S(\tau)$, the prevailing qualitative features are the following: (1) a fluctuating character of the sequence of observations that prevents any exact reproducibility of the time series and calls for stochastic modeling, (2) a statistical dependence between observations separated by a relative small time period, (3) a pronounced seasonal variability within the annual cycle, (4) a yearly statistical periodicity induced by the meteorological annual cycle, and, possibly, (5) an overyear trend, which is difficult to identify.

These characteristics are more or less common to many other environmental time series and have been extensively discussed, for example, in relation to hydrologic time series [see Kottegoda (1980, chapters 1-2); Salas et al. (1980, chapters 1-2); Salas (1993)]. However, they have not been widely investigated in relation to time series of sea state parameters (Athanassoulis and Stefanakos, 1995).

Since the sea state (spectral) characteristics are obtained by means of averaging procedures over short-term time intervals, the lower threshold (the infinitesimal time) for long-term considerations is of the order of some minutes. On the other hand, the sampling period required in studying (measuring, hindcasting, forecasting, or simulating) a time series of sea state parameters must not be greater than the mean value of the duration of stationarity of individual sea states, denoted by $T_{\rm stat}$. The statistical characteristics of $T_{\rm stat}$ have been studied by Labeyrie (1990; 1991), Tournadre (1993), Athanassoulis and Soukissian (1993a), and Mathiesen (1994). Labeyrie (1990; 1991) and Tournadre (1993), analyzing intensive in situ measurements, have found that the mean duration of stationarity is about 3 hours at Frigg oil field (60°N, 2°E, North Sea) and 6 hours at Palanca field (6°S, 12°30′E, Angola coast). Athanassoulis and Soukissian (1993a), analyzing hindcast data (from the same source as in this paper), have found a mean duration of about 6 hours at grid point (GP) 13 (55.9°N, 26.7°W, north Atlantic Ocean). Mathiesen (1994), analysing directional in situ measurements from Perranporth (50°21′N, 5°10′W, southeast England), Haltenbanken (65°5′N, 7°34′E, central Norway) and Tromsøflaket (71°30′N, 19°E, northern Norway), has found that duration of stationarity ranges from 2 to 4 hours. Accordingly, a sampling period of 3-6 hours can be considered satisfactory for long-term time series of significant wave height.

The most important mesoscale characteristic of an appropriately sampled time series of H_S is the dependence (serial correlation) between successive observations. Challenor et al. (1986) and Queffeulou (1983) have shown, based on satellite data, that, in general, mesoscale H_S variations can be split into a low-frequency part associated with meteorological synoptic events (weather systems) and a high-frequency noise component associated with instrumental and geophysical

noise. A similar point of view has also been developed by Tournadre (1993), who calculated the spectral densities of H_S time series in a period window ranging from 1 to 70 hours. The same figure for the characteristic correlation time has also been reported from other authors, based on in situ measurements (Oliveira Pires and Vasconcelos Pessanha, 1986) or hindcast data (Wyland and Thornton, 1991; Scheffner and Borgman, 1992). Monaldo (1988; 1990), analyzing satellite data, has studied the spatial mesoscale variability of H_S data, which exhibits very similar behavior to the temporal mesoscale variability of H_S.

In larger time intervals, but within the annual cycle, the seasonal variability becomes apparent, while a nonnegligible year-to-year variability is also present in the case of many-year data series. To detect overyear trends and other low-frequency characteristics, the long-term time series of significant wave height are subject to low-pass filtering. For example, the time series of mean annual H_S can be studied. Recent relevant studies have revealed increasing overyear trends (lasting for about 30 years). More precisely, Carter and Draper (1988) have found an increasing trend of 3.4 cm yr^{-1} by analyzing a 21-yearlong (1962-1983) data series obtained from 3-hour measurements of significant wave height at Sevenstones station (50^{o}N, 6^{o}W, north Atlantic Ocean). Bacon and Carter (1991), by analysing long series of measurements (1962-1986) from the same station and from the Ocean Weather Station Lima (57^{o}N, 20^{o}W, 1975-1988), have found an increasing trend of 2.2 cm yr^{-1}. More or less similar increasing trends have also been found by Neu (1984), who analyzed visual data. Also, T. Hirayama (personal communication with Professor G.A. Athanassoulis, 1993), based on visual, buoy, and hindcast data presented recently by Watanabe et al. (1992), concluded that an increasing trend of 5 cm yr^{-1} of the mean annual significant wave height is present in the north Pacific Ocean for a period of about 15 years. In more recent studies, Hogben (1995) concluded that increasing trends with a mean value ranging from 4 cm yr^{-1} (based on measured data) to 3.4 cm yr^{-1} (based on visual data) are present for the thirty-five-year period 1950-1985. The measured data used are from the Ocean Weather Stations India, Lima and from Sevenstones, and the visual from areas around Ocean Weather Stations Charlie, Juliett and around Sevenstones. Hogben (1995) also tried to explain increasing trends by relating wave heights with wind speeds in the same sea areas. WASA group (1995a) argued that the reported trends may have been caused by various non-physical reasons, such as, e.g., improvements and changes of instrumentation and observational practices; however Hogben (1995) replied that extensive validation of visual using buoy measurements and use of the same instruments for the whole time period studied assure the accuracy of his results. Kushnir et al. (1995), using model data, related mean air pressure field with mean wave field. Their results showed an increasing trend of 1.9 cm yr^{-1} using data from Sevenstones station. Athanassoulis and Stefanakos (1995) analysed hindcast data from various sites in the north Atlantic Ocean, and have hound increasing trends ranging from 1.6 to 4.1 cm yr^{-1}; see also Section 4.6.1, where these results are presented.

The question of climatic variability of significant wave height for periods of the order of a

century has been recently studied by some authors (Furnes and Reistad, 1993; WASA group, 1995b; WASA group, 1998). See also von Storch et al. (1994). Using existing databases for the winds over the northeastern Atlantic Ocean and the WINCH wave model of the Norwegian Meteorological Institute, Furnes and Reistad calculated a 109-yearlong time series of H_S from January 1881 to December 1989. From their results it is evident that the wave climate has been subject to significant variations over the last century. Periods of several years with rougher wave climate have been succeeded by periods of quieter conditions. Especially in recent years, there is a sign of an increasing trend starting in the early 1960s in accordance with the results of other studies mentioned above. The results of Furnes and Reistad (1993) for the North Sea are in agreement with results concerning time series of gale index presented by Lamb (1991).

The WASA project (1995b; 1998) has been set up to verify hypotheses of a worsening of the wave climate in the northeast Atlantic Ocean and the adjacent seas. A reconstruction of the wave climate for the last 40 years (1955-1994) was achieved by running the WAM model in two nested grids: a coarse one covering the whole north Atlantic Ocean, and a fine one covering only the northeast Atlantic Ocean. The analysis of wind and wave variability showed that variations in wave climate in the north Atlantic Ocean may be related to variations in the North Atlantic Oscillation (NAO) index, which is built from the pressure difference between Iceland and Azores and is a major mode of atmospheric variability (WASA group, 1998). Moreover, the regional variability of the trends computed from the WASA data shows that wind speed and wave height tendencies are closely correlated.

The aforementioned characteristics (serial correlation, seasonality, trends, and climatic variability) of many-yearlong data series of significant wave height should be taken into account in modeling and simulation studies, as well as in developing techniques for statistical (or stochastic) prediction of extreme and cumulative values. The main goal of this work is to study existing many-yearlong time series of wave data in order to assess a periodically correlated stochastic model, which reproduces serial correlation and seasonality and is capable of including lower-frequency climatic variability. Note, however, that the very long term climatic variability is not incorporated in the present model. Nevertheless, if enough data are available, the climatic time series of annual values can be also appropriately modeled, in which case a two-level stochastic structure can be constructed, with the periodically correlated model as the inner (fast time) one and the climatic model as the outer (slow time) one that drives the coefficients (parameters) of the inner model. For example, see Athanassoulis et al. (1992), who discuss thoroughly the concept of a two-level stochastic structure in the context of long-term wave climate modeling, although from a slightly different point of view.

Although, no similar results, up to present, have come to our knowledge for the wave period (zero-upcrossing, mean etc), analysis of time series of mean (energy) wave period T_e in the present work shows that wave period exhibits similar behaviour to the one exhibited by the significant wave height, and, thus, admits similar time series modelling. See also Figure 2.4, where decomposition (2.3) and model (2.4) for an eight-year long time series of mean wave period T_e are shown.

2.4 Tables and Figures

Table 2.1: Long-term time scales of wave-parameter variability

– Sampling subscale	3	hours (1–6 hours)
– Weather system	50	hours (1–5 days)
– Seasonal variability	1	year
– Interannual variability	10	years
– Climatic	100	years(?)

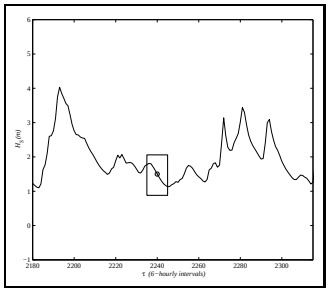
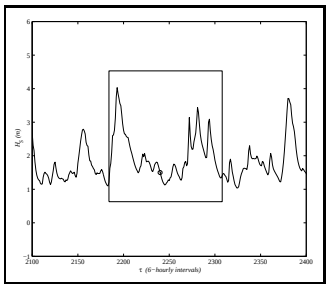
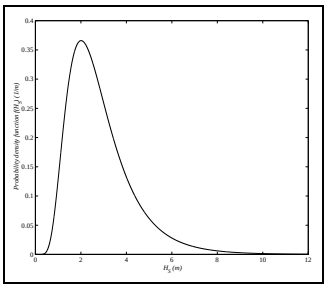
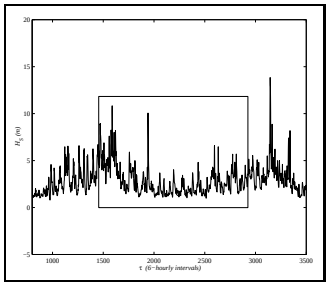
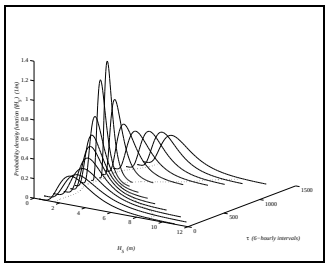
Time window	H_S-series	Model	$f(H_S)$
6 hours (Sea-state)		*Numerical value*	
1 Month		*Random variable*	
1 Year		*Random process*	

Figure 2.1: Modelling of significant wave height in various time scales

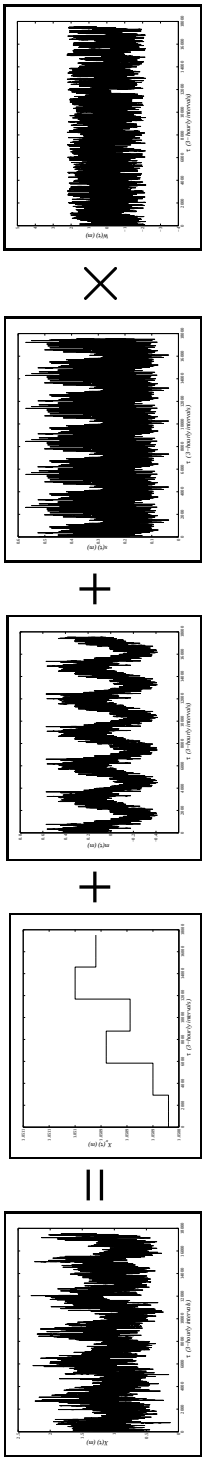


Figure 2.2: Direct numerical decomposition of $X(\tau)$ [Equ. (2.3)]

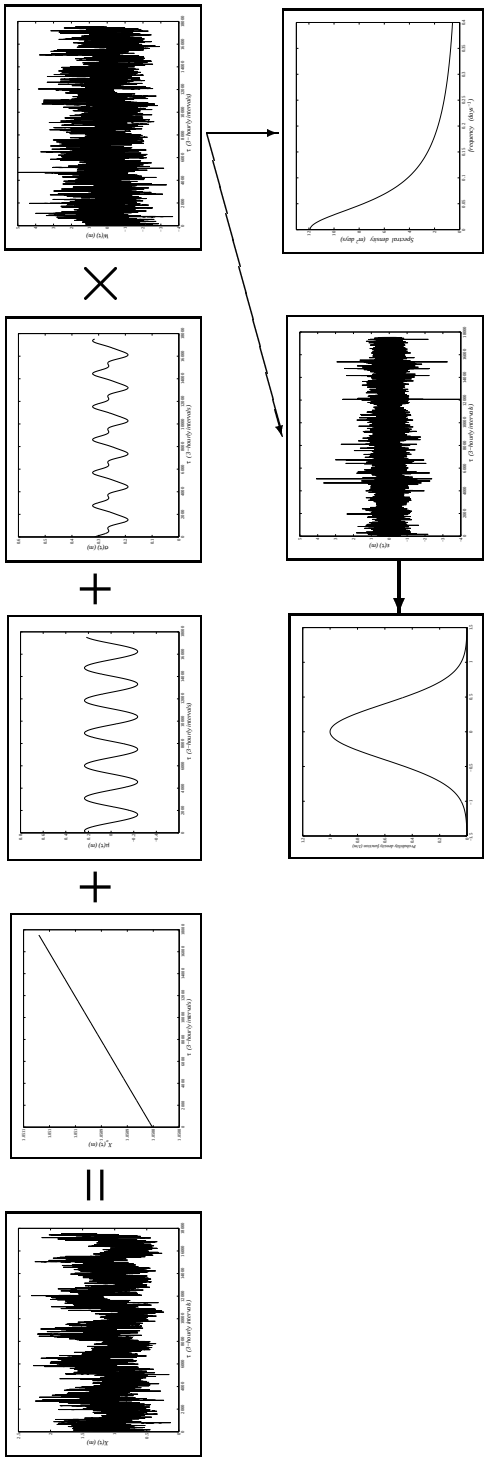


Figure 2.3: Modelling of $X(\tau)$ as a periodically correlated stochastic process [Equ. (2.4)]

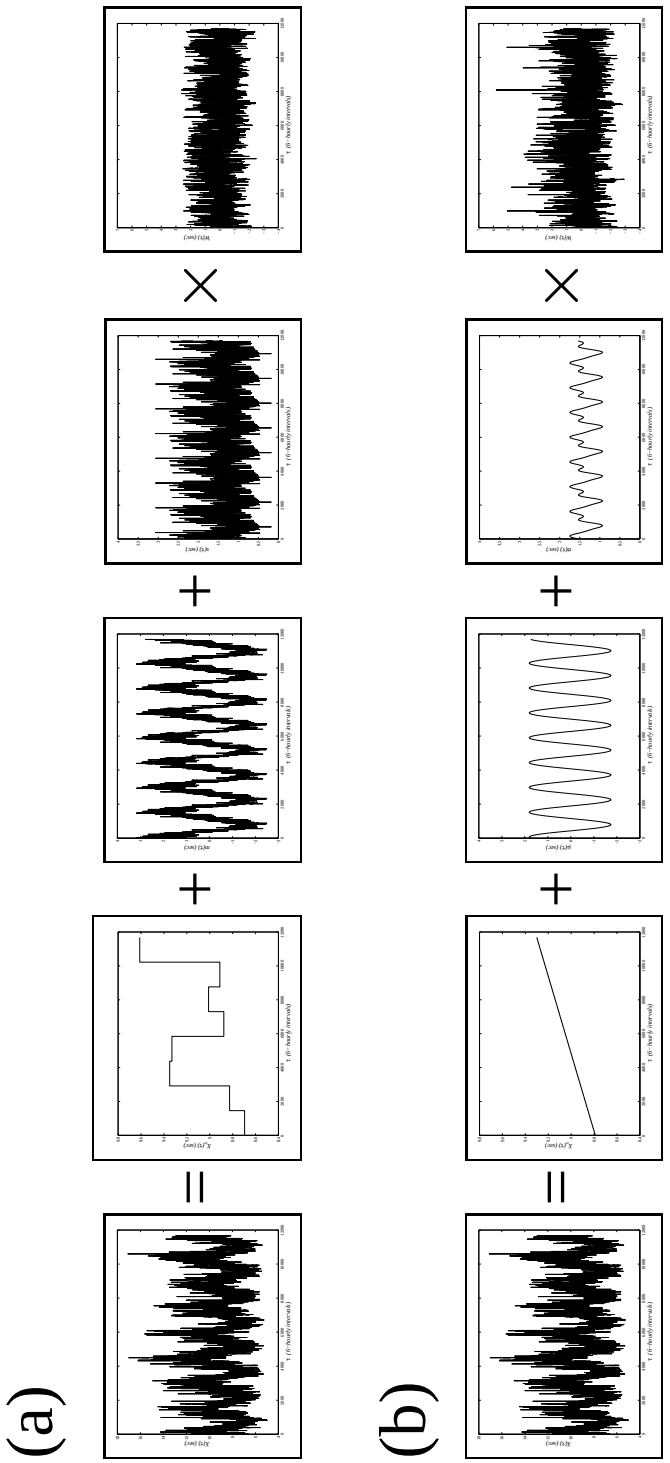


Figure 2.4: (a) Time series analysis of $T_e(\tau)$ by means of decomposition (2.3), (b) Modelling of $T_e(\tau)$ as a periodically correlated stochastic process, using equ. (2.4)

Chapter 3

Cyclostationary stochastic modeling

3.1 Cyclostationary stochastic processes

Consider the class of nonstationary stochatic processes $X(\tau)$ with finite second order moments, and whose the mean value and the aurocovariance functions have the following property:

$$\begin{aligned} m_X(\tau) &= m_X(\tau + T^c), \\ C_{XX}(\tau_1, \tau_2) &= C_{XX}(\tau_1 + T^c, \tau_2 + T^c), \end{aligned} \tag{3.1}$$

for some fixed T^c>0 and every τ_1, τ_2. These processes will be called cyclostationary in the wide sense, or simply cyclostationary stochastic processes, and T^c period of cyclostationarity (Hurd, 1969; Yaglom, 1987a; Gardner, 1990).

Cyclostationary processes can also be found in the literature with the following names: periodically nonstationary (Gudzenko, 1959; Rytov et al., 1988), periodically correlated (Gladyshev, 1961; Gladyshev, 1963; Hurd, 1969; Yaglom, 1987a), periodically stationary (Parzen and Pagano, 1979; Vecchia, 1985). The fundamentals of the theory of cyclostationary stochastic processes have been well developed and are given by Gardner (1990, Chapter 12); see also Yaglom (1987a, Section 26.5). Also the work by Gladyshev (1961; 1963) and the dissertations by Gladyshev (1962), Brelsford (1967), Hurd (1969), Gardner (1972) and Brown (1987) should be mentioned. For a recent survey of theoretical results of cyclostationary processes, see Dehay and Hurd (1994) and the Proceedings of a specialised Workshop (Loubaton and Hurd, 1996). For multivariate cyclostationary processes, see, e.g., Alekseev (1991), Dehay and Hurd (1996), Hurd, Kallianpur and Farshidi (1997).

Cyclostationary processes are natural stochastic models for many environmental phenomena exhibiting periodicities in various disciplines, such as, e.g., astronomy, radiophysics, meteorology, oceanography, hydrology, physiology, biology, economics, and electrical and mechanical

engineering, etc. Especially, for oceanographic applications, one can refer to, e.g., Dragan and Yavorskii (1982), Dragan et al. (1984), Ortiz and Ruiz de Elvira (1985), the discussion by Yaglom (1987a, Section 26.5) and the corresponding notes in Yaglom (1987b). However, to the author's knowledge, the first application of periodically correlated stochastic processes to the long-term description of wave climate has been given by Athanassoulis and Stefanakos (1995); see also Chapter 4, where the results of this work are presented.

Before proceeding to the following section, where general guidelines for the application of cyclostationary modeling to many-year long time series of wave parameters are given, let us first give some of the basic properties of cyclostationary processes:

- If the property of cyclostationarity (equs. (3.1)) holds true, then the following holds also true (Yaglom, 1987a)

$$\begin{aligned} m_X(\tau) &= m_X(\tau + kT^c), \\ C_{XX}(\tau_1, \tau_2) &= C_{XX}(\tau_1 + kT^c, \tau_2 + kT^c), \end{aligned} \tag{3.2}$$

  for any integer k, i.e. the cyclostationary process $X(\tau)$ possesses periodic mean value function $m_X(\tau)$ and doubly periodic aurocovariance function $C_{XX}(\tau_1, \tau_2)$. Note that, one should distinguish between cyclostationary and periodic stochastic processes, in which

$$C_{XX}(\tau_1, \tau_2) = C_{XX}(\tau_1, \tau_2 + T^c) \neq C_{XX}(\tau_1 + T^c, \tau_2 + T^c), \tag{3.3}$$

  for some fixed period T^c>0 and every τ_1, τ_2 (Stark and Woods, 1986).

- Stationary stochastic processes exhibit the cyclostationarity property (equs. (3.1)) for any arbitrary fixed T^c, and, thus, they are members of the class of cyclostationary processes (Yaglom, 1987a).

- Cyclostationary processes are asymptotically stationary, i.e.,

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} m_X(\tau) d\tau &= m_X, \\ \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} C_{XX}(\tau, \tau + r) d\tau &= C_{XX}(r). \end{aligned} \tag{3.4}$$

  This can be confirmed by means of cyclostationarity property (equs. (3.1)); see, e.g., Hurd (1969, p. 84), Yaglom (1987a, p.469) or Gardner (1990, p. 332).

- A random sequence $\{X_i, \quad i = 1, \ldots, I\}$ is a cyclostationary one with period K if and only if the vector random sequence $\{\boldsymbol{X}_j, \quad j = 1, \ldots, J\}$, where $\boldsymbol{X}_j = [X_{j,1}, X_{j,2}, \ldots,$

$X_{j,k}, \ldots, X_{j,K}]^{\mathrm{T}}$, and i=$(j-1)K+k$, is a second-order stationary vector process (Gladyshev, 1961; Vecchia, 1985). The same result for continuous-time parameter is given by Gardner (1990, p. 363), and is stated as follows: $X(\tau)$ is a cyclostationary (periodically correlated) process if and only if admits of the representation

$$X(\tau) = \sum_{k=-\infty}^{+\infty} W_k(\tau) \exp\left\{\frac{ik2\pi\tau}{T^c}\right\}, \tag{3.5}$$

where the processes $W_k(\tau)$ are all stationary and stationarily correlated and T^c is the period of cyclostationarity.

3.2 Cyclostationary modeling

In Section 2.2, an approach for the wave climate description is presented, by means of many-year long time series of a wave parameter $X(\tau)$; see equs. (2.3) and (2.4). Adopting the assumption of long-term stationarity of the residual component $W(\tau)$, we are in a position to introduce the following cyclostationary stochastic model for $X(\tau)$:

$$X(\tau;\beta) = G(\tau) + \sigma(\tau)W(\tau;\beta), \tag{3.6}$$

where $W(\tau;\beta)$ is a zero-mean stationary stochastic process, while $G(\tau)$ and $\sigma(\tau)$ are assumed to be deterministic functions. The argument β, appearing in (3.6), is a choice variable (stochastic argument), taking values in an appropriate sample space B, and is used to distinguish stochastic from deterministic functions. Function $\sigma(\tau)$ is assumed periodic with period 1 year, while function $G(\tau)$ may contain both periodic and nonperiodic components; compare with (2.4). See also (4.14) and (5.1).

Let us now examine what kind of stochastic process is $X(\tau;\beta)$, as defined through (3.6). Obviously, its mean-value function is given by

$$\mathsf{E}^{\beta}[X(\tau;\beta)] = G(\tau), \tag{3.7}$$

while, after some simple algebra we obtain the following formula for the autocovariance function of $X(\tau;\beta)$:

$$\begin{aligned} C_{XX}(\tau_1,\tau_2) &= \mathsf{E}^{\beta}[(X(\tau_1;\beta) - G(\tau_1)) \\ &\qquad (X(\tau_2;\beta) - G(\tau_2))] = \\ &= \sigma(\tau_1)\sigma(\tau_2)C_{WW}(\tau_1,\tau_2), \end{aligned} \tag{3.8}$$

where $C_{WW}(\tau_1, \tau_2)$ is the autocovariance function of the process $W(\tau; \beta)$. In accordance with the stationarity of $W(\tau; \beta)$,

$$C_{WW}(\tau_1, \tau_2) = C_{WW}(|\tau_1 - \tau_2|), \tag{3.9}$$

$$C_{WW}(\tau_1, \tau_2) \to 0, \quad |\tau_1 - \tau_2| \to \infty. \tag{3.10}$$

Furthermore, the correlation time of $W(\tau; \beta)$ is usually of the order of magnitude of some days (see, e.g., Figure 4.9). Thus, $C_{WW}(|\tau_1 - \tau_2|)$ is expected to become practically zero if $|\tau_1 - \tau_2|$ is greater than, say, 15 days.

Now, taking into account (3.9) and (3.10) and the periodicity of $\sigma(\tau)$, we easily find that

$$\begin{aligned} C_{XX}(\tau_1, \tau_2) &= C_{XX}(\tau_1 + T^\alpha, \tau_2 + T^\alpha) \neq \\ &\neq C_{XX}(\tau_1, T^\alpha + \tau_2), \end{aligned} \tag{3.11}$$

where T^α=1 year.

Equations (3.7) and (3.11) show that the stochastic process $X(\tau; \beta)$, as defined by (3.6), is a periodically correlated one (but not a periodic stochastic process). As it has been mentioned previously, periodically correlated (or cyclostationary) processes are natural stochastic models for many environmental phenomena exhibiting periodicities in various disciplines.

Equation (3.5) may explain why cyclostationary processes are so widely applicable in environmental applications; they are able to represent and appropriately mix periodicities (rhythms) and randomnesses of different characteristics. As Hasselmann and Barnett (1981) stated, “many parameters that measure climatic variability have nonstationary statistics, that is, they depend strongly on the phase of the annual cycle”. The use of cyclostationary processes opens new horizons in stochastic analysis, prediction, and simulation of quantities and events defined on long-term time series of wave spectral parameters. Various important directions of applications, as, e.g., simulation, filling in missing values, and extreme-value prediction, can be implemented within the context of cyclostationary modeling.

In the next chapters, this modeling is applied to many-year long time series of wave parameters for analysing them, for filling in of missing values in them, for describing the probability structure (first and second order) of them, and for deriving return-period predictions from them.

Chapter 4

Analysis procedure for complete time series (without missing values)

Let us denote by $X(\tau)$ the long-term time series of some spectral sea state parameter at a given location in the sea. In this work, only univariate time series will be studied, e.g., $X(\tau) = H_S(\tau)$.

In order to properly treat the yearly statistical periodicity, it is appropriate to reindex the time series $X(\tau)$ by using the Buys-Ballot double index (j, τ^α), where j is a year index ranging through the set of integers $\{1, 2, \ldots, J\}$ and τ^α is the time within the annual cycle ranging through the annual interval $0 \leq \tau < T^\alpha = 1$ year [cf. Kottegoda (1980); Salas et al. (1980, Section 2.2); Wei (1990, Section 8.2)]. In the subsequent analysis, τ^α and T^α, as well as τ, will be always expressed in hours ($T^\alpha = 365 \times 24 = 8760$ hours; the extra day of leap years is disregarded). In order to distinguish between τ and τ^α, we shall occasionally use the terms "the long-term time" for τ and "the annual time" for τ^α.

Using the double index (j, τ^α), any many-yearlong time series

$$\left\{X(\tau), \quad 0 \leq \tau \leq T^{LT}\right\}, \tag{4.1}$$

defined on the long-term interval $[0, T^{LT}]$, can be reindexed as follows:

$$\left\{X(j, \tau^\alpha), \quad j = 1, 2, \ldots, J, \quad 0 \leq \tau^\alpha < T^\alpha\right\}. \tag{4.2}$$

Assuming, without any loss of generality, that the value $\tau = 0$ of the single index τ corresponds to the value $(1, 0)$ of the double index (j, τ^α), a one-to-one correspondence between τ and (j, τ^α) can easily be established by means of equations

$$j = \left[\frac{\tau}{T^\alpha}\right] + 1, \quad \tau^\alpha = \tau - (j-1)T^\alpha, \tag{4.3}$$

$$\tau = \tau^\alpha + (j-1)T^\alpha, \tag{4.4}$$

where $[x]$ denotes the integer part of x. Let it be noted that the use of both single-index and double-index representations (4.1) and (4.2), respectively, is absolutely necessary in order to successfully carry out the analysis of time series exhibiting statistical annual periodicity.

Both long-term time τ and annual time τ^α can (and will) be considered either continuous or discrete, although in practice, they are always discrete. If we denote by $\Delta\tau^\alpha \equiv \Delta\tau$ the sampling interval (usually $\Delta\tau^\alpha$ has a value between 1 and 6 hours; see Chapter 2), then the discrete versions of the indices τ^α and τ have the form

$$\begin{aligned} \tau_k^\alpha &= (k-1/2)\Delta\tau^\alpha, \quad k = 1, 2, \ldots, K = \frac{T^\alpha}{\Delta\tau^\alpha}, \\ \tau_i &= (i-1/2)\Delta\tau, \qquad i = 1, 2, \ldots, I = JK, \end{aligned}$$

where τ_k^α and τ_i are taken to be the midvalues of consecutive sampling intervals.

In accordance with (4.3) and (4.4), there is a one-to-one correspondence between the discrete single index i and the discrete double index (j, k), which is realized by means of relations

$$j = \left[\frac{i-1/2}{K}\right] + 1, \quad k = i - (j-1)\,K. \tag{4.5}$$

After this preliminary discussion on notational aspects we are in a position to briefly describe the procedure of analysis that will be subsequently applied to many-yearlong time series of significant wave height.

4.1 Long-term trend

To bring to light year-to-year variability and possibly overyear trends, we first apply a yearly time average, obtaining the sequence of mean annual values

$$\overline{X}(j) = \frac{1}{K}\sum_{k=1}^{K} X(j, \tau_k^\alpha), \qquad j = 1, 2, \ldots, J. \tag{4.6}$$

The sequence $\{\overline{X}(j)\}$ exhibits a relatively small variation coefficient and, in some cases, a nonnegligible increasing trend. The statistical characteristics of this sequence are discussed in Section 4.6.

To eliminate the trend, the linear function

$$\overline{X}_{\rm tr}(\tau) = B_0 + B_1\frac{\tau}{T^\alpha} \tag{4.7}$$

is fitted to the data points $[\tau_j, \overline{X}_{\rm tr}(\tau_j)]$ = $[j\ T^\alpha, \overline{X}(j)]$, and the trend-free time series

$$Y(\tau) = X(\tau) - \overline{X}_{\rm tr}(\tau) \tag{4.8}$$

is obtained.

4.2 Seasonal characteristics

Using the yearly segments $Y(j, \tau^\alpha)$ of the detrended time series $Y(\tau)$, we can estimate the seasonal mean value $m(\tau^\alpha)$ and standard deviation $s(\tau^\alpha)$ by means of the formulas (Salas et al., 1980, Section 2.2); (Brockwell and Davis, 1991, Section 1.4):

$$\hat{m}(\tau_k^\alpha) \equiv \hat{m}(k) = \frac{1}{J}\sum_{j=1}^{J}\ Y(j, \tau_k^\alpha), \tag{4.9}$$

$$k = 1, 2, \ldots, K,$$

$$\hat{s}(\tau_k^\alpha) \equiv \hat{s}(k) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\ [Y(j, \tau_k^\alpha) - \hat{m}(\tau_k^\alpha)]^2}, \tag{4.10}$$

$$k = 1, 2, \ldots, K.$$

(The accent over a random quantity is used to denote values estimated from the raw data.)

In order to smooth out the estimated seasonal characteristics $\hat{m}(\tau^\alpha)$ and $\hat{s}(\tau^\alpha)$ and reduce the number of parameters involved in their description (which is essential for modeling purposes), it is postulated that $m(\tau^\alpha)$ and $s(\tau^\alpha)$ admit a low-order Fourier series representation denoted by $\mu(\tau^\alpha)$ and $\sigma(\tau^\alpha)$, respectively. Note that $\mu(\tau^\alpha)$ and $\sigma(\tau^\alpha)$ are estimated as low-order Fourier series of $\hat{m}(\tau^\alpha)$ and $\hat{s}(\tau^\alpha)$, respectively. This step is further described in Section 4.6.

Another useful seasonal characteristic of time series $Y(j, \tau_k^\alpha)$ is the skewness coefficient $g(\tau^\alpha)$, which can be estimated through the relation

$$\hat{g}(\tau_k^\alpha) \equiv \hat{g}(k) = \frac{J^{-1}\sum_{j=1}^{J}[Y(j, \tau_k^\alpha) - \hat{m}(\tau_k^\alpha)]^3}{[\hat{s}(\tau_k^\alpha)]^3}, \tag{4.11}$$

$$k = 1, 2, \ldots, K.$$

Similarly to seasonal mean value $m(\tau^\alpha)$ and standard deviation $s(\tau^\alpha)$, we assume that seasonal skewness coefficient $g(\tau^\alpha)$ admits a Fourier representation denoted by $\gamma(\tau^\alpha)$.

4.3 Time series decomposition

Now, having obtained $\mu(\tau^\alpha)$ and $\sigma(\tau^\alpha)$, we standardize the time series $Y(j,\tau^\alpha)$, obtaining the residual time series

$$W(j,\tau^\alpha) = \frac{Y(j,\tau^\alpha) - \mu(\tau^\alpha)}{\sigma(\tau^\alpha)}, \tag{4.12}$$

which will be synoptically called W residual component.

For some applications it is useful to consider the residual time series

$$Z(j,\tau^\alpha) = Y(j,\tau^\alpha) - \mu(\tau^\alpha) \quad [= \sigma(\tau^\alpha) W(j,\tau^\alpha)], \tag{4.13}$$

obtained by subtracting the estimated seasonal mean value $\mu(\tau^\alpha)$ from $Y(j,\tau^\alpha)$ but without standardizing the variance by dividing by $\sigma(\tau^\alpha)$, which will be synoptically called Z residual component.

A fundamental question here is whether the Z or the W residual component can be considered stationary. This question is discussed in Section 4.6.

Finally, defining $\mu(\tau)$ and $\sigma(\tau)$ to be the periodic extensions of $\mu(\tau^\alpha)$ and $\sigma(\tau^\alpha)$ in the long-term time τ and using $\overline{X}_{\rm tr}(\tau)$ and $W(\tau)$, we decompose the initial time series in the form

$$X(\tau) = \overline{X}_{\rm tr}(\tau) + \mu(\tau) + \sigma(\tau)\, W(\tau). \tag{4.14}$$

With the proviso that $W(\tau)$ can be considered stationary, (4.14) gives $Y(\tau)$=$X(\tau) - \overline{X}_{\rm tr}(\tau)$ the structure of a periodically correlated stochastic process. See also Chapter 3, where definition and main properties of these processes are given. Further, the consequences of this modelling are discussed in Chapter 7.

4.4 Autocovariance and autocorrelation functions

For the analysis of time series $Y(\tau)$, $Z(\tau)$, and $W(\tau)$ the following empirical autocovariance functions are of interest.

(1) The usual time-lag autocovariance function, obtained through a direct time averaging over a long-term sample:

$$\begin{aligned} \hat{C}^{(1)}_{\Psi\Psi}(r) &= \frac{1}{I}\sum_{i=1}^{I-r} [\Psi(\tau_i) - m_\Psi] \\ &\qquad [\Psi(\tau_{i+r}) - m_\Psi]\,, \\ &\qquad r = 0,1,2,\ldots,R < I, \end{aligned} \tag{4.15}$$

where Ψ stands for Y, Z, or W and m_Ψ is the overall time average of the series $\Psi(\tau)$, i.e., $m_\Psi = I^{-1}\sum_{i=1}^{I}\Psi(\tau_i)$.

(2) The two-time autocovariance function, obtained through an averaging over the various yearly segments of a many-year sample:

$$\begin{aligned}\hat{C}^{(2)}_{\Psi\Psi}(\tau^\alpha_k,\tau^\alpha_{k+r}) &\equiv \hat{C}^{(2)}_{\Psi\Psi}(k,r) = \\ &= \frac{1}{J}\sum_{j=1}^{J}\left[\Psi(j,\tau^\alpha_k)-m_\Psi(\tau^\alpha_k)\right] \\ &\qquad \left[\Psi(j,\tau^\alpha_{k+r})-m_\Psi(\tau^\alpha_{k+r})\right], \\ &\quad k=1,2,\ldots,K, \\ &\quad r=0,1,2,\ldots,\ r<K, \end{aligned} \tag{4.16}$$

where $m_\Psi(\tau^\alpha_k)$ denotes the mean value of $\Psi(j,\tau^\alpha_k)$, taken again over the various yearly segments, as in (4.9). Note that if $k+r>K$, then

$$m_\Psi(\tau^\alpha_{k+r}) = m_\Psi(\tau^\alpha_{k+r-K}), \tag{4.17}$$

$$\Psi(j,\tau^\alpha_{k+r}) = \Psi(j+1,\tau^\alpha_{k+r-K}), \tag{4.18}$$

where $\Psi(J+1,\cdot)$ is interpreted as $\Psi(1,\cdot)$. Note also that for $r=0$, (4.16) becomes

$$\hat{C}^{(2)}_{\Psi\Psi}(\tau^\alpha_k,\tau^\alpha_k) \equiv \hat{C}^{(2)}_{\Psi\Psi}(k,0) = \hat{s}^2_\Psi(k), \tag{4.19}$$

where $\hat{s}^2_\Psi(k)$ is the variance of the process $\Psi(\tau)$ (see (4.10)).

The exact meaning of each one of the autocovariance functions defined above is strongly dependent on the nature of the underlying process $\Psi(\tau)$. More precisely, if the process $\Psi(\tau)$ is stationary (and ergodic), then both $\hat{C}^{(1)}_{\Psi\Psi}(r)$ and $\hat{C}^{(2)}_{\Psi\Psi}(k,r)$ are estimators of the standard autocovariance function $C_{\Psi\Psi}(t-\tau)=\mathsf{E}\left[\Psi(t)-m_\Psi\right]\left[\Psi(\tau)-m_\Psi\right]$ [see Priestley (1981, Chapter 5); Brockwell and Davis (1991, Chapter 7)]. In this case, $\hat{C}^{(2)}_{\Psi\Psi}(k,r)$ is approximately k independent. On the contrary, if $\Psi(\tau)$ is a general nonstationary process, $\hat{C}^{(1)}_{\Psi\Psi}(r)$ and $\hat{C}^{(2)}_{\Psi\Psi}(k,r)$ have not any specific theoretical meaning. However, if $\Psi(\tau)$ exhibits some kind of periodic behavior, this is clearly reflected to both autocovariances (4.15) and (4.16). Especially if $\Psi(\tau)$ is periodically correlated (with period $T^\alpha = K\ \Delta\tau^\alpha$), as is expected to be the case for the detrended significant wave height time series examined herein, then $\hat{C}^{(1)}_{\Psi\Psi}(r)$ becomes nearly periodic (Brockwell and Davis, 1991, p. 29), while $\hat{C}^{(2)}_{\Psi\Psi}(\tau_k,\tau_{k+r})$ is a consistent estimator of the standard two-time autocovariance function $C_{\Psi\Psi}(t,\tau)=\mathsf{E}\left[\Psi(t)-m_\Psi(t)\right]\left[\Psi(\tau)-m_\Psi(\tau)\right]$ of the process $\Psi(\tau)$ (Hurd, 1969, Chapter 5).

Sometimes, instead of $\hat{C}^{(1)}_{\Psi\Psi}(r)$ and $\hat{C}^{(2)}_{\Psi\Psi}(k,r)$, it is preferable to use the normalized autocorrelation (coefficient) functions

$$\hat{\rho}^{(1)}_{\Psi\Psi}(r) = \frac{\hat{C}^{(1)}_{\Psi\Psi}(r)}{\hat{C}^{(1)}_{\Psi\Psi}(0)} \tag{4.20}$$

$$\begin{aligned} \hat{\rho}^{(2)}_{\Psi\Psi}(\tau^{\alpha}_{k}, \tau^{\alpha}_{k+r}) &\equiv \hat{\rho}^{(2)}_{\Psi\Psi}(k,r) = \\ &= \frac{\hat{C}^{(2)}_{\Psi\Psi}(k,r)}{\sqrt{\hat{C}^{(2)}_{\Psi\Psi}(k,0)}\sqrt{\hat{C}^{(2)}_{\Psi\Psi}(k+r,0)}} = \\ &= \frac{\hat{C}^{(2)}_{\Psi\Psi}(k,r)}{\hat{s}_{\Psi}(k)\hat{s}_{\Psi}(k+r)}. \end{aligned} \tag{4.21}$$

4.5 Spectral analysis of residual components

The spectrum $S_{\Psi\Psi}(f)$ of the Ψ residual component is estimated from the sample spectrum (periodogram) by applying spectral smoothing of order M_{sm}, and ensemble averaging of order M_{av}. More precisely, the many-yearlong time series $\Psi(\tau)$ is divided into M_{av} segments that are considered to be independent samples of the underlying process, and the periodogram of each segment is calculated using fast Fourier transform (FFT). Each periodogram is smoothed with the aid of an M_{sm} term moving averaging, and the final estimate $\hat{S}_{\Psi\Psi}(f)$ of the target spectrum $S_{\Psi\Psi}(f)$ is obtained by averaging the M_{av} smoothed periodograms. The estimated spectral ordinates $\hat{S}_{\Psi\Psi}(f)$ are distributed as

$$\hat{S}_{\Psi\Psi}(f) \sim S_{\Psi\Psi}(f)\frac{\chi^2(DF)}{DF}, \tag{4.22}$$

where DF is the degrees of freedom of χ^2 distribution. The product $M_{av}M_{sm}$ defines the number of degrees of freedom of the χ^2 distribution, which the sample spectrum follows, i.e., $DF = 2M_{sm}M_{av}$ (Bendat and Piersol, 1971; Papoulis, 1991).

The main reason for calculating the spectrum of the residual components is to examine their stationarity; see Section 4.6. Furthermore, if the stationarity hypothesis can be adopted, then the estimated spectrum provides us with the means of calculating various useful statistical properties of the wave climate. Such an application will be discussed in Chapter 7 and will be developed elsewhere.

4.6 Wave-climate assessement using nonstationary time-series modelling

In this section, the analysis presented in the previous sections (4.1– 4.5) will be applied to hindcast data. More precisely, 20-yearlong time series of hindcast significant wave height for five locations in the North Atlantic Ocean will be used (see Figure 4.1 and Table 4.1). Such time series are available from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). The generation of these data was part of a U.S. Navy project aimed at producing hindcasts of various wind and wave parameters for an extended grid of sea locations in the northern hemisphere (Lazanoff and Stevenson, 1975; Lazanoff and Stevenson, 1978). In the framework of this project, the spectral ocean wave model (SOWM) (Pierson et al., 1966) was implemented, and historical meteorological data, covering the period from January 1956 to December 1975, were used as input. Although SOWM is a first-generation spectral model that neglects wave-wave and wave-current interactions, it gives competent hindcasts for the deep ocean, provided that a strong current is not present (U.S. Navy, 1983). Various comparisons between hindcast data generated by SOWM and other types of data (visually observed or measured) have been performed, and the results have shown a satisfactory overall agreement [see Pierson (1982); U.S. Navy (1983); Bales et al. (1982); and references there in].

4.6.1 Interannual variability and detection of overyear trends

By applying a yearly time average we obtain the sequence of mean annual H_S values $\{\overline{H}_S(j)$, j=1,2, ... ,J=19$\}$ for each one of the five grid points in the North Atlantic Ocean studied herewith (GPs 2, 10, 21, 27, and 31; see also Figure 4.1). In Figure 4.2, mean annual H_S values are plotted, along with the least squares regression line, for GPs 10 and 31. The statistical characteristics (mean value, standard deviation, and variation coefficient) of each 19-term long sequence of mean annual H_S values are presented in Table 4.1. Although, as expected, the 19-year mean H_S value is clearly site dependent (see also Section 4.6.2), the corresponding variation coefficient shows a remarkable site independence, being approximately equal to 0.11. The small amount of data of mean annual values does not permit us to proceed to a deeper statistical analysis and modeling of the time series $\{\overline{H}_S(j),\ \ j = 1, 2, \ldots, J = 19\}$.

On the other hand, from Figure 4.3 it might be conjectured that an increasing trend is present in the data set, in accordance with the findings of other authors discussed in Section 2.3. This very important issue is studied here by introducing the linear regression model

$$\overline{H}_S(j) = \beta_0 + \beta_1 j + e_j, \qquad j = 1, 2, \ldots, J = 19, \tag{4.23}$$

where e_j are independently and identically distributed (i.i.d.) random variables, and testing the hypothesis

$$\mathcal{H}_0 : \quad \beta_1 = \beta_{10} \tag{4.24}$$

(specified), against the alternative $\mathcal{H}_1$: $\beta_1 \neq \beta_{10}$. For this purpose, the C statistic [Theil (1950a; 1950b; 1950c)]

$$C = \sum_{1 \leq i < j \leq J} \text{sign}(D_j - D_i) \tag{4.25}$$

is used, where $D_j = \overline{H}_S(j) - \beta_{10} j$ and sign(α) is defined by

$$\text{sign}(\alpha) = \frac{\alpha}{|\alpha|}, \quad \alpha \neq 0, \tag{4.26}$$

and sign(0)=0.

Theil's test is discussed, e.g., by Hollander and Wolfe (1973, Chapter 9), and Bhattacharyya (1984) and can be considered as an extension of the Mann-Kendall test for independence.

The value β_{10}, used in testing $\mathcal{H}_0$, is taken to be the slope of the linear (least squares) regression line. The value of C statistic is the median estimator of the slope β_1. The values of these (least squares and median) estimators are shown in Table 4.2. The results obtained are, in general, close, except for the GP 27. Slope estimates vary from 1 up to 4 cm yr^{-1} (for GP 10, near station India). Also in Table 4.2, the 95% confidence intervals of the slopes β_1, obtained through C statistic (4.25), are shown (Hollander and Wolfe, 1973). On the basis of the fact that C statistic lies always within the corresponding confidence intervals, i.e., hypothesis $\mathcal{H}_0$ is not rejected, we can conclude that the hypothesis that "an increasing trend is present in the examined data" is very likely true.

4.6.2 Seasonal variability

Apart from interannual variability, time series of H_S exhibits seasonal variability related to the annual cycle of the meteorological phenomena involved. This kind of variability is clearly seen in Figure 4.4a, where the estimated autocorrelation $\hat{\rho}^{(1)}_{YY}(r)$ is plotted against time lag r for GP 10. Disregarding the first hundred lags, $\hat{\rho}^{(1)}_{YY}(r)$ is nearly periodic with period 1 year (1460 lags). Recall that $\hat{\rho}^{(1)}_{YY}(r)$ is the estimated time-lag autocorrelation function of the detrended H_S time series, denoted by $Y(\tau)$. It is noteworthy that the autocorrelation $\hat{\rho}^{(1)}_{XX}(r)$ of the initial time series $X(\tau)$ is practically identical with $\hat{\rho}^{(1)}_{YY}(r)$.

The seasonal mean value $m(\tau^\alpha)$, standard deviation $s(\tau^\alpha)$, and skewness coefficient $g(\tau^\alpha)$ of $Y(\tau)$ are estimated by means of (4.9)–(4.11). Their low-order Fourier representations (denoted by $\mu(\tau^\alpha)$, $\sigma(\tau^\alpha)$, and $\gamma(\tau^\alpha)$, respectively) are of the form

$$\delta(\tau^\alpha) = A_0^\delta + \sum_{n=1}^{N_\delta} A_{cn}^\delta \cos\left[\frac{2\pi n\tau^\alpha}{T^\alpha}\right] + A_{sn}^\delta \sin\left[\frac{2\pi n\tau^\alpha}{T^\alpha}\right], \tag{4.27}$$

where δ stands for μ, σ, or γ. Fourier coefficients A_0^δ, A_{cn}^δ, and A_{sn}^δ are estimated by [see Anderson (1971)]

$$A_0^\delta = \frac{1}{T^\alpha}\sum_{k=1}^{K} \hat{d}(\tau_k^\alpha), \tag{4.28}$$

$$A_{cn}^\delta = \frac{2}{T^\alpha}\sum_{k=1}^{K} \hat{d}(\tau_k^\alpha)\cos\left[\frac{2\pi n\tau_k^\alpha}{T^\alpha}\right], \tag{4.29}$$

$$A_{sn}^\delta = \frac{2}{T^\alpha}\sum_{k=1}^{K} \hat{d}(\tau_k^\alpha)\sin\left[\frac{2\pi n\tau_k^\alpha}{T^\alpha}\right], \tag{4.30}$$

where $\hat{d}(\tau_k^\alpha)$ stands for $\hat{m}(\tau_k^\alpha)$, $\hat{s}(\tau_k^\alpha)$, or $\hat{g}(\tau_k^\alpha)$ (see (4.9)–(4.11)). The number of harmonics N_δ considered in (4.27) should be chosen in such a way so that Fourier series $\mu(\tau^\alpha)$, $\sigma(\tau^\alpha)$ and $\gamma(\tau^\alpha)$ represent sufficiently well the main seasonal pattern. Previous experience (Cummins and Bales, 1980; Medina et al., 1991; Frutuoso et al., 1993) shows that a small number of harmonics (one to five) suffices. In this work it has been chosen N_μ=1, N_σ=3, and N_γ=3.

In Figure 4.4b, estimated seasonal mean value $\hat{m}(\tau^\alpha)$, as well as its first-order Fourier representation $\mu(\tau^\alpha)$, are plotted against the annual time τ^α for GP 10. Similar plots for seasonal standard deviation and seasonal skewness coefficient for GP 10 are shown in Figures 4.4c and 4.4d. All seasonal characteristics shown in Figure 4.4 have been estimated by averaging over the total time period (J=19 years). It is very interesting to examine the effect of averaging over smaller time periods. Such an investigation has been conducted by averaging over smaller periods, i.e., J=5 and 10 years. The corresponding results for the low-order Fourier representations of seasonal mean value and standard deviation at GPs 10 and 31 are shown in Figure 4.5. As the time period of averaging increases from J=5 to J=10 years, the "fluctuations" of the estimated functions $\mu(\tau^\alpha)$ and $\sigma(\tau^\alpha)$ around the corresponding estimates for J=19 (solid line) become smaller, as expected. From these results it can also be conjectured that 10-yearlong wave data

characterize sufficiently well the seasonal variability at a given site. It is interesting, however, to note here that satellite data could help in estimating the seasonal characteristics by replacing time averaging with spatial averaging over appropriately selected ranges.

The spatial variability of mean seasonal characteristics is also of interest. This issue has also been studied, and some results are shown in Figures 4.5 and 4.6. Especially in Figure 4.6, the seasonal mean value and standard deviation for five locations in the North Atlantic Ocean (see Figure 4.1) are plotted. If we exclude GP 2, which is located in a different sea area (off the Norwegian coast), the other four locations form an arc from north to south on which a spatial trend seems present. This trend seems to be multiplicative for seasonal mean value and additive for seasonal standard deviation. Clearly, wave characteristics become more and more severe as we move from south to north. However, spatial variability should be more extensively studied in order to be able to decide the space scale in which a geographic parameterization of mean seasonal characteristics would be meaningful.

4.6.3 Analysis of residual components

In order to study its stationarity, time series of the Ψ residual component (Ψ stands for either Z or W residual components; see Section 4.5) is segmented in J ($\equiv M_{av}$=19) yearly records that are considered mutually independent. Using DFT (2048-point FFT), the one-sided spectral densities

$$S_{\Psi\Psi}^{(j)}(f_n) = 2\Delta\tau^{\alpha} \left| \sum_{k=1}^{K} \Psi(j, \tau_k^{\alpha}) \exp(i2\pi f_n \tau_k^{\alpha}) \right|^2 , \qquad (4.31)$$

$$j = 1, 2, \ldots, J \equiv M_{av}$$

are calculated for each yearly segment. Recall that the sampling period $\Delta\tau^{\alpha}$=6 hours.

If we denote by f_c=1/(2$\Delta\tau^{\alpha}$) the Nyquist frequency and by L the total number of sampling points, the frequency range that can be examined varies from $f_{\min}$= Δf=2f_c/L to $f_{\max}$=f_c. In our case the number of sampling points (number of 6-hour intervals) within a year is 1460, and thus we choose L=2048(=2^{11}) by adding zeroes to the basic record (Geçkinli and Yavuz, 1983). Accordingly, the frequency range varies from 0.002 to 2 days^{-1}. This frequency range is separated into two parts (1) from 0 to 0.4 days^{-1} called the low-frequency part and (2) from 0.4 to 2 days^{-1} called the medium-frequency part. Unfortunately, the sampling period ($\Delta\tau$=6 hours) is such that the high-frequency part (from 2 to 48 days^{-1}) cannot be observed.

Each spectral density estimate is smoothed by means of an M_{sm} term moving average (M_{sm}=9). The final spectral density estimate is the average of the above estimates over the yearly segments

given by

$$S_{\Psi\Psi}(f) = \frac{1}{M_{av}} \sum_{j=1}^{M_{av}} S_{\Psi\Psi}^{(j)}(f). \tag{4.32}$$

The main question to be checked here is whether (or to what extent) the residual time series can be considered free of seasonal effects, i.e., stationary. For this purpose, the following test has been conducted: the spectral densities $S_{ZZ}(f)$ and $S_{WW}(f)$ have been estimated by using (1) 19 complete-year segments, (2) 19 "winter" segments (November through February), and (3) 19 "summer" segments (May through August).

In Figure 4.7 the low-frequency part (Figure 4.7a) and the medium-frequency part (Figure 4.7b) of the spectral densities of Z and W residual components are plotted against frequency f (days^{-1}). The low-frequency part of spectral densities is plotted in linear axes, while the medium-frequency part is plotted in logarithmic axes. Estimated spectral density $S_{WW}(f)$ exhibits a remarkable season independence. More precisely, "winter" and "summer" versions of $S_{WW}(f)$ lie within the 95% confidence limits of the complete-year version of it. This strongly supports the assumption of stationarity of the W residual component. On the other hand, estimates of $S_{ZZ}(f)$ coming from different-season segments are very different. This is a clear indication of the nonstationarity of the Z residual component, which can be attributed to the seasonality of variance $\sigma(\tau^{\alpha})$ of Z component (see (4.13)).

In Figure 4.8 the low-frequency part (Figure 4.8a) and the medium-frequency part (Figure 4.8b) of the spectral densities $S_{ZZ}(f)$ and $S_{WW}(f)$ are plotted against frequency f (days^{-1}) for the five locations examined in the North Atlantic Ocean (see Figure 4.1). Spectral density $S_{WW}(f)$ shows a remarkable site independence. (See also Figure 4.9, where the corresponding autocorrelation function is shown). On the other hand, $S_{ZZ}(f)$ exhibits a large variability and a clear site dependence. The fact that spectral density $S_{WW}(f)$ seems to be approximately site--independent suggests that there would be a kind of global spectrum characterizing the residual dependence structure of wave data time series throughout large sea areas (e.g., the North Atlantic Ocean). More exhaustive analysis, using measured data also, is needed in order to confirm or disprove this suggestion.

4.7 Conclusions

In this chapter, many-yearlong time series of significant wave height have been analyzed and modeled. It has been found that these time series can be modeled as periodically correlated (cyclostationary) stochastic processes of the form $X(\tau,\beta) = \overline{X}_{\mathrm{tr}}(\tau) + \mu(\tau) + \sigma(\tau)W(\tau,\beta)$, where $W(\tau,\beta)$ is a zero-mean stationary (in general non-Gaussian) stochastic process, $\mu(\tau)$

and $\sigma(\tau)$ are deterministic, yearly periodic functions representing the mean seasonal effects, and $\overline{X}_{\rm tr}(\tau)$ is a deterministic function representing the overyear climatic trend. The analysis of hindcast data of significant wave height for five sea sites in the North Atlantic Ocean led to the following conclusions:

1. It is very likely that an increasing overyear trend (1-4 cm yr^{-1}) is present in the examined period of time (1956-1975). This is in accordance with the findings of other authors (Carter and Draper, 1988; Neu, 1984; Bacon and Carter, 1991; Watanabe et al., 1992),

2. The interannual variability of the studied H_S time series exhibits a remarkable site independence. It has been found that the variation coefficient of annual mean values is approximately 0.11 (0.106–0.118, dependent on the site),

3. The seasonal variability is well characterized by a 10-yearlong record or longer (see Figure 4.5),

4. The time-lag autocorrelation coefficient function $\rho^{(1)}_{WW}(r)$ and the spectral density $S_{WW}(f)$ of the seasonally standardized time series also show a remarkable site independence (see Figures 4.8 and 4.9).

Apart from the above results, the proposed model can also be used for various applications in the context of probabilistic wave climate prediction and engineering design. See also Chapter 8, where the general conclusions of the work are presented.

The present model can be extended to various directions. One of the most important is studied in the next chapter, and it concerns the extension of the estimation procedure to the case where a significant amount of missing values is present.

Finally, let it be noted that this model can also be applied to many other ocean and/or atmospheric parameters, such as, e.g., wave period, wave and wind direction, wind speed, sea temperature, air temperature, humidity, sea level, etc., which are expected to exhibit seasonally varying randomness.

4.8 Tables and Figures

Table 4.1: Statistical characteristics of annual mean values of significant wave height H_S for various locations in the North Atlantic Ocean

Grid point	Latitude (degrees)	Longitude (degrees)	Mean value (meters)	Standard deviation (meters)	Variation coefficient
2	66.6 N	6.8 E	2.173	0.293	0.135
10	58.6 N	18.2 W	3.639	0.386	0.106
21	50.0 N	11.7 W	2.996	0.347	0.116
27	44.6 N	12.9 W	2.680	0.316	0.118
31	39.3 N	13.9 W	2.263	0.256	0.113

Table 4.2: Point and interval estimation of slopes of the linear regression model for the mean annual values of significant wave height H_S

Grid point	Least-square slope, (m yr^{-1})	Median slope, (m yr^{-1})	95% Confidence intervals
2	0.033	0.027	[0.004, 0.055]
10	0.041	0.039	[0.007, 0.070]
21	0.027	0.016	[-0.001, 0.059]
27	0.024	0.009	[-0.009, 0.043]
31	0.016	0.012	[-0.008, 0.033]

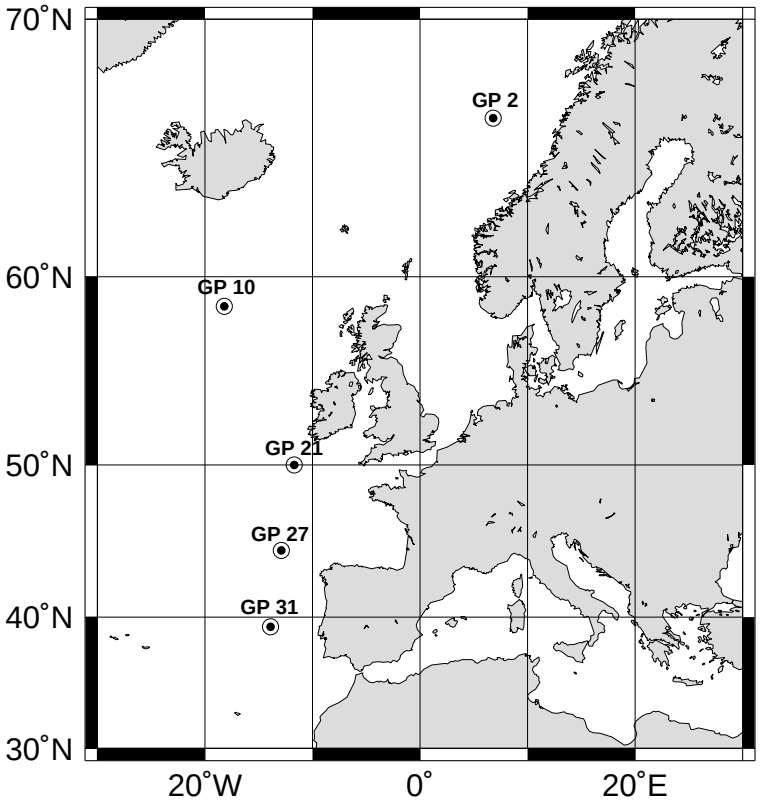


Figure 4.1: Examined sites in the North Atlantic Ocean. Grid points (GPs) 2, 10, 21, 27, and 31.

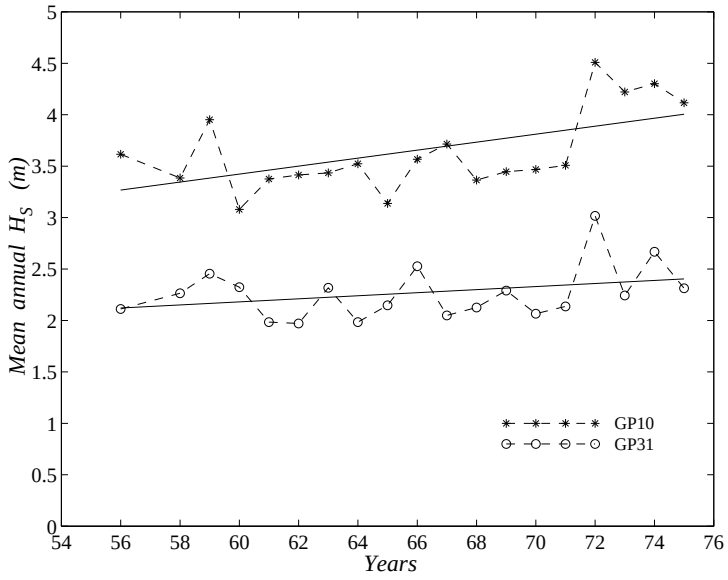


Figure 4.2: Interannual variability and overyear trend of mean annual significant wave height H_S for GPs 10 and 31.

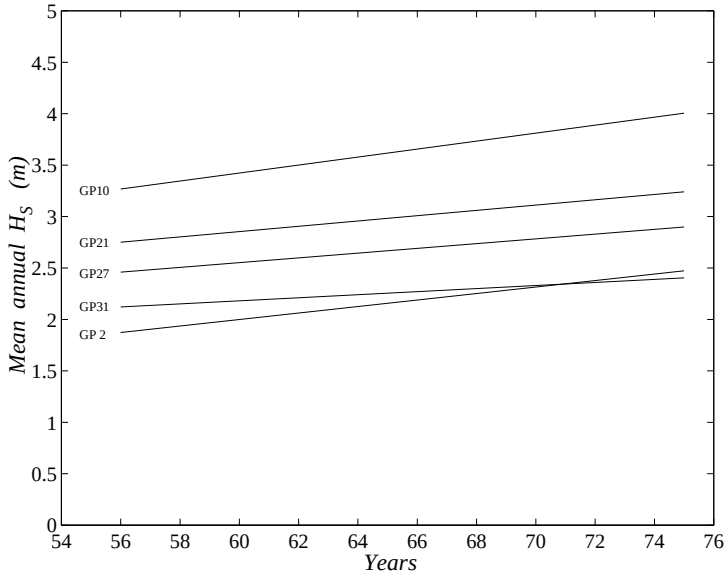


Figure 4.3: Overyear trend of mean annual significant wave height for various locations in the North Atlantic Ocean (GPs 2, 10, 21, 27, 31).

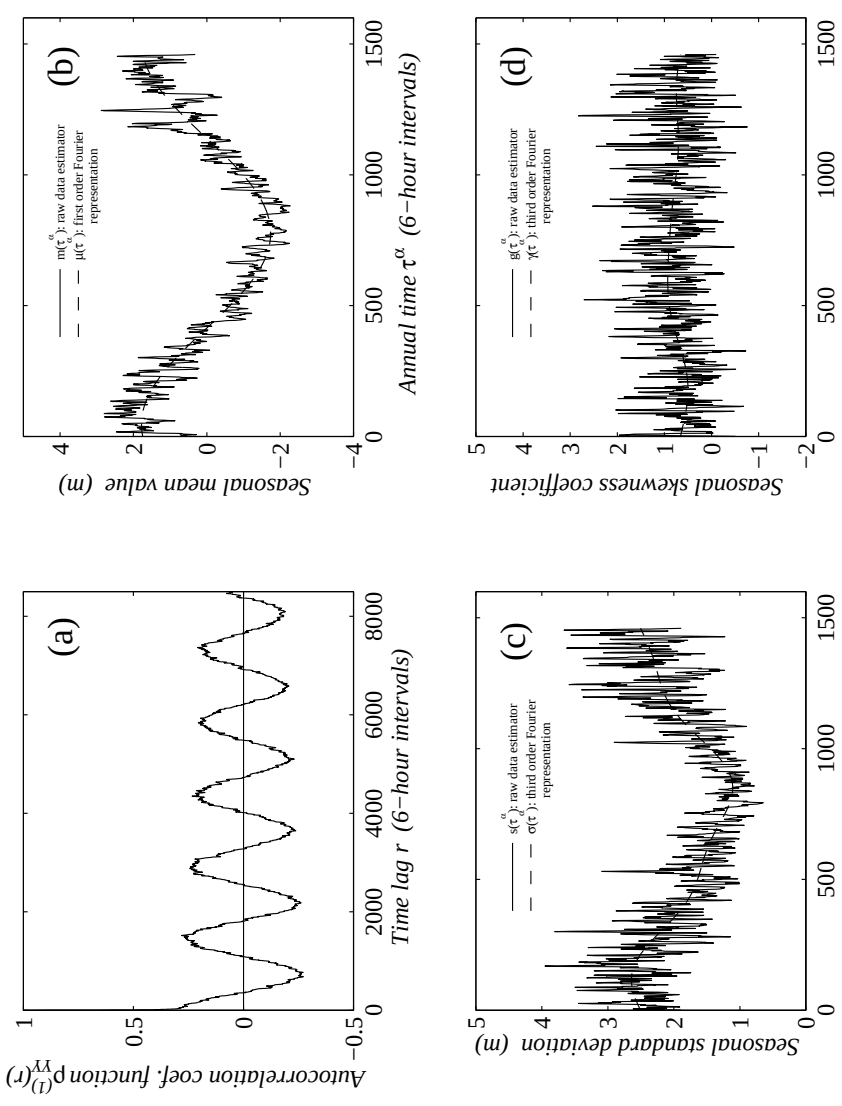


Figure 4.4: Seasonal variability of H_S time series for GP 10. (a) Time-lag autocorrelation function $\hat{\rho}_{YY}^{(1)}(r)$. (b) Seasonal mean value $\hat{m}(\tau^\alpha)$ and its Fourier representation $\mu(\tau^\alpha)$. (c) Seasonal standard deviation $\hat{s}(\tau^\alpha)$ and its Fourier representation $\sigma(\tau^\alpha)$. (d) Seasonal skewness coefficient $\hat{g}(\tau^\alpha)$ and its Fourier representation $\gamma(\tau^\alpha)$.

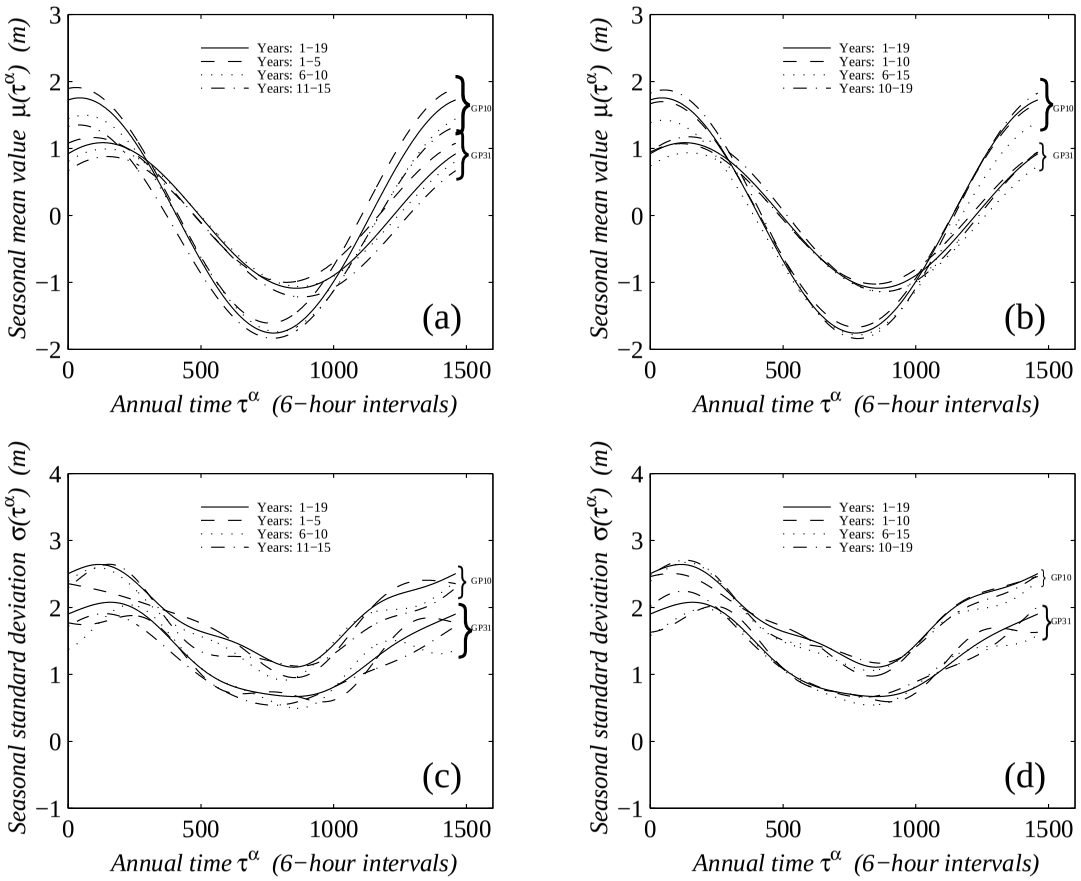


Figure 4.5: Seasonal mean value and standard deviation of H_S time series for different wave data length (J=5 years, J=10 years) and two different locations (GPs 10 and 31). (a) Mean value, J=5. (b) Mean value, J=10. (c) Standard deviation, J=5. (d) Standard deviation, J=10. (Solid lines are obtained throughout by using J=19 years).

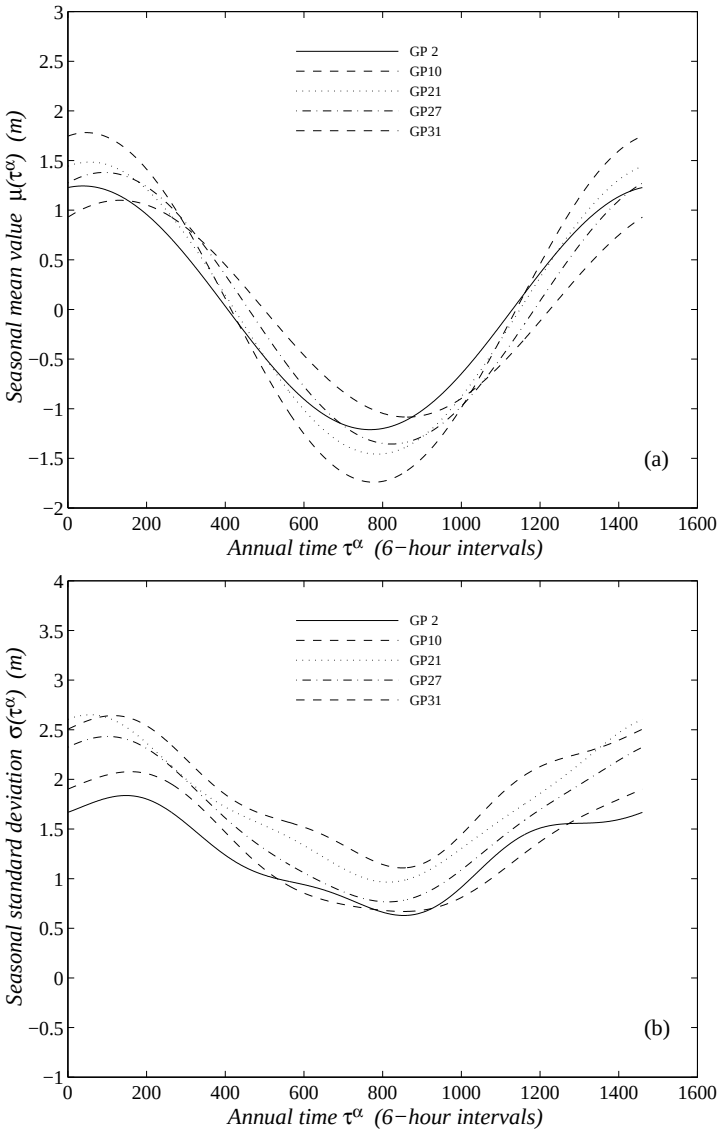


Figure 4.6: Seasonal variability of H_S time series for various locations in the North Atlantic Ocean (GPs 2, 10, 21, 27, and 31). (a) Seasonal mean value. (b) Seasonal standard deviation. (Fourier representations were estimated using J=19 years).

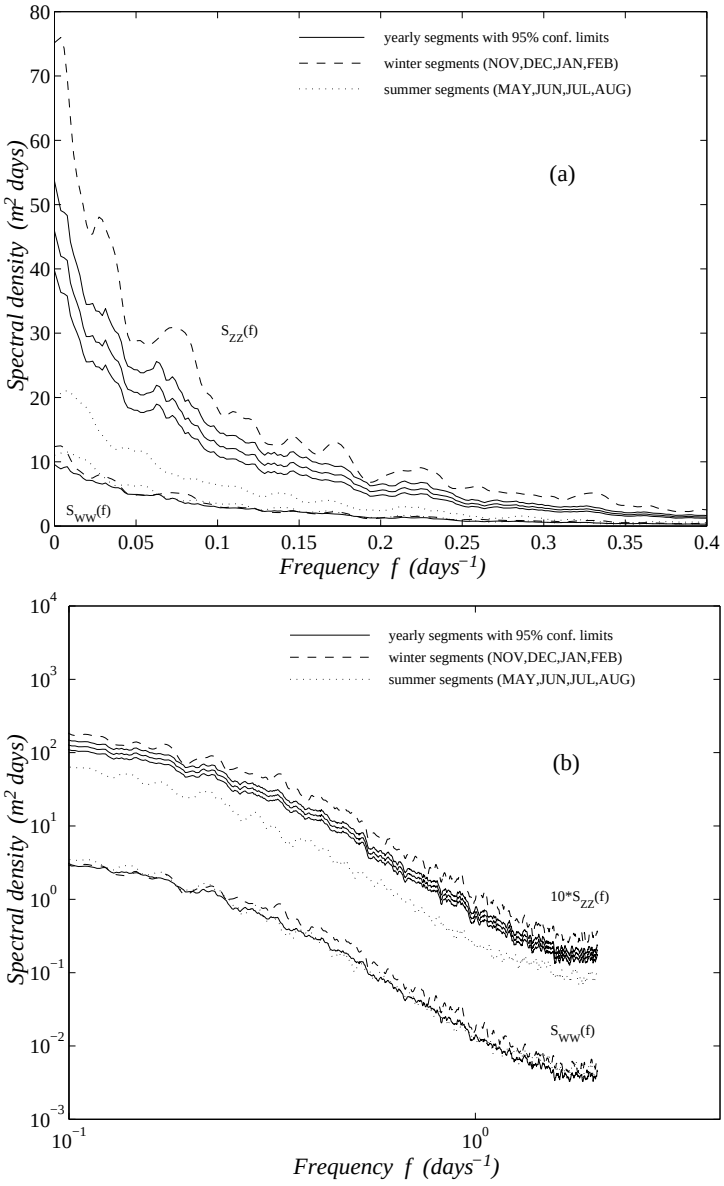


Figure 4.7: Estimated spectral densities $\hat{S}_{ZZ}(f)$ and $\hat{S}_{WW}(f)$ of H_S time series for GP 10 (order of averaging M_{av}=19, order of smoothing M_{sm}=9), showing the (a) low-frequency part and (b) medium-frequency part.

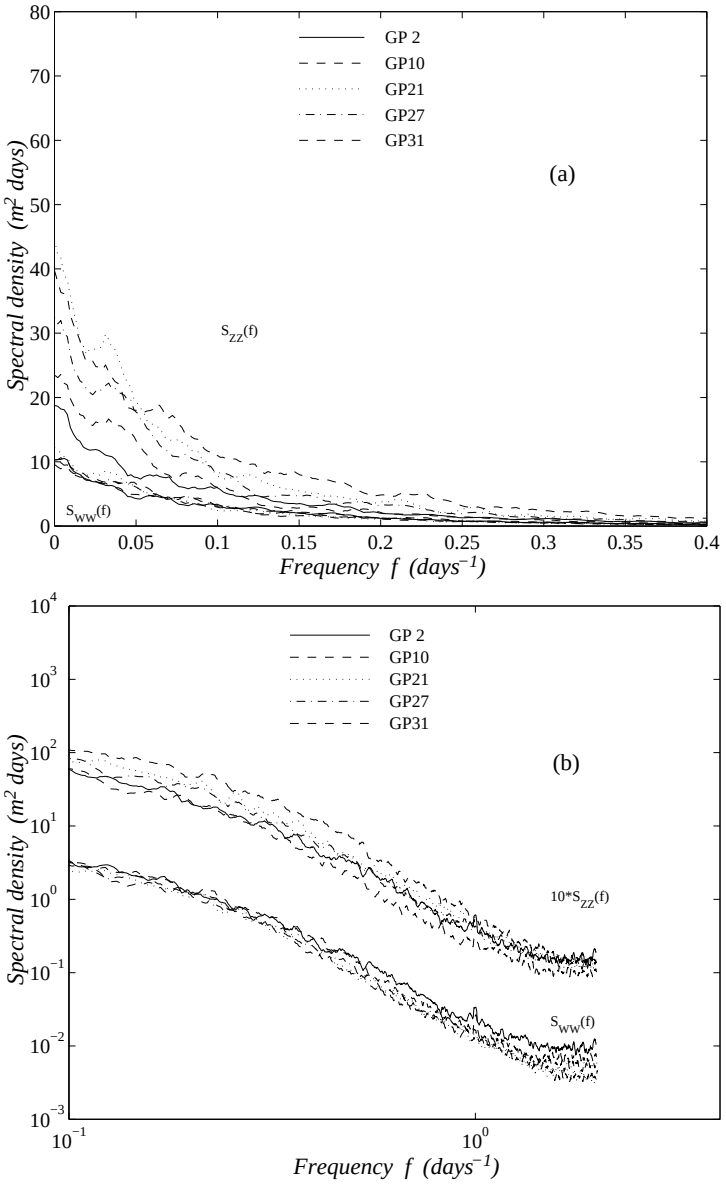


Figure 4.8: Estimated spectral densities $\hat{S}_{ZZ}(f)$ and $\hat{S}_{WW}(f)$ of H_S time series for various locations in the North Atlantic Ocean (GPs 2, 10, 21, 27, and 31), showing the (a) low-frequency part and (b) medium-frequency part.

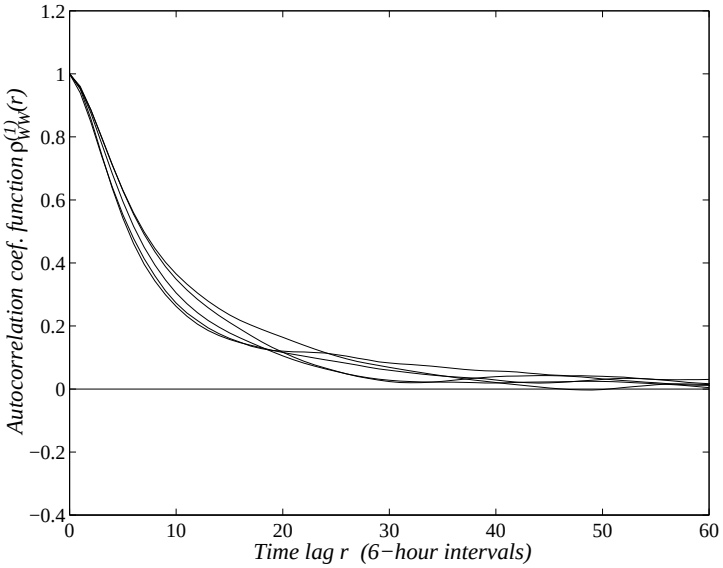


Figure 4.9: Estimated time-lag autocorrelation function $\hat{\rho}_{WW}^{(1)}(r)$ for various locations in the North Atlantic Ocean (GPs 2, 10, 21, 27, and 31).

Chapter 5

Analysis procedure for time series with missing values

5.1 Time series decomposition

Consider again the long-term time series of some spectral sea-state parameter at a given location in the sea, denoted by $X(\tau)$. According to the modelling presented in the previous chapter; see also Parzen and Pagano (1979), Salas (1993), Athanassoulis and Stefanakos (1995), the time series $X(\tau)$ is considered to be nonstationary and, thus, is decomposed in the following form

$$X(\tau) = \overline{X}_{\rm tr}(\tau) + \mu(\tau) + \sigma(\tau)\, W(\tau), \tag{5.1}$$

where $\overline{X}_{\rm tr}(\tau)$, $\mu(\tau)$, $\sigma(\tau)$ are deterministic functions. The function $\overline{X}_{\rm tr}(\tau)$ is linear and it represents an overyear trend, whereas the functions $\mu(\tau)$, $\sigma(\tau)$ are periodic with period one year and represent the seasonal mean value and seasonal standard deviation, respectively.

Assume further that we have at our disposal a discrete set of values, denoted by $X_{\rm e.v.}(\tau)$, which is an incomplete version of time series $X(\tau)$, i.e.

$$X_{\rm e.v.}(\tau) = \Big\{ X(\tau_i), \quad i \in \boldsymbol{I}_{\rm e.v.} = \{i_1, i_2, \ldots, i_{I_{\rm e.v.}}\} \Big\}, \tag{5.2}$$

where the set $\boldsymbol{I}_{\rm e.v.}$ contains the indices of existing observations (e.v.= existing value). Evidently, $\boldsymbol{I}_{\rm e.v.} \subset \boldsymbol{I} = \{1, 2, \ldots, I\}$ with I being the total number of observations. Let us also define the complement of $\boldsymbol{I}_{\rm e.v.}$ as the set $\boldsymbol{I}_{\rm m.v.}$ that contains the indices of missing observations (m.v.=missing value).

In the following subsection, we present a methodology for the estimation and modelling of the deterministic functions of decomposition (5.1), appropriately modified to take into account the missing values.

5.2 Estimation of the deterministic components

Before proceeding with the estimation procedure, let us first re-index time series $X_{\text{e.v.}}(\tau)$ using the Buys-Ballot double index (j, τ_k^α) (Parzen and Pagano, 1979; Wei, 1990) as follows

$$\{X(\tau_i), \quad i \in \boldsymbol{I}_{\text{e.v.}}\} = \bigcup_{j=1}^{J} \Big\{ X(j, \tau_k^\alpha), \quad k \in \boldsymbol{K}_{\text{e.v.}}(j) \Big\}, \tag{5.3}$$

where J is the total number of years, τ^α is the time within the annual cycle (also called annual time), and $\boldsymbol{K}_{\text{e.v.}}(j)$ are the sets of indices of the available time series values for each year j.

In a similar way, the sets $\boldsymbol{J}_{\text{e.v.}}(k)$ are defined, by means of which the incomplete time series is written

$$\{X(\tau_i), \quad i \in \boldsymbol{I}_{\text{e.v.}}\} = \bigcup_{k=1}^{K} \Big\{ X(j, \tau_k^\alpha), \quad j \in \boldsymbol{J}_{\text{e.v.}}(k) \Big\}, \tag{5.4}$$

where K is the total number of time instances within the annual time, and $\boldsymbol{J}_{\text{e.v.}}(k)$ are the sets of indices of the available time series values for each time instance τ_k^α of the annual time.

Evidently,

$$\boldsymbol{I}_{\text{e.v.}} = \bigcup_{j=1}^{J} \boldsymbol{K}_{\text{e.v.}}(j) = \bigcup_{k=1}^{K} \boldsymbol{J}_{\text{e.v.}}(k). \tag{5.5}$$

Note that, the Buys-Ballot double index has been extensively used in hydrology (Kottegoda, 1980; Salas et al., 1980) and recently introduced in ocean engineering (Athanassoulis and Stefanakos, 1995).

According to the Section 4.1 (see also Athanassoulis and Stefanakos (1995)), an estimate of the function $\overline{X}_{\text{tr}}(\tau)$ is obtained by means of the mean annual values $\overline{X}(j)$, which describe the exhibiting year-to-year variability of the time series. In the case of missing values, we calculate the mean annual values

$$\overline{X}_{\text{e.v.}}(j) = \frac{1}{K_{\text{e.v.}}(j)} \sum_{k \in \boldsymbol{K}_{\text{e.v.}}(j)} X(j, \tau_k^\alpha), \qquad j = 1, 2, \ldots, J. \tag{5.6}$$

These mean values are representative of $\overline{X}(j)$ as long as the numbers of existing values per year $K_{\text{e.v.}}(j)$, $j = 1, 2, \ldots, J$, are closely enough to the total number of annual time points K, i.e. $K_{\text{e.v.}}(j) \approx K$ for every j. If this is the case, a linear function of the form

$$\overline{X}_{\text{tr}}(\tau) = B_0 + B_1 \frac{\tau}{T^\alpha} \tag{5.7}$$

is fitted to the data points $(\tau_j, \overline{X}_{\mathrm{tr}}(\tau_j))$ = $(j\ T^{\alpha}, \overline{X}_{\mathrm{e.v.}}(j))$, where T^{α}=1 year.

In this way, the trend-free time series

$$Y_{\mathrm{e.v.}}(\tau) = X_{\mathrm{e.v.}}(\tau) - \overline{X}_{\mathrm{tr}}(\tau) \tag{5.8}$$

is obtained. Note that, time series $Y_{\mathrm{e.v.}}(\tau)$ can be written in the same form as time series $X_{\mathrm{e.v.}}(\tau)$ does; see equs. (5.3), (5.4).

The seasonal characteristics (mean value and standard deviation) of the detrended time series $Y_{\mathrm{e.v.}}(j, \tau^{\alpha})$, are calculated using the existing amount of values, i.e.

$$m_{\mathrm{e.v.}}(\tau_k^{\alpha}) = \frac{1}{J_{\mathrm{e.v.}}(k)} \sum_{j \in \boldsymbol{J}_{\mathrm{e.v.}}(k)} Y(j, \tau_k^{\alpha}), \qquad k = 1, 2, \ldots, K, \tag{5.9}$$

and

$$s_{\mathrm{e.v.}}(\tau_k^{\alpha}) = \sqrt{\frac{1}{J_{\mathrm{e.v.}}(k)} \sum_{j \in \boldsymbol{J}_{\mathrm{e.v.}}(k)} \left[Y(j, \tau_k^{\alpha}) - m_{\mathrm{e.v.}}(\tau_k^{\alpha})\right]^2}, \qquad k = 1, 2, \ldots, K. \tag{5.10}$$

Note that $J_{\mathrm{e.v.}}(k),\ k = 1, 2, \ldots, K$ must be non-zero, so that the seasonal characteristics are meaningful. Moreover, if at a specific time instance τ_k^{α} only one year j is available, then, $s_{\mathrm{e.v.}}(\tau_k^{\alpha})$ is identically zero. That is, the value of $J_{\mathrm{e.v.}}(k)$ should be different from unity.

In fact, the quantities $m_{\mathrm{e.v.}}(\tau^{\alpha})$ and $s_{\mathrm{e.v.}}(\tau^{\alpha})$ are statistical characteristics of the series $Y_{\mathrm{e.v.}}(\tau)$ rather than of the original series $Y(\tau)$. However, and under the condition that $J_{\mathrm{e.v.}}(k) \approx J$ for every k, the above mentioned quantities can be regarded as representative of the true statistical characteristics of time series $Y(\tau)$ also.

Nevertheless, even in the extreme case where $\{m_{\mathrm{e.v.}}(\tau_k^{\alpha})\} \subset \{m(\tau_k^{\alpha})\}$, i.e., the value of $m(\tau_k^{\alpha})$ is missing for some k, a low-order Fourier series can always be fitted to $m_{\mathrm{e.v.}}(\tau^{\alpha})$ (Anderson, 1971; Athanassoulis and Stefanakos, 1995). This Fourier series, denoted by $\mu(\tau^{\alpha})$, is then assumed to be representative of the true $m(\tau^{\alpha})$. In the same way, low-order Fourier series is also fitted to $s_{\mathrm{e.v.}}(\tau^{\alpha})$, denoted by $\sigma(\tau^{\alpha})$. Finally, we define $\mu(\tau)$ and $\sigma(\tau)$ to be the periodic extensions of $\mu(\tau^{\alpha})$ and $\sigma(\tau^{\alpha})$ in the long-term time τ.

The use of Fourier representations is advantageous for several reasons. First, seasonal variability is now described by a small number of parameters. Secondly, they are smoothed versions of the estimators of the seasonal characteristics, and, as a result, the sampling variability is significantly reduced.

5.3 Autocorrelation and spectral estimation

A fundamental question here is whether time series $W_{\text{e.v.}}(\tau)$ can be considered stationary. For this, one has to study the autocorrelation function as well as the spectral density of $W_{\text{e.v.}}(\tau)$. The estimation procedure for the autocorrelation is due to Parzen (1963), who modelled the existing-value time series as an "amplitude-modulated" version of the "true" stochastic process (Parzen, 1962)

$$W_{\text{e.v.}}(\tau) = u(\tau)W(\tau), \tag{5.11}$$

where $u(\tau)$ is a deterministic bounded function and $W(\tau)$ a stationary and Gaussian process. More specifically, the following theorem and proposition hod true.

Theorem 5.1 (Parzen, 1962, 1963). *Let $W(\tau;\beta)$ be a zero mean stationary, autocorrelation ergodic, Gaussian stochastic process. Let also $u(\tau)$ be a deterministic, bounded function for which the following limit*

$$\widehat{R}_{uu}(r) \equiv \lim_{T\to\infty} \widehat{R}_{uu}(r,T) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} u(\tau)u(\tau+r)d\tau \tag{5.12}$$

exists. Then, the amplitude-modulated version of $W(\tau;\beta)$

$$W_{\text{e.v.}}(\tau;\beta) = u(\tau)\, W(\tau;\beta) \tag{5.13}$$

is (i) asymptotically stationary, and (ii) ergodic. ∎

Proof of Theorem 5.1
(i) According to Parzen, a process $W_{\text{e.v.}}(\tau;\beta)$ is said to be asymptotically stationary, if its mean value and autocorrelation function are given by the following relations

$$\begin{aligned} m_{W_{\text{e.v.}}} &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} \mathsf{E}\Big[W_{\text{e.v.}}(\tau;\beta)\Big]d\tau = \\ &= \lim_{T\to\infty} \mathsf{E}\Big[\hat{m}_{W_{\text{e.v.}}}(T;\beta)\Big], \end{aligned} \tag{5.14a}$$

$$\begin{aligned} R_{W_{\text{e.v.}}W_{\text{e.v.}}}(r) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} \mathsf{E}\Big[W_{\text{e.v.}}(\tau;\beta)W_{\text{e.v.}}(\tau+r;\beta)\Big]d\tau = \\ &= \lim_{T\to\infty} \mathsf{E}\Big[\widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta)\Big], \end{aligned} \tag{5.14b}$$

where

$$\hat{m}_{W_{\text{e.v.}}}(T;\beta) = \frac{1}{T}\int_{-T/2}^{T/2} W_{\text{e.v.}}(\tau;\beta)d\tau, \tag{5.15a}$$

$$\widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta) = \frac{1}{T}\int_{-T/2}^{T/2} W_{\text{e.v.}}(\tau;\beta)W_{\text{e.v.}}(\tau+r;\beta)d\tau \tag{5.15b}$$

are the time average mean value and autocorrelation functions of $W_{\text{e.v.}}(\tau;\beta)$, respectively.

Since $u(\tau)$ is a deterministic function

$$\begin{aligned}\mathsf{E}\Big[W_{\text{e.v.}}(\tau;\beta)W_{\text{e.v.}}(\tau+r;\beta)\Big] &= u(\tau)u(\tau+r)\mathsf{E}\Big[W(\tau;\beta)W(\tau+r;\beta)\Big] = \\ &= u(\tau)u(\tau+r)R_{WW}(r).\end{aligned} \tag{5.16}$$

Substituting (5.16) to (5.14b) we obtain

$$R_{W_{\text{e.v.}}W_{\text{e.v.}}}(r) = R_{WW}(r)\left(\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} u(\tau)u(\tau+r)d\tau\right). \tag{5.17}$$

Both members of the product in the right-hand-side of (5.17) exist, and, thus, the limit in (5.14b) also exists.

(ii) The process $W_{\text{e.v.}}(\tau;\beta)$ is ergodic if and only if

$$\lim_{T\to\infty}\ \mathcal{C}_{W_{\text{e.v.}}}(T;r) = 0, \tag{5.18}$$

where

$$\begin{aligned}\mathcal{C}_{W_{\text{e.v.}}}(T;r) = \frac{1}{T}\int_{-T/2}^{T/2}\Bigg(&\mathsf{E}\Big[W_{\text{e.v.}}(\tau)W_{\text{e.v.}}(\tau+r)W_{\text{e.v.}}(\tau+s)W_{\text{e.v.}}(\tau+r+s)\Big] - \\ &- R_{W_{\text{e.v.}}W_{\text{e.v.}}}(\tau,\tau+r)R_{W_{\text{e.v.}}W_{\text{e.v.}}}(\tau+s,\tau+s+r)\Bigg)ds\ .\end{aligned} \tag{5.19}$$

Using definition (5.13) and the fact that $W(\tau;\beta)$ is Gaussian, $\mathcal{C}_{W_{\text{e.v.}}}(T;r)$ can be written as follows

$$\begin{aligned}\mathcal{C}_{W_{\text{e.v.}}}(T;r) = \frac{1}{T}\int_{-T/2}^{T/2} &u(\tau)u(\tau+r)u(\tau+s)u(\tau+r+s) \\ &\Big[R_{WW}^2(s) + R_{WW}(s+r)R_{WW}(s-r)\Big]ds.\end{aligned} \tag{5.20}$$

Let $\mathcal{G}$ be the bound of $u(\tau)$, then

$$|\mathcal{C}_{W_{\text{e.v.}}}(T;r)| \quad \leq \quad \mathcal{G}^4 \frac{1}{T} \int_{-T/2}^{T/2} \Big[R_{WW}^2(s) \; + \; |R_{WW}(s \; + \; r) R_{WW}(s \; - \; r)| \Big] ds. \quad (5.21)$$

Since $W(\tau;\beta)$ is ergodic the right-hand-side limit is zero, and, thus, relation (5.18) holds also true, and the proof of the theorem is completed. ∎

Proposition 5.1 *The function*

$$\widetilde{R}_{WW}(r,T;\beta) = \frac{\widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta)}{\widehat{R}_{uu}(r)}, \quad (5.22)$$

with $\widehat{R}_{uu}(r) \neq 0$, is a consistent, in the mean square sense, estimator of the autocorellation function $R_{WW}(r)$. ∎

Proof of Proposition 5.1
Since $W_{\text{e.v.}}(\tau;\beta)$ is ergodic,

$$\underset{T\to\infty}{\text{l.i.m.}} \, \widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta) = \lim_{T\to\infty} \mathsf{E}\Big[\widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta)\Big]. \quad (5.23)$$

From (5.14b),

$$\lim_{T\to\infty} \mathsf{E}\Big[\widehat{R}_{W_{\text{e.v.}}W_{\text{e.v.}}}(r,T;\beta)\Big] = R_{W_{\text{e.v.}}W_{\text{e.v.}}}(r). \quad (5.24)$$

Combining (5.17), (5.23) and (5.24), we obtain the proof of the proposition. ∎

Parzen (1963) considered the following deterministic function $u(\tau)$

$$u(\tau) = \begin{cases} 0, & \text{if } W(\tau) \text{ is missing at time } \tau, \\ 1, & \text{if } W(\tau) \text{ is existent at time } \tau, \end{cases} \quad (5.25)$$

and examined in detail the case of periodically missing values of Gaussian time series. However, his method for the calculation of the autocorrelation function and the corresponding spectral density can be applied under some very general conditions (Dunsmuir and Robinson, 1981a; Giannakis and Zhou, 1994) to any pattern of missing observations without any particular request for the distribution of the underlying process.

Although $\widetilde{R}_{WW}(r,T;\beta)$ is a consistent estimator of $R_{WW}(r)$, its Fourier transform is not (Parzen, 1961, p. 174),(Wei, 1990, p. 267). In order to obtain a consistent estimate of the spectral density an appropriate lag window or spectral window (or a combination of both) should be chosen (Parzen, 1957, Theorem 4). For the numerical results of Section 5.6, this procedure is implemented in two steps. First, a rectangular (truncated) lag window (Wei, 1990) is used, that is,

$$\widetilde{S}_{WW}(f) = \int_{-T/2}^{T/2} \mathsf{k}^{\mathrm{L}}(r,T_M)\widetilde{R}_{WW}(r,T;\beta)\exp[-j2\pi fr]dr, \tag{5.26}$$

where

$$\mathsf{k}^{\mathrm{L}}(r,T_M) = \begin{cases} 1, & |r| \le T_M, \\ 0, & |r| > T_M, \end{cases} \tag{5.27}$$

with $T_M \le T$ being the truncation point. Second, the Daniell spectral window (Daniell, 1946; Wei, 1990) is applied, and the resulted spectrum is written

$$\widehat{S}_{WW}(f_s) = \sum_{h=s-m}^{s+m} \mathsf{K}^{\mathrm{S}}(f_h)\widetilde{S}_{WW}(f_s - f_h), \tag{5.28}$$

where $\mathsf{K}^{\mathrm{S}}(f)$=1/(2$m$+1). Attention should be paid on (i) the truncation point T_M so that $\widetilde{R}_{WW}(r,T;\beta)$ will have vanished until then (Priestley, 1981, p. 539), and (ii) on smoothing parameter m, so that the form of the spectrum will not be distorted (Priestley, 1981, p. 541). Further study on the sensitivity of $\widehat{S}_{WW}(f)$ due to the smoothing parameter m, will be presented in Section 5.6. Moreover, the smoothed spectrum is χ^2-distributed with degrees of freedom the product of the degrees of freedom resulted from each window separately, i.e. $2\,[T/T_M]\,(2m+1)$, where $[\cdot]$ stands for the integer part of a number.

5.4 ARMA modelling of the stationary part

Considering that time series $W_{\text{e.v.}}(\tau)$ is stationary, and assuming further that is also Gaussian, it can then be represented by a linear ARMA model of order (P,Q) (ARMA(P,Q)) of the form (Box and Jenkins, 1976; Priestley, 1981; Brockwell and Davis, 1991)

$$W_{\mathsf{ARMA}}(\tau_\ell) = \sum_{p=1}^{P} a_p W_{\mathsf{ARMA}}(\tau_{\ell-p}) + \varepsilon(\tau_\ell) + \sum_{q=1}^{Q} b_q \varepsilon(\tau_{\ell-q}), \qquad \ell = 1,2,\dots, \tag{5.29}$$

where $\varepsilon(\tau)$ is a time series of uncorrelated Gaussian random variables with zero mean and standard deviation σ_ε.

If $W_{\text{e.v.}}(\tau)$ is found to be non-Gaussian, it should, then, be transformed to a Gaussian one before a linear ARMA model is fitted. One frequently used method is the logarithmic transformation, especially useful when the original time series follows a lognormal distribution (Salas, 1993). There are various techniques to detect Gaussianity; see, e.g., Snedecor and Cochran (1980), Hinich (1982). However, usually the plot of the empirical propability distribution function on a normal probability paper suffices (Wilk and Gnanadesikan, 1968; Gnanadesikan, 1977).

In the simple case, where no missing values exist, ARMA coefficients are estimated either in the time domain or in the frequency domain (Box and Jenkins, 1976; Priestley, 1981; Brockwell and Davis, 1991). In the missing-value case, special techniques have been developed to form modified versions of the usual methods of Yule-Walker and Maximum Likelihood equations (Dunsmuir and Robinson, 1981a; Dunsmuir and Robinson, 1981b; Dunsmuir and Robinson, 1981c; Jones, 1980) or other specialized techniques (Porat and Friedlander, 1984; Rosen and Porat, 1989b; Giannakis and Zhou, 1994). In this work, a simpler frequency domain procedure is implemented. For this, the consistent estimation of the autocorrelation function, as well as the estimation of the corresponding spectral density presented in the previous section is needed.

Then, the spectrum of a low-order ARMA model is fitted to the raw spectrum of $W_{\text{e.v.}}(\tau)$. The goodness-of-fit can be checked using a least-square criterion of the form (Spanos, 1983)

$$\int \left| S_{\text{ARMA}}(f) - \widehat{S}_{WW}(f) \right|^2 = \min, \tag{5.30}$$

or

$$\int \left| \frac{S_{\text{ARMA}}(f) - \widehat{S}_{WW}(f)}{\widehat{S}_{WW}(f)} \right|^2 = \min, \tag{5.31}$$

where the spectrum of the ARMA(P,Q) model $S_{\text{ARMA}}(f)$ is defined by (Jenkins and Watts, 1968)

$$S_{\text{ARMA}}(f_n) = 2\Delta\tau\ \sigma_\varepsilon^2 \frac{\left| 1 + \sum_{q=1}^{Q} b_q \exp(-j2\pi f_n \Delta\tau\ q) \right|^2}{\left| 1 - \sum_{p=1}^{P} a_p \exp(-j2\pi f_n \Delta\tau\ p) \right|^2}, \tag{5.32}$$

with $f_n = \frac{n}{N\Delta\tau},\ n = 0, 1, \ldots, \frac{N}{2},\ N = r_{\max}$, and $\Delta\tau$ be the sampling period.

Equations (5.30) and (5.31) are nonlinear on the parameters $\{a_p,\ p = 1, \ldots, P\}$, $\{b_q,\ q = 1, \ldots, Q\}$, and σ_ε^2. To solve them, the Levenberg-Marquardt method for solving nonlinear

least-square equations is used (Levenberg, 1944; Marquardt, 1963). This method needs an initial estimate to start, which can be obtained according to the following procedure. First, consider the complete time series $W^c_{\rm e.v.}(\tau)$ defined as

$$W^c_{\rm e.v.}(\tau) = \{W_{\rm e.v.}(\tau_i), \quad i = 1, 2, \ldots, I_{\rm e.v.}\}, \tag{5.33}$$

which has been produced from $W_{\rm e.v.}(\tau)$ by replacing the set of indices $\boldsymbol{I}_{\rm e.v.}$ with the set of successive indices $\{1, 2, \ldots, I_{\rm e.v.}\}$. Now, we are able to fit a linear ARMA model to $W^c_{\rm e.v.}(\tau)$ using any of the usual methods; see, e.g., Brockwell and Davis (1991). This estimate can be used as an initial guess for the solution of (5.30) or (5.31).

Further, if time series $W_{\rm e.v.}(\tau)$ is very long and the percentage of missing values is moderate, the complete version $W^c_{\rm e.v.}(\tau)$ can then be also used to compute first estimates for the autocorrelation function and the spectral density. Otherwise, its behaviour does not reflect the behaviour of the original one, and, thus, we need a different approach (see Section 5.3).

5.5 Methodology for missing-value completion

In this section, we present a simple yet powerful procedure for the reconstruction of an incomplete nonstationary, Gaussian time series $X_{\rm e.v.}(\tau)$.

1. Analyse time series $X_{\rm e.v.}(\tau)$ and calculate deterministic components $\overline{X}_{\rm tr}(\tau)$, $\mu(\tau)$ and $\sigma(\tau)$ (see Section 5.2, and Stefanakos and Athanassoulis (1997)).

2. Using equ. (5.1), form the stationary series $W_{\rm e.v.}(\tau)$.

3. Calculate the spectrum $S_{WW}(f)$ of the unobserved (complete) series $W(\tau)$ by using a consistent estimator of the corresponding autocovariance function (Parzen, 1963; Stefanakos and Athanassoulis, 1997). See also Section 5.3 of the present work.

4. Model stationary series $W_{\rm e.v.}(\tau)$ as a low-order linear ARMA(P,Q) process. Estimate the ARMA parameters and the variance of the residuals σ^2_ε by least-square fitting of the ARMA spectrum to the raw spectrum $\widehat{S}_{WW}(f)$ (Spanos, 1983; Stefanakos and Athanassoulis, 1997). See also Section 5.4 of the present work.

5. Generate by numerical simulation a sequence $\varepsilon_{\rm sim}(\tau)$= $\{\varepsilon_{\rm sim}(\tau_i), \quad i = 1, 2, \ldots, I_{\rm m.v.}\}$, where $\varepsilon_{\rm sim}(\tau_i)$ are uncorrelated Gaussian random variables with zero mean and variance σ^2_ε.

6. Form the following (complete) series

$$W^c(\tau) = \Big\{ W^c(\tau_i), \quad i = 1, \ldots, I \Big\}, \tag{5.34}$$

as follows

$$W^c(\tau_i) = \begin{cases} W_{\text{e.v.}}(\tau_i), & \text{if } i \in \boldsymbol{I}_{\text{e.v.}}, \\ W_{\text{sim}}(\tau_i), & \text{if } i \in \boldsymbol{I}_{\text{m.v.}}. \end{cases} \tag{5.35}$$

The quantities $W_{\text{sim}}(\tau_i)$ are calculated by means of the recurrence relation

$$W_{\text{sim}}(\tau_i) = \varepsilon_{\text{sim}}(\tau_i) + \left[\sum_{p=1}^{P} a_p W^c(\tau_{i-p}) + \sum_{q=1}^{Q} b_q \varepsilon^c(\tau_{i-q}) \right], \tag{5.36}$$

where the $\varepsilon^c(\tau_i)$'s are defined as

$$\varepsilon^c(\tau_i) = \begin{cases} \varepsilon_{\text{e.v.}}(\tau_i), & \text{if } i \in \boldsymbol{I}_{\text{e.v.}}, \\ \varepsilon_{\text{sim}}(\tau_i), & \text{if } i \in \boldsymbol{I}_{\text{m.v.}}. \end{cases} \tag{5.37}$$

The simulated values $\varepsilon_{\text{sim}}(\tau_i)$ have been produced in the previous step, whereas $\varepsilon_{\text{e.v.}}(\tau_i)$ are calculated by means of

$$\varepsilon_{\text{e.v.}}(\tau_i) = W_{\text{e.v.}}(\tau_i) - \left[\sum_{p=1}^{P} a_p W^c(\tau_{i-p}) + \sum_{q=1}^{Q} b_q \varepsilon^c(\tau_{i-q}) \right]. \tag{5.38}$$

In this way, the sequence $\varepsilon^c(\tau) = \{\varepsilon^c(\tau_i), \quad i = 1, 2, \ldots, I\}$, has been produced. Note that, the starting values of the series $W^c(\tau)$, i.e. for i=1,2, ... ,$\max(P, Q)$, are obtained using equ. (5.35) and equations similar to (5.36), where the sums are performed up to a value lesser than P and Q, respectively. See also Box and Jenkins (1976) and Priestley (1981).

7. Form the (complete) time series $X^c(\tau)$, using $\overline{X}_{\text{tr}}(\tau)$, $\mu(\tau)$, $\sigma(\tau)$ and $W^c(\tau)$ [cf. equ. (5.1)]

$$X^c(\tau) = \overline{X}_{\text{tr}}(\tau) + \mu(\tau) + \sigma(\tau)\ W^c(\tau). \tag{5.39}$$

5.6 Completion of missing values in numerically simulated time series

In order to validate the procedure presented in the previous section, we apply it to time series produced by numerical simulation. For this purpose, a sequence of uncorrelated Gaussian random variables $\{\varepsilon(\tau_i), \ i = 1, 2, \ldots, I\}$ is generated with zero mean and variance σ_ε^2=0.108.

To accomplish this task, the routine NORMRND of the MATLAB® language is used, which is based on the algorithm presented in Forsyth et al. (1977). Then, letting P=Q=2, a linear ARMA sequence is produced by means of equ. (5.29) with parameters a_1=1.7093, a_2=−0.726, b_1=−0.6277, b_2=−0.0321, and starting values calculated as described in item 6 of Section 5.5. The roots of the polynomials

$$1-\sum_{p=1}^{P} a_p \lambda^p = 0, \qquad \text{and} \qquad 1+\sum_{q=1}^{Q} b_q \lambda^q = 0, \tag{5.40}$$

are outside the unit circle, and, thus, the series $W(\tau)\equiv W_{\text{ARMA}}(\tau)$ is stationary and invertible (Priestley, 1981; Jones, 1985). The values of σ_ε^2, a_1, a_2, b_1, b_2 are typical for sea-wave applications; and they have been taken by fitting ARMA models to measured time series of wave data.

Further, the time series (5.1) is formed by using deterministic functions $\overline{X}_{\text{tr}}(\tau)$, $\mu(\tau)$ and $\sigma(\tau)$ also typical for sea-wave applications. The function $\overline{X}_{\text{tr}}(\tau)$ is defined by equ. (5.7) and is linear, whereas the functions $\mu(\tau)$ and $\sigma(\tau)$ are periodic with period one year, and their parameters are shown in Table 5.1 and 5.2, respectively. In this way, $X(\tau)$ is given the structure of a nonstationary (cyclostationary) stochastic process and the only thing that remains to do is to make it incomplete. Before this, let us first give the definition of the α% missing-value pattern.

Definition 5.1 *A missing-value pattern associated with an indicator function $u(\tau)$, defined in (5.25), with mean value*

$$\mu_u = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} u(\tau)\, d\tau, \tag{5.41}$$

will be called α% missing-value pattern, *where α=100×(1−μ_u).* ■

Now, two different incomplete time series are produced from the complete time series $X(\tau)$, using two different α% missing-value patterns, α_1% and α_2%, with $\alpha_1<\alpha_2$. In particular, two six-year long measured time series of significant wave height are processed. The first one is from Figueira da Foz, a site off Portuguese coasts (40.19^{o}N, 9.15^{o}W), while the second one is from Palamós, a site off Spanish coasts in the Mediterranean Sea (41.82^{o}N, 3.18^{o}E). The first one has a β_1% missing-value pattern, with β_1=16.5 and an associated indicator function $u^{\beta_1}(\tau)$. The second one has a β_2% missing-value pattern, with β_2=22.3 and an associated indicator function $u^{\beta_2}(\tau)$. Then, the α_1% missing-value pattern is chosen to have an associated indicator function $u^{\alpha_1}(\tau)$=$u^{\beta_1}(\tau)$, and the α_2% missing-value pattern to have an associated indicator function $u^{\alpha_2}(\tau)$=$u^{\beta_1}(\tau)u^{\beta_2}(\tau)$. In this way, the two α% missing-value patterns have been produced with α_1=16.5 and α_2=33.

Using the indicator function $u(\tau)$, one can obtain from the complete time series $X(\tau)$ the incomplete time series $X_{\text{e.v.}}(\tau)$ as follows

$$X_{\text{e.v.}}(\tau_i) = \begin{cases} missvalcode, & \text{if } u(\tau_i)=0, \\ X(\tau_i), & \text{if } u(\tau_i)=1, \end{cases} \tag{5.42}$$

where "$missvalcode$" is usually put a very large negative number.

Having generated a nonstationary time series with missing values $X_{\text{e.v.}}(\tau)$, we are in a position to perform the analysis of it as this is described in Sections 5.1– 5.4. Then, the time series is reconstructed, according to the procedure presented in Section 5.5, and, thus, a new (complete) time series $X^c(\tau)$ is produced. In the two subsequent sections, the analysis results of the following three time series will be examined and compared for α_1% and α_2% missing-value patterns: (i) the initial $X(\tau)$, (ii) the incomplete $X_{\text{e.v.}}(\tau)$, and (iii) the reconstructed $X^c(\tau)$.

Before proceeding, let us first fix some parameters concerning the spectral estimation procedure described in Section 5.3. For this, the spectrum of $W(\tau)$, denoted by $S_{\text{ARMA}}(f)$, is first computed using the analytic form (5.32) with the initial ARMA parameters. Then, the spectrum of $W(\tau)$, denoted by $\widehat{S}_{WW}(f)$, is calculated using estimation procedure described in equs. (5.26)–(5.28) with truncation point T_M=1024$\Delta\tau$. In Figure 5.1, the estimated spectrum $\widehat{S}_{WW}(f)$ is plotted for three different values of smoothing parameter m (m=1,5,9; see equ. (5.28)) along with the analytic one $S_{\text{ARMA}}(f)$. To sum up, we have seen that it suffices to calculate autocorrelation $\widetilde{R}_{WW}(r)$ up to lag r=1024, and to use a small smoothing parameter; say m=5.

5.6.1 Results concerning a 16.5% missing-value pattern

Let us now turn to the analysis of the three time series $X(\tau)$, $X_{\text{e.v.}}(\tau)$ and $X^c(\tau)$. First, the indicator function $u(\tau)$ associated with the 16.5% missing-value pattern is examined. In Figure 5.2, $u(\tau)$ is shown along with its autocovariance $C_{uu}(r)$=$R_{uu}(r)-\mu_u^2$ and its spectral density $S_{uu}(f)$, where $R_{uu}(r)$ and μ_u are defined in (5.12) and (5.41), respectively, and $S_{uu}(f)$ is calculated as the Fourier transform of $C_{uu}(r)$. By re-indexing function $u(\tau)$, the distribution of existing values within the anual time can also be examined. This is done using the Buys-Ballot double index; see also equs. (5.3) and (5.4), by which the following "existing-value count" is obtained

$$\widetilde{u}(\tau_k^\alpha) = \sum_{j=1}^{J} u(j, \tau_k^\alpha), \qquad k = 1, \ldots, K. \tag{5.43}$$

The values of this function in each time location within the annual time depicts the amount of existing values available throughout the whole time series. For example, for the 3rd of January

at 3a.m. only 4 values are avalaible instead of the 6, which is the upper-limit for a six-year long time series. Thus, plotting $\widetilde{u}(\tau^{\alpha})$ against annual time τ^{α} an "existing-value diagram" is produced, which shows the distribution of existing values along the annual time; see Figure 5.3a. In Figure 5.3b, the seasonal mean value $\mu(\tau)$ of $X(\tau)$ (dotted line) and $X_{\text{e.v.}}(\tau)$ (solid line) is shown for one yearly period. Accordingly, in Figure 5.3c, the seasonal standard deviation $\sigma(\tau)$ of $X(\tau)$ (dotted line) and $X_{\text{e.v.}}(\tau)$ (solid line) is shown. In Figures 5.4a–c, the same quantities as in Figures 5.3a–c, are depicted, but now the dotted-line results have been produced by means of the reconstructed time series $X^c(\tau)$ instead of $X(\tau)$. The parameters of functions $\mu(\tau)$ and $\sigma(\tau)$ are depicted in Tables 5.1 and 5.2, respectively, for the time series $X_{\text{e.v.}}(\tau)$ and $X^c(\tau)$, and can be compared with the corresponding ones of $X(\tau)$. Especially, the parameters for $X^c(\tau)$ have been estimated twice since $X^c(\tau)$ has been produced by means of both citeria (5.30) and (5.31). Results are in very good agreement and small discrepancies are due to the estimation in the presence of missing values. Moreover, the mean value and the standard deviation of $X(\tau)$, $X_{\text{e.v.}}(\tau)$ and $X^c(\tau)$, shown in Table 5.3, are also in a very good agreement.

In the sequel, $X_{\text{e.v.}}(\tau)$ is decomposed following (5.1) and the correlation structure of stationary part $W_{\text{e.v.}}(\tau)$ is examined. For this, the consistent estimators of the autocorrelation coefficient function (acf) $\rho_{WW}(r)$=$\widetilde{R}_{WW}(r)/\widetilde{R}_{WW}(0)$ and the spectral density $\widehat{S}_{WW}(f)$ have been calculated according to Section 5.3. In Figure 5.5, the acf of $W_{\text{e.v.}}(\tau)$ is plotted along with the 95% confidence limits, and compared with the corresponding acf's of $W(\tau)$ and $W^c(\tau)$. In Figure 5.6a, spectral densities $\widehat{S}_{WW}(f)$ and $\widehat{S}_{W^cW^c}(f)$ are depicted along with the 95% confidence limits of $\widehat{S}_{WW}(f)$, and they are compared to the analytic density $S_{\mathsf{ARMA}}(f)$. The same spectral densities are plotted in logarithmic scale in Figure 5.6b in order to study the behaviour of the high frequency part. It appears that the three spectral densities are also in good agreement between themshelves. Following the procedure of Section 5.5, Step 4, the parameters of ARMA model are estimated using both criteria (5.30) and (5.31), the parameters shown in Table 5.4 along with the initial ones.

Further, the time series of residuals $\varepsilon(\tau)$ is examined. First, the empirical probability density function (pdf) of $\varepsilon(\tau)$, $\varepsilon_{\text{e.v.}}(\tau)$ and $\varepsilon^c(\tau)$ are shown in Figure 5.8a. Moreover, these empirical pdf's are depicted two by two in Figures 5.8b–d and compared with an analytic Gaussian pdf with zero mean and standard deviation σ_{ε}. In these figures, the empirical pdf's exhibit a very good goodness-of-fit with the corresponding theoretical pdf. That is, our residuals are with very good approximation Gaussian. The correlation structure of residuals is also studied; see Figures 5.8 and 5.9, where the acf and the spectral density are depicted, respectively. Especially, acf becomes zero after the first few lags, i.e. residuals of the three time series are uncorrelated. Moreover, spectral density is almost horizontal, as the spectral density of a white noise time series should be.

5.6.2 Results concerning a 33% missing-value pattern

Following the same analysis procedure as in the previous section, results are obtained for the 33% mising-value pattern. Although the amount of missing values is twice as big than the amount of the 16.5% missing-value pattern, the obtained results are satisfactory enough and very close to each other. Thus, there is no need to present results from of the whole analysis procedure. Instead, some indicative figures are shown.

First, the associated indicator function $u(\tau)$ along with its autocovariance $C_{uu}(r)$ and its spectral density $S_{uu}(f)$ are depicted in Figure 5.10. The values of $u(\tau)$ exhibit a more persistent correlation in comparison with the corresponding ones of the 16.5% missing-value pattern.

Further, the seasonal analysis results presented in Figure 5.11 present very good accordance between these of the initial time series and these of the incomplete one. They are also compared with the results presented in Figure 5.3 and found to be in agreement.

The estimated spectral density from the incomplete time series $\widehat{S}_{WW}(f)$ seems to overestimate a bit the region around the peak value of the analytic one $S_{\mathsf{ARMA}}(f)$. However, the ARMA estimation procedure does not fail, and the obtained ARMA parameters lead to an acceptable reconstruction of $W(\tau)$. This can be justified by considering the estimated density $\widehat{S}_{W^cW^c}(f)$ also depicted in Figure 5.12. See also Table 5.5, where the initial as well as the estimated, by means of both criteria, ARMA coefficients are given.

Finally, in Figure 5.13, spectral density $\widehat{S}_{WW}(f)$ for both missing-value patterns (16.5% and 33%) are compared with the analytic one $S_{\mathsf{ARMA}}(f)$. Considering the fact that $\widehat{S}_{WW}(f)$ is estimated when a significant amount of data is missing, the results are very satisfactory.

5.7 Completion of missing values in wave-parameter time series

The presented procedure for the completion of nonstationary time series with missing values will now be applied to measured time series of wave data. The methodology has already been applied to synthetic data in the previous section (see also Stefanakos and Athanassoulis (1997)), and the results were very satisfactory. The data sets that will be used are: (i) one from the Portuguese site Figueira da Foz (40.19^{o}N, 9.15^{o}W, off Portuguese coasts), and (ii) one from the Spanish site Palamós (41.82^{o}N, 3.18^{o}E, off Mediterranean Spanish coasts). See also Figure 5.14 in the end of the present chapter.

5.7.1 Figueira da Foz

The data set used is a six-year long (1983, 1985–1989) time series of significant wave height H_S. The in-between time series of 1984 was excluded, because 56% of its values is missing. In the remaining data set, the 16.5% of the values is missing.

First, a time series analysis is performed to the series $X_{\text{e.v.}}(\tau)$=ln$[H_S(\tau) + 1\text{m}]$, following the guidelines of the Section 5.5. In Figure 5.15a, the existing-value diagram shows the distribution of existing values along the annual time. For a description of the existing-value diagram, see subsection 5.6.1, where this diagram is explained by means of an example. In Figure 5.15b and 5.15c, the seasonal mean value and the seasonal standard deviation, respectively, are shown (solid lines). After the seasonal standardization of the detrended time series, the stationary series $W_{\text{e.v.}}(\tau)$ is obtained, which will be used for the estimation of the spectral density $S_{WW}(f)$ of the unobserved series $W(\tau)$.

According to Section 5.4, the spectrum of an ARMA(2,2) is fitted to the spectral density $S_{WW}(f)$. The estimated ARMA coefficients are a_1=1.8128, a_2=−0.8193, b_1=−0.2288, b_2=−0.6165, and the variance of the residual series is σ_ε^2=0.0841. See also Figures 5.16a,b, where the estimated raw spectral density $S_{WW}(f)$ is depicted along with the analytic one $S_{\text{ARMA}}(f)$ using the estimated parameters.

In the sequel, the time series of residuals is formed, which can be considered Gaussian. See also Figure 5.17, where the empirical probability density function of the residuals is plotted (dashed line), along with the corresponding analytic Gaussian one (solid line). Time series of residuals is filled in by means of a simulated sequence of uncorrelated Gaussian random variables with zero mean and variance σ_ε^2. However, the residuals exhibit a small (of some lags) remaining correlation. An AR(2) model is now fitted to the residuals, and the obtained parameter values are c_1=−0.4386 and c_2=0.1568. The new residuals are uncorrelated, see Figure 5.18, where the autocorrelation coefficient function is depicted.

Using the completed time series of the residuals and the estimated AR and ARMA coefficients, we proceed with the construction of the stationary series $W^c(\tau)$. Furthermore, using also the deterministic components, we finally construct the series $X^c(\tau)$. Comparisons showed that reconstructed series has the same statistical characteristics and dependence structure as the initial incomplete series does. Compare, e.g., solid with dotted lines in Figures 5.15b,c, solid with dashed lines in Figures 5.16a,b, or dashed with dotted lines in Figure 5.17. Further, in Figure 5.19, the series $X^c(\tau)$ along with $X_{\text{e.v.}}(\tau)$ are depicted for a time window with duration of two and a half months (600 three-hourly intervals=75 days). One can observe that the two patterns are in very good agreement.

5.7.2 Palamós

The same procedure is now applied to a measured time series with a larger amount of missing values, namely 22.3%. The data set used is again a six-year long (1990–1995) time series of significant wave height H_S.

In the same way as in the previous section, first an analysis is performed to the series $X(\tau)$=ln[$H_S(\tau)$ + 2m]. In Figure 5.20a, the existing-value diagram shows the distribution of existing values along the annual time. In Figures 5.20b and 5.20c the seasonal mean value and seasonal standard deviation, respectively, are shown (solid lines). After the seasonal standardization of the detrended time series, time series $W_{\text{e.v.}}(\tau)$ is obtained.

Again an ARMA(2,2) model is fitted, and the estimated ARMA coefficients are a_1=1.8677, a_2=−0.8694, b_1=−0.6976, b_2=−0.2774, and the variance of residuals is σ_ε^2=0.1987. In Figure 5.21, the spectral density $S_{WW}(f)$ is plotted along with the analytic $S_{\text{ARMA}}(f)$ using the estimated ARMA parameters and the exhibiting fit is very good.

The residual series is again considered Gaussian, and they are filled in by means of a simulated sequence of uncorrelated Gaussian random variables with zero mean and variance σ_ε^2. See also Figure 5.22, where the empirical probability density function of the residuals is plotted (dashed line), along with the corresponding analytic Gaussian one (solid line). By contrast with the residuals of Figueira da Foz, the residuals of Palamós do not exhibit any remaining correlation as one can observe in Figure 5.23, where the autocorrelation coefficient function of the residuals is depicted.

Using the completed time series of the residuals and the estimated ARMA coefficients, the stationary series $W^c(\tau)$ is formed. Using furthermore the deterministic components, the series $X^c(\tau)$ is constructed. Again the two time series (incomplete and reconstructed) are shown in Figure 5.24 for a time window of duration two and a half months. Other comparisons concerning the various characteristics of the two series can be made by inspecting solid and dotted lines in Figures 5.20b,c, solid with dashed in Figure 5.21a,b, solid with dashed in Figure 5.23. All are in very good agreement.

5.8 Conclusions

In the present chapter, a new methodology for the missing-value completion of an incomplete nonstationary time series of a certain structure is presented and applied to simulated and to measured wave data. The method is based on the modelling of long-term time series of wave data as a nonstationary stochastic process with yearly-long periodic mean value and standard deviation (periodically correlated stochastic process), introduced in the previous chapter (see

also Athanassoulis and Stefanakos (1995; 1997)). The key idea is to perform the completion after having removed not only the deterministic components of the series, but also the correlation structure of the stationary series. After a detrending and seasonal standardization, a low-order ARMA model is fitted to the (incomplete) residual stationary series using appropriate estimation techniques. The raw spectrum, calculated as the Fourier transform of a consistent estimate of the corresponding autocovariance function [Step 3 of the missing-value procedure; see Section 5.5], is used for the estimation of the ARMA coefficients and the variance of the residuals. The incomplete time series of uncorrelated residuals is then completed by means of simulated data with the same first-order probability structure, and used, along with the ARMA model and the estimated deterministic components, to construct a new time series of the same structure without missing values.

The above procedure has been applied to two simulated time series with different percentage of missing values (16.5% and 33%, respectively), and to two measured time series of wave parameters with different percentage of missing values (16.5% and 22.3%, respectively). Comparisons of various statistical characteristics of the initial (incomplete) and reconstructed (completed) time series are satisfactory in both cases. Some general conclusions, concerning these results, are the following:

1. The estimates of the seasonal characteristics calculated from the incomplete and the reconstructed time series are in a very good agreement with the initial one. Even in the case of the 33% missing-value pattern (see Figures 5.3, 5.4, 5.11). Also, estimates from the incomplete and the reconstructed measured time series are in a very good agreement (see Figures 5.15, 5.20).

2. Spectral density estimate $\widehat{S}_{WW}(f)$ is more sensitive to an increase of missing values. However, it manages to describe satisfactorily the analytic density $S_{\mathsf{ARMA}}(f)$ (see Figures 5.13, 5.16,5.21).

3. The residuals of the measured time series are found to be in both cases (after appropriate transformation) Gaussian (or nearly ones) (see Figures 5.17, 5.22). This fact advocates for the ARMA modelling of the stationary residual part $W_{\mathrm{e.v.}}(\tau)$.

4. If $W_{\mathrm{e.v.}}(\tau)$ is found to be non-Gaussian, it should be then (see also Section 5.4) transformed to a Gaussian one before a linear ARMA model is fitted. The most popular method (and the one used in the present work) is the logarithmic transformation (Salas, 1993), especially useful when the original time series follows a lognormal distribution. Salas (Salas et al., 1980; Salas, 1993) give an excellent survey of the various possible transformations of a hydrologic time series, which is very similar to a wave data time series, in order to obtain a stationary (nearly) Gaussian time series before fitting an ARMA model. It remains to each user to decide which trasformation to chose.

The reconstructed time series can substitute for the initial one in a number of applications where the existense of missing values causes serious problems. Examples are the use of wave data time series as input to coastal morphodynamics models, the derivation of long-term simulation of short-term (wave elevation) time series, or the direct calculation of sea-state duration statistics. There, the probability density function of sea-state duration should be produced, derived both from univariate and bivariate time series of spectral parameters, as well as the probability density function of storms (level excursion) duration.

5.9 Tables and Figures

Table 5.1: Parameters of seasonal mean value $\mu(\tau)$ for time series $X(\tau)$, $X_{\text{e.v.}}(\tau)$ and $X^c(\tau)$ [16.5% missing-value pattern]

	$X(\tau)$	$X_{\text{e.v.}}(\tau)$	$X^c(\tau)$	
			Crit. (30)	Crit. (31)
A_0	0.01118	0.01316	-0.00022	-0.00027
A_{c1}	0.20784	0.21815	0.21807	0.21802
A_{s1}	0.09010	0.10632	0.10039	0.10048

Table 5.2: Parameters of seasonal standard deviation $\sigma(\tau)$ for time series $X(\tau)$, $X_{\text{e.v.}}(\tau)$ and $X^c(\tau)$ [16.5% missing-value pattern]

	$X(\tau)$	$X_{\text{e.v.}}(\tau)$	$X^c(\tau)$	
			Crit. (30)	Crit. (31)
A_0	0.24339	0.25473	0.26466	0.26423
A_{c1}	0.04130	0.04611	0.05349	0.05319
A_{s1}	-0.00104	0.00305	-0.00103	-0.00108
A_{c2}	0.00447	0.00039	0.00618	0.00597
A_{s2}	-0.00507	-0.00361	-0.00415	-0.00427
A_{c3}	-0.01579	-0.01866	-0.02339	-0.02324
A_{s3}	0.00404	0.00495	-0.00474	-0.00435

Table 5.3: Mean value and standard deviation of $X(\tau)$, $X_{\mathrm{e.v.}}(\tau)$ and $X^c(\tau)$ [16.5% missing-value pattern]

	$X(\tau)$	$X_{\mathrm{e.v.}}(\tau)$	$X^c(\tau)$	
			Crit. (30)	Crit. (31)
Mean Value	1.0538	1.0512	1.0510	1.0509
Standard Deviation	0.3465	0.3492	0.3512	0.3507

Table 5.4: ARMA parameters of $W(\tau)$ and $W_{\mathrm{e.v.}}(\tau)$ [16.5% missing-value pattern]

	$W(\tau)$	$W_{\mathrm{e.v.}}(\tau)$	
		Crit. (30)	Crit. (31)
a_1	1.7093	1.6178	1.6776
a_2	-0.7260	-0.6423	-0.6970
b_1	-0.6277	-0.7287	-0.5911
b_2	-0.0321	0.2036	0.0103
σ_ε^2	0.1080	0.1067	0.0863

Table 5.5: ARMA parameters of $W(\tau)$ and $W_{\mathrm{e.v.}}(\tau)$ [33% missing-value pattern]

	$W(\tau)$	$W_{\mathrm{e.v.}}(\tau)$	
		Crit. (30)	Crit. (31)
a_1	1.7093	1.4994	1.5077
a_2	-0.7260	-0.5298	-0.5349
b_1	-0.6277	-0.6146	-0.5212
b_2	-0.0321	0.2165	0.0715
σ_ε^2	0.1080	0.1278	0.1341

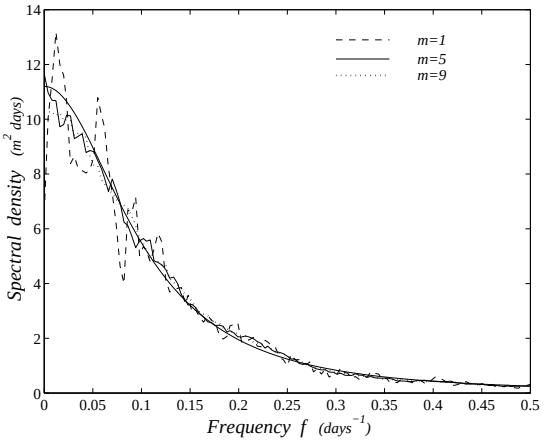


Figure 5.1: Investigation of goodness-of-fit of $\widehat{S}_{WW}(f)$ to $S_{\mathsf{ARMA}}(f)$ using various values of smoothing parameter m.

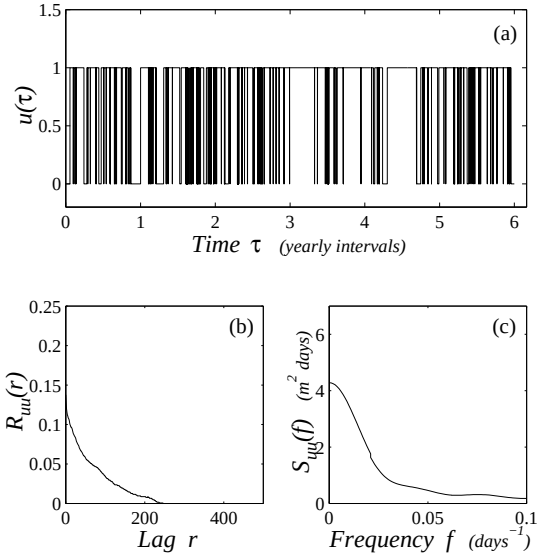


Figure 5.2: (a) Indicator function $u(\tau)$, (b) autocovariance function $C_{uu}(r)$, (c) spectral density $S_{uu}(f)$. [16.5% missing-value pattern].

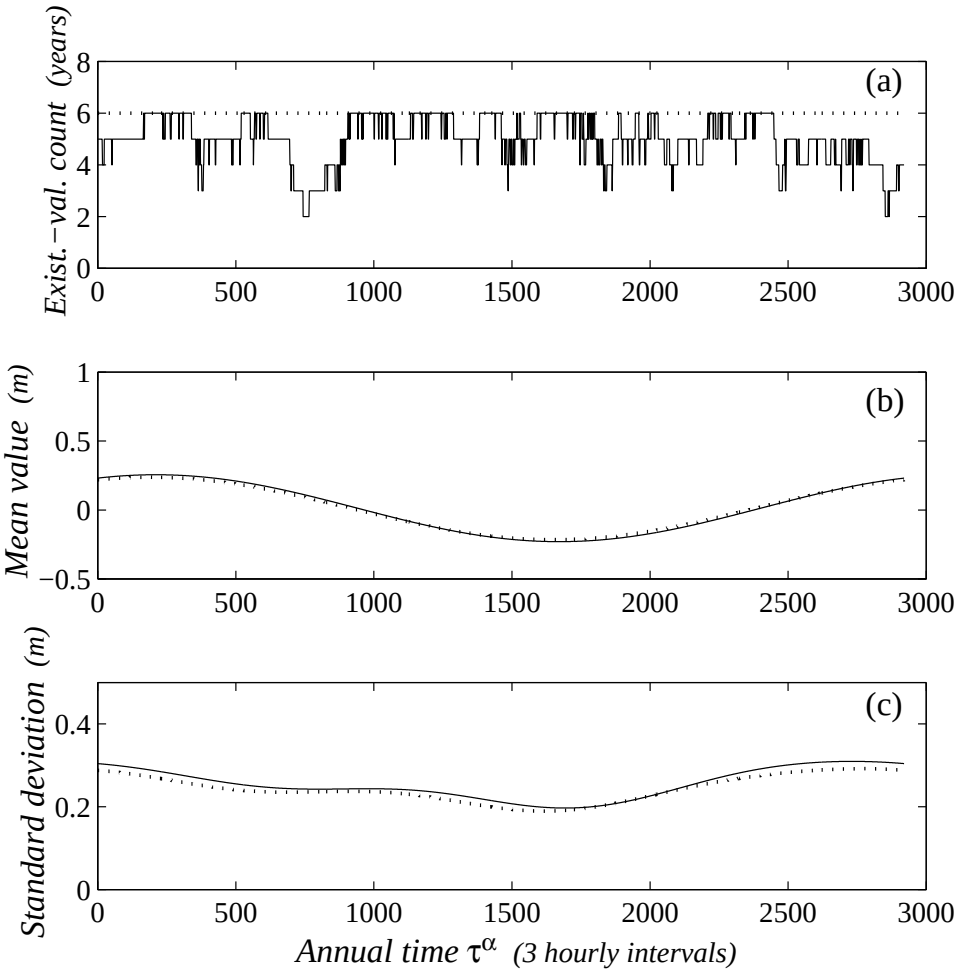


Figure 5.3: (a) Existing-value diagram, (b) seasonal mean value, (c) seasonal standard deviation. (Dotted line: initial time series, solid line: incomplete time series). [16.5% missing-value pattern].

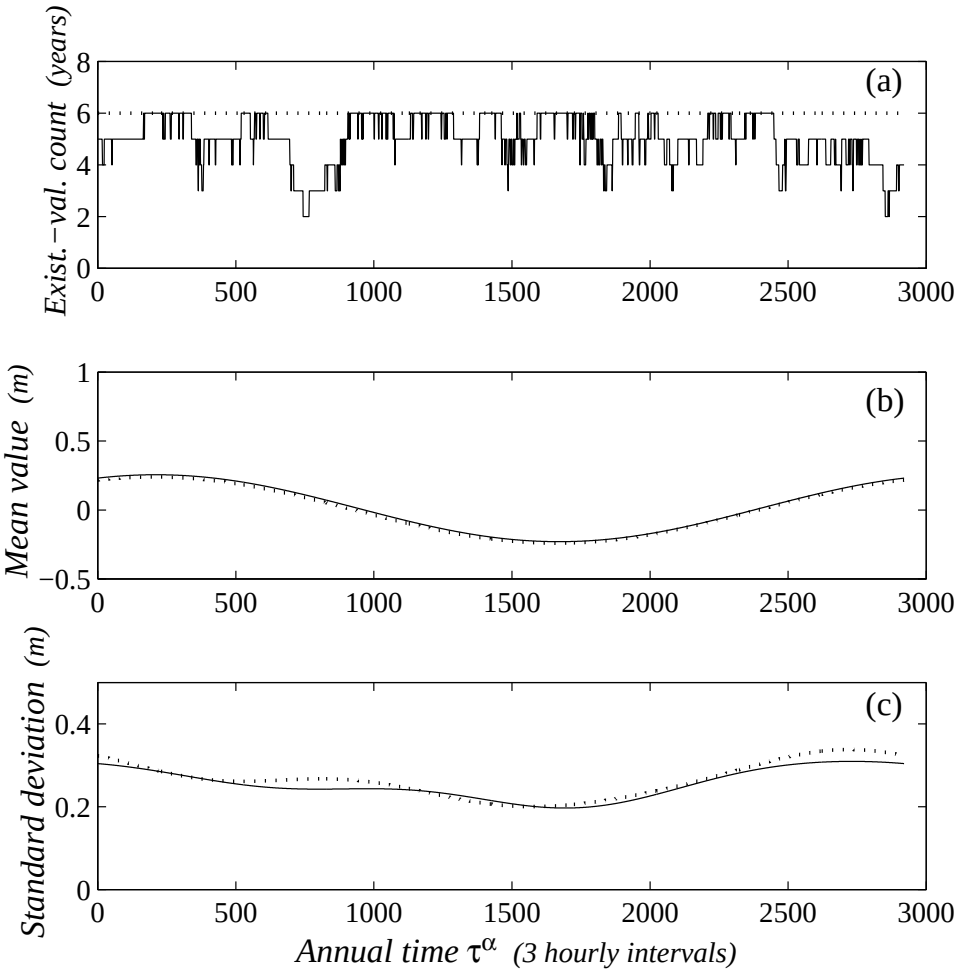


Figure 5.4: (a) Existing-value diagram, (b) seasonal mean value, (c) seasonal standard deviation. (Dotted line: reconstructed time series, solid line: incomplete time series). [16.5% missing-value pattern].

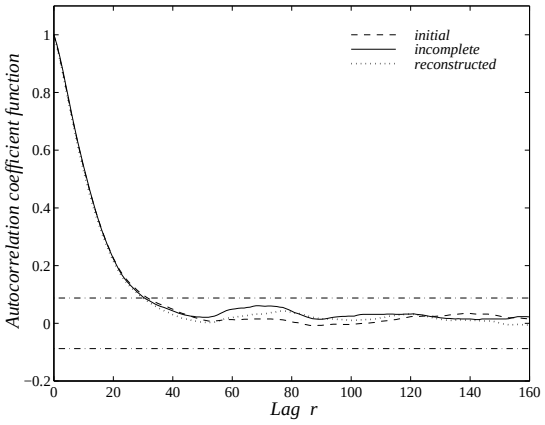


Figure 5.5: Autocorrelation coefficient function of $W(\tau)$ (dashed line), $W_{\mathrm{e.v.}}(\tau)$ (solid line) and $W^c(\tau)$ (dotted line) along with the 95% confidence limits. [16.5% missing-value pattern].

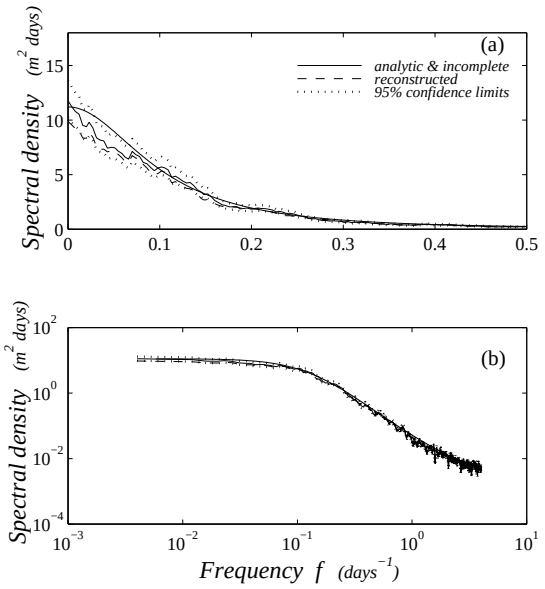


Figure 5.6: Estimated spectral densities $\widehat{S}_{WW}(f)$ (solid line) and $\widehat{S}_{W^cW^c}(f)$ (dashed line) along with the analytic one $S_{\mathsf{ARMA}}(f)$ (solid line) and the 95% confidence limits (dotted line). (a) Linear scale, (b) logarithmic scale. [16.5% missing-value pattern].

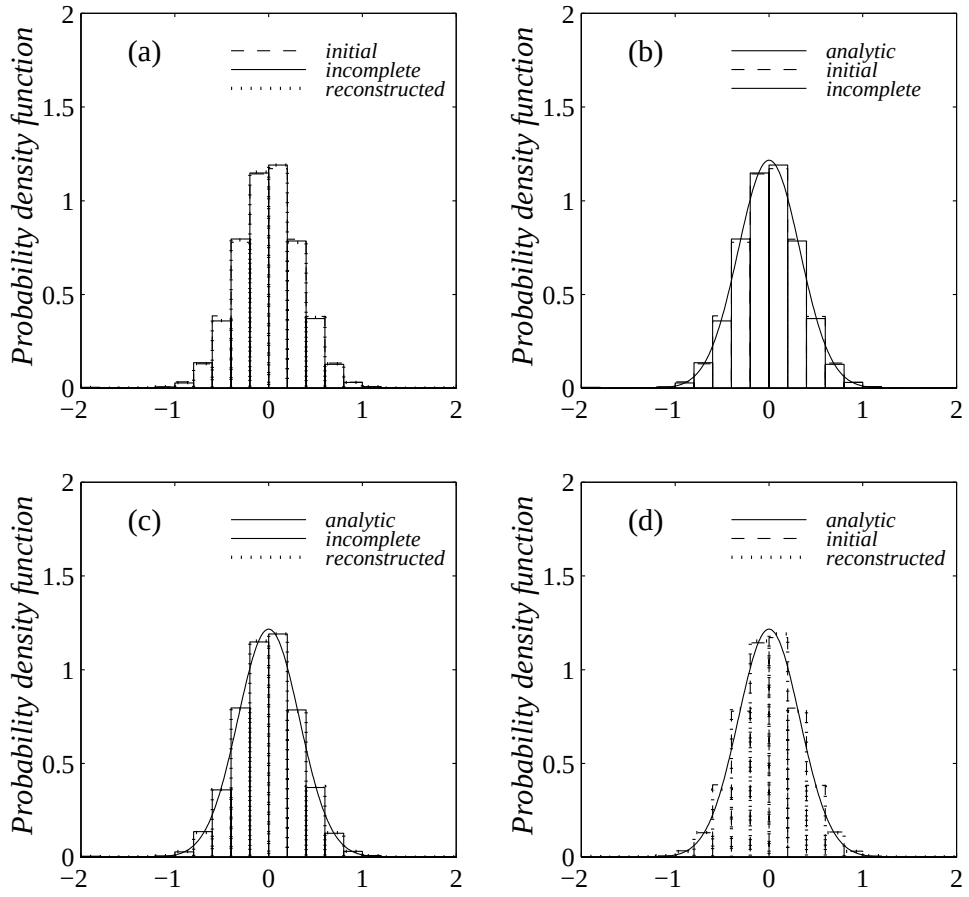


Figure 5.7: Probability density functions (pdf's) of time series of residuals $\varepsilon(\tau)$, $\varepsilon_{\mathrm{e.v.}}(\tau)$ and $\varepsilon^c(\tau)$. (a) Empirical pdf's of $\varepsilon(\tau)$, $\varepsilon_{\mathrm{e.v.}}(\tau)$ and $\varepsilon^c(\tau)$, (b) empirical pdf's of $\varepsilon(\tau)$ and $\varepsilon_{\mathrm{e.v.}}(\tau)$ along with the corresponding Gaussian pdf, (c) empirical pdf's of $\varepsilon_{\mathrm{e.v.}}(\tau)$ and $\varepsilon^c(\tau)$ along with the corresponding Gaussian pdf, (d) empirical pdf's of $\varepsilon(\tau)$ and $\varepsilon^c(\tau)$ along with the corresponding Gaussian pdf. [16.5% missing-value pattern].

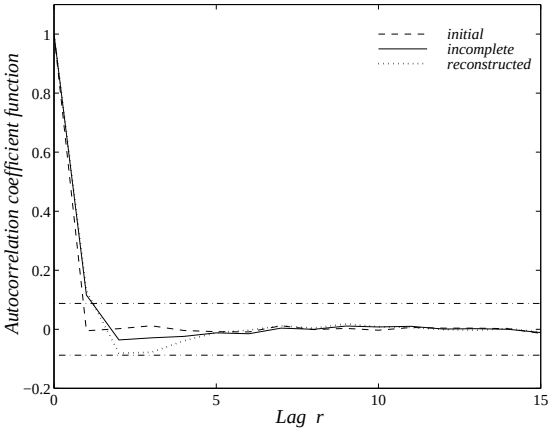


Figure 5.8: Autocorrelation coefficient function of time series of residuals $\varepsilon(\tau)$ (dashed line), $\varepsilon_{\mathrm{e.v.}}(\tau)$ (solid line) and $\varepsilon^c(\tau)$ (dotted line) along with the 95% confidence limits. [16.5% missing-value pattern].

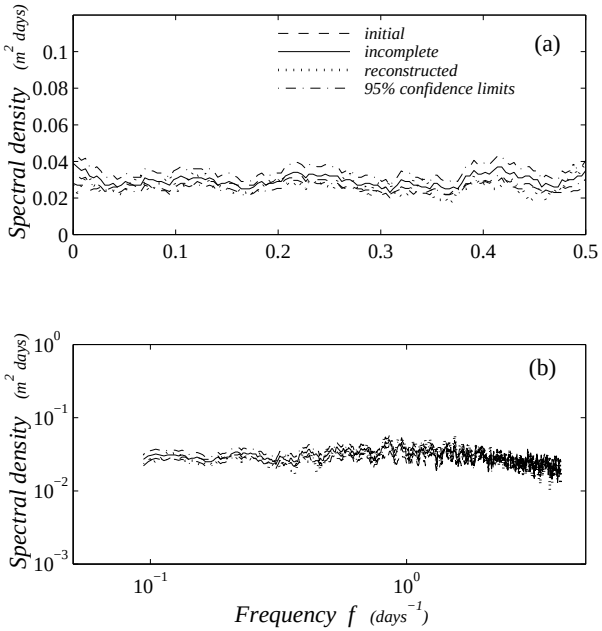


Figure 5.9: Estimated spectral densities of time series of residuals $\varepsilon(\tau)$ (dashed line), $\varepsilon_{\mathrm{e.v.}}(\tau)$ (solid line) and $\varepsilon^c(\tau)$ (dotted line) along with the 95% confidence limits. (a) Linear scale, (b) logarithmic scale. [16.5% missing-value pattern].

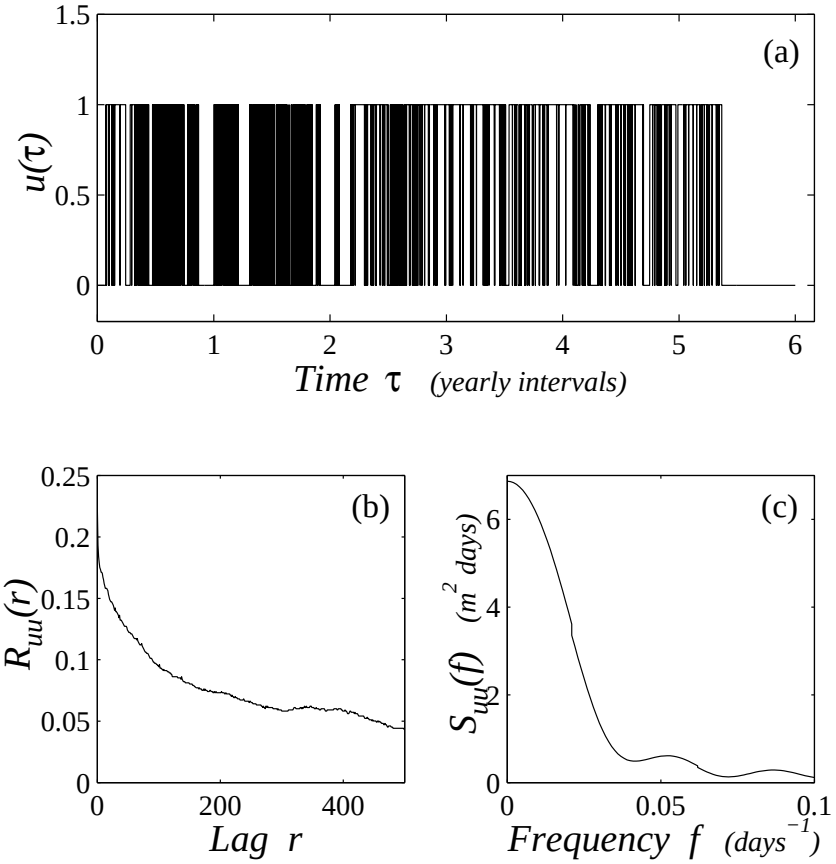


Figure 5.10: (a) Indicator function $u(\tau)$, (b) autocovariance function $C_{uu}(r)$, (c) spectral density $S_{uu}(f)$. [33% missing-value pattern].

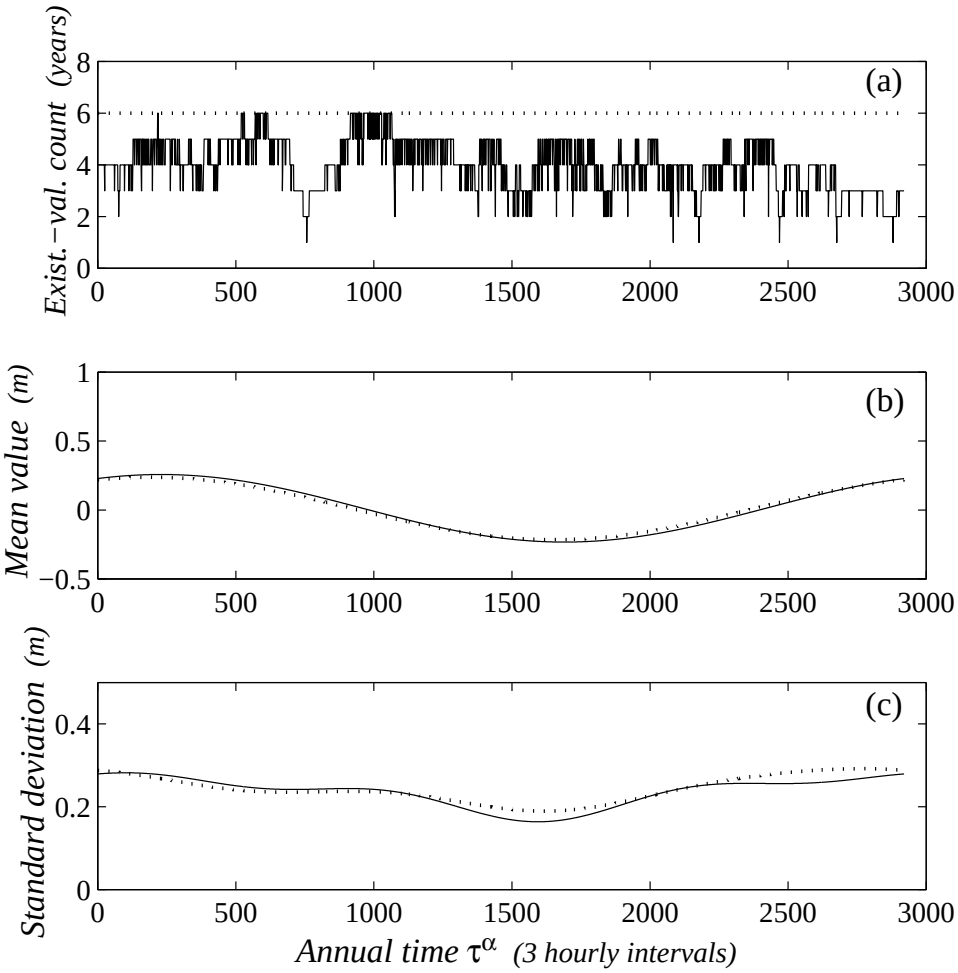


Figure 5.11: (a) Existing-value diagram, (b) seasonal mean value, (c) seasonal standard deviation. (Dotted line: initial time series, solid line: incomplete time series). [33% missing-value pattern].

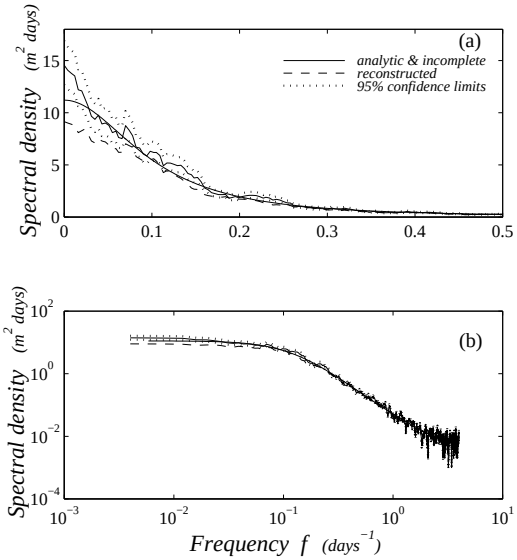


Figure 5.12: Estimated spectral densities $\widehat{S}_{WW}(f)$ (solid line) and $\widehat{S}_{W^cW^c}(f)$ (dashed line) along with the analytic one $S_{\mathsf{ARMA}}(f)$ (solid line) and the 95% confidence limits (dotted line). (a) Linear scale, (b) logarithmic scale. [33% missing-value pattern].

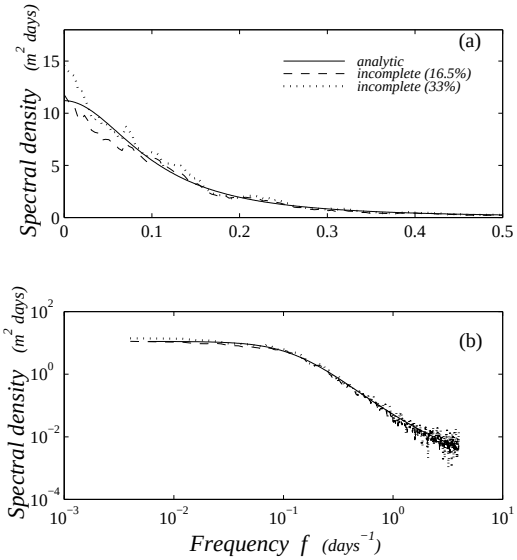


Figure 5.13: Analytic spectral density $S_{\mathsf{ARMA}}(f)$, estimated spectral densities $\widehat{S}_{WW}(f)$ for 16.5% and 33% missing-value patterns.

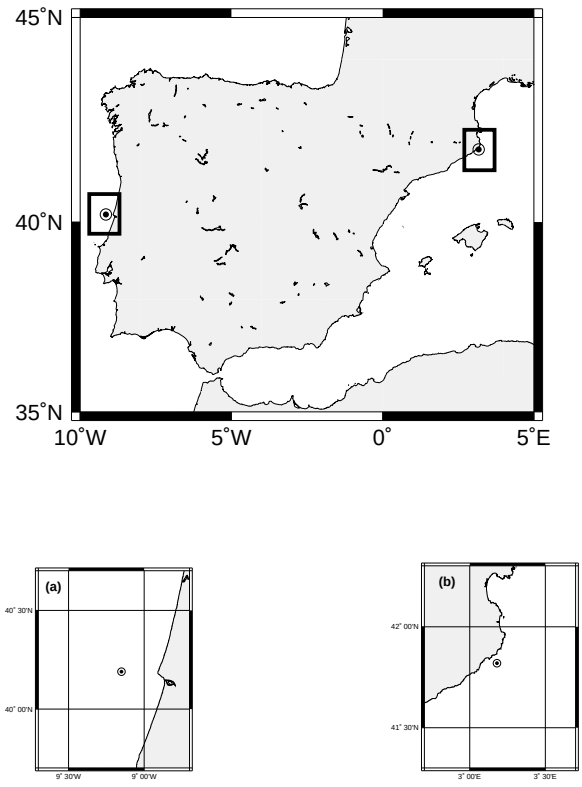


Figure 5.14: Examined sites: (a) Figueira da Foz, (b) Palamós

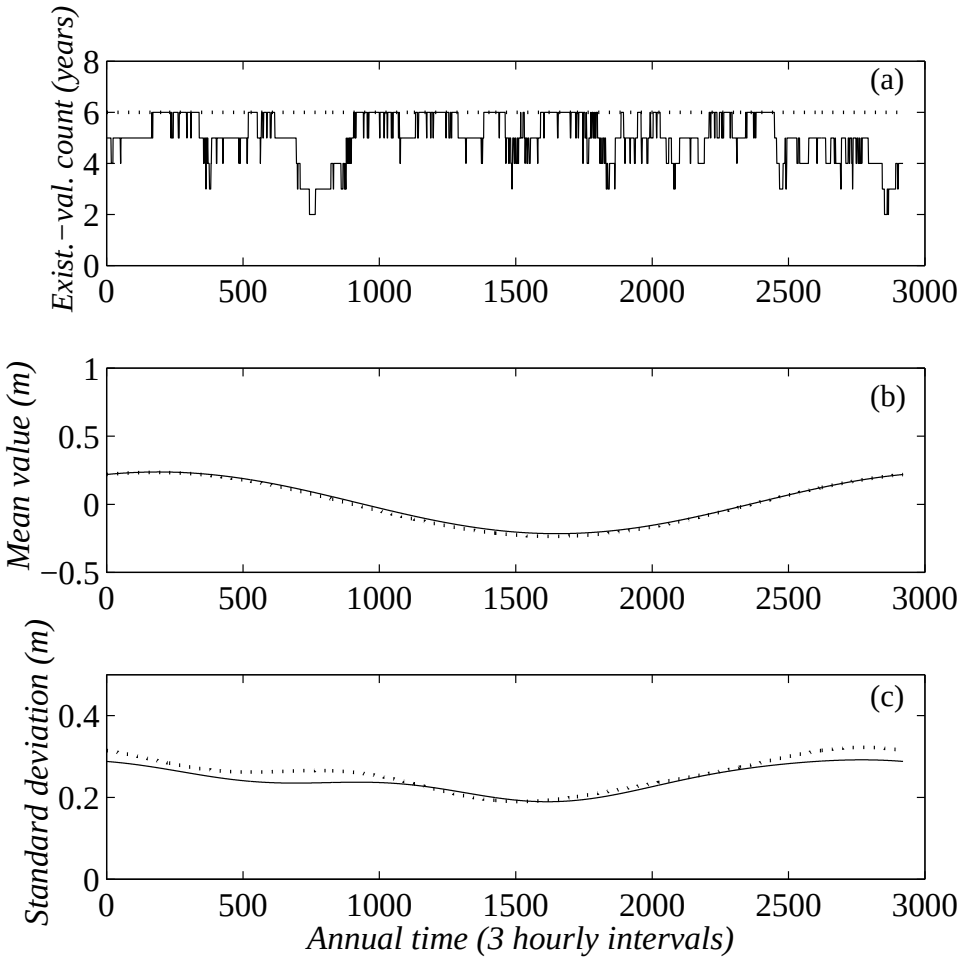


Figure 5.15: Seasonal characteristics of H_S. (a) Existing-value diagram, (b) Seasonal mean value, (c) Seasonal standard deviation. (Solid line: incomplete time series, dotted line: reconstructed time series) Site: Figueira da Foz

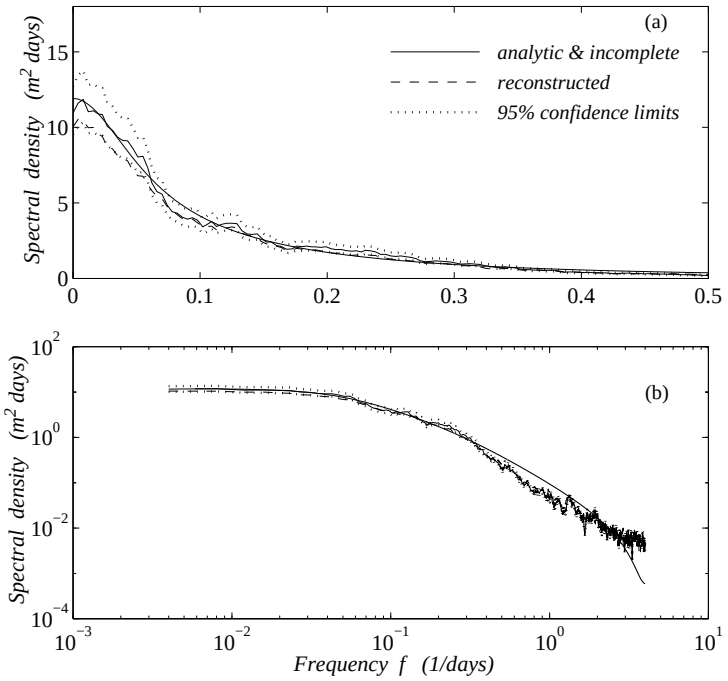


Figure 5.16: Estimated spectral densities $S_{WW}(f)$ (solid line) and $S_{W^cW^c}(f)$ (dashed line), along with the analytic one (solid line) and the 95% confidence limits (dotted line). (a) Linear scale, (b) Logarithmic scale. Site: Figueira da Foz.

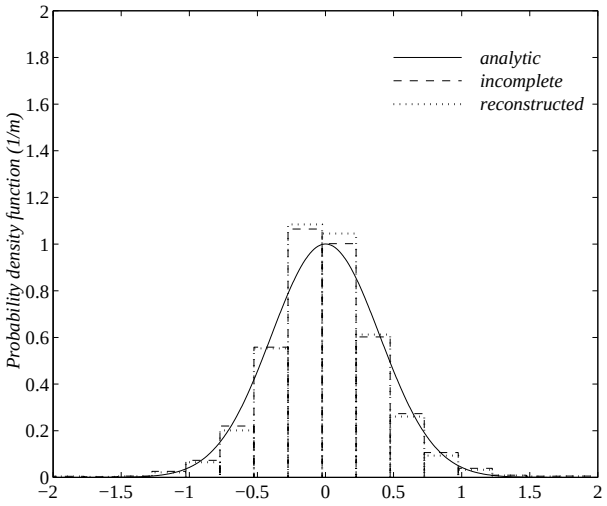


Figure 5.17: Probability density function of residual time series. (i) Analytic (solid line), (ii) Incomplete (dashed line), (iii) Reconstructed (dotted line). Site: Figueira da Foz.

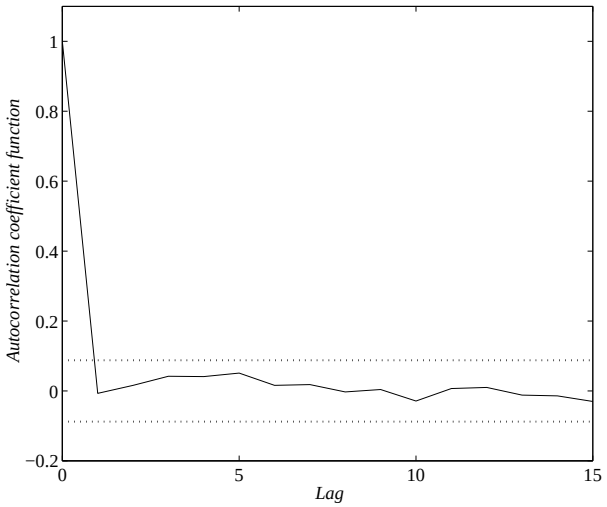


Figure 5.18: Autocorrelation coefficient function of the final residuals along with the 95% confidence limits. Site: Figueira da Foz.

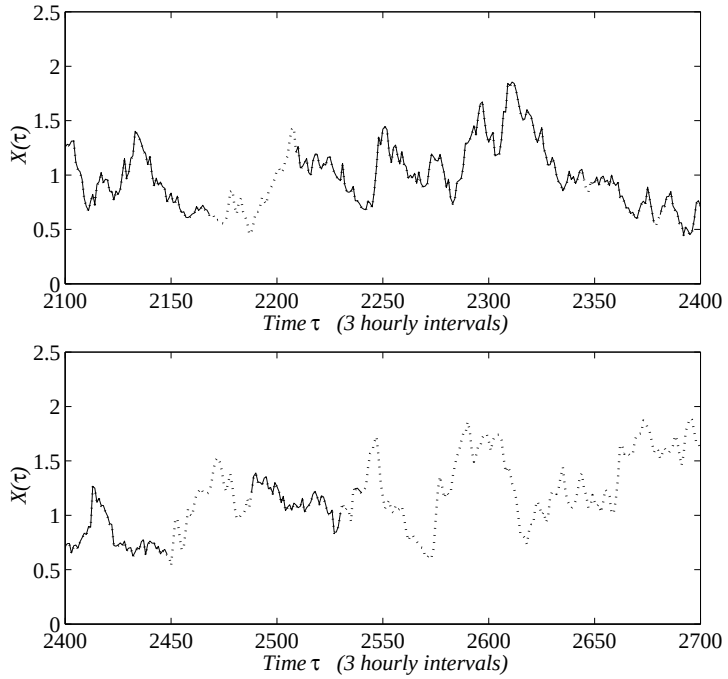


Figure 5.19: Initial incomplete time series (solid line) and reconstructed one (dotted line) for a time window of duration two and a half months. Site: Figueira da Foz.

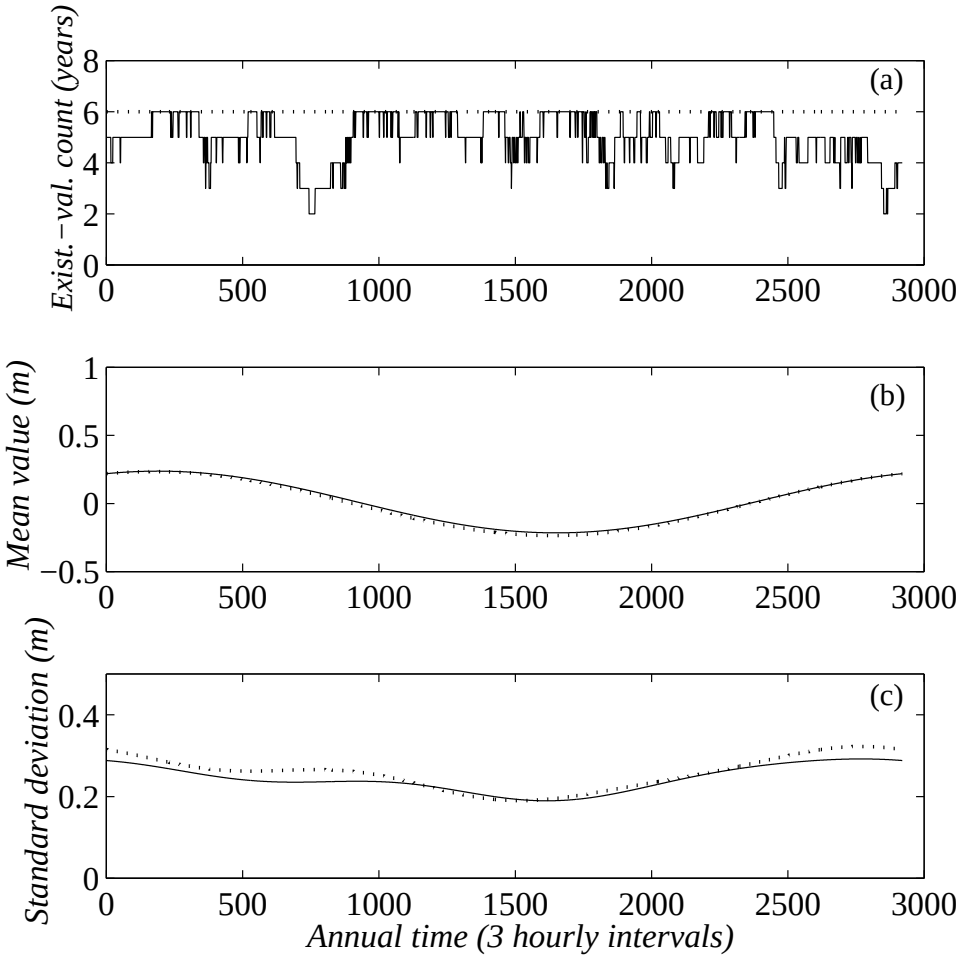


Figure 5.20: Seasonal characteristics of H_S. (a) Existing-value diagram, (b) Seasonal mean value, (c) Seasonal standard deviation. (Solid line: incomplete time series, dotted line: reconstructed time series). Site: Palamós.

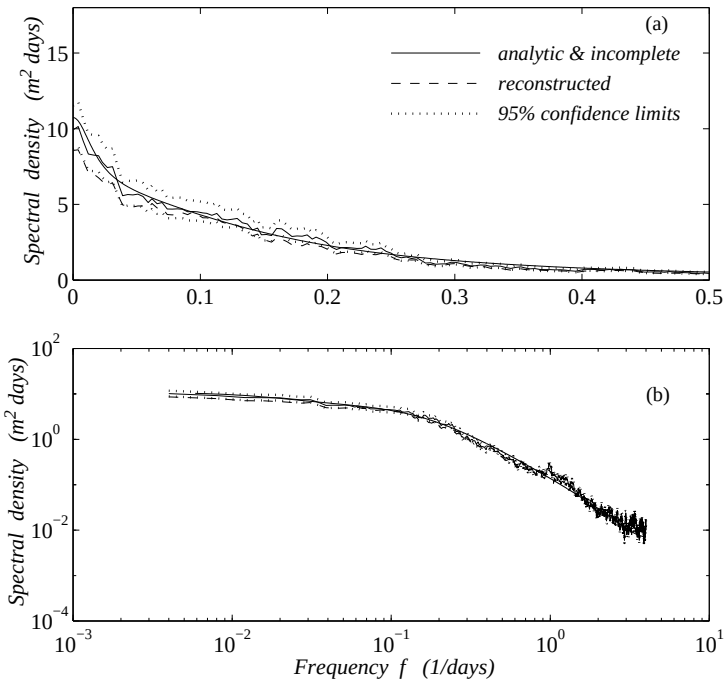


Figure 5.21: Estimated spectral densities $S_{WW}(f)$ (solid line) and $S_{W^cW^c}(f)$ (dashed line), along with the analytic one (solid line) and the 95% confidence limits (dotted line). (a) Linear scale, (b) Logarithmic scale. Site: Palamós.

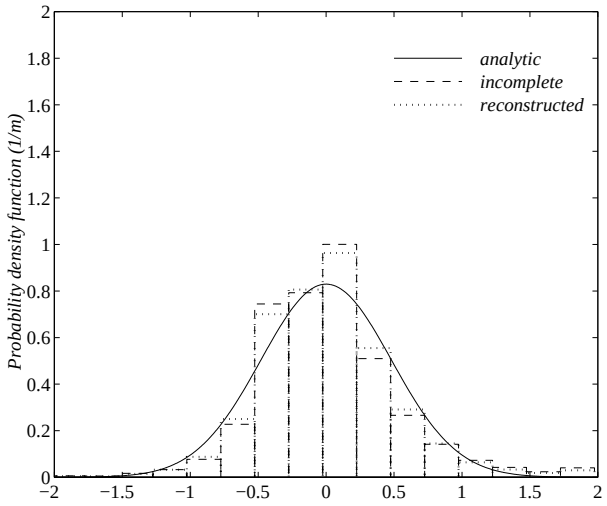


Figure 5.22: Probability density function of residual time series. (i) Analytic (solid line), (ii) Incomplete (dashed line), (iii) Reconstructed (dotted line). Site: Palamós.

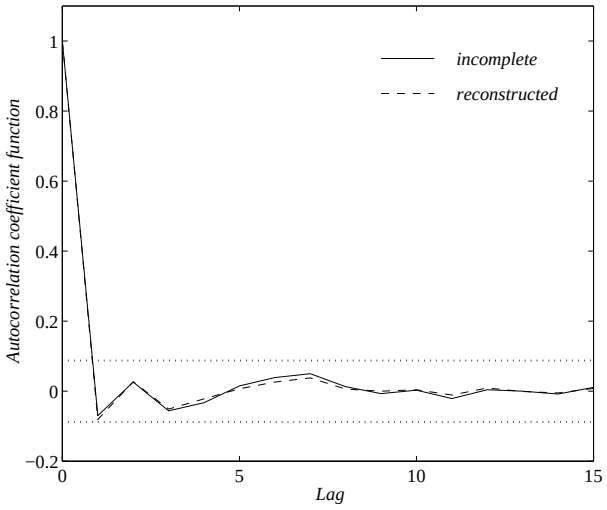


Figure 5.23: Autocorrelation coefficient function of the residuals series: (i) incomplete (solid line), (ii) reconstructed (dashed line), along with the 95% confidence limits (dotted line). Site: Palamós.

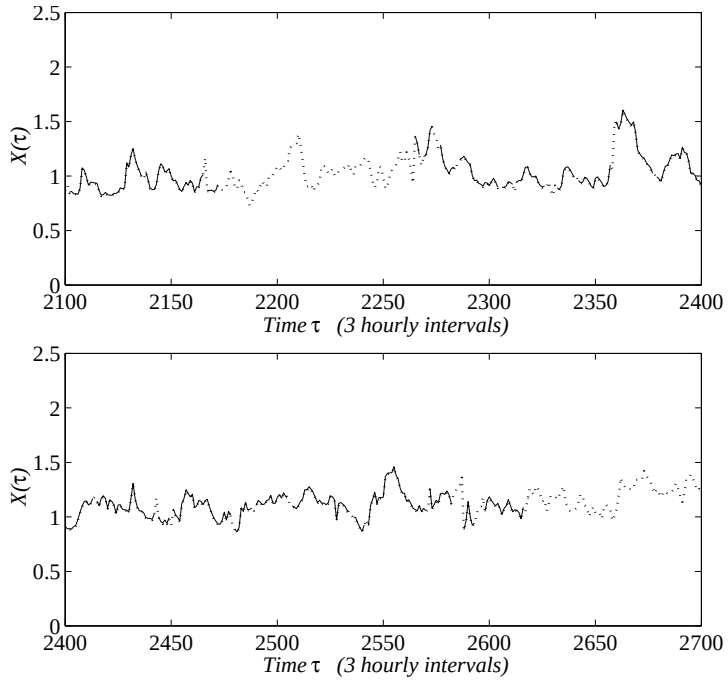


Figure 5.24: Initial incomplete time series (solid line) and reconstructed one (dotted line) for a time window of duration two and a half months. Site: Palamós.

Chapter 6

Study of the second-order probability structure of wave-parameter time series

6.1 Introduction. Statement of the problem

In this chapter, the first- and second-order probability structure of the nonstationary time series of significant wave height $X(\tau)$ is studied. In this connection, appropriate univariate and bivariate probability models are needed. There is a wide variety of multivariate (in general) probability models, the construction of which is a difficult problem in its generality; see, e.g., Hutchinson and Lai (1990), Joe (1997). In the case where no underlying theoretical evidence is available for their form, the most reliable information concerns the univariate marginals. Accordingly, it is very important that the multivariate distribution can incorporate any marginal model. In the statistical literature, multivariate models with given (prescribed) marginals are known as the Fréchet class (Dall' Aglio et al., 1991; Rüschendorf et al., 1996; Beneš and Štěpán, 1997). A brief description of this class is given in Sections 6.2 and 6.3, while the maximum likelihood estimation procedure for a specific member of this class in Sections 6.4–6.6. The main advantages of probability models from the Fréchet class are: a) the marginal models can be chosen and estimated in advance, b) the dependence structure is controlled by an appropriate set of parameters, c) various general alternative representations are available.

In the present work, a special bivariate model, first constructed by Plackett (1965) is applied to the representation of the second-order probability distribution of the stationary part $W(\tau)$ of the initial series $X(\tau)$; see Section 6.7. For comparison purposes, the standard bivariate lognormal distribution is also examined. The Plackett model seems to offer a better performance regarding the description of the empirical population, and an increased flexibility regarding the shape of the distribution, at the same computational cost. Then, by means of the modelling presented in

Chapter 4, the first- and second-order probability distributions of $X(\tau)$ are easily derived from the corresponding ones of $W(\tau)$, by simply considering the relation between $X(\tau)$ and $W(\tau)$ as a linear time-varying transformation.

6.2 Bivariate probability models. General considerations

In proceeding to the modelling and estimation of bivariate probability distributions, it is important to incorporate in the bivariate model the univariate marginal distributions $F_1(x_1)$ and $F_2(x_2)$, which have already been studied in the previous stage, i.e., in the stage of univariate analysis. This almost obvious requirement is not satisfied in the case of the conditional probability approach, i.e. in the case where the bivariate density is represented by means of the equation

$$f(x_1, x_2) = f_1(x_1)\ f_{2|1}(x_2|x_1). \tag{6.1}$$

In this case, only one marginal can be *a priori* introduced in the representation, the other one being obtained by integration after estimating the conditional densities. The compatibility of the second marginal distribution with the directly estimated one (by means of the univariate data), although well-established theoretically, is often poor in practice. In addition, the amount of bivariate data required for estimating the conditional densities $f_{2|1}(x_2|x_1)$ is large.

To overcome these difficulties, we decided to work within the so-called **Fréchet class** $\mathcal{F}(F_1, F_2)$, that is, the class of all (sufficiently smooth) bivariate distributions with predetermined marginal distributions and . This class was first introduced and studied in 1951 by Maurice Fréchet (1951), the famous French mathematician who had important contributions to functional analysis and probability theory. See also Fréchet (1956; 1958). Since then, many authors have worked on the concept of bivariate Fréchet class and various multivariate generalisations of it. See, e.g., Mardia (1970), Rüschendorf (1985), Dall' Aglio et al. (1991), Cuadras (1992), Marco and Ruiz-Rivas (1992), Joe (1997).

From the very definition of the Fréchet class it is clear that, the marginals being fixed, only the correlation (or association) structure is remaining (in the context of the bivariate treatment) to be dealt with. This fact is exploited by a, simply proven yet important, theorem that follows. This theorem provides a way of characterization of a Fréchet class by means of a two-variable deterministic function satisfying some conditions. Before proceeding we give the exact definition of this function, which is called copula.

Definition 6.1 *A bivariate* ***copula*** *is a function* $G(y_1, y_2)$ *defined on* $[0,1]^2$*:*

$$[0,1] \times [0,1] \ni (y_1, y_2) \longrightarrow G(y_1, y_2) \in [0,1], \tag{6.2}$$

which satisfies the following conditions:

(i) For each $y_1 \in [0,1]$ *and* $y_2 \in [0,1]$, $G(y_1, 1) = y_1$, *and* $G(1, y_2) = y_2$,

(ii) For each $y_1 \in [0,1]$ *and* $y_2 \in [0,1]$, $G(y_1, 0) = G(0, y_2) = 0$,

(iii) For every pair (a_1, b_1), (a_2, b_2), *satisfying* $a_1 \leq a_2$ *and* $b_1 \leq b_2$, *the inequality*

$$G(a_2, b_2) - G(a_1, b_2) - G(a_2, b_1) + G(a_1, b_1) \geq 0 \tag{6.3}$$

holds true. ■

Alternatively, a bivariate copula can be defined as a bivariate probability distribution function with uniform marginals in [0,1].

From condition (iii) it follows that the copula $G(y_1, y_2)$ is a nondecreasing function in each of its arguments, and satisfies the Lipschitz condition

$$|G(b_2, b_1) - G(a_2, a_1)| \leq |b_2 - a_2| + |b_1 - a_1|. \tag{6.4}$$

The concept and terminology of copula was introduced by Sklar (1959). See also Sklar (1973), Schweizer (1991). The notion of bivariate copula can be generalised to n dimensions; see e.g. Schweizer (1991). For a recent survey of the theory of copulas, see, e.g., Dall' Aglio et al. (1991), Rüschendorf et al. (1996), Beneš and Štěpán (1997).

We are now in a position to state and prove the following

Theorem 6.1 *Let* $F_1(x_1)$, $x_1 \in I_1 \subset \mathbb{R}$ *and* $F_2(x_2)$, $x_2 \in I_2 \subset \mathbb{R}$ *be two given, absolutely continuous, univariate probability distribution functions.*

Consider a bivariate probability distribution $F(x_1, x_2) \in \mathcal{F}(F_1, F_2)$. *Then the function*

$$G(y_1, y_2) = F\big(F_1^{-1}(y_1), F_2^{-1}(y_2)\big) \tag{6.5}$$

is a bivariate copula. Inversely, if $G(y_1, y_2)$ *is a bivariate copula, then the function*

$$F(x_1, x_2) = G\big(F_1(x_1), F_2(x_2)\big) \tag{6.6}$$

is a bivariate probability distribution with univariate marginals $F_1(x_1)$ *and* $F_2(x_2)$.

Equations (6.5) and (6.6) define a one-to-one correspondence between the members of the Fréchet class $\mathcal{F}(F_1, F_2)$ *and the bivariate copulas.* ■

Proof of the Theorem 6.1
The proof of the theorem is straightforward and, thus, is omitted. ■

In terms of random variables the above result can be stated as follows:

$$\left[\begin{array}{c}(X_1, X_2) \text{ is distributed} \\ \text{in accordance to } F(x_1, x_2)\end{array}\right] \text{ if and only if } \left[\begin{array}{c}(Y_1, Y_2) \text{ is distributed in accordance to} \\ G(y_1, y_2) = F\big(F_1^{-1}(y_1), F_2^{-1}(y_2)\big)\end{array}\right].$$

The joint probability density function corresponding to $F(x_1, x_2)$ is given by

$$f(x_1, x_2) = g\big(F_1(x_1), F_2(x_2)\big)\ f_1(x_1)\ f_2(x_2), \tag{6.7}$$

where $g(y_1, y_2) = \dfrac{\partial^2 G(y_1, y_2)}{\partial y_1 \partial y_2}$.

There are two main directions to implement the Fréchet-class concept for applications: Either by using general representation theorems, or by using simple closed-form parametric models of copulas.

- General representation theorems

Let us first take a quick look at some general representation theorems for the elements of a given Fréchet class:

- **Lancaster** representation (1958; 1963; 1969):

$$f(x_1, x_2) = f_1(x_1)\ f_2(x_2)\ \left[1 + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \rho_{ij} q_{1i}(x_1) q_{2j}(x_2)\right], \tag{6.8}$$

  where $\{q_{1i}(x_1), \quad i{=}1, 2, \dots\}$ and $\{q_{2j}(x_2), \quad i{=}1, 2, \dots\}$ are complete orthonormal systems with weights $f_1(x_1)$ and $f_2(x_2)$, respectively, and ρ_{ij} are generalised correlation coefficients.

  Lancaster representation (6.8) can be considered as a generalisation of the bivariate Hermite polynomial expansion with univariate marginals given by Gram-Charlier expansions (Mardia, 1970; Johnson and Kotz, 1972). Some special cases of this representation have been studied by Sarmanov (1966), and Sarmanov and Bratoeva (1967), who also gave conditions ensuring that a simplified version of (6.8) is strictly positive.

  Despite its general character and possible extension to the multivariate case, Lancaster representation shows some disadvantages in practice:

(a) The truncated series may become negative in some (x_1, x_2) regions, and

(b) It is not easy to find the functions $q_{1i}(x_1)$, $q_{2j}(x_2)$ they are known only in some special cases).

- **Abazaliev** representation (1968):

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) + \frac{1}{\pi^2}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty}\frac{\lambda_{F_1F_2}\big(k, s, F(x_1, x_2)\big)}{ks}\sin(2\pi k z_1)\sin(2\pi k z_2), \tag{6.9}$$

where $z_1 = F_1(x_1) - \dfrac{1}{2}$ and $z_2 = F_2(x_2) - \dfrac{1}{2}$.

- Closed-form parametric models of copulas

We are now turning to the second direction of exploitation of Fréchet-class concept, presenting some important special families of copulas, providing us with special but handy bivariate models:

- **Morgenstern** model (1956)

$$f_{\boldsymbol{Y}}(y_1, y_2; \alpha) = 1 + \alpha(2y_1 - 1)(2y_2 - 1). \tag{6.10}$$

The model is valid only for *weakly correlated* Y_1, Y_2.

- **Huang and Kotz** model (1984)

$$f_{\boldsymbol{Y}}(y_1, y_2; \alpha, \beta) = 1 + \alpha(2y_1 - 1)(2y_2 - 1) + \beta y_1 y_2(2 - 3y_1)(2 - 3y_2). \tag{6.11}$$

The model is better than the previous one, but again is appropriate only for *weakly correlated* Y_1, Y_2.

- **Ali, Mikhail and Haq** model (1978)

$$F_{\boldsymbol{Y}}(y_1, y_2; \alpha) = \frac{y_1 y_2}{1 - \alpha(1 - y_1)(1 - y_2)}. \tag{6.12}$$

- **Plackett** model (1965), which will be extensively studied in the following sections. This model will be used for modeling the bivariate data in the present work. Its applicability to wave data has been established by Athanassoulis et al. (1994). See also Athanassoulis, Skarsoulis and Stefanakos (1995a), Athanassoulis and Stefanakos (1996b), where this method is extensively applied to bivariate samples of significant wave height and zero upcrossing period, significant wave height and mean period, significant wave height and peak period.

If the initial model $G(\cdot,\cdot)$, in equation (6.6), is the bivariate normal, the family of models obtained is known as a translation family or translation system. It was first introduced by Edgeworth (1898) and extensively studied by Johnson (1949). For a detailed description, see Mardia (1970) and Johnson (1987).

6.3 The bivariate Plackett model

The **standard Plackett model (Plackett copula)**, introduced by Plackett (1965), is a specific copula, i.e. a specific bivariate probability distribution function with both *marginals uniformly distributed* in [0,1], which is given by the equation

$$F_{\boldsymbol{Y}}(y_1,y_2;\psi)=\begin{cases}\dfrac{S(y_1,y_2)-\left[S^2(y_1,y_2)-4\psi(\psi-1)y_1y_2\right]^{1/2}}{2(\psi-1)}, & \text{for } \psi\neq 1,\\ y_1\ y_2, & \text{for } \psi=1,\end{cases} \tag{6.13}$$

where $S(y_1,y_2)=1+(\psi-1)(y_1+y_2)$.

The corresponding bivariate density function is easily obtained by differentiation of equ. (6.13):

$$f_{\boldsymbol{Y}}(y_1,y_2;\psi)=\frac{\psi(\psi-1)(y_1+y_2-2y_1y_2)+\psi}{\left\{\left[1+(\psi-1)(y_1+y_2)\right]^2-4\psi(\psi-1)y_1y_2\right\}^{1/2}}. \tag{6.14}$$

In both (6.13) and (6.14), the independent variable (y_1,y_2) and the parameter ψ range over the following sets of values:

$$(y_1,y_2)\in[0,1]^2 \quad \text{and} \quad \psi\in\mathbb{R}^+=(0,\infty).$$

The association parameter ψ is related to the correlation coefficient ρ of the bivariate variable (Y_1,Y_2) by means of the equation (Johnson and Kotz, 1985)

$$\rho=\frac{\psi^2-1-2\psi\log\psi}{\psi^2-1}. \tag{6.15}$$

From the above equation it can be seen that: for $0<\psi<1$, the random variables Y_1, Y_2 are negatively correlated, i.e. $\mathsf{Cov}(Y_1, Y_2)<0$, while for $1<\psi<\infty$, Y_1, Y_2 are positively correlated, i.e. $\mathsf{Cov}(Y_1, Y_2)>0$. Especially, for ψ=1, the random variables Y_1, Y_2 are uncorrelated (in fact, independent), since

$$f_{\boldsymbol{Y}}(y_1, y_2; 1) = 1 = 1 \cdot 1 = f_U(y_1)\ f_U(y_2),$$

where $f_U(\cdot)$ is the density function of a random variable uniformly distributed in [0,1].

Also, it can easily be shown that

$$f_{\boldsymbol{Y}}(y_1, y_2; \psi) \neq 0, \quad \text{for all} \quad (y_1, y_2) \in [0,1]^2 \quad \text{and all} \quad \psi \in \mathbb{R}^+ = (0, \infty).$$

Consider now a general bivariate quantity (X_1, X_2) with joint probability distribution function $F_{\boldsymbol{X}}(x_1, x_2)$=$F_{\boldsymbol{X}}(x_1, x_2; \boldsymbol{\theta})$ and marginal probability distribution functions $F_1(x_1)$=$F_1(x_1; \boldsymbol{\theta}_1)$ and $F_2(x_2)$=$F_2(x_2; \boldsymbol{\theta}_2)$. Here $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameter vectors of the univariate marginals and $\boldsymbol{\theta}$ is the parameter vector of the joint distribution $F_{\boldsymbol{X}}(x_1, x_2)$. It is assumed that $\boldsymbol{\theta}$=($\boldsymbol{\theta}_1$,$\boldsymbol{\theta}_2$,$\boldsymbol{\theta}_{12}$), where $\boldsymbol{\theta}_{12}$ is a parameter vector related to the correlation structure of (X_1, X_2).

Definition 6.2 *We shall say that the bivariate random variable (X_1, X_2) follows the general Plackett distribution (or model), if and only if the bivariate random variable (Y_1, Y_2)=$(F_1(X_1), F_2(X_2))$ follows the standard Plackett model (6.13).* ■

By direct substitution, we find

$$F_{\boldsymbol{X}}(x_1, x_2; \boldsymbol{\theta}) = F_{\boldsymbol{Y}}\big(F_1(x_1; \boldsymbol{\theta}_1), F_2(x_2; \boldsymbol{\theta}_2); \psi\big), \tag{6.16}$$

or

$$F_{\boldsymbol{X}}(x_1, x_2; \boldsymbol{\theta}) = \begin{cases} \dfrac{S(x_1, x_2) - \big[S^2(x_1, x_2) - 4\psi(\psi-1)F_1(x_1)F_2(x_2)\big]^{1/2}}{2(\psi-1)}, & \text{for } \psi \neq 1, \\ F_1(x_1)\ F_2(x_2), & \text{for } \psi = 1, \end{cases} \tag{6.17}$$

where $S(x_1, x_2) = 1 + (\psi-1)\big(F_1(x_1) + F_2(x_2)\big)$ and $\boldsymbol{\theta}$=$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \psi)$.

The corresponding density function is easily obtained by differentiating equation (6.16):

$$f_{\boldsymbol{X}}(x_1, x_2; \boldsymbol{\theta}) = f_{\boldsymbol{Y}}\big(F_1(x_1; \boldsymbol{\theta}_1), F_2(x_2; \boldsymbol{\theta}_2); \psi\big) f_1(x_1; \boldsymbol{\theta}_1) f_2(x_2; \boldsymbol{\theta}_2), \tag{6.18}$$

or equation (6.17):

$$f_{\boldsymbol{X}}(x_1, x_2; \psi) = \frac{\psi\Big\{(\psi-1)\big[F_1(x_1) + F_2(x_2) - 2F_1(x_1)F_2(y_2)\big] + 1\Big\} f_1(x_1) f_2(x_2)}{\Big\{\big[1 + (\psi-1)\big(F_1(x_1) + F_2(x_2)\big)\big]^2 - 4\psi(\psi-1)F_1(x_1)F_2(x_2)\Big\}^{3/2}}. \tag{6.19}$$

6.4 Maximum likelihood estimation of parameters. General considerations

The log-likelihood function of a bivariate probability model with given marginals is given by the relation

$$\ell(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}) = \log L(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}) = \sum_{n=1}^{N} \log f(X_1^{(n)}, X_2^{(n)}; \boldsymbol{\theta}), \tag{6.20}$$

where

$$(\boldsymbol{X}_1, \boldsymbol{X}_2) = \Big((X_1^{(1)}, X_2^{(1)}), (X_1^{(2)}, X_2^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)}), \ldots, (X_1^{(N)}, X_2^{(N)})\Big) \tag{6.21}$$

is the bivariate sample of size N from a bivariate population following the above bivariate model.

In order to calculate this log-likelihood function, use is made of equ. (6.18). Then, the log-likelihood function takes the form:

$$\ell(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}) = \ell_1(\boldsymbol{X}_1; \boldsymbol{\theta}_1) + \ell_2(\boldsymbol{X}_2; \boldsymbol{\theta}_2) + \ell_{12}(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{12}), \tag{6.22}$$

where

$$\ell_\alpha(\boldsymbol{X}_\alpha; \boldsymbol{\theta}_\alpha) = \sum_{n=1}^{N} \log f_\alpha(X_\alpha^{(n)}; \boldsymbol{\theta}_\alpha), \qquad \alpha = 1, 2, \tag{6.23}$$

and

$$\ell_{12}(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{12}) = \sum_{n=1}^{N} \log f_{\boldsymbol{Y}}\Big(F_1(X_1^{(n)}; \boldsymbol{\theta}_1), F_2(X_2^{(n)}; \boldsymbol{\theta}_2); \boldsymbol{\theta}_{12}\Big). \tag{6.24}$$

That is, $\ell_\alpha(\boldsymbol{X}_\alpha; \boldsymbol{\theta}_\alpha)$, α=1,2 are the univariate log-likelihood functions of the marginal models, and $\ell_{12}(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{12})$ is the bivariate log-likelihood function of the standard Plackett model.

The following strategy is applied to estimating the parameter vector $\boldsymbol{\theta}$=$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{12})$: First, the parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ of the univariate marginal models $f_1(x_1; \boldsymbol{\theta}_1)$, $f_2(x_2; \boldsymbol{\theta}_2)$ are estimated from the corresponding univariate marginal data. This is made by applying the maximum likelihood (ML) method separately to each univariate model. This step, concerning the univariate

marginals, is described in detail in Athanassoulis and Stefanakos (1996a). See also the specialized monographs by Crow and Shimizu (1988), Bowman and Shenton (1988), and references cited therein. Secondly, the association parameter vector $\boldsymbol{\theta}_{12}$ is estimated from the bivariate data, keeping $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ fixed to their values estimated in the previous step, and applying again the ML method to the bivariate model.

The above procedure can also be described as follows:

1. Find $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ by solving the two optimisation problems

$$\ell_\alpha(\boldsymbol{X}_\alpha;\hat{\boldsymbol{\theta}}_\alpha) \geq \ell_\alpha(\boldsymbol{X}_\alpha;\boldsymbol{\theta}_\alpha), \quad \text{for every} \quad \boldsymbol{\theta}_\alpha \in \boldsymbol{\Theta}_\alpha, \quad \alpha = 1,2. \tag{6.25}$$

2. Find $\hat{\boldsymbol{\theta}}_{12}$ by solving the optimisation problem

$$\ell_{12}(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{\theta}_1=\hat{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2=\hat{\boldsymbol{\theta}}_2,\hat{\boldsymbol{\theta}}_{12}) \geq \ell_{12}(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{\theta}_1=\hat{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2=\hat{\boldsymbol{\theta}}_2,\boldsymbol{\theta}_{12}),$$
$$\text{for every} \quad \boldsymbol{\theta}_{12} \in \boldsymbol{\Theta}_{12}. \tag{6.26}$$

In the case where $\ell_{12}(\boldsymbol{X}_1,\boldsymbol{X}_2;\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\boldsymbol{\theta}_{12})$ is continuously differentiable with respect to $\boldsymbol{\theta}_{12}$, the ML-estimate $\hat{\boldsymbol{\theta}}_{12}$ should satisfy the equations

$$\frac{\partial \ell_{12}(\boldsymbol{X}_1,\boldsymbol{X}_2;\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\boldsymbol{\theta}_{12})}{\partial \boldsymbol{\theta}_{12}} = 0, \tag{6.27}$$

which may also be used for finding $\hat{\boldsymbol{\theta}}_{12}$. In the next section, equation (6.27) for the standard Plackett model will be thoroughly investigated both theoretically and numerically.

The two-step estimation procedure described above will be called the *hierarchical ML estimation* of parameter vectors $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_{12}$.

One could think of a third step of the ML estimation procedure, in which the global log-likelihood function $\ell(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_{12})$ would be maximised simultaneously with respect to all parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_{12}$. Such a step is indeed possible and it has been studied in Athanassoulis et al. (1994). The main conclusions drawn by this study are that:

i) The *global ML estimation* of parameters leads to almost the same values as the hierarchical ML estimation.

ii) The implementational and computational effort is much greater for the global ML estimation, than for the hierarchical ML estimation.

In accordance to the above conclusions, in the present work we restrict ourselves to hierarchical ML-estimation, avoiding to consider a global ML-estimation step.

6.5 Maximum likelihood estimation for the standard Plackett model

We shall now focus on the second step of the hierarchical ML estimation of the association parameter $\boldsymbol{\theta}_{12}$=ψ for the standard Plackett model. ML estimation of parameter ψ will be based on the numerical solution of the non-linear algebraic equation (6.27), which can also be written in the form

$$\sum_{n=1}^{N} \frac{1}{f_{\boldsymbol{Y}}\left(Y_1^{(n)}, Y_2^{(n)}; \hat{\psi}\right)} \frac{\partial f_{\boldsymbol{Y}}\left(Y_1^{(n)}, Y_2^{(n)}; \hat{\psi}\right)}{\partial \psi} = 0, \tag{6.28}$$

where $Y_1^{(n)}$=$F_1\left(X_1^{(n)}; \hat{\boldsymbol{\theta}}_1\right)$ and $Y_2^{(n)}$=$F_2\left(X_2^{(n)}; \hat{\boldsymbol{\theta}}_2\right)$.

After a course of algebraic manipulations, equ. (6.28) is transformed to

$$\frac{1}{\psi} + \frac{1}{N}\sum_{n=1}^{N} \frac{A^{(n)}}{(\psi-1)A^{(n)}+1} = 3\frac{1}{N}\sum_{n=1}^{N} \frac{(\psi-1)B^{(n)} + A^{(n)}}{(\psi-1)^2 B^{(n)} + 2(\psi-1)A^{(n)} + 1}, \tag{6.29}$$

where

$$A^{(n)} = Y_1^{(n)} + Y_2^{(n)} - 2Y_1^{(n)}Y_2^{(n)}, \tag{6.30}$$

$$B^{(n)} = \left[Y_1^{(n)}\right]^2 + \left[Y_2^{(n)}\right]^2 - 2Y_1^{(n)}Y_2^{(n)} = \left(Y_1^{(n)} - Y_2^{(n)}\right)^2. \tag{6.31}$$

Both derived samples $\boldsymbol{A}$=$\{A^{(n)}, \quad n$=$1, 2, \ldots, N\}$ and $\boldsymbol{B}$=$\{B^{(n)}, \quad n$=$1, 2, \ldots, N\}$ range over $[0, 1]$. Besides, the following important inequalities hold true for $A^{(n)}$ and $B^{(n)}$:

$$0 \le \left[B^{(n)}\right]^2 \le A^{(n)}B^{(n)} \le B^{(n)} \le \left[A^{(n)}\right]^2 \le A^{(n)} \le 1. \tag{6.32}$$

These inequalities are proved in Section 6.A (Lemma 6.2).

To facilitate the analysis related to equ. (6.29), we denote by $h_1(\psi; \boldsymbol{A})$ the left-hand side of (6.29) and by $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ the right-hand side of the same equation, i.e.

$$h_1(\psi; \boldsymbol{A}) = \frac{1}{\psi} + \frac{1}{N}\sum_{n=1}^{N} \frac{A^{(n)}}{(\psi-1)A^{(n)}+1}, \tag{6.33}$$

$$h_3(\psi;\boldsymbol{A},\boldsymbol{B})=3\frac{1}{N}\sum_{n=1}^{N}\frac{(\psi-1)B^{(n)}+A^{(n)}}{(\psi-1)^2B^{(n)}+2(\psi-1)A^{(n)}+1}. \tag{6.34}$$

As it is shown in Section 6.A (devoted to various proofs), for all possible positive non-degenerate samples $\boldsymbol{A}$, $\boldsymbol{B}$, both functions $h_1(\psi;\boldsymbol{A})$ and $h_3(\psi;\boldsymbol{A},\boldsymbol{B})$ are non-negative on the whole range of ψ:

$$h_1(\psi;\boldsymbol{A})\geq 0,\quad h_3(\psi;\boldsymbol{A},\boldsymbol{B})\geq 0,\qquad \begin{array}{l}\text{for all } \psi>0, \text{ and}\\ \text{for every positive, non-degenerate}\\ \text{samples } \boldsymbol{A},\boldsymbol{B}.\end{array} \tag{6.35}$$

We shall now examine the limiting behaviour of functions $h_1(\psi;\boldsymbol{A})$ and $h_3(\psi;\boldsymbol{A},\boldsymbol{B})$ as $\psi\to 0$ and $\psi\to+\infty$.

(i) The limiting case $\psi\to 0$:

$$\lim_{\psi\to 0}h_1(\psi;\boldsymbol{A})=\lim_{\psi\to 0}\frac{1}{\psi}+\frac{1}{N}\sum_{n=1}^{N}\frac{A^{(n)}}{1-A^{(n)}}=+\infty, \tag{6.36}$$

Thus, as $\psi\to 0$, the function $h_1(\psi;\boldsymbol{A})$ tends to $+\infty$, being asymptotically equivalent to the function $\dfrac{1}{\psi}$.

On the other hand,

$$\lim_{\psi\to 0}h_3(\psi;\boldsymbol{A},\boldsymbol{B})=3\frac{1}{N}\sum_{n=1}^{N}\frac{A^{(n)}-B^{(n)}}{B^{(n)}-2A^{(n)}+1}=h_3(0;\boldsymbol{A},\boldsymbol{B})\geq 0. \tag{6.37}$$

The last inequality is proved in Section 6.A.

(ii) The limiting case $\psi\to+\infty$:

$$\begin{aligned}\lim_{\psi\to\infty}h_1(\psi;\boldsymbol{A})&=\lim_{\psi\to\infty}\frac{1}{\psi}+\lim_{\psi\to\infty}\frac{1}{\psi}\left[\frac{1}{N}\sum_{n=1}^{N}\frac{A^{(n)}}{\left(1-\frac{1}{\psi}\right)A^{(n)}+\frac{1}{\psi}}\right]=\\&=\lim_{\psi\to\infty}\frac{1}{\psi}\left[1+\frac{1}{N}\sum_{n=1}^{N}\left(1+0\Big(\frac{1}{\psi}\Big)\right)\right]=\lim_{\psi\to\infty}\left(\frac{2}{\psi}+0\Big(\frac{1}{\psi}\Big)\right),\end{aligned}$$

from which we obtain

$$\lim_{\psi\to\infty} h_1(\psi;\boldsymbol{A}) = \lim_{\psi\to\infty}\left(\frac{2}{\psi}+0\Big(\frac{1}{\psi}\Big)\right) = h_1(+\infty;\boldsymbol{A}) = 0. \tag{6.38}$$

That is, as $\psi\to+\infty$, the function $h_1(\psi;\boldsymbol{A})$ tends to zero, being asymptotically equivalent to the function $\dfrac{2}{\psi}$.

The asymptotic behaviour of $h_3(\psi;\boldsymbol{A},\boldsymbol{B})$ is found similarly:

$$\lim_{\psi\to\infty} h_3(\psi;\boldsymbol{A},\boldsymbol{B}) = \lim_{\psi\to\infty}\frac{1}{\psi}\left[3\frac{1}{N}\sum_{n=1}^{N}\frac{\Big(1-\dfrac{1}{\psi}\Big)B^{(n)}+\dfrac{A^{(n)}}{\psi}}{\Big(1-\dfrac{1}{\psi}\Big)^2 B^{(n)}+2\Big(\dfrac{1}{\psi}-\dfrac{1}{\psi^2}\Big)A^{(n)}+\dfrac{1}{\psi^2}}\right] =$$

$$= \lim_{\psi\to\infty}\frac{3}{\psi}\left[\frac{1}{N}\sum_{n=1}^{N}\left(1+0\Big(\frac{1}{\psi}\Big)\right)\right] = \lim_{\psi\to\infty}\left(\frac{3}{\psi}+0\Big(\frac{1}{\psi^2}\Big)\right),$$

from which we obtain

$$\lim_{\psi\to\infty} h_3(\psi;\boldsymbol{A},\boldsymbol{B}) = \lim_{\psi\to\infty}\left(\frac{3}{\psi}+0\Big(\frac{1}{\psi^2}\Big)\right) = h_3(+\infty;\boldsymbol{A},\boldsymbol{B}) = 0. \tag{6.39}$$

That is, as $\psi\to+\infty$, the function $h_3(\psi;\boldsymbol{A},\boldsymbol{B})$ tends to zero, being asymptotically equivalent to the function $\dfrac{3}{\psi}$.

On the basis of the limiting results (6.38) and (6.39), we can easily deduce the following simple, but important, lemma:

Lemma 6.1 *For any two* $[0,1]$*-valued samples* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, and any* $\alpha>0$*, the following inequality holds true:*

$$0 < h_1(\psi;\boldsymbol{A}) < h_3(\psi;\boldsymbol{A},\boldsymbol{B}) < \frac{3+\alpha}{\psi}, \quad \textit{for all} \quad \psi > \psi_* = \psi_*(\alpha,\boldsymbol{A},\boldsymbol{B}), \tag{6.40}$$

where $\psi_*=\psi_*(\alpha,\boldsymbol{A},\boldsymbol{B})$ *is a sufficiently large positive number, dependent on* α *and the two samples* $\boldsymbol{A}$*,* $\boldsymbol{B}$*.* ■

We are now proceeding to the calculation of the first and second derivatives of the functions $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$:

$$\frac{\partial h_1(\psi; \boldsymbol{A})}{\partial \psi} = -\left[\frac{1}{\psi^2} + \frac{1}{N}\sum_{n=1}^{N} \frac{\left[A^{(n)}\right]^2}{\left[(\psi-1)A^{(n)}+1\right]^2}\right] \leq 0, \quad \text{for all} \quad \psi > 0, \tag{6.41}$$

$$\frac{\partial h_3(\psi; \boldsymbol{A}, \boldsymbol{B})}{\partial \psi} = -3\frac{1}{N}\sum_{n=1}^{N} \frac{\left[(\psi-1)B^{(n)}+A^{(n)}\right]^2 + \left[A^{(n)}\right]^2 - B^{(n)}}{\left[(\psi-1)^2 B^{(n)} + 2(\psi-1)A^{(n)}+1\right]^2} \leq 0,$$
$$\text{for all} \quad \psi > 0, \tag{6.42}$$

$$\frac{\partial^2 h_1(\psi; \boldsymbol{A})}{\partial \psi^2} = 2\left[\frac{1}{\psi^3} + \frac{1}{N}\sum_{n=1}^{N} \frac{\left[A^{(n)}\right]^3}{\left[(\psi-1)A^{(n)}+1\right]^3}\right] \geq 0, \quad \text{for all} \quad \psi > 0, \tag{6.43}$$

$$\frac{\partial^2 h_3(\psi; \boldsymbol{A}, \boldsymbol{B})}{\partial \psi^2} =$$
$$= 6\frac{1}{N}\sum_{n=1}^{N} \frac{\left[(\psi-1)B^{(n)}+A^{(n)}\right]^3 + 3\left[(\psi-1)B^{(n)}+A^{(n)}\right]\left(\left[A^{(n)}\right]^2 - B^{(n)}\right)}{\left[(\psi-1)^2 B^{(n)} + 2(\psi-1)A^{(n)}+1\right]^3} \geq 0,$$
$$\text{for all} \quad \psi > 0. \tag{6.44}$$

All inequalities, which are included in relations (6.41) to (6.44) above, are proved in Section 6.A.

From the limiting behaviour of $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ as $\psi \to 0$ and $\psi \to +\infty$, the following conclusions are drawn:

(i) As $\psi \to 0$, for every positive, non-degenerate bivariate sample $(\boldsymbol{X}_1, \boldsymbol{X}_2)$:

$$h_1(\psi; \boldsymbol{A}) > h_3(\psi; \boldsymbol{A}, \boldsymbol{B}), \tag{6.45}$$

(ii) As $\psi \to +\infty$, for every sample $(\boldsymbol{X}_1, \boldsymbol{X}_2)$:

$$h_1(\psi; \boldsymbol{A}) < h_3(\psi; \boldsymbol{A}, \boldsymbol{B}), \tag{6.46}$$

Inequalities (6.45) and (6.46), in combination with the fact that the two functions $h_1(\psi; \boldsymbol{A})$, $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ are continuous, lead to the following important result:

"For every positive, non-degenerate bivariate sample $(\boldsymbol{X}_1, \boldsymbol{X}_2)$,
there exists at least one solution $\hat{\psi} \in (0, \infty)$ of equation (6.29)."

Numerical experience shows that this solution is also unique. In Figure 6.1a, plots of $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ are shown, which are typical for the Mediterranean Sea data. The unique solution of equ. (6.29) and its maximum likelihood character are evident in Figures 6.1a,b,c.

Let us now focus on the practical problem of finding an efficient algorithm for the numerical solution of equ. (6.29). Although this can be safely performed by, e.g., a simple bisection method, the procedure may become heavily time consuming if the sample is large (N=some or many thousands) and, at the same time, the functions $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ should be evaluated for a great number of different values of ψ. However, the simple asymptotic forms of $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ as $\psi \to +\infty$, in conjunction with the simple geometric shape of the graph of these functions, suggests that the corresponding graphs in a log-log scale will be nearly linear. And this is indeed the case, as it can be seen in Figure 6.2. Accordingly, in the log-log domain, we can approximate the two graphs by simple parabolas, obtaining an excellent agreement over the whole range of interest of ψ (i.e. in the range $1 < \psi < 500$, where all solutions happen to lie). Numerical experience suggests that our method of solution ensures an accuracy for ψ of the order of 0.1. In Figure 6.3, two contour plots are shown for two Plackett models with the same marginals and values of ψ differing by 1 (ψ=11 and ψ=12). The closeness of the two contour plots shows that the accuracy of ± 0.1 for ψ parameter is very satisfactory.

In this way, to solve equ. (6.29), we need to calculate $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ only for three different values of ψ, which is what we need in order to calculate the parameters of the two parabolas. In Section 6.6, the above procedure for the numerical solution of equ. (6.29) is algorithmically formulated.

6.6 Implementation of parameter estimation procedure

In this section, the whole procedure for the estimation of the parameters of the general Plackett model (6.16) is presented in algorithmic form:

1. Consider that a bivariate sample is given

$$(\boldsymbol{X}_1, \boldsymbol{X}_2) = \Big((X_1^{(1)}, X_2^{(1)}), (X_1^{(2)}, X_2^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)}), \ldots, (X_1^{(N)}, X_2^{(N)}) \Big),$$

of large size N from a *continuously distributed* population.

2. Consider two univariate probability models $F_\alpha(x_\alpha;\boldsymbol{\theta}_\alpha)$, α=1,2, corresponding to the two univariate samples $\boldsymbol{X}_1$, $\boldsymbol{X}_2$.

3. Find the ML estimators $\hat{\boldsymbol{\theta}}_\alpha$ of the parameter vectors $\boldsymbol{\theta}_\alpha$, using univariate samples $\boldsymbol{X}_\alpha$ (as well as any additional univariate data).

4. Compute the derived samples $\boldsymbol{A}$ and $\boldsymbol{B}$ by using equs. (6.30) and (6.31), where $Y_1^{(n)}$ and $Y_2^{(n)}$ are defined by

$$Y_1^{(n)} = F_1(X_1^{(n)};\hat{\boldsymbol{\theta}}_1) \quad \text{and} \quad Y_2^{(n)} = F_2(X_2^{(n)};\hat{\boldsymbol{\theta}}_2), \qquad n = 1, 2, \ldots, N.$$

5. Solve equ. (6.29) and obtain the ML estimate $\hat{\psi}$ of the association parameter ψ.

Step 5 can be performed as follows (see Section 6.5 for details):

a) Select three distinct points $\psi_1<\psi_2<\psi_3$, such that the interval $[\psi_1, \psi_3]$ to approximately cover the range of interest of ψ. (In the present work we have taken ψ_1=5, ψ_2=50, ψ_3=500).

b) By means of the three selected points ψ_1, ψ_2, ψ_3, define the two set of points

$$\big\{(x_i, y_{1,i}),\ \ i = 1, 2, 3\big\} \quad \text{and} \quad \big\{(x_i, y_{3,i}),\ \ i = 1, 2, 3\big\} \tag{6.47}$$

in the log-log plane, where

$$x_i = \log_{10}\psi_i, \quad y_{1,i} = \log_{10}\big[h_1(\psi_i;\boldsymbol{A})\big], \quad y_{3,i} = \log_{10}\big[h_3(\psi_i;\boldsymbol{A},\boldsymbol{B})\big], \quad i = 1, 2, 3. \tag{6.48}$$

c) Calculate the coefficients of the parabolas

$$P_1(x) = a_{11}x^2 + a_{12}x + a_{13}, \quad \text{and} \quad P_3(x) = a_{31}x^2 + a_{32}x + a_{33},$$

that pass from the corresponding points (6.47) in the log-log plane. These two parabolas are approximants of the two functions $\log_{10}\big[h_1(\psi_i;\boldsymbol{A})\big]$, $\log_{10}\big[h_3(\psi_i;\boldsymbol{A},\boldsymbol{B})\big]$, respectively.

d) Find the solution x=$\hat{x}$ of the equation

$$P_1(x) = P_3(x), \tag{6.49}$$

which is an approximate version of equ. (6.29). Equ. (6.49) is quadratic and can be solved in closed form. Numerical experience shows that only one of the two solutions is acceptable.

e) Inverse the logarithmic transformation (6.48) to obtain the solution $\hat{\psi}$ of equ. (6.29) in terms of the solution $\hat{x}$ of equation (6.49):

$$\hat{\psi} = 10^{\hat{x}}. \tag{6.50}$$

6.7 Second-order statistics of wave-parameter time series

In this section, the first- and second-order statistics of many-year long time series of significant wave height H_S is examined and appropriate analytic probability laws are fitted to them. The procedure is quite general, and can be applied to any other wave (or environmental) parameter having the same behaviour as H_S. Especially, for the modelling of the second-order probability density, the Plackett model, presented in the previous sections, is used.

Two different kind of data are examined: (i) hindcast data, (ii) *in situ* measurements. For the first kind of data, results of the well-known third-generation wave model WAM are used, as this is implemented at European Centre for Medium-range Weather Forecasts (ECMWF) in Reading, England. In the present work, an eight-year long dataset is analyzed from a site with the code name ATL_15, located off the Spanish Coasts in the Atlantic Ocean (45^{o}N, 12^{o}W). For the second kind of data, the six-year long dataset from the Portuguese site Figueira da Foz (see Figure 5.14) is examined. See also Section 5.7, where this dataset is utilised to fill in its missing values. For more details on data used in the present work, see also Appendix A.

Let us now denote by $X(\tau)$ the long-term time series of significant wave height. Performing the time series analysis presented in Chapter 4, the residual time series $W(\tau)$ is obtained; see Section 4.3. As we have seen in Chapter 4, $W(\tau)$ is modelled as a stationary (in general, non-Gaussian) stochastic process; see also Figures 6.6 and 6.7, where the autocorrelation coefficient function of $W(\tau)$ (defined by equ. (4.20)) is depicted. In these figures, the stationary character of the process is more than apparent. It is well known (Yaglom, 1987a; Papoulis, 1991) that, the first-order probability density of a stationary stochastic process is independent of time τ, and it is uniquely defined by the density $f_\tau^W(w) \equiv f^W(w)$. Similarly, its second-order probability density function depends only on the time differences $(\tau+r\Delta\tau)-\tau$, and it is written as $f^{WW}_{\tau,\tau+r\Delta\tau}(w_1, w_2)$, where r=1,2, . . . , and $\Delta\tau$ is the sampling interval.

In Figures 6.1 and 6.2, the calculated first-order empirical probability density function of $W(\tau)$ for the two datasets is shown. Then, the following three-parameter lognormal density (Johnson and Kotz, 1970, Chapter 14) is fitted to it,

$$f_\tau^W(w) \equiv f^W(w;\lambda,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}(w-\lambda)} \exp\left\{-\frac{\left[\ln(w-\lambda)-\mu\right]^2}{2\,\sigma^2}\right\}, \qquad (6.51)$$

$$\lambda < w < +\infty, \quad \sigma > 0.$$

The parameters are estimated by means of the maximum likelihood method; see, e.g., Stedinger (1980), Cohen (1988). The fit is excellent to the hindcast data, and very good to the measured ones; see Figures 6.1 and 6.2.

Let us now proceed to the study of the second-order probability structure of $W(\tau)$. For this purpose, the Plackett model (6.19) is implemented using for both univariate marginals the log-

normal density $f_\tau^W(w)$, as this has been estimated by means of the first-order empirical probability density. The parameters of Plackett model are estimated using the maximium likelihood method, implementing the procedure described in Sections 6.4–6.6. As it has been stated earlier in this section, the second-order ptobability density $f^{WW}_{\tau,\tau+r\Delta\tau}(w_1,w_2)$ depends only on time differences $(\tau+r\Delta\tau)-\tau=r\Delta\tau$, r=1,2, ... Thus, in principle, the density $f^{WW}_{\tau,\tau+r\Delta\tau}(w_1,w_2)$ should be calculated for all possible values of r. Fortunately, we will see that this is not necessary for our datasets.

Let us examine the behaviour of association parameter ψ of the Plackett model as r increases. Recall that, parameter ψ is another measure of the correlation between $W(\tau)$ and $W(\tau+r\Delta\tau)$. In Figures 6.8a,b, ψ is plotted for r=1(1)1000. For the first few lags, where the correlation is very strong, ψ takes very large values. After these lags, ψ descreases rapidly to a small, almost constant, value very close to the value ψ=1, where the two variables are uncorrelated. In view of this result, density $f^{WW}_{\tau,\tau+r\Delta\tau}(w_1,w_2)$ is calculated for some indicative lags, namely for r=1, 2, 3, 10, 50, 100. The first three of these have been chosen to better describe the region of strong correlation, whereas the last three of them as examples of the (gradual) independence of the process. Apart from the Plackett model, and for comparison purposes, the bivariate lognormal density is also fitted to the empirical probability density function. In Figures 6.9 and 6.10, contour plots of the two fitted models are shown for the hindcast dataset. In these figures, contours of the empirical probability density function are also depicted. In Figures 6.11 and 6.12, the same results are depicted for the measured dataset. Both analytic densities (Plackett and lognormal) model well the second-order density of the process for large values of r, where the two random variables $W(\tau)$ and $W(\tau+r\Delta\tau)$ become uncorrelated and the density $f^{WW}_{\tau,\tau+r\Delta\tau}(w_1,w_2)$ is written as the product of the two univariate ones $f_\tau^W(w_1)$ and $f^W_{\tau+r\Delta\tau}(w_2)$. However, Plackett model fits better the second-order density of the process for small values of r; compare, e.g., Figure 6.11a with Figure 6.12a. Thus, it seems that Plackett model with lognormal marginals is more appropriate for the representation of the second-order structure of the stationary part $W(\tau)$ of time series $X(\tau)$.

Let us now proceed to the calculation of the first- and second-order probability density function of $X(\tau)$. For every time instance τ, $X(\tau)$ can be thought of as a transformation of $W(\tau)$, cf. equ. (4.14),

$$X(\tau) = G(\tau) + \sigma(\tau)\, W(\tau) = h\big(W(\tau)\big), \quad \text{for every } \tau, \tag{6.52}$$

where $G(\tau)$=$\overline{X}_{\rm tr}(\tau)+\mu(\tau)$ and $\sigma(\tau)$ are time-dependent deterministic functions, defined in Sections 4.1–4.3.

Then, the first-order density of $X(\tau)$ is obtained by means of the first-order density of $W(\tau)$ (Larson and Shubert, 1979; Ochi, 1990)

$$f_\tau^X(x) = \frac{1}{\sigma(\tau)} f_\tau^W\left(\frac{x-\overline{X}_{\rm tr}(\tau)-\mu(\tau)}{\sigma(\tau)}\right), \quad \tau = 1,2,\ldots . \tag{6.53}$$

Thus, due to the transformation (6.52), the first-order probability density function of $X(\tau)$ is time-dependent; see also Figures 6.13 and 6.14, where this density is depicted for various time instances during the annual time. Note that, this pattern of densities is repeated for each yearly segment, a fact that justifies the cyclostationary modelling proposed in Chapter 3.

Working similarly, the second-order probability density function of $X(\tau)$ is calculated for various lags r.

Let us first define the inverse of the transformation (6.52)

$$W(\tau) = h^{-1}\big(X(\tau)\big) = \frac{X(\tau) - \overline{X}_{\mathrm{tr}}(\tau) - \mu(\tau)}{\sigma(\tau)}, \quad \text{for every } \tau. \tag{6.54}$$

Then, the second-order density of $X(\tau)$ is defined by means of the second-order density of $W(\tau)$, as follows

$$f^{XX}_{\tau,\tau+r\Delta\tau}(x_1, x_2) = |J| \;\; f^{WW}_{\tau,\tau+r\Delta\tau}\Big(h^{-1}\big(x_1\big), h^{-1}\big(x_2\big)\Big), \tag{6.55}$$

where $|J|$ is the Jacobian of the transformation, defined by

$$|J| = \begin{vmatrix} \dfrac{\partial h^{-1}\big(x_1\big)}{\partial x_1} & 0 \\ 0 & \dfrac{\partial h^{-1}\big(x_2\big)}{\partial x_2} \end{vmatrix} = \frac{1}{\sigma(\tau)\sigma(\tau+r\Delta\tau)}. \tag{6.56}$$

Note that, density $f^{XX}_{\tau,\tau+r\Delta\tau}(x_1, x_2)$ is dependent on time τ and on the time differences $(\tau+r\Delta\tau)-\tau$. In Figures 6.15 and 6.16, the second-order density of $X(\tau)$ is depicted for the hindcast and measured datasets, respectively. The results are given for r=1 and twelve values of τ corresponding to the 0h on the 1st of each calendar month. As in the first-order density, the seasonal variation of the density is apparent.

Finally, it is important to say that, in principle, the m-th order density of $X(\tau)$ can be calculated, provided that the corresponding m-th order density of $W(\tau)$ is available. Then, the m-th order density is given by (Larson and Shubert, 1979)

$$f^{XX\ldots X}_{\tau,\tau+r_1\Delta\tau\ldots,\tau+r_{m-1}\Delta\tau}(x_1, x_2, \ldots, x_m) = |J_m| \; f^{WW\ldots W}_{\tau,\tau+r_1\Delta\tau\ldots,\tau+r_{m-1}\Delta\tau}\Big(h^{-1}\big(x_1\big), h^{-1}\big(x_2\big), \ldots, h^{-1}\big(x_m\big)\Big), \tag{6.57}$$

where

$$|J_m| = \left|\prod_{\ell=1}^{m} \frac{\partial h^{-1}\big(x_\ell\big)}{\partial x_\ell}\right| = \prod_{\ell=0}^{m-1} \frac{1}{\sigma(\tau+r_\ell\Delta\tau)}, \tag{6.58}$$

with r_0=0.

6.8 Conclusions

In this chapter, the first- and second-order probability density functions of nonstationary time series of significant wave height are produced. Two datasets are used for this study: one with hindcast data, and one with *in situ* measurements.

The key-idea is to take advantage of the modelling presented in Chapter 4, grace to which the initial nonstationary series is written in the form $X(\tau)$=$G(\tau)+\sigma(\tau)W(\tau)$, where $W(\tau)$ is a zero-mean stationary (in general non-Gaussian) stochastic process, and $G(\tau)$, $\sigma(\tau)$ are appropriate deterministic functions.

Then, the first- and second-order densities of $W(\tau)$ are calculated. The first-order density is modelled by fitting the three-parameter lognormal to the first-order empirical density. Further, an analytic (non-Gaussian) probability model, the so-called Plackett model, is used for the representation of the second-order density. This analytic form is a versatile tool, which can incorporate any form of univariate marginals and the correlation structure is controlled by a reasonable number of parameters. For comparison purposes, the standard bivariate lognormal density is also examined. The Plackett model seems to offer a better performance regarding the description of the initial data, and an increased flexibility regarding the shape of the distribution, at the same computational cost.

In the sequel, the first- and second-order densities of $X(\tau)$ are easily derived, by simply considering the relation between $X(\tau)$ and $W(\tau)$ as a linear time-varying transformation. In Figures 6.13 and 6.14, where the first-order, and in Figures 6.15 and 6.16, where the second-order densities of $X(\tau)$ are depicted, one can easily observe their seasonal variability.

Concluding, we would like to stress that the applied probability modelling (i) is fairly general, (ii) is non-Gaussian, (iii) can incorporate information acquired in previous stages (e.g., univariate marginals), and (iv) is considered in ocean engineering applications for the first time.

6.9 Tables and Figures

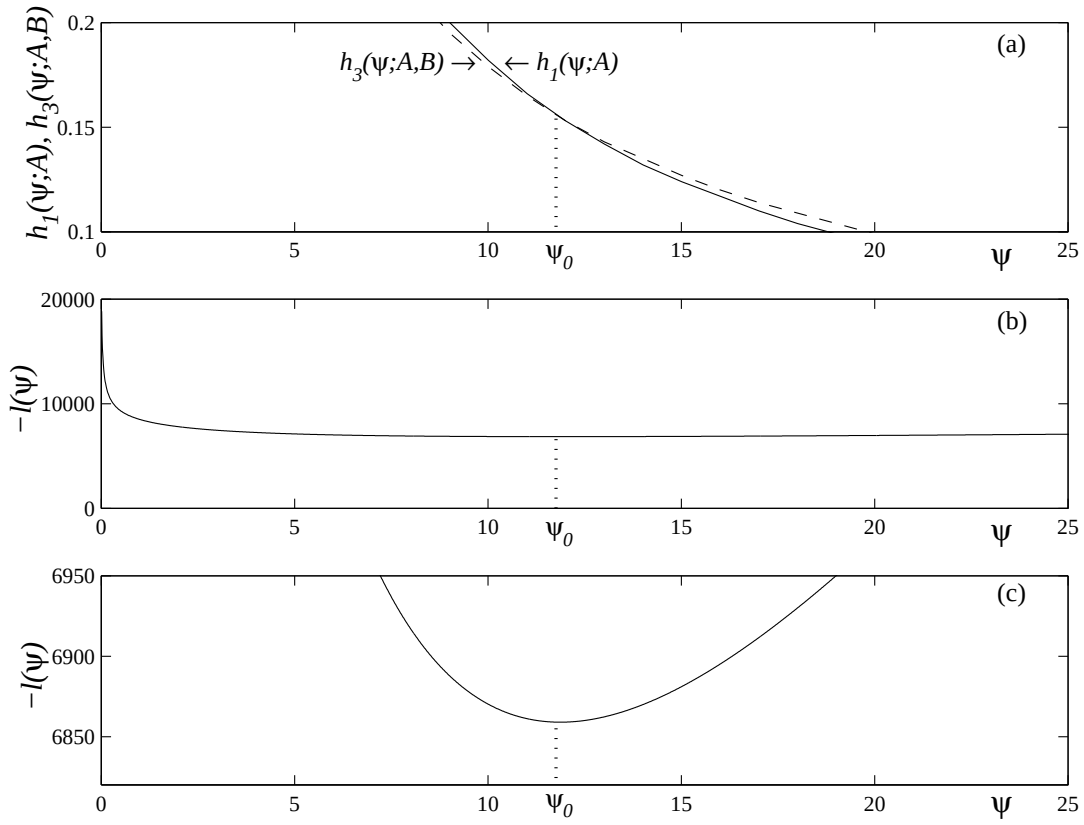


Figure 6.1: (a) Plots of the functions $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ for a typical sample from Mediterranean wave data. (b) Plot of the minus log-likelihood function in the same range of ψ-values. (c) A vertical-axis rescaled version of (b), helping to locate the value of ψ minimising the minus log-likelihood function.

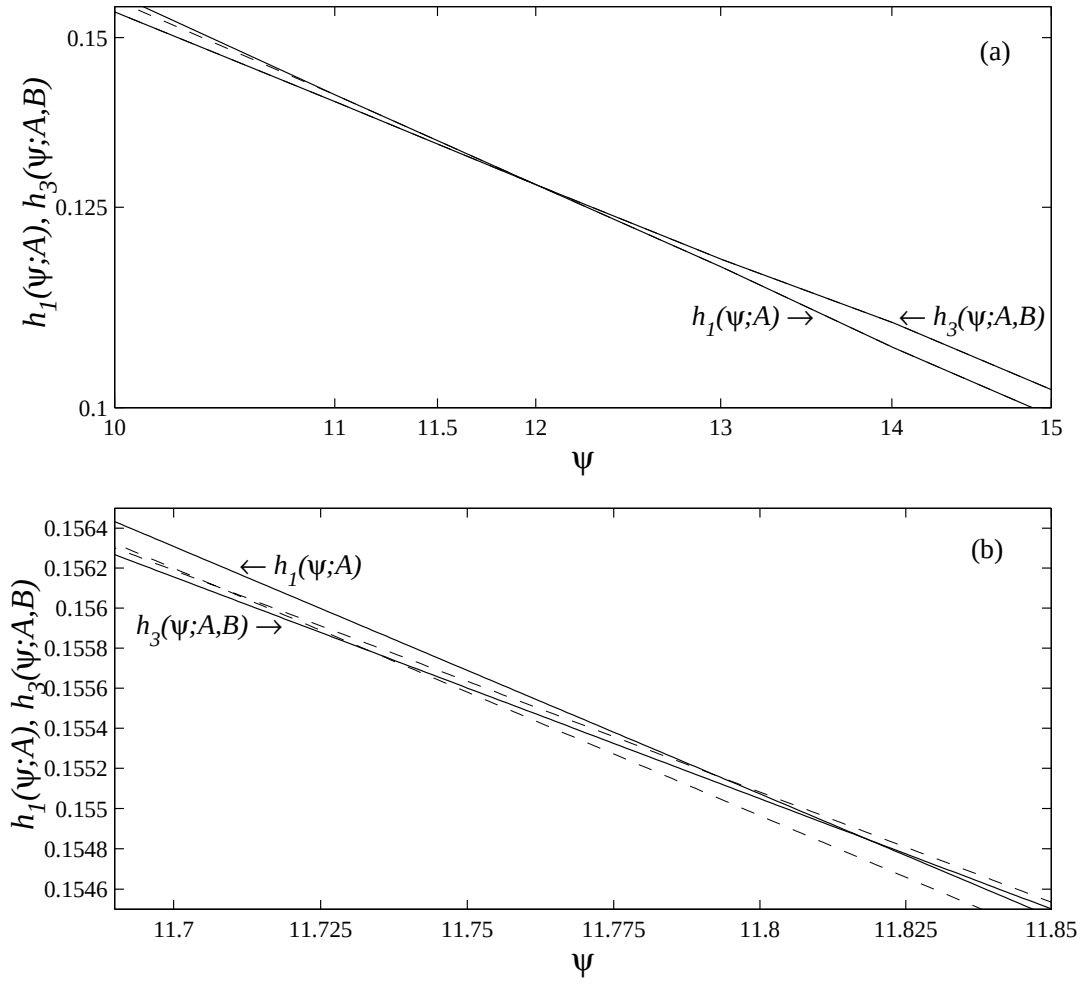


Figure 6.2: Plots of the functions $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ (solid lines) and their parabolic approximations (dashed lines) in the log-log plane. Figure (b) is a rescaling of (a) in the neighborhood of solution of equation (7.2). From (b) it is seen that the accuracy of the approximate solution is ± 0.1.

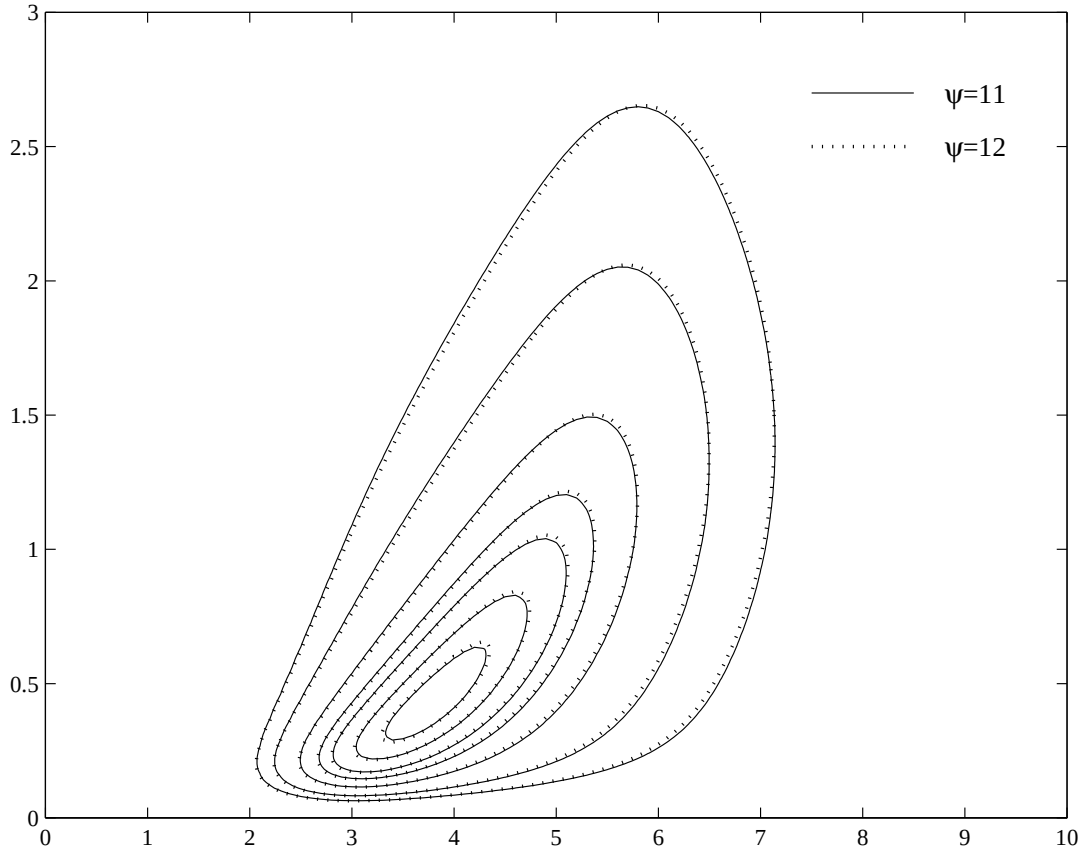


Figure 6.3: Contour plots of two bivariate Plackett models with the same univariate marginals and different values of ψ (solid line: ψ=11, and dotted line: ψ=12).

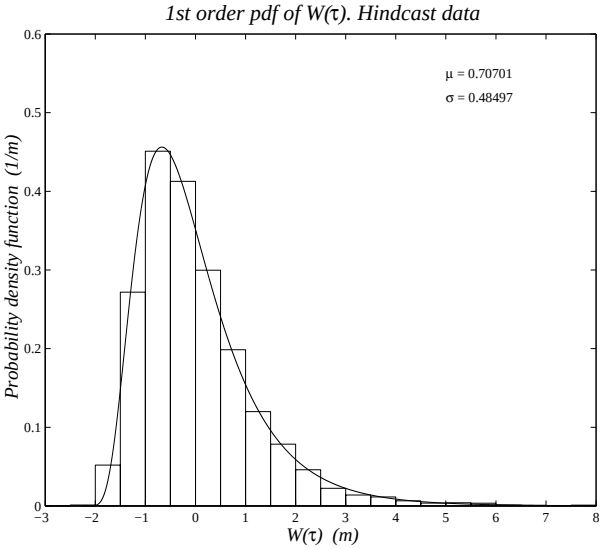


Figure 6.4: First-order probability density function of $W(\tau)$. Analytic model: lognormal. Site: ATL_15, Charcot. Data: Hindcast.

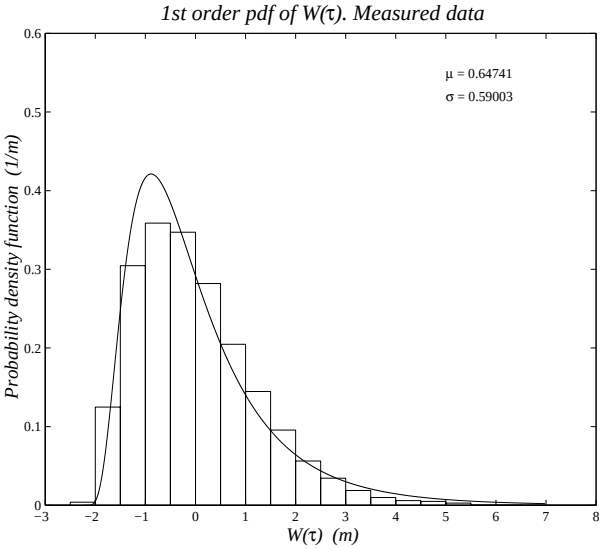


Figure 6.5: First-order probability density function of $W(\tau)$. Analytic model: lognormal. Site: Figueira da Foz. Data: Measured.

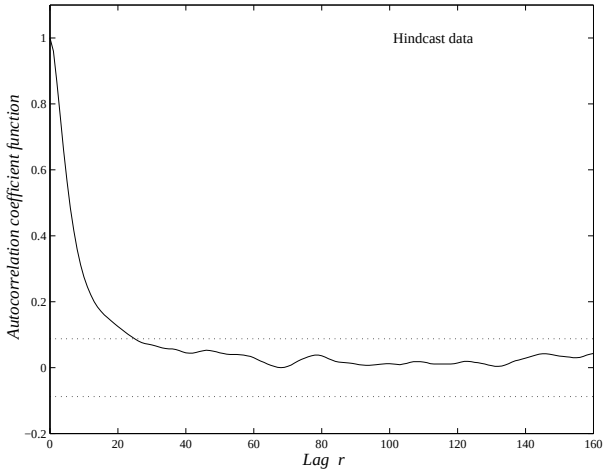


Figure 6.6: Autocorrelation coefficient function of $W(\tau)$ along with the 95% confidence limits. Site: ATL_15, Charcot. Data: Hindcast.

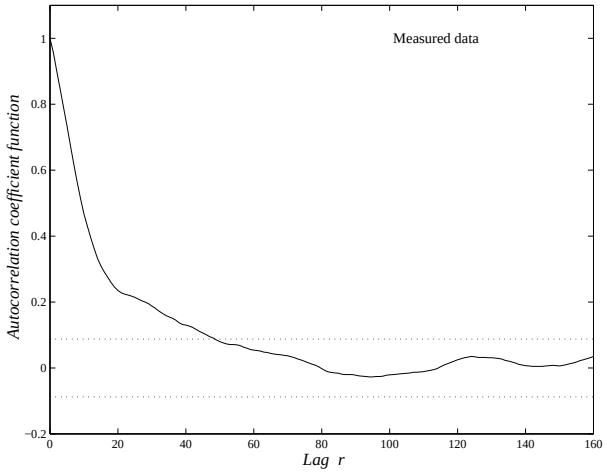


Figure 6.7: Autocorrelation coefficient function of $W(\tau)$ along with the 95% confidence limits. Site: Figueira da Foz. Data: Measured.

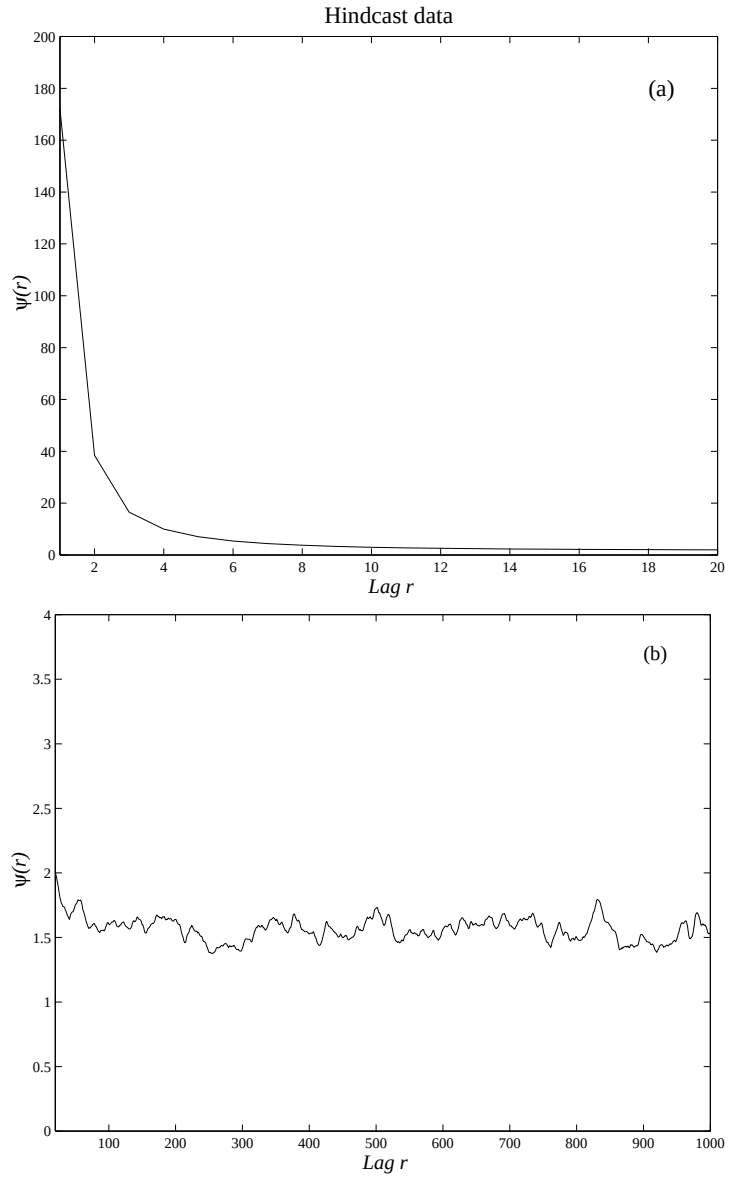


Figure 6.8: Association parameter ψ for various lags r. Site: ATL_15, Charcot. Data: Hindcast.

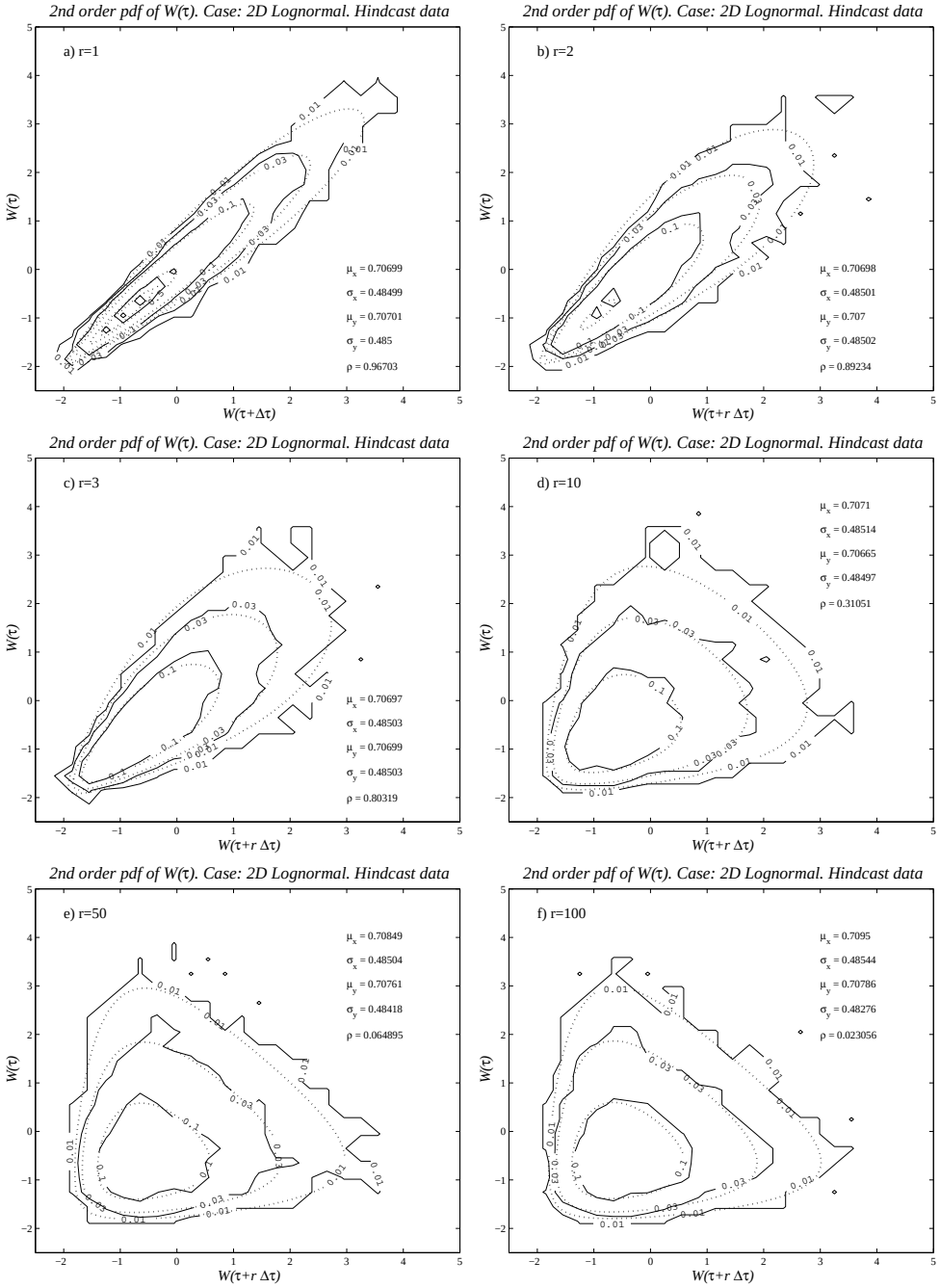


Figure 6.9: Contour plots of the second-order probability density function of $W(\tau)$ for lags r=1, 2, 3, 10, 50, 100. Continuous line: empirical pdf, Dotted line: bivariate lognormal. Contour levels: 0.01, 0.03, 0.10, 0.50, 0.75. Site: ATL_15, Charcot. Data: Hindcast.

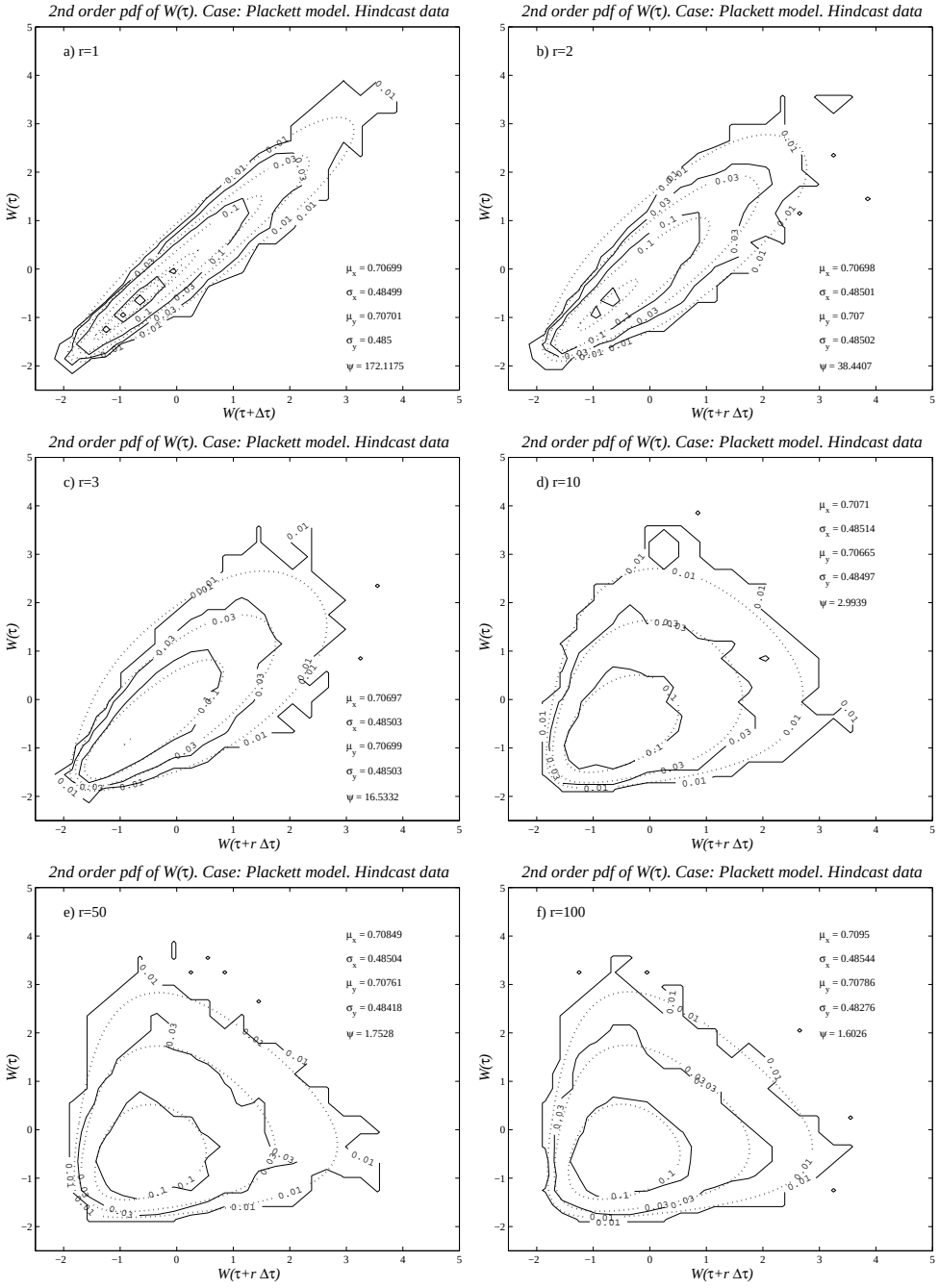


Figure 6.10: Contour plots of the second-order probability density function of $W(\tau)$ for lags r=1, 2, 3, 10, 50, 100. Continuous line: empirical pdf, Dotted line: Plackett model. Contour levels: 0.01, 0.03, 0.10, 0.50, 0.75. Site: ATL_15, Charcot. Data: Hindcast.

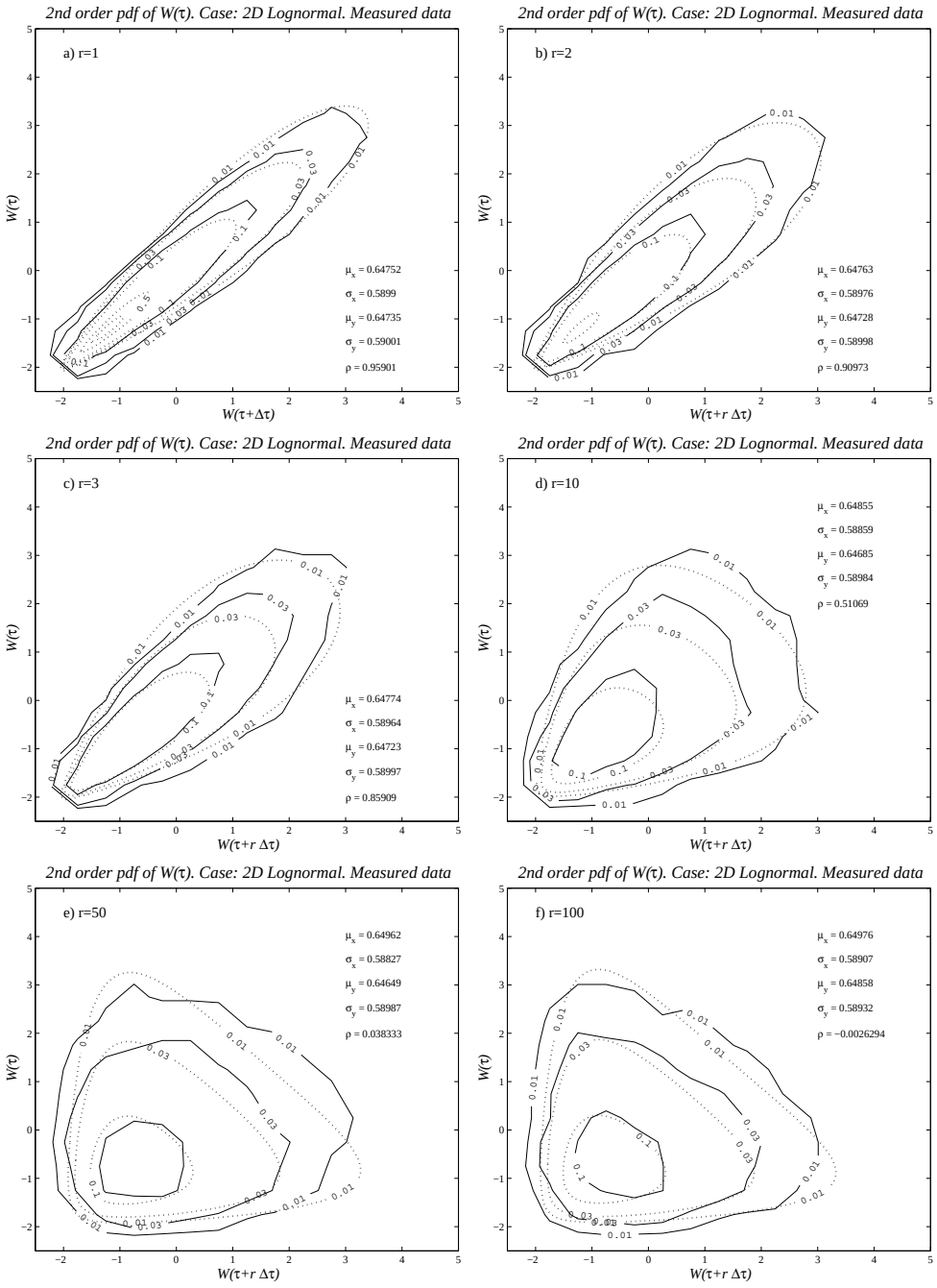


Figure 6.11: Contour plots of the second-order probability density function of $W(\tau)$ for lags r=1, 2, 3, 10, 50, 100. Continuous line: empirical pdf, Dotted line: bivariate lognormal. Contour levels: 0.01, 0.03, 0.10, 0.50, 0.75. Site: Figueira da Foz. Data: Measured.

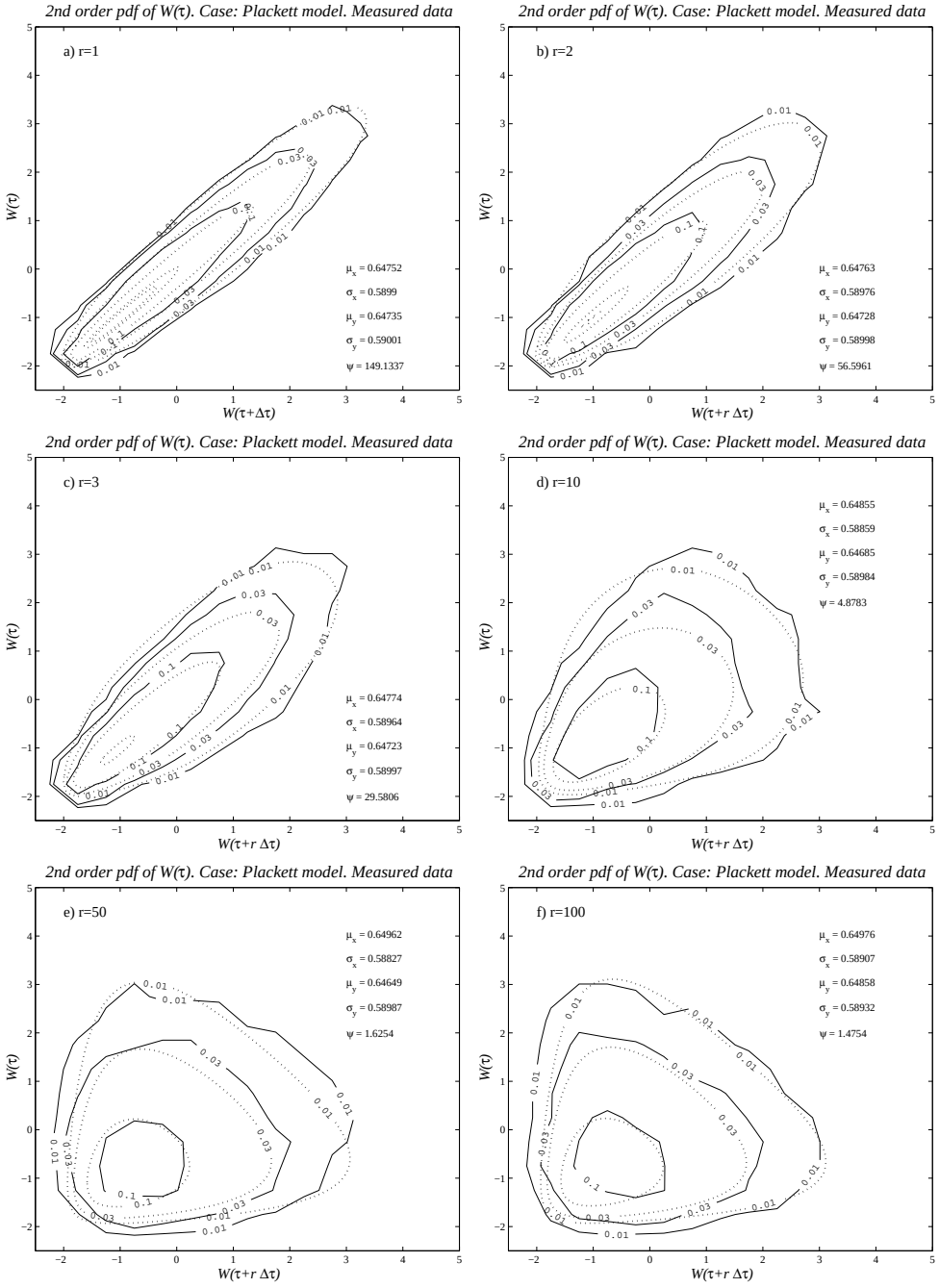


Figure 6.12: Contour plots of the second-order probability density function of $W(\tau)$ for lags r=1, 2, 3, 10, 50, 100. Continuous line: empirical pdf, Dotted line: Plackett model. Contour levels: 0.01, 0.03, 0.10, 0.50, 0.75. Site: Figueira da Foz. Data: Measured.

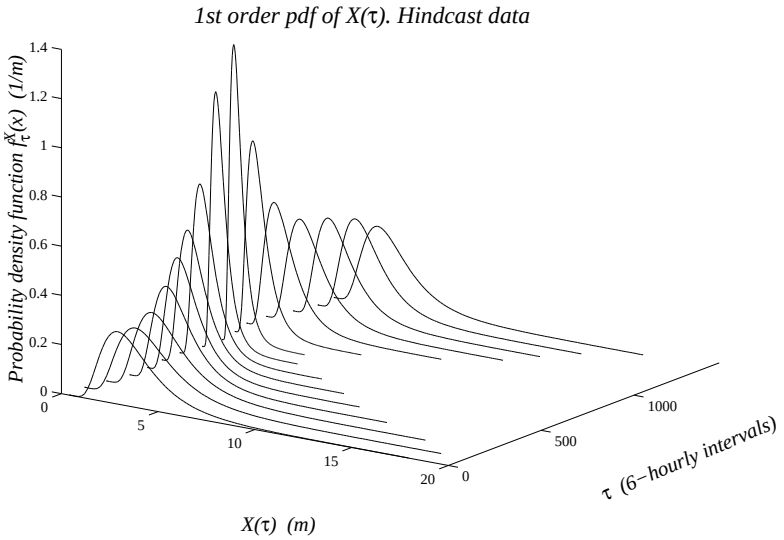


Figure 6.13: First-order probability density function of $X(\tau)$. Site: ATL_15, Charcot. Data: Hindcast.

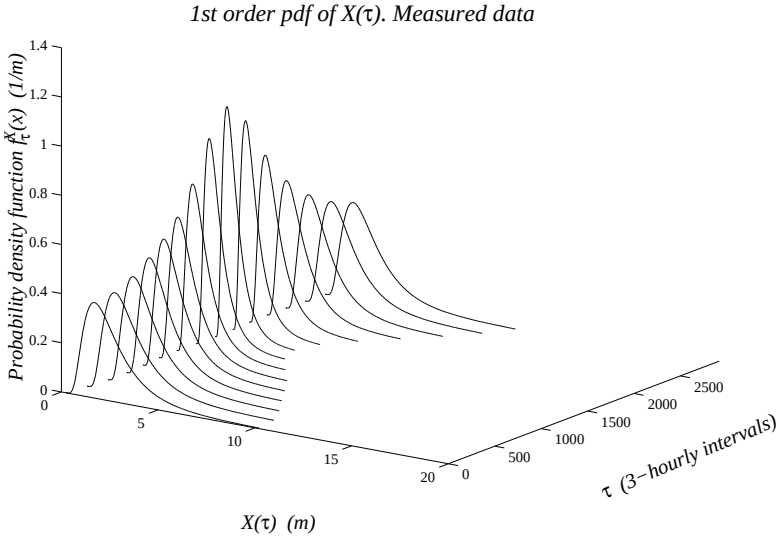


Figure 6.14: First-order probability density function of $X(\tau)$. Site: Figueira da Foz. Data: Measured.

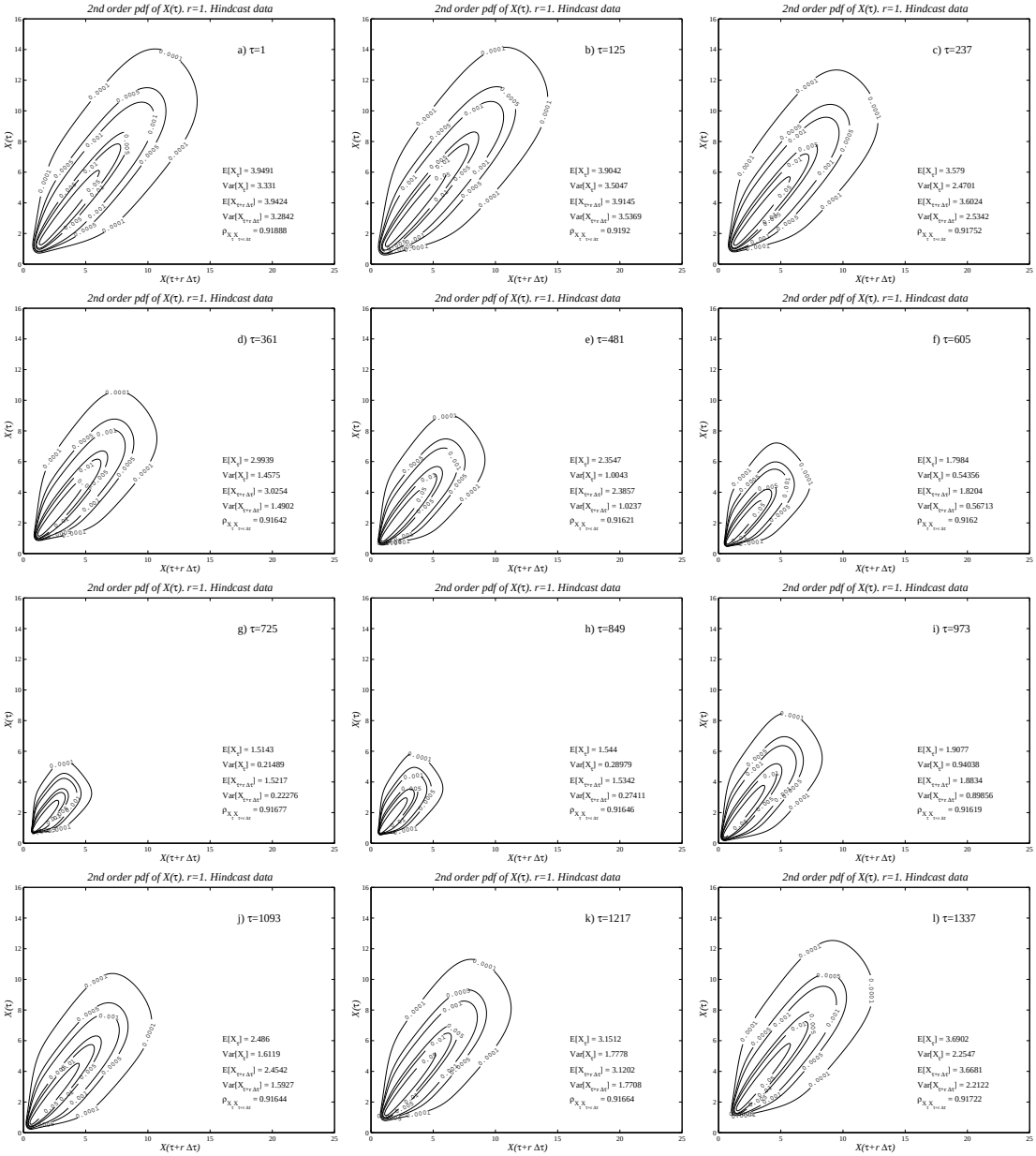


Figure 6.15: Second-order probability density function of $X(\tau)$ for lag r=1, for various τ-values within the year, as follows: a) τ=1, 1st January, b) τ=125, 1st February, c) τ=237, 1st March, d) τ=361, 1st April, e) τ=481, 1st May, f) τ=605, 1st June, g) τ=725, 1st July, h) τ=849, 1st August, i) τ=973, 1st September, j) τ=1093, 1st October, k) τ=1217, November, l) τ=1337, 1st December. Contour levels: 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05. Site: ATL_15, Charcot. Data: Hindcast.

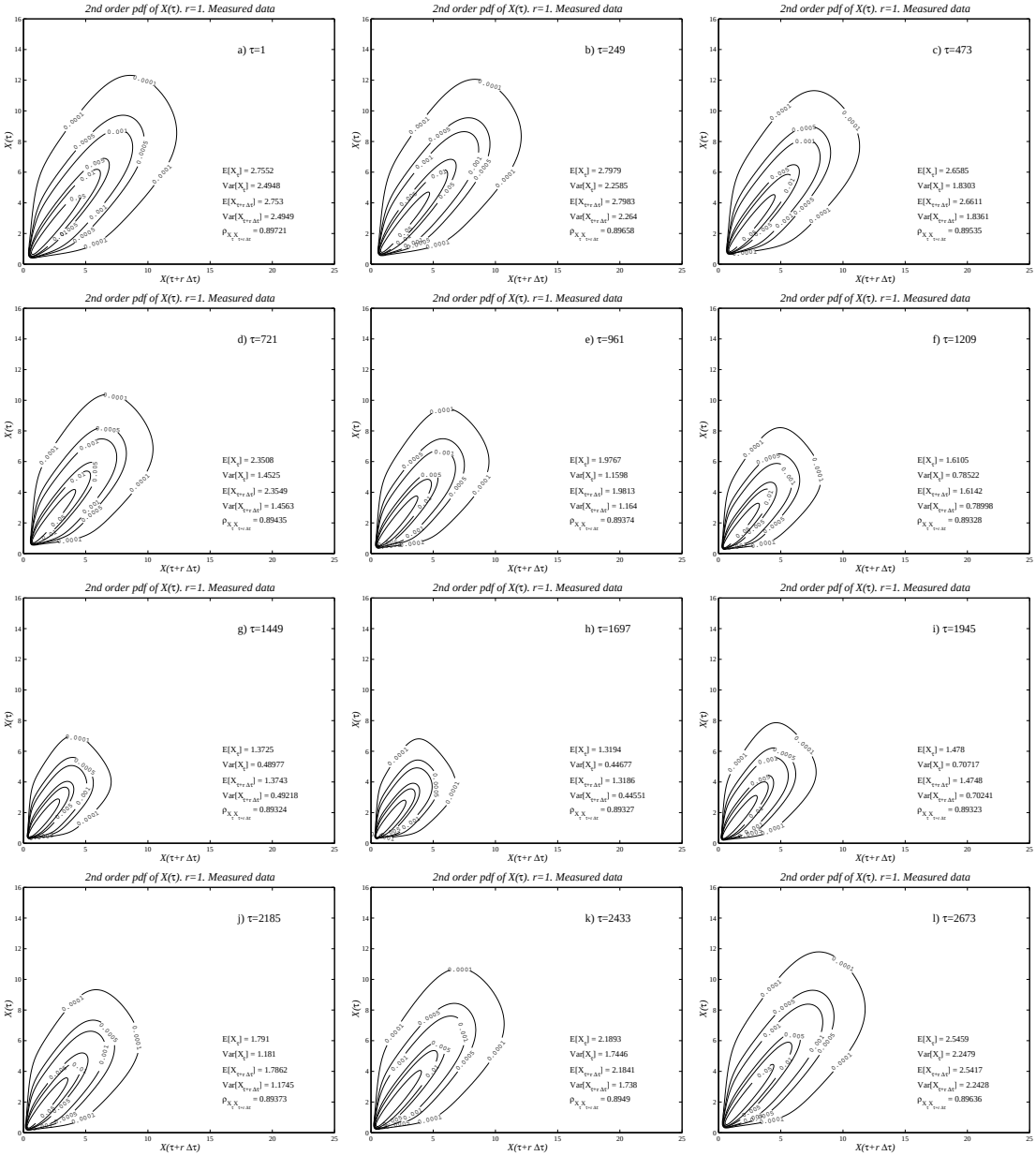


Figure 6.16: Second-order probability density function of $X(\tau)$ for lag r=1, for various τ-values within the year, as follows: a) τ=1, 1st January, b) τ=249, 1st February, c) τ=473, 1st March, d) τ=721, 1st April, e) τ=961, 1st May, f) τ=1209, 1st June, g) τ=1449, 1st July, h) τ=1697, 1st August, i) τ=1945, 1st September, j) τ=2185, 1st October, k) τ=2433, 1st November, l) τ=2673, 1st December. Contour levels: 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05. Site: Figueira da Foz. Data: Measured.

6.A Proofs of inequalities

In Section 6.5 a number of inequalities concerning the values and the first and second derivatives of functions $h_1(\psi; \boldsymbol{A})$ and $h_3(\psi; \boldsymbol{A}, \boldsymbol{B})$ have been given and used for proving various results. The proofs of these inequalities are presented in the this section.

Before starting presenting the proofs, we recall the following obvious inequalities for the elements of $[0, 1]$-valued samples $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$:

$$0 \leq \left[Y_\alpha^{(n)}\right]^2 \leq Y_\alpha^{(n)} \leq 1, \quad \alpha = 1, 2. \tag{6.59}$$

Everything which will be proved in this section is a simple (or a not-so-simple) consequence of the above inequalities.

Lemma 6.2 *The quantities $A^{(n)}$, $B^{(n)}$ defined by equs. (6.30) and (6.31), satisfy the inequalities*

$$0 \leq \left[B^{(n)}\right]^2 \leq A^{(n)}B^{(n)} \leq B^{(n)} \leq \left[A^{(n)}\right]^2 \leq A^{(n)} \leq 1. \tag{6.60}$$

■

Proof of Lemma 6.2
In the course of the proof of inequalities (6.60) and all other inequalities subsequently in this section, we shall simplify the notation, omitting the sample-index (n). That is, we shall write Y_1, Y_2, A, B, instead of $Y_1^{(n)}$, $Y_2^{(n)}$, $A^{(n)}$, $B^{(n)}$, respectively.

We first prove

$$0 \overset{(a)}{\leq} B^{(n)} \overset{(b)}{\leq} A^{(n)} \overset{(c)}{\leq} 1. \tag{6.61}$$

(a) is obvious from equation (6.31) of Section 6.5.

(b) is proved as follows:

$$A = Y_1 + Y_2 - 2Y_1Y_2 \geq Y_1^2 + Y_2^2 - 2Y_1Y_2 = B.$$

(c) is proved as follows:

From (6.59) it follows that $(1 - Y_1)(1 - Y_2) \geq 0$, which is written $Y_1 + Y_2 - Y_1Y_2 \leq 1$, and from the latter we get

$$A \equiv Y_1 + Y_2 - 2Y_1Y_2 \leq 1 - Y_1Y_2 \leq 1.$$

Thus, all inequalities (6.61) have been proved.

Now, we shall prove $B \leq A^2$:

$$A^2 - B = (Y_1 + Y_2 - 2Y_1Y_2)^2 - (Y_1 - Y_2)^2 = 4Y_1Y_2(1 - Y_1)(1 - Y_2) \geq 0,$$

which is the same as

$$B \leq A^2. \tag{6.62}$$

Combining (6.61) and (6.62), we easily obtain all inequalities (6.60). The proof of Lemma 6.2 is now completed. ■

We are now in a position to prove all inequalities appearing in Section 6.5. (We recall once again the simplification in the notation introduced at the beginning of the proof of the above lemma):

(i) Inequality (6.37)

$A-B>0$ follows immediately from (6.60), and $A-B=0$, if and only if (Y_1, Y_2) is one of the points (0,0) and (1,1).

Moreover, $B - 2A + 1 \geq 0$, because

$$B - 2A + 1 = (Y_1 - Y_2)^2 - 2(Y_1 + Y_2 - 2Y_1Y_2) + 1 = Y_1^2 + Y_2^2 - 2Y_1Y_2 - 2(Y_1 + Y_2) + 1 =$$
$$= (Y_1 + Y_2)^2 - 2(Y_1 + Y_2) + 1 = \Big[(Y_1 + Y_2) - 1\Big]^2 \geq 0. \tag{6.63}$$

Especially, $B - 2A + 1=0$, if and only if $Y_1+Y_2=1$, i.e. lies on the second diagonal of the unit square.

Thus, $h_3(0; \boldsymbol{A}, \boldsymbol{B})$ is always non-negative. Moreover, it may be zero only for trivial samples, and becomes infinite if there are sample values (Y_1, Y_2) on the second diagonal $Y_1+Y_2=1$.

(ii) Inequality (6.41)

In this case, we have to study the denominator

$$D_1(\psi; A) = (\psi - 1)A + 1. \tag{6.64}$$

Clearly, if $\psi \geq 1$, then

$$D_1(\psi; A) > 0. \tag{6.65}$$

Moreover, inequality (6.65) remains valid for $0<\psi<1$, because, in this case, it is equivalent to $(1 - \psi)A<1$, which is always true.

(iii) Inequality (6.42)

We shall prove that the denominator

$$D_3(\psi; A, B) = (\psi - 1)^2 B + 2(\psi - 1)A + 1 = B\psi^2 + 2(A - B)\psi + (B - 2A + 1), \quad (6.66)$$

and all numerators appearing in relation (6.42) are positive.

Assume first that $B \neq 0$. Then the two roots ψ_1, ψ_2 of the trinomial (6.66) have product $\psi_1\psi_2 = (B - 2A + 1)/B \geq 0$ and sum $\psi_1 + \psi_2 = -2(A - B)/B < 0$. Thus, they are always non-positive. From this fact, we deduce at once that the sign of trinomial (6.66) does not change for all $\psi > 0$, which gives

$$D_3(\psi; A, B) \geq 0. \quad (6.67)$$

If $B = 0$, then

$$D_3(\psi; A, 0) = 2(\psi - 1)A + 1 = D_1(\psi; A) > 0,$$

as it has been already proved previously in *(ii)*.

Let us split the numerator in two terms:

$$N_3(\psi; A, B) = \Big[N_{31}(\psi; A, B)\Big]^2 + N_{32}(\psi; A, B), \quad (6.68)$$

where

$$N_{31}(\psi; A, B) = (\psi - 1)B + A, \quad \text{and} \quad N_{32}(\psi; A, B) = A^2 - B.$$

From (6.60), $N_{32}(\psi; A, B) \geq 0$. Further, for $\psi > 1$ $N_{31}(\psi; A, B) \geq 0$. For $0 < \psi < 1$, by appropriately treating inequalities (6.60), it follows again that $N_{31}(\psi; A, B) > 0$. Thus, $N_3(\psi; A, B) > 0$ for all $\psi \geq 0$.

(iv) Inequality (6.43)

The proof is based on inequality (6.65), proved previously.

(v) Inequality (6.44)

The denominator is again $D_3(\psi; A, B)$, which has been studied previously.

The numerator can be written in the form

$$N_3^*(\psi; A, B) = N_{31}(\psi; A, B) \left\{ \left[N_{31}(\psi; A, B) \right]^2 + 3N_{32}(\psi; A, B) \right\}, \tag{6.69}$$

where $N_{31}(\psi; A, B)$ and $N_{32}(\psi; A, B)$ were defined earlier in (6.68) and found to be positive for all $\psi \geq 0$. Thus, also $N_3^*(\psi; A, B) > 0$ for all $\psi \geq 0$.

Chapter 7

Application to extreme-value prediction

7.1 General

The cyclostationary modeling of H_S time series permits us to improve stochastic predictions of extreme values of significant wave height. Traditionally, in wave climate analysis (as well as in other areas of hydrography and oceanography) there are two different approaches for describing extreme value characteristics, either through the prediction of return values or through the prediction of asymptotic probability distribution of extremes.

These two approaches have been extensively studied in the simplest (but widely used up to now) case of an i.i.d. underlying process (sequence) [see Ochi (1982); Labeyrie (1991)]. However, if the dependence and/or nonstationarity of the underlying process is to be taken into account, the two approaches become quite difficult and the corresponding analysis is still in its infancy.

A method for calculating return values for periodically correlated time series, corresponding to predefined level, starting time, and return period, has been proposed by Middleton and Thompson (1986) and elaborated further by Hamon and Middleton (1989). These authors developed their methodology in the context of sea level prediction. Recently, Athanassoulis and Soukissian (1993b) developed an approach inspired by the above works and applied it to the prediction of H_S return values. The analysis of this method is presented in Athanassoulis and Soukissian (1995), and Soukissian (1995); see also Athanassoulis, Soukissian and Stefanakos (1995b). In subsequent sections, this method is used, along with the modelling presented in the previous chapter, to calculate return periods from nonstationary time series of H_S.

The asymptotic probability distribution of the maximum value $\max\{X(\tau;\beta),\ \tau \in [0,T]\}$, $T\to\infty$, of a cyclostationary process $X(\tau;\beta)$ has been recently obtained by Konstant and Piter-

barg (1993) for a Gaussian cyclostationary process. Their result reads as follows:

$$\lim_{T\to\infty} P\left[\left(\max_{[0,T]} X(\tau) - \ell_T\right)\ell_T < x\right] = \exp(-e^{-x}), \tag{7.1}$$

where $\ell_T = \mathrm{O}(\sqrt{\log T})$ is dependent on the second-order properties of the Gaussian process $X(\tau;\beta)$ and it can be analytically expressed in terms of $\sigma(\tau)$ and $C_{WW}(|\tau_1 - \tau_2|)$. Such a result can be used for H_S extreme-value predictions only after transforming the initial time series to a Gaussian one, e.g., by using the standard normal-scores or a logarithmic transformation.

The same extreme-value problem, but for the case of a discrete-time stochastic process of the form (3.6), has been studied by Ballerini and McCormick (1989). Their results are independent of the probability law of the underlying process and can be directly applied to H_S extreme value predictions. The use of cyclostationary modeling of H_S time series as a means to improve extreme-value prediction seems to be very promising, and its numerical implementation, although difficult, appears feasible.

7.2 The return period concept

In this section, the traditional as well as some improved techniques for return period calculations are briefly reviewed, and some new ideas for the return period concept are introduced. These ideas are mainly based on the more elaborated stochastic modelling of the time series of significant wave height presented in Chapters 3 and 4 [see also Athanassoulis and Stefanakos (1995; 1998)], and on a level crossing interpretation of the return period concept (Athanassoulis and Soukissian, 1995; Soukissian, 1995).

The concept of return period is traditionally defined by means of a sequence of random variables

$$X_1, X_2, \ldots, X_i, \ldots, \tag{7.2}$$

which are:

A1 independent, and

A2 identically distributed (i.i.d.).

Definition 7.1 *The return period associated with a level value x^* is defined as the mean value of the index i for which the event*

$$X_i > x^*, \quad i \geq 1,$$

happens for the first time. See, e.g., Borgman (1975). ■

This quantity is converted to time by assuming (arbitrarily, more or less) that

A3 $\Delta\tau$ = time between successive events (e.g., duration of sea states) is constant.

The above definition essentially realizes the simplest means for modeling and calculating the mean first passage time of the level value x^*. It admits of the well known closed-form solution

$$T_R(x^*) = \frac{1}{1 - F(x^*)}, \tag{7.3}$$

where $F(x)$ denotes the (common) cumulative distribution function of $\{X_i,\ \ i{=}1, 2, \dots\}$. If $T_R(x^*){=}N$years, then the level value x^* is also known as the N-year return value and, talking about H_S, the N-year wave height. For large values of N, the probability that the level value x^* has been exceeded at least once is approximately 63% and is independent of N.

In ocean engineering practice, return periods are calculated using methods based on either initial or extreme population data. More recently, alternative methods have been developed that use not only the maxima over a fixed time period but all the (moderately) large values above a threshold, the most well-known of these being the so-called POT method. Furthermore, methods that take into account the dependence structure as well as the second-order probability structure of data are also in use. In the following subsections, all these different approaches are briefly presented.

7.2.1 Methods based on initial population data

First, let us suppose that the stochastic sequence (7.2) represents a sequence of significant wave height observations H_S, i.e.

$$\left\{H_{S_n},\ \ n \in \mathbb{N}\right\} \equiv \left\{H_{S_n}\right\}. \tag{7.4}$$

If the sample size is large, the frequency table (histogram) of the sample values can be obtained and the empirical distribution of H_S is estimated. Then the return periods associated with various levels x^* is calculated by simply applying equ. (7.3). The use of an appropriate probability paper (e.g., Weibull probability paper) permits one to extrapolate and calculate return periods at level values greater than the observed values of H_S. See e.g., Ochi (1982).

Examining closely the compatibility of theoretical assumptions A1, A2, A3, on which the return period concept is based, and the specific (statistical) characteristics of the stochastic sequence (7.4), we cannot disregard that:

- Successive or neighboring H_{S_n} are not independent. See e.g., Chapters 4 and 6 of the present work, as well as Athanassoulis and Stefanakos (1995). This can be partly confronted by considering appropriate modifications of the cumulative distribution function $F(x)$, so that the dependence structure of the time series is taken into account. See, e.g., Pugh and Vassie (1978; 1980), Tawn and Vassie (1989). This technique is based on the concept of extremal index introduced by Loynes (1965) and elaborated by Leadbetter (1974).

- H_{S_n} cannot be considered identically distributed. A decisive reason for this is the seasonal variability. (A modification of the standard approach taking into account the fact that monthly H_{S_n} distributions are different is presented by Carter and Challenor (1981)).

- In any case, the exact distribution(s) of H_S is (are) not known. Of course one can always assume an analytical probability law, e.g., log-normal, Weibull or Gamma distribution, and estimate the parameters. See, e.g., Ochi (1993), Teng et al. (1994), Athanassoulis and Stefanakos (1996a). However, such an assumption is more or less arbitrary and its effect on the return period calculations is strong at high level values.

- Finally, it is well known that the time interval bertween specific events is not constant and its value seriously affects the return period calculations. (A partial solution of this difficulty is to use as $\Delta\tau$ the mean or the most probable value of the random variable $\Delta\tau$, usually ranging from 3 to 6 hours. See, e.g., Haring and Heideman (1978), Athanassoulis and Soukissian (1991).

7.2.2 Methods based on extreme population data

In this case, the stochastic sequence (7.2) corresponds to a sequence of successive maximum values of significant wave height corresponding to fixed time intervals (e.g., months, years, etc). This sequence is denoted by

$$\left\{H_{S_{\max,n}},\ \ n \in \mathbb{N}\right\} \equiv \left\{H_{S_{\max,n}}\right\}. \tag{7.5}$$

Predicting return periods and the corresponding design values by using data of maxima is known as the Gumbel's approach. This approach, when based on annual maxima, essentially overcomes all difficulties enumerated above concerning the validity of assumptions A1, A2, A3, as well as the fact that the distribution of the initial population is unknown. For, now A1,A2,A3 are more or less valid and, moreover and most importantly, the distribution of the population of maxima, denoted here by $G(x)$, is "almost" known. Clearly, we are referring to the well known theroretical result that the limiting (as $n\to\infty$) form of $G(x)$ can be only one of the

following three (extreme-type) distribution: (i) FT-I (Gumbel distribution), (ii) FT-II (Fréchet distribution), or (iii) FT-III (reversed Weibull distribution). This result is due to Fisher and Tippet (1928). See also Gnedenko (1943), Galambos (1978). The three types can be included in an single analytic form, called the von Mises form, which is also known as the Generalized Extreme Value (GEV) distribution

$$G(x;\lambda,\delta,k)=\begin{cases}\exp\left\{-\left[1-k\,\dfrac{x-\lambda}{\delta}\right]^{1/k}\right\}, & k\neq 0,\\[2ex] \exp\left\{-\exp\left[-\dfrac{x-\lambda}{\delta}\right]\right\}, & k=0,\end{cases} \tag{7.6}$$

where $\lambda,\delta>0$, and $-\infty<k<+\infty$. Note that, for k<0, the above equation holds true for $x>\lambda+\delta/k$, while for k>0, it holds for $x<\lambda+\delta/k$. The Generalized Extreme Value distribution provides a convenient unifying representation of the three extreme-type distributions, which are obtained as follows: (i) for k=0, the FT-I distribution is obtained, (ii) for k<0, λ=1, δ=$-k$, the FT-II distribution, (iii) for k>0, λ=-1, δ=k, the FT-III distribution.

The estimation of return periods for high levels x^* is again based on the formula (7.3),

$$T_R(x^*)=\frac{\Delta\tau}{1-G(x^*)}, \tag{7.7}$$

but now the extreme population distribution $G(x)$ is used instead of the initial population distribution $F(x)$.

The use of $G(x)$ instead of $F(x)$ in equation (7.7), is justified by the fact that the two distributions $G(x)$ and $F(x)$ are right-tailed equivalent. See, e.g., Resnick (1971), Castillo (1988). In ocean and coastal engineering practice, equation (7.7) is considered as a milestone for return-period calculations of various wave and wind parameters (Ochi, 1982; Ochi, 1990; Bishop, 1984).

Obviously, the application of the above method to the calculation of return periods goes through the following two steps: (i) Decide upon the appropriate type of the asymptotic extreme-value distribution $G(x)$, and (ii) Estimate the parameters of the asymptotic distribution $G(x)$, selected in step (i). See, e.g., Castillo and Sarabia (1992), Soukissian (1995), Carter et al. (1998). Practically, the most expedient method to accomplish step (i) is to calculate $G_{emp}(x)$ by means of appropriate plotting formulas. Then, the selection of the type of $G(x)$ should be based on the behaviour of $G_{emp}(x)$ at the right tail only. This task can be processed by adopting a variety of short-cut procedures or by analytical methods either of purely statistical nature or based on a combination of physical observations and statistical arguments. Some serious problems arising during the selection procedure are well referenced in Muir and El-Shaarawi (1986), Castillo (1988), Castillo and Sarabia (1992) and Soukissian (1995).

If we focus on physical arguments, the most "reasonable" choice of the type of the asymptotic distribution of $H_{S_{\max}}$ seems to be the FT-III, which is upper limited, as $H_{S_{\max}}$ is expected to be. (Recall that steep waves break). However, working with the FT-III distribution, one faces the difficult problem of estimating the position parameter λ, which is the cut-off value of $H_{S_{\max}}$. This estimation is so sensitive to data pecularities that renders the results unreliable (Muir and El-Shaarawi, 1986; Soukissian, 1995).

Step (ii) can be performed by a variety of methods, most of them well-known and widely used: the maximum likelihood method (MLM), the method of moments, as well as various types of linear tail-weighted least-squares methods (LSM). An application of these methods is presented analytically in Section 7.4.1. Another method for estimating the parameters of the asymptotic type distributions is the one of the r-largest values introduced by Weissman (1978) (see also Leadbetter et al. (1983)) and subsequently used by Smith (1986), Tawn (1988).

In principle, Gumbel's approach forms a complete methodology for predicting long-term extreme values and the corresponding return periods. The most serious, and practically unresolved (up to now), problem of this approach is the lack of sufficiently large extreme population data samples that would permit the type of distribution to be safely selected and its parameters to be reliably estimated.

7.2.3 The Peaks-Over-Threshold method

Apart from the traditional approaches described above, various alternative methods for return period prediction have also been developed. Most of them are primarily developed in the context of stochastic hydrology, in order to overcome the deficiencies of the traditional approach. A common characteristic of these methods is that they use quite different data from those used by the traditional approach. The most important of them is the so-called Peaks-Over-Threshold (POT) method, which has been originally introduced by hydrologists to describe flooding events (NERC, 1975). This method has been introduced in ocean engineering (although in a simplified manner) the last decade. See e.g., Rosbjerg and Knudsen (1984), van Vledder and Zitman (1992), van Vledder et al. (1994).

In this method, the events exceeding a moderately high threshold level are considered instead of maxima over fixed time periods, i.e. instead of yearly or monthly maxima storms with H_S greater than a threshold value are used. The exceedances over a threshold have been shown by Pickands (1975) that can be modelled sufficiently well by means of the Generalized Pareto Distribution (GPD) with parameters k and a such that $-\infty<k<+\infty$, $a>0$, and cumulative distribution given by

$$F_{GPD}(x;k,a)=\begin{cases}1-\left(1-\dfrac{kx}{a}\right)^{1/k}, & k\neq 0,\\[2ex] 1-\exp\left(-\dfrac{x}{a}\right), & k=0.\end{cases} \tag{7.8}$$

The range for x is $x>0$ for $k\leq 0$, and $0<x<a/k$ for $k>0$. See also Hosking and Wallis (1987) and Embrechts et al. (1997), where examples of GPD are given for various parameter values. Parameter estimation of GPD can be made by standard statistical techniques, such as the method of moments, the method of probability-weighted moments and the maximum likelihood method (Hosking and Wallis, 1987; Grimshaw, 1993, and references cited therein).

Let us now suppose that N random variables from the initial sequence (7.2) exceed a threshold value X_u. Then, the exceedances $\{Y_n=X_n-X_u,\ n=1,2,\ldots,N\}$ follow GPD, where the number N is another random variable, which is independent of Y_n's and follows Poisson distribution. Then, the distrution of the maximum $Y_{\max}=\max\{Y_n,\ n=1,2,\ldots,N\}$ is found to be (Embrechts et al., 1997; Carter et al., 1998)

$$\begin{aligned} F(Y_{\max}\leq y) &= \exp\Big[-\nu(X_u)\ \big(1-F_{GPD}(y;k,a)\big)\Big] = \\ &= G(y;\lambda,\delta,k), \end{aligned} \tag{7.9}$$

where $\nu(X_u)$ is the expected number of exceedances of the threshold X_u and $G(\cdot)$ is the Generalized Extreme Value (GEV) distribution (see equ. (7.6)) with parameters $\lambda=a(1-\nu(X_u)^{-k})/k$ and $\delta=a/\nu(X_u)^k$. In the special case where $k=0$, the parameters of GEV are $\lambda=a\ln\nu(X_u)$ and $\delta=a$.

A complete description of the POT method and its interrelation with classical extreme-value theory can be found in Smith (1984) and in Embrechts et al. (1997). Interesting applications of it are presented in Rosbjerg and Knudsen (1984); see also van Vledder and Zitman (1992), van Vledder et al. (1994). In Carter et al. (1998), a description of the procedure for POT analysis of data is given, which is summarised as follows:

1. Collect observations of the variable of interest (e.g. the significant wave height) over many years; denoted by $X_1, X_2, \ldots, X_i, \ldots, X_I$; cf. equ. (7.2).

2. Chose a moderately high threshold X_u and select those values of the $X-$sequence that exceed X_u. Form the sequence of excesses $\{Y_1, Y_2, \ldots, Y_n, \ldots, Y_N\}$, with $Y_n= X_n - X_u>0$.

3. Estimate the expected number of exceedances per year $\nu(X_u)$.

4. By means of the sequence $\{Y_1, Y_2, \ldots, Y_N\}$, fit a Generalized Pareto Distribution with parameters k and a.

5. Calculate the distribution of maxima of X_i, by means of equ. (7.6), with parameters $\lambda=X_u + a(1-\nu(X_u)^{-k})/k$ and $\delta=a/\nu(X_u)^k$, or, if $k=0$, $\lambda=X_u + a\ln\big(\nu(X_u)\big)$ and $\delta=a$.

6. Calculate the associated return period by means of equ. (7.7).

In Section 7.4, an application of POT method to wave data is presented, using the above procedure.

Finally, it should be noted that in the implementation of the POT method, special attention should be paid during the sampling procedure of the large values in order to avoid clustering. The selection of intensities of the extreme events must be done in a way ensuring their statistical independence. A weak point of the method is that it is entirely based on ad hoc statistical models which cannot be justified theoretically (Smith, 1984). However, a significant advantage of it, is its potentiality to deal, in some way, with seasonality and serial dependence, which are always present in many-year long time series of wave and wind parameters. See, e.g., Smith (1984), Carter et al. (1998).

7.2.4 Approaches based on the stochastic modeling of long-term time series of data

The motivation for developing such kind of methods comes from the fact that long-term time series of wave parameters (as well as of many other environmental parameters) are nonstationary time series, exhibiting a year-to-year statistical variability and longer-term climatic variability. Under these circumstances, it is natural to try to predict extreme values by using an appropriate stochastic modelling of the basic process (see representations (7.19) and (7.20) of Section 7.4).

These methods of extreme-value prediction are very recently introduced in ocean engineering applications by Athanassoulis and Soukissian (1995) and Soukissian (1995). Their basic characteristics are:

a) Replace assumptions A1 and A2 by less restrictive ones,

b) Disregard assumption A3,

c) Model and treat the stochastic character and the dynamic nature of the underlying phenomenon (e.g., the various time scales involved and the correlation structure of the wave parameters).

Return period associated with the level value x^* is calculated as the time period in which the MEan Number of Upcrossings of the level x^* becomes equal to unity (MENU method). The theoretical background of this method has been developed in the pioneering work by Rice (1944/1945). However, a version of MENU method as a technique for extreme value predictions first appeared in 1986 (Middleton and Thompson, 1986; Hamon and Middleton, 1989), in the context of sea-level extreme-value prediction. A detailed description of the method along with

its complete implementaion for the case the underlying process is Gaussian is given in Athanassoulis and Soukissian (1995) and in Soukissian (1995). See also Athanassoulis, Soukissian and Stefanakos (1995b), where an application with Gaussian synthetic data is presented. The implementation is heavily based on the nonstationary modelling of the stochastic process $H_S(\tau)$ presented in Chapters 3 and 4; see also Athanassoulis and Stefanakos (1995). In Section 7.4, the MENU method is applied to a case where the underlying process is not Gaussian, and the results are compared with the ones obtained by Athanassoulis, Soukissian and Stefanakos (1995b). The present application takes advantage of the appropriate modelling of the second-order probability structure of the stochastic process presented in the previous chapter.

7.3 Return periods for nonstationary stochastic processes

In this section, a new definition of the notion of return period associated with a given level value is introduced, which works well even if the underlying stochastic process is a general nonstationary one. This definition will be based on an appropriate crossing problem. Crossing problems are a major part of stochastic geometry which deals, among others things, with the stochastic description of the number of crossings of a given level by the graph of a (nonstationary) stochastic process, as well as the stochastic description of the time intervals between crossings or until a stochastic process reaches (crosses) a given upper or lower level (barrier). One of the most interesting problems is the so-called "one-sided barrier problem" or "one-sided first passage problem" which can be stated as follows:

> Given that at a time instant τ_0 the stochastic process $X(\tau;\beta)$ has a known value $X(\tau_0;\beta)$, find the stochastic structure of the time interval $T(\tau_0;\beta)$ until the process reaches for the first time the level x^*.

A detailed review of the stochastic crossing problems can be found in Blake and Lindsey (1973) and Abrahams (1986). See also Berman (1992).

It is clear that the time interval $T(\tau_0;\beta)$ is, in general, a random variable. Therefore, its complete stochastic characterization requires the knowledge of its cumulative distribution function. The problem, in its greatest generality, remains unsolved except for some very special cases. Significant theoretical progress has been made for stationary and nonstationary Gaussian processes but, in any case, the problem is still in its infancy. Fortunately, however, for the applications we are interested in, it suffices to know only some "gross" statistical characteristics of $T(\tau_0;\beta)$ as the mean value and the variance. In fact, our new definition of the return period (Definition 7.3, below) is taylored in such a way that the knowledge of the first and second moments of the underlying process to be sufficient for the corresponding calculation.

Let us now state the general definition of return period which is applicable whatever the nature of the underlying process may be.

Definition 7.2 *Assume that* $\{X(\tau;\beta),\ \beta \in B\}$ *is a nonstationary stochastic process with mean-square differentiable path functions (*β *is a choice variable, used for distinguishing different path functions). Let* t *be a given time instant and* x^* *a given level value. Further, let us denote by* $\tau_1(x^*;t;\beta)$ *the first passage time of level* x^* *by the path function* $X(\tau;\beta)$ *occurring after the time instant* t*. Then, the mean value of the quantity* $T_R(x^*;t;\beta)=\tau_1(x^*;t;\beta)-t$*, is called the return period of* $X(\tau;\beta)$ *associated with the level value* x^* *and the starting time* t*, and it is denoted by* $\overline{T}_R(x^*;t)$*:*

$$\overline{T}_R(x^*;t) = \mathsf{E}^{\beta}\big[T_R(x^*;t;\beta)\big]. \tag{7.10}$$

■

This definition is quite general and, clearly, free of any restrictive assumptions as regards the underlying stochastic process $\{X(\tau;\beta),\ \beta \in B\}$. However, to facilitate the numerical calculations, an alternative definition will now be introduced, which applies equally well to general stochastic processes, but treats the first passage on the basis of the number of upcrossings of the level x^*.

Definition 7.3 *Assume that* $\{X(\tau;\beta),\ \beta \in B\}$ *is a nonstationary stochastic process with mean-square differentiable path functions. Let* $M(x^*;t,t+T)$ *be the mean number of upcrossings of the level* x^* *by the process* $\{X(\tau;\beta),\ \beta \in B\}$ *in the time interval* $(t,t+T)$*. When* $M(x^*;t,t+T)$ *becomes equal to unity, the time lag (interval)* $T=(t+T)-t$ *will be called the return period of* $X(\tau;\beta)$ *associated with the level value* x^* *and the starting time* t*, and it will be denoted by* $T_R(x^*,t)$*.*

■

The second definition has been first used by Middleton and Thompson (1986) (see also Hamon and Middleton (1989)), in the context of statistical prediction of sea-level extremes. Although it seems likely the above two definitions of return period to be equivalent, this has not yet been established. However, on intuitive basis, one can conjecture that the numerical results obtained by means of these two definitions should be either the same or very near.

Now, in order to implement Definition 7.3, we have to calculate the mean number of upcrossings of the level x^* by the nonstationary process $X(\tau;\beta)$. As it is well known (Rice, 1944/1945), an upcrossing of the level x^* by the process $X(\tau;\beta)$ occurs when

$$X(\tau;\beta) = x^*, \tag{7.11}$$

and

$$\frac{dX(\tau;\beta)}{d\tau} > 0. \tag{7.12}$$

The total number of the upcrossings of the level x^* within the time interval (t_1, t_2) is given by the equation

$$C_R^+(x^*; t_1, t_2; \beta) = \frac{1}{2}\int_{t_1}^{t_2} |\dot{X}(\tau;\beta)|\delta\big(X(\tau;\beta) - x^*\big)d\tau, \tag{7.13}$$

where $\delta(\cdot)$ is the Dirac delta function. Equation (7.13) was first derived by Rice (1944/1954). Recalling now that, for any (possibly generalized) function $G(x, y)$

$$\mathsf{E}^{\beta}\Big[G\Big(X(\tau_1;\beta), Y(\tau_2;\beta)\Big)\Big] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} G(x,y) f_{\tau_1,\tau_2}(x,y)dxdy, \tag{7.14}$$

where $f_{\tau_1,\tau_2}(x, y)$ is the joint probability density of the random vector $\big(X(\tau_1;\beta), Y(\tau_2;\beta)\big)$, and applying the ensemble average operator $\mathsf{E}^{\beta}[\cdot]$ in both sides of equation (7.13), we obtain

$$\mathsf{E}^{\beta}\Big[C_R^+(x^*; t_1, t_2; \beta)\Big] = \frac{1}{2}\int_{t_1}^{t_2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\dot{x}|\delta(x - x^*) f_{\tau,\tau}(x, \dot{x})dxd\dot{x}d\tau. \tag{7.15}$$

Integrating equ. (7.15) with respect to x, and observing that $\mathsf{E}^{\beta}\big[C_R^+(x^*; t_1, t_2; \beta)\big]$ is not but the quantity $M(x^*; t_1, t_2)$ appearing in Definition 7.3 above, we easily obtain

$$M(x^*; t_1, t_2) = \frac{1}{2}\int_{t_1}^{t_2}\int_{-\infty}^{+\infty} |\dot{x}|\ f_{\tau,\tau}(x^*, \dot{x})d\dot{x}d\tau. \tag{7.16}$$

Since $f_{\tau,\tau}(x^*, \dot{x})$ is an even function with respect to the second argument $\dot{x}$ (see, e.g., Levine 1973, Vol. 1, pp. 201, 424), equ. (7.16) can be written in the following more convenient form

$$M(x^*; t_1, t_2) = \int_{t_1}^{t_2}\int_{0}^{+\infty} \dot{x}\ f_{\tau,\tau}(x^*, \dot{x})d\dot{x}d\tau. \tag{7.17}$$

It should be noted that the above relation is totally independent from the specific type of the underlying bivariate distribution considered. Clearly, the deterministic quantity $M(x^*; t_1, t_2)$ has the following analytical properties:

(i) For any $t_1 > t_2$, and any $x^* > 0$, $M(x^*; t_1, t_2) \geq 0$ and $M(x^*; t, t) = 0$,

(ii) For constant t_1 and $x^*>0$, $M(x^*;t_1,t_2)$ is an increasing function of t_2,

(iii) For constant $t_1>t_2$, $M(x^*;t_1,t_2)$ is a decreasing function of x^*.

It should be emphasized that $M(x^*;t_1,t_2)$ is dependent on both time instants t_1 and t_2, since the underlying process $\{X(\tau;\beta),\ \beta\in B\}$ may be nonstationary.

In accordance with Definition 7.3, the return period $T_R(x^*,t)$ associated with the level x^* and the starting time t is calculated as the unique value T for which

$$M(x^*;t,t+T)=1. \tag{7.18}$$

Let it be noted that the above definition of return period $T_R(x^*,t)$ can be applied for any level value x^* (high or low) and any starting time t. Accordingly, seasonal weather windows can be obtained in this manner.

7.4 Application to non-Gaussian stochastic processes

In this section, the mean number of upcrossings $M(x^*,t,t+T)$ of a periodically-correlated stochastic process $X(\tau;\beta)$, as well as the associated return periods for various level values x^*, is calculated. Assume that $X(\tau;\beta)$ admits of the representation (3.6), i.e.

$$X(\tau;\beta)=G(\tau)+\sigma(\tau)W(\tau;\beta), \tag{7.19}$$

where $G(\tau)$ and $\sigma(\tau)$ are deterministic time-dependent periodic functions and $W(\tau;\beta)$ is a zero-mean stationary stochastic process, which will be called hereafter the residual stochastic process.

The complete stochastic characterization of a general stochastic process $X(\tau;\beta)$ is a very complicated task, consisting in finding the sequence of cumulative distribution functions of all orders. In this work, we are interested solely in the first-order joint stochastic characterization of $X(\tau;\beta)$ and $\dot{X}(\tau;\beta)$, since for the numerical implementation of the MENU method use is only made of the joint probability density function $f_{\tau,\tau}^{X\dot{X}}(s_1,s_2)$.

In order to evaluate the building elements of equ. (7.18) of Section 7.3, equ. (7.19) is differentiated with respect to time, obtaining an analytical representation of the process $\dot{X}(\tau;\beta)$:

$$\dot{X}(\tau;\beta)=\dot{G}(\tau)+\dot{\sigma}(\tau)W(\tau;\beta)+\sigma(\tau)\dot{W}(\tau;\beta). \tag{7.20}$$

Now, either an *a priori* assumption for the joint probability density function $f_{\tau,\tau}^{X\dot{X}}(s_1,s_2)$ should be made, or it should be obtained by means of $f_{\tau,\tau}^{W\dot{W}}(v_1,v_2)$ by considering the bivariate linear

transformation defined by equs. (7.19) and (7.20). The former case has already been investigated by Athanassoulis, Soukissian and Stefanakos (1995b). In the present work, the latter option will be studied and compared with the former. In this direction, a bivariate analytic model for $f_{\tau,\tau}^{W\dot{W}}(v_1, v_2)$ is needed.

From now on, for brevity and without loss of generality, the chance variable β will be omitted from the argument of the stochatsic processes $X(\tau)$, $\dot{X}(\tau)$, $W(\tau)$, $\dot{W}(\tau)$. For the exterme-value calculation of the present section, the dataset introduced and studied in Section 6.7 will be used. It is an eight-year long time series of significant wave height in a location in the Atlantic Ocean (45^{o}N, 12^{o}W). See also Section 6.7 and Appendix A, where a description of this dataset is given.

Considering the density of $\big(W(\tau), \dot{W}(\tau)\big)$, there are numerous ways to model it, some of which are the following:

1. Use the bivariate Plackett model (6.19) with univariate marginal densities estimated from the univariate samples of $W(\tau)$ and $\dot{W}(\tau)$, respectively.

2. Use the bivariate Plackett model with both univariate marginals lognormal for the joint density of $\big(W(\tau), W(\tau + \Delta\tau)\big)$ (see Section 6.7), and by means of the bivariate linear transformation

$$\begin{aligned} W(\tau) &= W(\tau), \\ \dot{W}(\tau) &= \frac{W(\tau + \Delta\tau) - W(\tau)}{\Delta\tau}, \end{aligned} \tag{7.21}$$

   calculate the $f_{\tau,\tau}^{W\dot{W}}(v_1, v_2)$, where $\Delta\tau$=6h is the fixed sampling interval of the time series.

3. Assume that the joint density $f_{\tau,\tau+\Delta\tau}^{WW}(u_1, u_2)$ of $\big(W(\tau), W(\tau+\Delta\tau)\big)$ is a bivariate normal density, and, by means of transformation (7.21), obtain the density $f_{\tau,\tau}^{W\dot{W}}(v_1, v_2)$.

In the sequel, all the above three approaches will be examined and discussed.

Let us first obtain a population of $\dot{W}(\tau)$ in order to (i) estimate the parameters of the various analytic models examined, (ii) check the goodness-of-fit of each analytic model. This task is accomplished using the population of $W(\tau)$ and the fact that the derivative is approximated well by means of the difference presented in the second of the equations (7.21). In Figure 7.1, the time series of $\dot{W}(\tau)$ is depicted, as this has been obtained from the time series of $W(\tau)$ for the hindcast dataset described in Section 6.7). The mean value and the standard deviation of the series are 0 and 0.0548, respectively. Now, the first-order empirical density function of $\dot{W}(\tau)$ is obtained (see Figure 7.2), and various analytic models are tested for fit. It follows that none of the known analytic probability models fits well to it. To surconvent this difficulty, various

alternative modellings can be attempted, such as, e.g., a mixture of known shapes (densities), a kernel density approximation etc. We will not further discuss these alternatives, and the first approach proposed above will not further be examined.

In the sequel, we proceed with the second approach, i.e., a bivariate Plackett model with univariate lognormal marginals (see Section 6.7) is fitted to the sample of $\big(W(\tau), W(\tau+\Delta\tau)\big)$; see Figure 7.3. In this way, the density $f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)$ is obtained.

Then, consider the bivariate linear transformation

$$v_1 = u_1 = h_1(u_1,u_2), \qquad v_2 = \frac{u_2-u_1}{\Delta\tau} = h_2(u_1,u_2), \tag{7.22}$$

where $\Delta\tau$=$const$, with inverse transformation

$$u_1 = v_1 = h_1^{-1}(v_1,v_2), \qquad v_2 = v_1 + v_2\,\Delta\tau = h_2^{-1}(v_1,v_2), \tag{7.23}$$

and corresponding Jacobian

$$|J_1| = \begin{vmatrix} \dfrac{\partial h_1^{-1}(v_1,v_2)}{\partial v_1} & \dfrac{\partial h_1^{-1}(v_1,v_2)}{\partial v_2} \\ \dfrac{\partial h_2^{-1}(v_1,v_2)}{\partial v_1} & \dfrac{\partial h_2^{-1}(v_1,v_2)}{\partial v_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & \Delta\tau \end{vmatrix} = \Delta\tau. \tag{7.24}$$

The density $f^{W\dot{W}}_{\tau,\tau}(v_1,v_2)$ is obtained by means of transformation (7.22) from the analytic density $f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)$, as (Larson and Shubert, 1979)

$$\begin{aligned} f^{W\dot{W}}_{\tau,\tau}(v_1,v_2) &= |J_1|\;\; f^{WW}_{\tau,\tau+\Delta\tau}\big(h_1^{-1}(v_1,v_2), h_2^{-1}(v_1,v_2)\big) = \\ &= \Delta\tau\;\; f^{WW}_{\tau,\tau+\Delta\tau}(v_1, v_1+\Delta\tau\; v_2). \end{aligned} \tag{7.25}$$

In Figure 7.4, contours of the above transformed density are depicted, along with the contours of the corresponding empirical joint probability density function of $\big(W(\tau), \dot{W}(\tau)\big)$. The fit between the two densities is satisfactory. For this, one can compare the mean value and the standard deviation of the two variables, as well as their correlation coefficient, as they derived from the empirical density and from the analytic one, depicted in Table 7.1. We chose not to use one of the well-known statistical goodness-of-fit tests (χ^2, Kolmogorov-Smirnov etc.), because they all involve the total number of observations in their calculations, a fact that makes statistical tests to reject the null hypothesis for any significance level if the total number has a very large value. Following the experience by Athanassoulis et al. (1994), a nonprobabilistic measure is used instead, namely the Euclidian distance

$$D^2 = \sum_{i=1}^{I} (p_i - \pi_i)^2, \tag{7.26}$$

where π_i's are the empirical probabilities, and p_i's the corresponding analytic ones. This distance for the examined dataset gave D^2=0.0017, which we think that is a very good fit.

Also, by integrating over all v_1-values we obtain the marginal density of $\dot{W}(\tau)$; see Figure 7.2. Comparing with the empirical density, it is found in a very good agreement.

Let us now proceed to the calculation of the density $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$. Consider equs. (7.19) and (7.20) as a bivariate time-dependent linear transformation of the form

$$\begin{aligned} s_1 &= g_\tau + \sigma_\tau \, v_1 = \ell_1(v_1, v_2), \\ s_2 &= dg_\tau + d\sigma_\tau \, v_1 + \sigma_\tau \, v_2 = \ell_2(v_1, v_2), \end{aligned} \tag{7.27}$$

where $g_\tau \equiv G(\tau)$, $\sigma_\tau \equiv \sigma(\tau)$, $dg_\tau \equiv \dot{G}(\tau)$, $d\sigma_\tau \equiv \dot{\sigma}(\tau)$. The function $G(\tau)$ is defined as

$$G(\tau) = \overline{X}_{mean} + \mu(\tau), \tag{7.28}$$

where $\overline{X}_{mean}$ is the overall mean value of the process $X(\tau)$ and $\mu(\tau)$ is the seasonal mean value; see also Chapter 4. Note that, sometimes a linear trend is used instead of a fixed mean value. However, since its existence is still questionable, it has been decided to use for this study the fixed mean value $\overline{X}_{mean}$. Similarly, the function $\sigma(\tau)$ is the seasonal standard deviation of the process. The functions $\dot{G}(\tau)$ and $\dot{\sigma}(\tau)$ are the derivatives of $G(\tau)$ and $\sigma(\tau)$ with respect to τ. Note also that, according to their definition, all the functions $G(\tau)$, $\sigma(\tau)$, $\dot{G}(\tau)$ and $\dot{\sigma}(\tau)$ are periodic.

The inverse transformation of (7.27) is

$$\begin{aligned} v_1 &= \frac{s_1 - g_\tau}{\sigma_\tau} = \ell_1^{-1}(s_1, s_2), \\ v_2 &= \frac{(s_2 - dg_\tau)\,\sigma_\tau - (s_1 - g_\tau)\,d\sigma_\tau}{\sigma_\tau^2} = \ell_2^{-1}(s_1, s_2), \end{aligned} \tag{7.29}$$

and the corresponding Jacobian is

$$|J_2| = \begin{vmatrix} \dfrac{\partial \ell_1^{-1}(s_1, s_2)}{\partial s_1} & \dfrac{\partial \ell_1^{-1}(s_1, s_2)}{\partial s_2} \\ \dfrac{\partial \ell_2^{-1}(s_1, s_2)}{\partial s_1} & \dfrac{\partial \ell_2^{-1}(s_1, s_2)}{\partial s_2} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{\sigma_\tau} & 0 \\ -\dfrac{d\sigma_\tau}{\sigma_\tau^2} & \dfrac{1}{\sigma_\tau} \end{vmatrix} = \frac{1}{\sigma_\tau^2}. \tag{7.30}$$

Thus, the density $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ is obtained by means of the above transformation and the density $f_{\tau,\tau}^{W\dot{W}}(v_1, v_2)$ as follows:

$$\begin{aligned} f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) &= |J_2| \;\; f_{\tau,\tau}^{W\dot{W}}\left(\ell_1^{-1}(s_1, s_2), \ell_2^{-1}(s_1, s_2)\right) = \\ &= \frac{1}{\sigma_\tau^2} \;\; f_{\tau,\tau}^{W\dot{W}}\left(\frac{s_1 - g_\tau}{\sigma_\tau}, \frac{(s_2 - dg_\tau)\,\sigma_\tau - (s_1 - g_\tau)\,d\sigma_\tau}{\sigma_\tau^2}\right). \end{aligned} \tag{7.31}$$

In Figure 7.5, contours of the bivariate density (7.31) are depicted for various τ-values within the year, namely for the first six-hourly interval of the first day of each month. On each plot, the statistical characteristics (mean value and standard deviation) of each marginal, as well as the correlation coefficient are given. Here, we cannot obtain reliable samples, since only eight values (the number of the years covered by the specific dataset) are available for each six-hourly interval!

Also, some remarks on the physical interpretation of the depicted densities should be made. Recall first that $X(\tau)$ is a time series of significant wave height, and, thus, $\dot{X}(\tau)$ represents the time series of the slopes of the wave height. In this connection, the probability functions of $\dot{X}(\tau)$ and $\dot{W}(\tau)$ should be "almost" symmetric about zero. That is, the probability that a wave height will increase should be "almost" the same with the probability that a wave height will decrease. We quote the word "almost", because in fact the probability is dependent on the value of wave height we refer to. If, for example, we choose a relative small value for significant wave height (low waves), it is more probable for an increasing wave height to occur (positive slope) rather than for a decreasing one to occur (negative slope). The contrary holds true for the highest waves. A second remark is that, in summer, the probability mass is located in lower wave heights than in the winter. A fact that is more or less expected, still intesifying our analysis.

Let us now turn to the calculation of the quantity (7.16). First, let us examine the integrand in the right-hand side of equ. (7.16). This is the (time-dependent) density $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ for fixed its first argument s_1=s_1^*=$fixed$, multiplied by the absolute value of $\dot{X}(\tau)$. The integrand reveals a periodic character, which is due to the periodic (with period one year) functions included in $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$. Taking into account this character, the integral of the right-hand side of equ. (7.16) is simplified as follows

$$\begin{aligned} M(x^*; t_1, t_2) &= \int_{t_1}^{t_2} J(x^*; \tau) d\tau = \\ &= n \int_{t_1}^{t_1+T} J(x^*; \tau) d\tau + \Re(t_1, t_2; T), \end{aligned} \tag{7.32}$$

where

$$J(x^*; \tau) = \frac{1}{2} \int_{-\infty}^{+\infty} |\dot{x}| \; f_{\tau,\tau}(x^*, \dot{x}) d\dot{x}, \tag{7.33}$$

$$n = \left[\frac{t_2 - t_1}{T} \right], \tag{7.34}$$

and

$$\Re(t_1, t_2; T) = M(x^*; t_1, t_2) - n \int_{t_1}^{t_1+T} J(x^*; \tau) d\tau. \tag{7.35}$$

Note that, $[\cdot]$ stands for the integer part. In practice, the value t_2=nT is selected, so that $\Re(t_1, t_2; T)$=0. For example, in Figure 7.6, $M(x^*; t_1, t_2)$ is calculated and plotted up to three periods (years). Finally, according to Definition 7.3, the time interval T_R, which corresponds to the value $M(x^*, t, t+T)$=1, will be called the return period of $X(\tau)$ associated with the level value x^*.

The above described calculations are repeated for various fixed values of $X(\tau)$, namely x^*=5, 10, 12.5, 13, 14, 15, 16, 17, 18, 19, 20m. In this way, we obtain the return-value diagram shown in Figure 7.7. In the following two subsections, these results will be compared with results from classical methods, as well as with results of MENU method if the density $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ is *a priori* assumed Gaussian.

7.4.1 Comparison with classical methods

First, the so-called Gumbel approach is examined. The method consists in fitting the FT-I (Gumbel) extreme-type distribution to the yearly maxima of the process $X(\tau)$; see also Section 7.2.2, where this method is described.

There are numerous different approaches for the estimation of the parameters of FT-I distribution; see, e.g., Castillo (1988), Embrechts et al. (1997), and references cited therein. The following four have been selected for the present study:

(i) *The probability paper method*.
In this method a straight line is fitted by the least-squares method to the data on probability paper (Castillo, 1988).

(ii) *The least-squares return period relative error method*.
A weighted least-squares method is used. The distance (error) to be minimized is the relative return-period error (Castillo, 1988).

(iii) *The probability paper method with different plotting position* (Gringorten, 1963).
The method is the same with (i), except that it uses the plotting positions proposed by Gringorten (1963). These positions are defined as π_i=$(i - 0.44)/(n + 1)$ instead of the classical ones π_i=$i/(n+1)$. The method is popular in ocean engineering applications; see, e.g., Carter et al. (1983).

(iv) *The maximum likelihood method.*
The well-known method consisting in the maximization of the likelihood function, which is done by solving the so-called maximum-likelihood (ML) equations. In our case, no explicit solution exists to these equations. Prescott and Walden (1980; 1983) suggest variants of the Newton-Rapson scheme. For the numerical implementation of this approach, a Fortran algorithm, published by Hosking (1985) and supplemented by Macleod (1989), is used.

In Table 7.2, the estimated parameters are given for the above four methods. It should be noted that the parameters derived from the maximum likelihood method are very sensitive to the selection of the initial parameters. Although we have tried as initial guess the ones proposed by Hosking (1985), the best results are obtained by inserting the mean value of the estimated parameters by the other three methods.

In Figure 7.7, the associated return periods are compared with the results obtained by the MENU method. Following Figure 7.7, the results of MENU method are in very good agreement with the results from the other approaches, especially if the ML method is excluded.

Let us now turn to another popular approach, namely the POT method. This method is briefly described in Section 7.2.3, where a procedure for its numerical implementation is also given. According to this procedure, a population above a threshold value X_u is formed and the parameters of the associated Generalized Pareto distribution (GPD) are estimated. For this, the maximum likelihood method, fully described by Grimshaw (1993), is applied to the derived population.

This task is accomplished for various threshold values X_u, and the results are shown in Table 7.3. Then, the parameters of GPD are used to calculate associated parameters of Generalized Extreme-Value distribution needed for the calculation of return periods. It is important to note here, that the return-period calculations of the POT method are threshold dependent. One should expect that, for moderately high thresholds, the return-period diagrams will coincide. In practice, however, they exhibit a small dispersion. See Figure 7.8, where return-period diagrams are shown for threshold values X_u=8, 9, 10, 11, 12m, and they are compared with the results of MENU method.

7.4.2 Comparison with the Gaussian case

As we have already mentioned, Athanassoulis, Soukissian and Stefanakos (1995b) applied the MENU method by assuming *a priori* that the density $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ is Gaussian. This approach, will be compared in the present subsection with the one previously described in the main part of Section 7.4.

The dataset we have at hand is not Gaussian; see Section 6.7. In order to obtain an equivalent Gaussian population, a logarithmic transformation of the form $X(\tau)$=ln$[H_S(\tau)+c]$ is applied to the initial time series, where c is appropriate constant. In the specific case, it has been chosen c=1m.

Using the above obtained population, the time series analysis presented in Chapter 4 is performed and the residual time series $W(\tau)$ is modelled as a Gaussian process; see Figures 7.9 and 7.10, where the first- and second-order probability density function of $W(\tau)$ is presented. Especially for the second-order, it should be noted that the bivariate normal fits better to our data than the Plackett model with univariate marginals normal.

Thus, the density $f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)$ is written

$$f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)=\frac{1}{2\pi\sigma_{u_1}\sigma_{u_2}\sqrt{1-\rho^2_{u_1u_2}}}\ \exp\Bigg\{-\frac{1}{2(1-\rho^2_{u_1u_2})}\Bigg[\Big(\frac{u_1-m_{u_1}}{\sigma_{u_1}}\Big)^2-$$

$$-\,2\,\rho_{u_1u_2}\,\Big(\frac{u_1-m_{u_1}}{\sigma_{u_1}}\Big)\,\Big(\frac{u_2-m_{u_2}}{\sigma_{u_2}}\Big)+\Big(\frac{u_2-m_{u_2}}{\sigma_{u_2}}\Big)^2\Bigg]\Bigg\},\quad(7.36)$$

$$-\infty<u_1,u_2<+\infty.$$

Since the process $W(\tau)$ is stationary, m_{u_1}=m_{u_2}=m_u, σ_{u_1}=σ_{u_2}=σ_u, and $\rho_{u_1u_2}$=$\dfrac{\sigma_{u_1u_2}}{\sigma^2_u}$=$\dfrac{C_{WW}(1)}{C_{WW}(0)}$=$\rho_u$.

Thus, equ. (7.36) is rewritten, in more compact form, as

$$f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)=\frac{1}{2\pi\sigma^2_u\sqrt{1-\rho^2_u}}\ \exp\Bigg\{-\frac{1}{2(1-\rho^2_u)}\Bigg[\Big(\frac{u_1-m_u}{\sigma_u}\Big)^2-$$

$$-\,2\,\rho_u\,\Big(\frac{u_1-m_u}{\sigma_u}\Big)\,\Big(\frac{u_2-m_u}{\sigma_u}\Big)+\Big(\frac{u_2-m_u}{\sigma_u}\Big)^2\Bigg]\Bigg\}.\quad(7.37)$$

By applying the transformation (7.22), the density $f^{W\dot{W}}_{\tau,\tau}(v_1,v_2)$ is obtained from the density $f^{WW}_{\tau,\tau+\Delta\tau}(u_1,u_2)$. Using equs. (7.25) and (7.37), we can write

$$f^{W\dot{W}}_{\tau,\tau}(v_1,v_2)=\frac{\Delta\tau}{2\pi\sigma^2_u\sqrt{1-\rho^2_u}}\ \exp\Bigg\{-\frac{1}{2(1-\rho^2_u)}\Bigg[\Big(\frac{v_1-m_u}{\sigma_u}\Big)^2-$$

$$-\,2\,\rho_u\,\Big(\frac{v_1-m_u}{\sigma_u}\Big)\,\Big(\frac{v_1+\Delta\tau\ v_2-m_u}{\sigma_u}\Big)+\Big(\frac{v_1+\Delta\tau\ v_2-m_u}{\sigma_u}\Big)^2\Bigg]\Bigg\}.\quad(7.38)$$

In Figure 7.11, density (7.38) is plotted along with the empirical joint probability density in order to check its fit to the data. Its behaviour is very good and the distance (7.26) is found D^2=0.0018.

In the sequel, the time-dependent transformation (7.27) is applied in order to obtain $f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ from the density $f_{\tau,\tau}^{W\dot{W}}(v_1, v_2)$. Using equs. (7.31) and (7.38), we find

$$
{}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) = \frac{\Delta\tau}{2\pi\sigma_u^2\sigma_\tau^2\sqrt{1-\rho_u^2}}
$$

$$
\exp\Bigg\{-\frac{1}{2(1-\rho_u^2)}\Bigg[\Big(\frac{s_1 - g_\tau - m_u\sigma_\tau}{\sigma_u\,\sigma_\tau}\Big)^2 - 2\,\rho_u\,\Big(\frac{s_1 - g_\tau - m_u\,\sigma_\tau}{\sigma_u\,\sigma_\tau}\Big)
$$

$$
\Big(\frac{(s_1 - g_\tau)\sigma_\tau + \Delta\tau\,\big[(s_2 - dg_\tau)\sigma_\tau - (s_1 - g_\tau)d\sigma_\tau\big] - m_u\sigma_\tau^2}{\sigma_u\,\sigma_\tau^2}\Big) +
$$

$$
+\Big(\frac{(s_1 - g_\tau)\sigma_\tau + \Delta\tau\,\big[(s_2 - dg_\tau)\sigma_\tau - (s_1 - g_\tau)d\sigma_\tau\big] - m_u\sigma_\tau^2}{\sigma_u\,\sigma_\tau^2}\Big)^2\Bigg]\Bigg\}. \quad (7.39)
$$

Athanassoulis, Soukissian and Stefanakos (1995b) assume directly that the density of $\big(X(\tau), \dot{X}(\tau)\big)$ is bivariate normal, i.e.

$$
{}^{(2)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) = \frac{1}{2\pi\sigma_{s_1}\sigma_{s_2}\sqrt{1-\rho_{s_1s_2}^2}}\;\exp\Bigg\{-\frac{1}{2(1-\rho_{s_1s_2}^2)}\Bigg[\Big(\frac{s_1 - m_{s_1}}{\sigma_{s_1}}\Big)^2 -
$$

$$
-\,2\,\rho_{s_1s_2}\,\Big(\frac{s_1 - m_{s_1}}{\sigma_{s_1}}\Big)\,\Big(\frac{s_2 - m_{s_2}}{\sigma_{s_2}}\Big) + \Big(\frac{s_2 - m_{s_2}}{\sigma_{s_2}}\Big)^2\Bigg]\Bigg\}, \quad (7.40)
$$

and find its parameters m_{s_1}, m_{s_2}, σ_{s_1}, σ_{s_2}, $\rho_{s_1s_2}$ in terms of g_τ, dg_τ, σ_τ, $d\sigma_\tau$, and of the zeroth and second moments of the spectrum of $W(\tau)$. For details, see Athanassoulis, Soukissian and Stefanakos (1995b).

In Figure 7.12, the densities ${}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ and ${}^{(2)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ are depicted for τ=1, i.e., for 0h on 1st January. Although ${}^{(2)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ does not seem to model equally well the correlation structure of $\big(X(\tau), \dot{X}(\tau)\big)$, it will be shown that this has not serious effect on the calculation of return periods.

Before proceeding with these calculations, let us first write equ. (7.39) in a more compact form, i.e.

$$ {}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) = Q(\tau)\ \exp[-P(s_1, s_2; \tau)], \tag{7.41} $$

where

$$ Q(\tau) = \frac{\Delta\tau}{2\pi\sigma_u^2\sigma_\tau^2\sqrt{1-\rho_u^2}}, \tag{7.42} $$

and

$$ P(s_1, s_2; \tau) = C(\tau)s_2^2 + D(s_1, \tau)s_2 + E(s_1, \tau). \tag{7.43} $$

The functions $C(\tau)$, $D(s_1, \tau)$, $E(s_1, \tau)$ are defined as follows

$$ C(\tau) = \frac{\Delta\tau}{2(1-\rho_u^2)\sigma_u^2\sigma_\tau^2}, \tag{7.44} $$

$$ D(s_1, \tau) = \frac{\Delta\tau\big[(E_2 + E_3) - \rho_u E_1\big]}{(1-\rho_u^2)\sigma_u^2\sigma_\tau^3}, \tag{7.45} $$

(7.46)

$$ E(s_1, \tau) = \frac{E_1^2 + (E_2 + E_3)^2 - 2\rho_u E_1(E_2 + E_3)}{2(1-\rho_u^2)\sigma_u^2\sigma_\tau^4}, \tag{7.47} $$

where

$$ \left.\begin{aligned} E_1 &= (s_1 - g_\tau - m_u\sigma_\tau)\sigma_\tau, \\ E_2 &= (\sigma_\tau - \Delta\tau d\sigma_\tau)s_1, \\ E_3 &= \Delta\tau\ (dg_\tau\sigma_\tau - g_\tau d\sigma_\tau) - g_\tau\sigma_\tau - m_u\sigma_\tau^2. \end{aligned}\right\} \tag{7.48} $$

Note that the function $P(s_1, s_2; \tau)$ is rather regarded as a function of s_2, since we want finally to integrate over all values of s_2. The variables s_1 and τ will be treated as parameters.

Let us now turn to the return-period calculations. Introducing equ. (7.43) into (7.41), the probability density ${}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ is written

$$ {}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) = Q(\tau)\ \exp\big[-C(\tau)s_2^2 - D(s_1, \tau)s_2 - E(s_1, \tau)\big], \tag{7.49} $$

or, equivalently,

$$ {}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2) = B(s_1, \tau)\ \exp\big[-C(\tau)s_2^2 - D(s_1, \tau)s_2\big], \tag{7.50} $$

where

$$ B(s_1, \tau) = Q(\tau)\ \exp\big[-E(s_1, \tau)\big]. \tag{7.51} $$

Using equ. (7.50), relation (7.17) is written

$$ M(s_1; t_1, t_2) = \int_{t_1}^{t_2}\int_{0}^{+\infty} s_2\ B(s_1, \tau)\ \exp\big[-C(\tau)s_2^2 - D(s_1, \tau)s_2\big] ds_2 d\tau. \tag{7.52} $$

Since $C(\tau)>0$ (see equation (7.44)), the integral in the right-hand side of (7.52) is convergent and integration with respect to s_2 can be performed analytically, by using the formula (see Gradshteyn and Ryzhik (1980, p. 338))

$$ \begin{aligned} &\int_0^{+\infty} s_2\ \exp\big[-C(\tau)s_2^2 - D(s_1, \tau)s_2\big]\ ds_2 = \\ &\quad = \frac{1}{2C(\tau)} - \frac{D(s_1, \tau)}{4C(\tau)}\sqrt{\frac{\pi}{C(\tau)}}\exp\left[\frac{D(s_1, \tau)^2}{4C(\tau)}\right]\left[1 - \Phi\left(\frac{D(s_1, \tau)}{2\sqrt{C(\tau)}}\right)\right] \equiv I(s_1, \tau), \end{aligned} \tag{7.53} $$

where $\Phi(\cdot)$ denotes the standard error function. Using this result, $M(s_1; t_1, t_2)$ is finally written in the form

$$ M(s_1; t_1, t_2) = \int_{t_1}^{t_2} B(s_1, \tau)\ I(s_1, \tau)\ d\tau. \tag{7.54} $$

We recall here that $B(s_1, \tau)$ is defined by equ. (7.51), while $I(s_1, \tau)$ is defined by equ. (7.53), the quantities $C(\tau)$ and $D(s_1, \tau)$ being defined by equs. (7.44) and (7.46).

Now, given a starting time t_1 and a level value x^*, the corresponding return period $T_R(x^*, t_1)$ is simply calculated by considering $M(x^*; t_1, t)$ as a function of t (the upper limit of integration), and determining $T_R(x^*, t_1)$ so that

$$ \int_{t_1}^{t_1+T_R(x^*,t_1)} B(s_1, \tau)\ I(s_1, \tau)\ d\tau = 1. \tag{7.55} $$

See also Figure 7.6, where an illustration of this calculation is given.

Performing similar analysis, Athanassoulis et al. (1995b) obtained expressions for $B(s_1, \tau)$ and $I(s_1, \tau)$ using density ${}^{(2)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$. We will refer to the results based on (i) the density

${}^{(1)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ as the "transformed Gaussian approach", (ii) the density ${}^{(2)}f_{\tau,\tau}^{X\dot{X}}(s_1, s_2)$ as the "*a priori* Gaussian approach".

In Figure 7.13, the return periods based on these two approaches are depicted. Note that the vertical scale is not the same with the one in Figures 7.11 and 7.12, because the transformation x^*=ln(H_S^*+1m) has been performed. Although there is a small overestimation of the return periods based on the *a priori* Gaussian approach, the results of the two variants of MENU method are too close, if they are compared to, e.g., the dispersed results obtained by the four variants of Gumbel's method.

In any case, the data, we have at hand, are eight-year long only. In order to validate our results, we should have at least 50 or, even better, 100 years of data. Unfortunately, there are no such a long series of data in ocean engineering. In the next section, we perform a sensitivity analysis of the MENU method using an artificially produced time series with length 50 years. Return-period calculations are performed for five 10-year long parts of the (50-year long) initial time series, and results are compared with the ones produced by means of the initial time series.

7.5 Sensitivity analysis of MENU method

In order to validate the robustness of MENU method, a sensitivity analsyis is performed, the results of which are compared with the corresponding ones obtained by means of classical methods. For this, very long-term time series of, say, 50 (or 100) years should be analyzed. To the author's knowledge, there are no such large datasets available in ocean engineering community.

Instead, a simulation procedure is set up, in order to produce a very long time series. Special effort is made to retain the nonstationary character exhibited by real time series. See also Section 5.6, where a similar procedure is used for the production of a long-term incomplete time series.

It has been decided to produce a 50-year long nonstationary Gaussian population, since it is well known (Castillo, 1988; Soukissian, 1995) that the normal distribution belongs to the domain of attraction of the FT-I (Gumbel) extreme-type distribution. That is, the distribution of maxima of a Gaussian population is the FT-I distribution.

Further, a sequence of uncorrelated Gaussian random variables $\{\varepsilon(\tau_i),\ i = 1, 2, \ldots, I\}$ is generated with zero mean and variance σ_ε^2=0.108. To accomplish this task, the routine NORM-RND of the MATLAB® language is used. Then, letting P=Q=2, a linear ARMA sequence is produced by means of equ. (5.29) of Section 5.4 with parameters a_1=1.7093, a_2=−0.726, b_1=−0.6277, b_2=−0.0321. The obtained series $W(\tau)\equiv W_{\text{ARMA}}(\tau)$ is stationary, and the values of σ_ε^2, a_1, a_2, b_1, b_2 are taken so that to be typical for sea-wave applications. Further, a time series $X(\tau)$, as this is defined by equ. (7.19), is formed by using deterministic functions $G(\tau)$

and $\sigma(\tau)$ also typical for sea-wave applications. Recall that, the function $G(\tau)$ is defined by equ. (7.28), and the functions $\mu(\tau)$ and $\sigma(\tau)$ are periodic with period one year. In this way, $X(\tau)$ is given the structure of a nonstationary stochastic process.

In the sequel, return-period calculations are performed by means of a) the MENU method, b) the variants of Gumbel approach examined in Section 7.4.1, and c) the POT method for various threshold values. The methods are applied to (i) the initial 50-year long time series, (ii) five 10-year long subseries of the initial series (i). Then, the results obtained from datasets (ii) are compared with the ones produced by dataset (i), which are considered to be more accurate. Plots from all methods have been standardized, i.e., they have same format and same scaling, for comparison purposes. Results based on the 50-year long series are plotted by continuous lines, whereas results based on the various decades are drawn by dotted lines and are numbered consecutively (1–5).

In Figures 7.14 and 7.15, return-period results are shown for the two variants of MENU method presented in Section 7.4.2, namely the transformed Gaussian approach and the *a priori* Gaussian approach. The dispersion of the results based on the various decades is not significant, and they are very close also to the ones based on the 50-year long series.

In Figures 7.16–7.18, results from the following three variants of Gumbel approach are given: (i) the probability paper method, (ii) the least-squares return period relative error method, (iii) the probability paper method with Gringorten's plotting position. Note that, the fourth method (maximum likelihood) failed to give reasonable results, and, thus, it is excluded from the final presentation. In the above mentioned figures, the variation between the five decades is much greater than the one observed in the MENU's results; compare with Figures 7.14 and 7.15. The same can be concluded by observing the estimates of the parameters of FT-I (Gumbel) distribution given in Tables 7.4 and 7.5. Especially, scale parameter δ, although it exhibits remarkable stability with respect to the chosen estimation method, it presents a significant variation between the various decades.

In Figure 7.19, results of the POT method are depicted. Here, we have one more complexity: for each dataset, results are obtained for a number of threshold values X_u. In the present study, the following threshold values have been chosen: X_u=1.7703, 1.8099, 1.8520, 1.8838, 1.9299, 1.9726, 1.9919. In this method too, the large variation is more than apparent. In order to make easier the comparison with the other mentioned methods, for each decade and for the 50-year series, we take average over the results obtained for all threshold values. In Figure 7.20, these averaged return-period results are shown, retaining the variability depicted by the initial ones in Figure 7.19. Further, the same is true by considering the estimated parameters of the Generalized Pareto Distribution ginen in Tables 7.6 and 7.7.

Concluding this section, we are in a position to say that MENU method, in spite of the large amount of data and the computational burden it requires, gives comparable results indifferently

of what dataset is used. On the other hand, both Gumbel approach and POT method exhibit a much greater variability between the results based on the decades and the ones based on the initial 50-year one. That is, if we have at hand a 10-year long time series, which is the usual case in ocean engineering applications, we cannot be certain that our predictions of return periods based on classical approaches are reliable enough. The contrary holds true for the MENU method.

7.6 Conclusions

In this chapter, a new method for calculating return periods of various level values from nonstationary time series is used, along with the nonstationary modelling presented in Chapters 3, 4 and 5. The key idea of the method is a new definition of the return period concept based on the MEan Number of Upcrossings of a level value x^* (MENU method). This method is applied for the first time to non-Gaussian data, and return-period calculations for various level values are produced by means of real data.

Results are compared with results obtained from a) Gumbel approach, and b) POT method; see Figures 7.7 and 7.8. Further, the method is also tested against an earlier variant of its own, the so-called *a priori* Gaussian approach (Athanassoulis et al., 1995b); see Figure 7.13. All comparison results show that predictions based on MENU method are in agreement with predictions from traditional methods.

Finally, a sensitivity analysis of the method is performed using an artificially produced very long (50 years) time series. Return-period calculations are performed by means of this series, as well as of five 10-year long subseries of it. Results are again compared with the ones produced by traditional methods (Gumbel approach and POT method). One may conclude that MENU method is much more stable than classical methods. Cf. Figures 7.14 and 7.15 with Figures 7.16, 7.17, 7.18 and 7.20.

7.7 Tables and Figures

Table 7.1: Moments of the empirical and the analytic (Plackett with both marginals lognormal) joint probability density function of $\left(W(\tau), \dot{W}(\tau)\right)$

	Empirical pdf	Analytic pdf
Mean value of $W(\tau)$	0.0029	-0.0647
Standard deviation of $W(\tau)$	1.1578	1.0273
Mean value of $\dot{W}(\tau)$	0.0001	0.0001
Standard deviation of $\dot{W}(\tau)$	0.0548	0.0464
Correlation coefficient $\rho_{W\dot{W}}$	-0.1418	-0.1292

Table 7.2: Estimated parameters of FT-I (Gumbel) distribution applying (i) the probability paper method, (ii) the least-squares return period relative error method, (iii) the probability paper method with Gringorten's plotting position, (iv) the maximum likelihood method. [Data: population of the annual maxima]

Method	δ	λ
(i)	3.3826	9.1281
(ii)	3.6721	8.7150
(iii)	2.7817	9.2817
(iv)	1.0000	9.0416

Table 7.3: Estimated parameters of the Generalized Pareto Distribution for various threshold values X_u. [Data: population of the annual maxima]

Threshold value X_u (m)	k	a
8	-0.0058	1.7996
9	-0.0519	1.6946
10	0.2047	2.3851
11	0.4652	3.0546
12	0.5204	2.7850

Table 7.4: Estimated values of parameter δ of FT-I (Gumbel) distribution (see equ. (7.6)) applying (i) the probability paper method, (ii) the least-squares return period relative error method, (iii) the probability paper method with Gringorten's plotting position, (iv) the maximum likelihood method, to the annual maxima of a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a).

Method	50-year	1st decade	2nd decade	3rd decade	4th decade	5th decade
(i)	0.0555	0.0725	0.0524	0.0609	0.0871	0.0254
(ii)	0.0546	0.0719	0.0530	0.0523	0.0885	0.0246
(iii)	0.0519	0.0617	0.0444	0.0517	0.0728	0.0214

Table 7.5: Estimated values of parameter λ of FT-I (Gumbel) distribution (see equ. (7.6)) applying (i) the probability paper method, (ii) the least-squares return period relative error method, (iii) the probability paper method with Gringorten's plotting position, (iv) the maximum likelihood method, to the annual maxima of a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a).

Method	50-year	1st decade	2nd decade	3rd decade	4th decade	5th decade
(i)	1.8401	1.8659	1.8072	1.8356	1.8539	1.8423
(ii)	1.8384	1.8610	1.8058	1.8443	1.8417	1.8420
(iii)	1.8410	1.8684	1.8092	1.8378	1.8576	1.8433

Table 7.6: Estimated values of parameter k of the Generalized Pareto Distribution (see equ. (7.8)) for various threshold values X_u. The data used are the annual maxima of a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a).

Threshold value X_u	50-year	1st decade	2nd decade	3rd decade	4th decade	5th decade
1.7703	0.0662	0.2117	0.0567	0.2957	0.0149	0.4083
1.8099	-0.0295	0.3165	-0.0971	0.2580	-0.0636	0.3238
1.8520	0.0491	0.6687	1.0000	0.2392	0.0840	0.2169
1.8838	0.2885	0.8391	1.0000	0.1002	0.5337	1.0000
1.9299	0.2277	0.7874	—	1.0000	0.0000	—
1.9726	-0.0276	0.0176	—	—	1.0000	—
1.9919	1.0000	1.0000	—	—	1.0000	—

Table 7.7: Estimated values of parameter a of the Generalized Pareto Distribution (see equ. (7.8)) for various threshold values X_u. The data used are the annual maxima of a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a).

Threshold value X_u	50-year	1st decade	2nd decade	3rd decade	4th decade	5th decade
1.7703	0.0563	0.0827	0.0329	0.0661	0.0600	0.0574
1.8099	0.0463	0.0853	0.0249	0.0508	0.0530	0.0355
1.8520	0.0528	0.1113	0.0754	0.0378	0.0687	0.0182
1.8838	0.0723	0.1107	0.0436	0.0237	0.1186	0.0197
1.9299	0.0533	0.0679	—	0.0361	0.0518	—
1.9726	0.0314	0.0140	—	—	0.1143	—
1.9919	0.0950	0.0230	—	—	0.0950	—

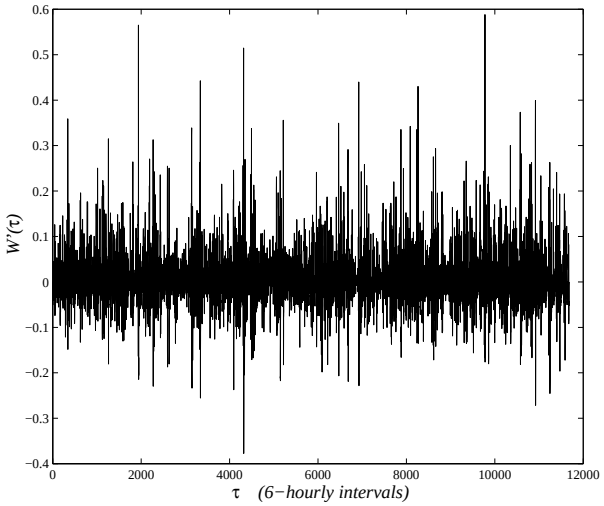


Figure 7.1: Time series of $\dot{W}(\tau)$.

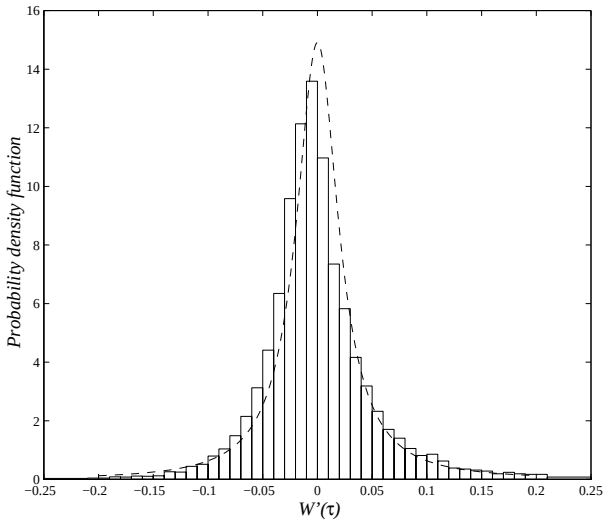


Figure 7.2: Empirical probability density function of $\dot{W}(\tau)$ along with an analytic representation resulting from the joint probability density function of $\left(W(\tau), \dot{W}(\tau)\right)$ by intergating over all $W(\tau)$'s.

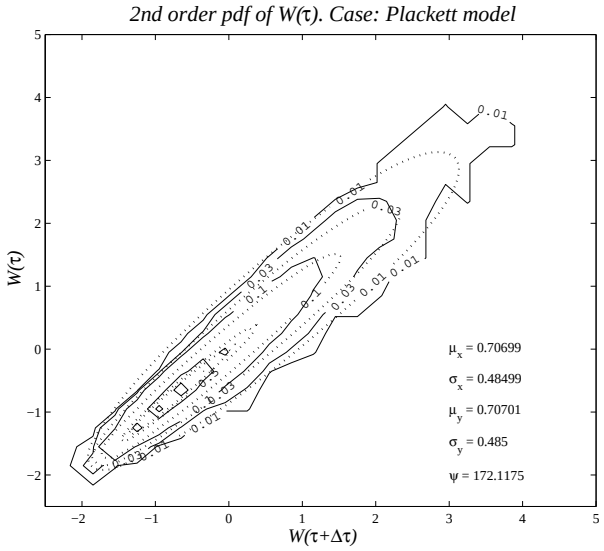


Figure 7.3: Contour plot of a) the empirical joint probability density of $\big(W(\tau), W(\tau+\Delta\tau)\big)$, b) the bivariate Plackett model with both univariate marginals lognormal. [Contour levels: 0.01, 0.03, 0.10, 0.50, 0.75]

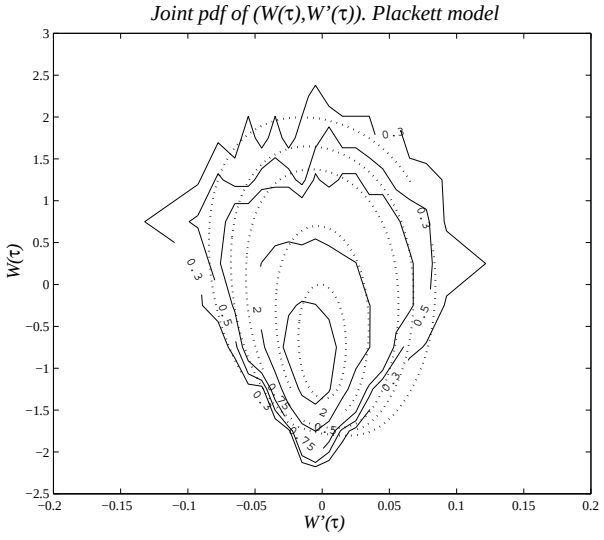


Figure 7.4: Contour plot of a) the empirical joint probability density of $\big(W(\tau), \dot{W}(\tau)\big)$, b) the derived (transformed) analytic model. [Contour levels: 0.30, 0.50, 0.75, 2.00, 5.00]

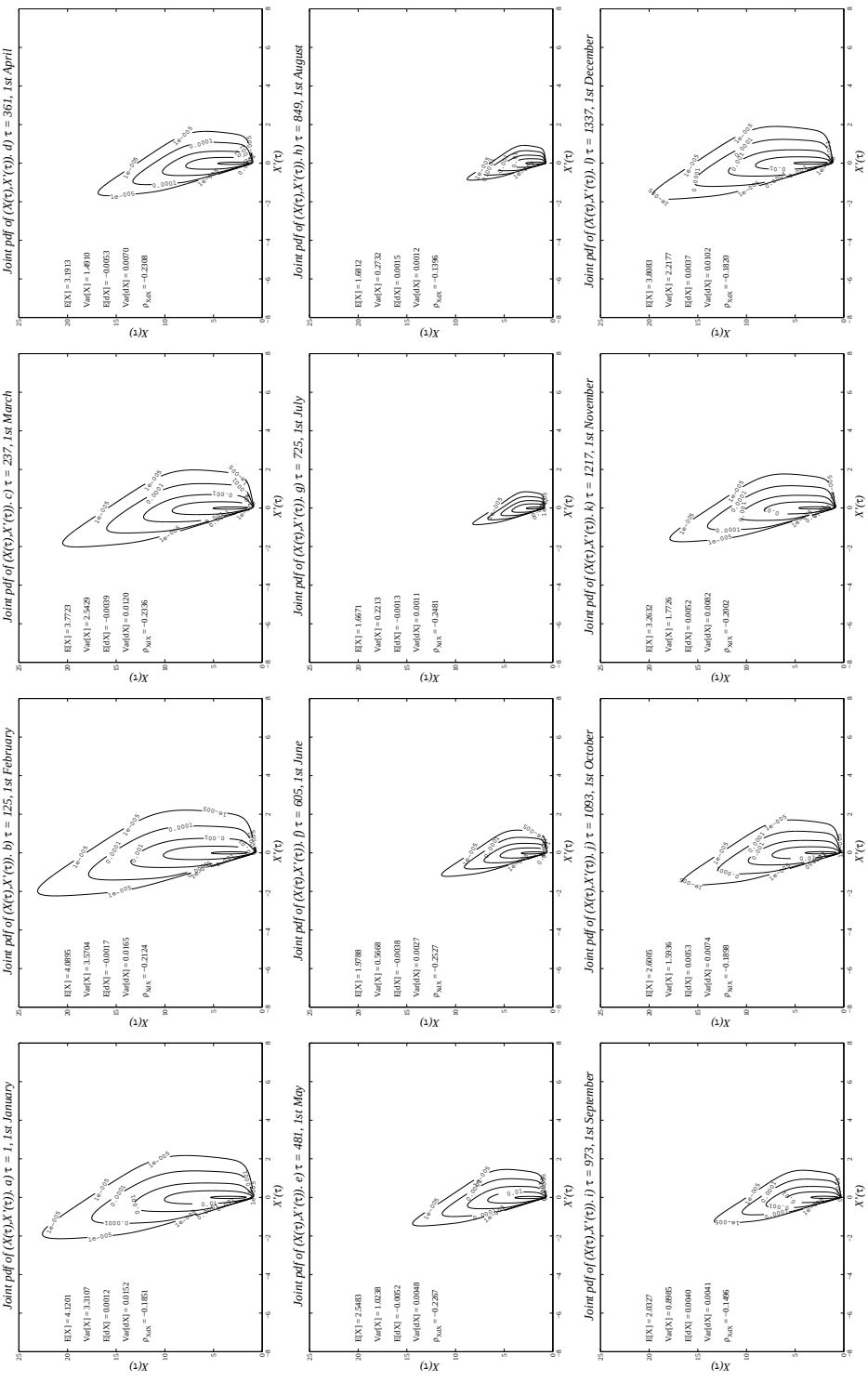


Figure 7.5: Contour plots of the time-dependent joint probability density of $\left(X(\tau), \dot{X}(\tau)\right)$ for various τ-values within the year, as follows: a) τ=1, 1st January, b) τ=125, 1st February, c) τ=237, 1st March, d) τ=361, 1st April, e) τ=481, 1st May, f) τ=605, 1st June, g) τ=725, 1st July, h) τ=849, 1st August, i) τ=973, 1st September, j) τ=1093, 1st October, k) τ=1217, 1st November, l) τ=1337, 1st December. [Contour levels: 0.00001, 0.0001, 0.001, 0.01, 0.75]

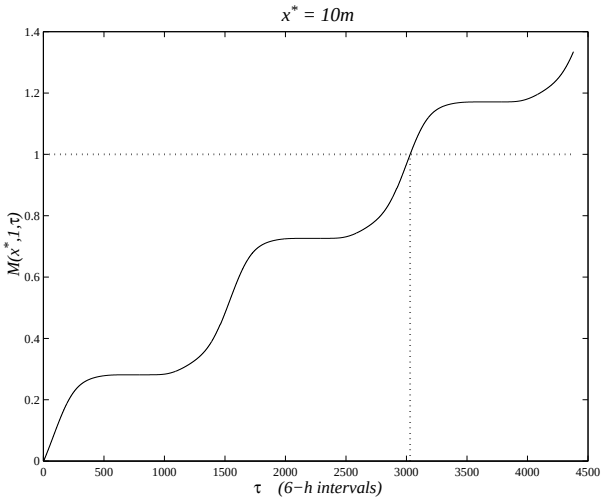


Figure 7.6: Calculation of the quantity $M(x^*, 1, \tau)$ for three yearly periods and level value $X(\tau)$=x^*=10m.

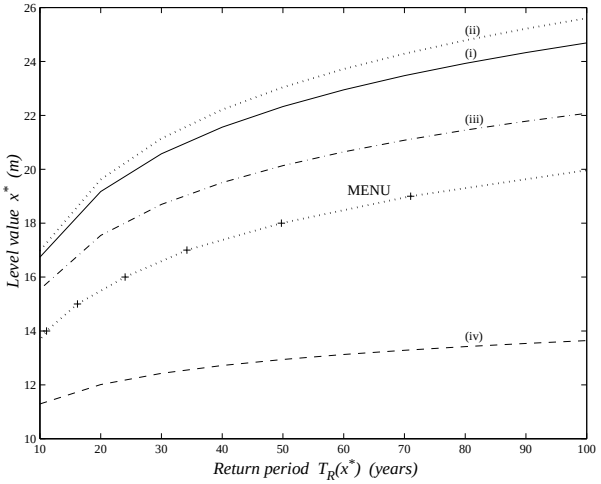


Figure 7.7: Return periods calculated by means of (i) the probability paper method, (ii) the least-squares return period relative error method, (iii) the probability paper method with Gringorten's plotting position, (iv) the maximum likelihood method, applied to the annual maxima. The results are compared with the return periods obtained by means of the MENU method.

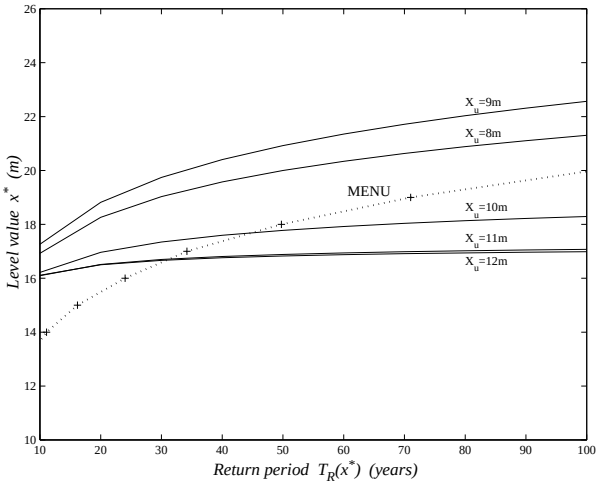


Figure 7.8: Return periods by means of POT method for various threshold values X_u=8, 9, 10, 11, 12m, compared with the results of MENU method.

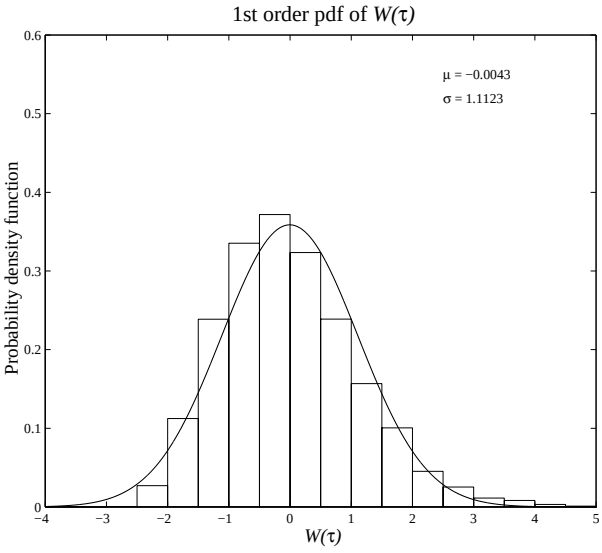


Figure 7.9: Empirical probability density function of $W(\tau)$ after the logarithmic transformation along with the corresponding normal.

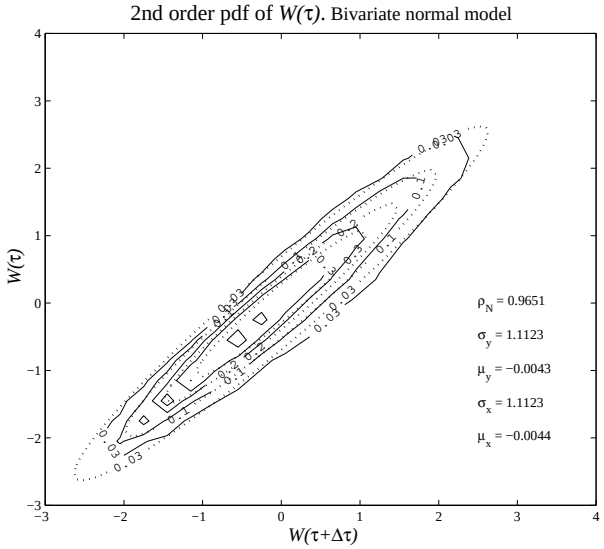


Figure 7.10: Contour plot of a) the empirical joint probability density of $\big(W(\tau), W(\tau+\Delta\tau)\big)$ after the logarithmic transformation, b) the bivariate normal. [Contour levels: 0.03, 0.10, 0.20, 0.30, 0.50, 0.75]

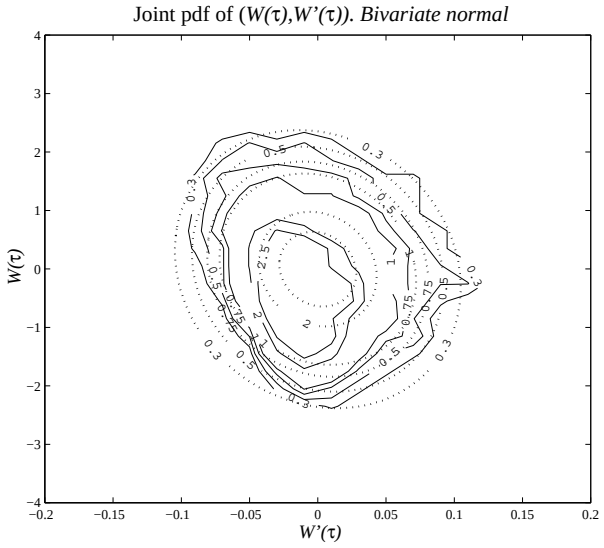


Figure 7.11: Contour plot of a) the empirical joint probability density of $\big(W(\tau), \dot{W}(\tau)\big)$ after the logarithmic transformation, b) the derived (transformed) analytic model. [Contour levels: 0.30, 0.50, 0.75, 1.00, 2.00, 2.50]

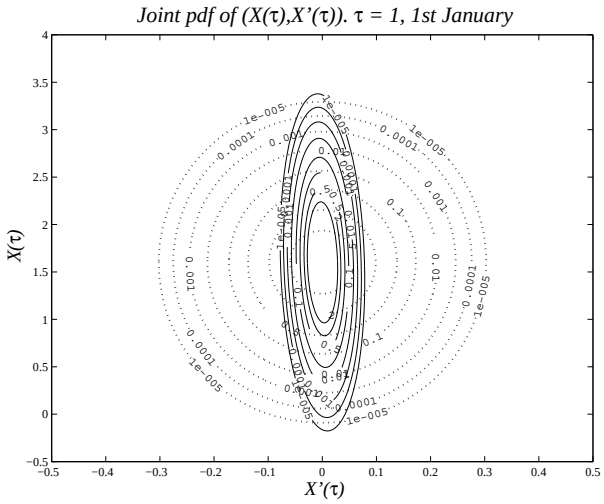


Figure 7.12: Contour plot of the time-dependent joint probability density of $\left(X(\tau), \dot{X}(\tau)\right)$ after the logarithmic transformation for τ=1 six-hourly interval, i.e., for 0h on 1st January. [Contour levels: 0.00001, 0.0001, 0.001, 0.01, 0.10, 0.50, 2.00, 5.00]

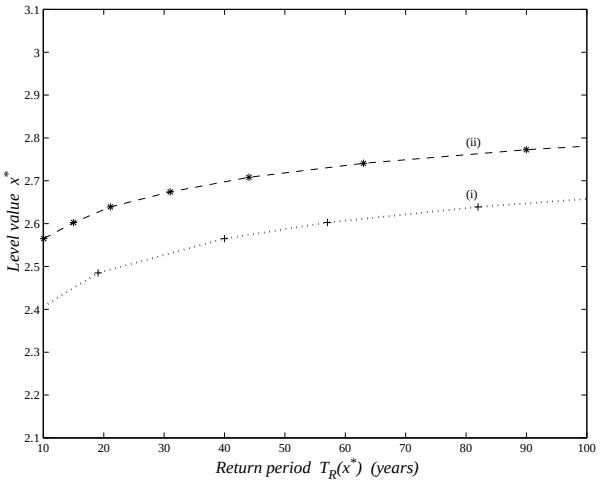


Figure 7.13: Return periods calculated by means of the MENU method: (i) based on the "transformed Gaussian approach", (ii) based on the "*a priori* Gaussian approach".

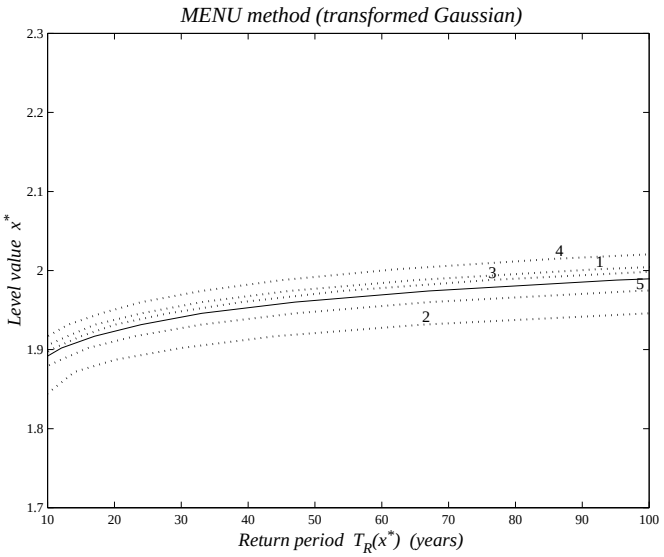


Figure 7.14: Return periods calculated by means of the MENU method (transformed Gaussian approach), applied to (i) the 50-year long time series (continuous line), (ii) the five 10-year long subseries of the series (i) (dotted lines).

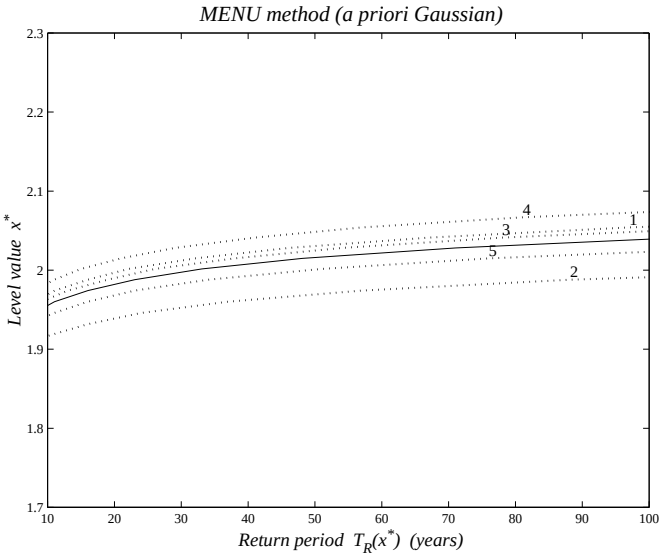


Figure 7.15: Return periods calculated by means of the MENU method (*a priori* Gaussian approach), applied to (i) the 50-year long time series (continuous line), (ii) the five 10-year long subseries of the series (i) (dotted lines).

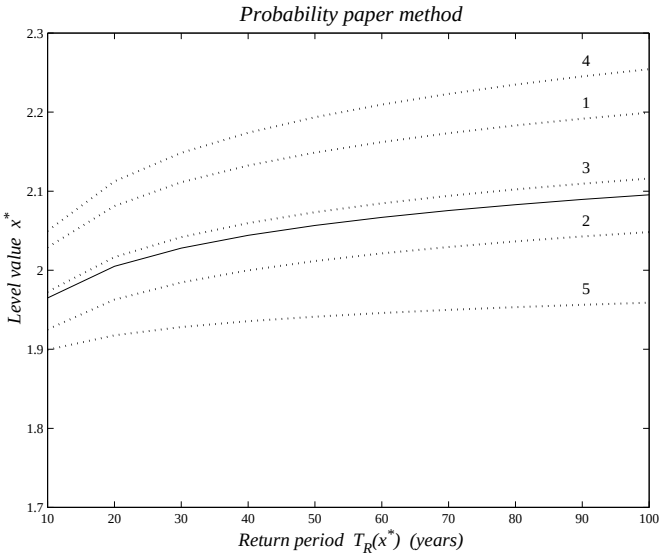


Figure 7.16: Return periods calculated by means of the probability paper method, applied to the annual maxima of (i) the 50-year long time series (continuous line), (ii) the five 10-year long subseries of the series (i) (dotted lines).

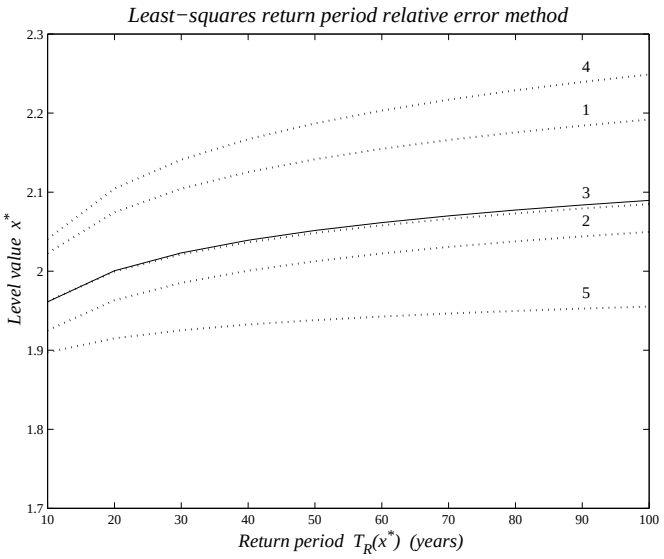


Figure 7.17: Return periods calculated by means of the least-squares return period relative error method, applied to the annual maxima of (i) the 50-year long time series (continuous line), (ii) the five 10-year long subseries of the series (i) (dotted lines).

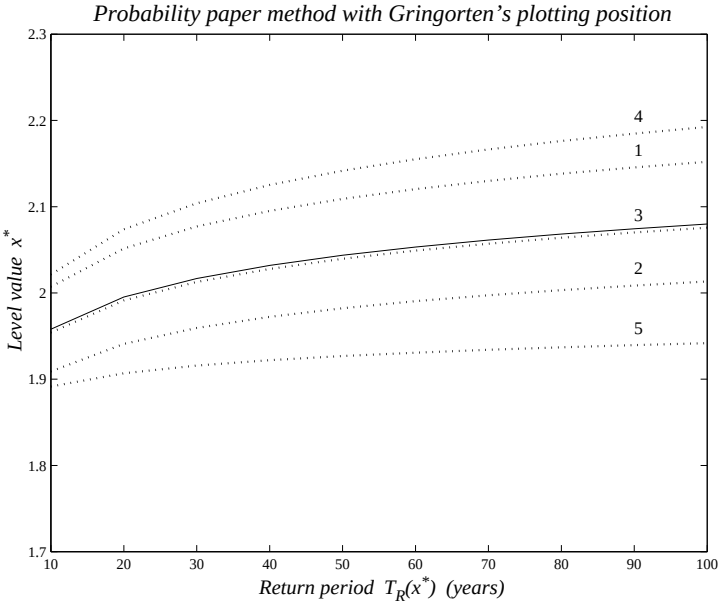


Figure 7.18: Return periods calculated by means of the probability paper method with Gringorten's plotting position, applied to the annual maxima of (i) the 50-year long time series (continuous line), (ii) the five 10-year long subseries of the series (i) (dotted lines).

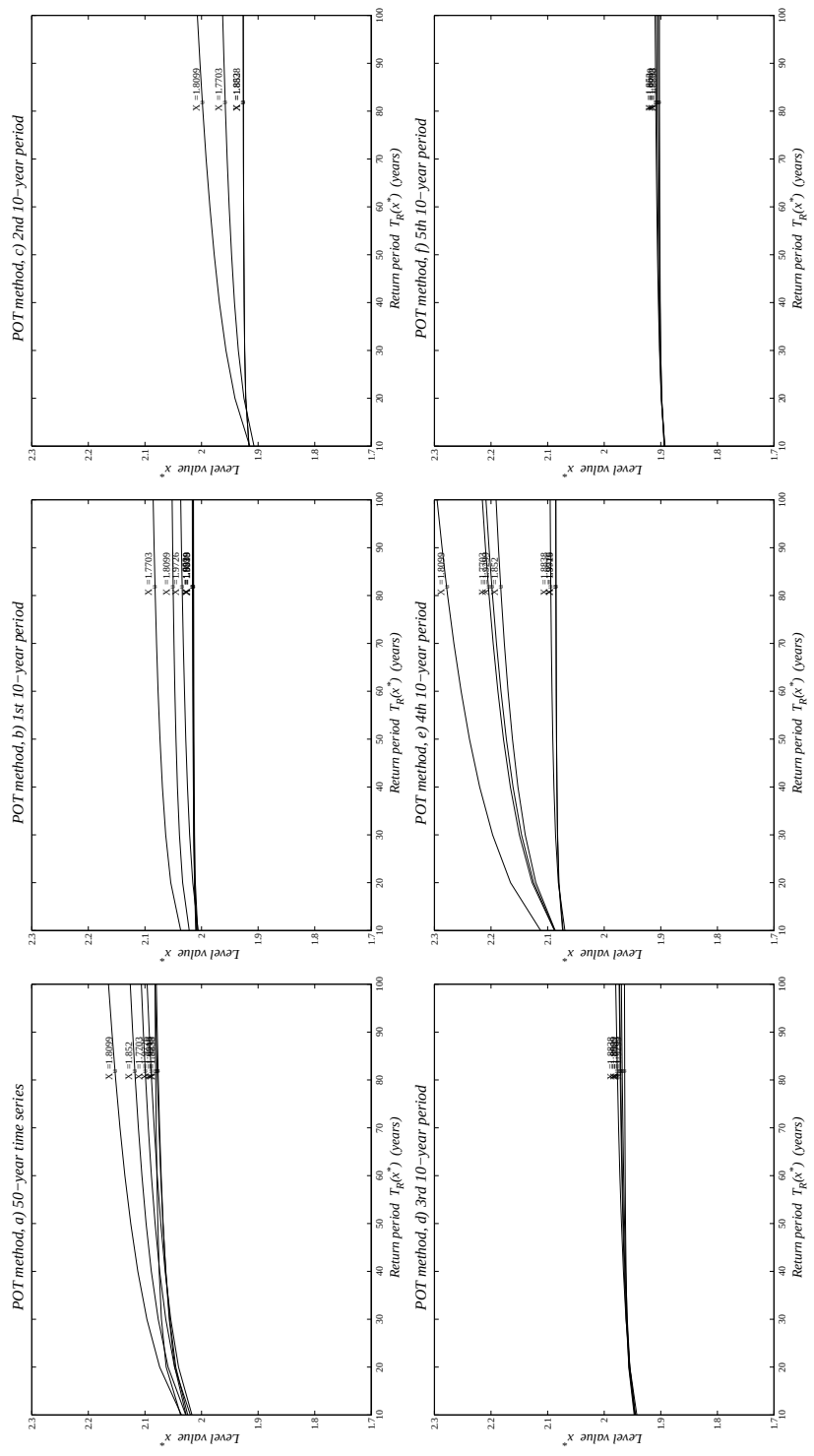


Figure 7.19: Return periods calculated by means of the POT method, applied to a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a), for threshold values X_u= 1.7703, 1.8099, 1.852, 1.8838, 1.9299, 1.9726, 1.9919.

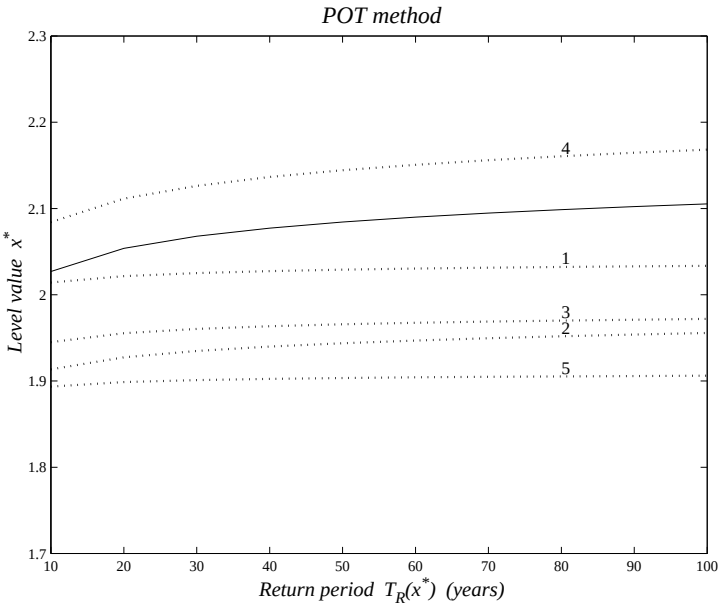


Figure 7.20: Return periods calculated by means of the POT method, applied to a) the 50-year long time series, b) the 1st 10-year period of the series a), c) the 2nd 10-year period of the series a), d) the 3rd 10-year period of the series a), e) the 4th 10-year period of the series a), f) the 5th 10-year period of the series a), for the mean value of the threshold values examined.

Chapter 8

Conclusions and proposal for further studies

In the present work, an attempt is initiated to model the various characteristics exhibited by a many-yearlong time series of a wave parameter. These features can be summarized as follows: (i) a fluctuating character of the sequence of observations that prevents any exact reproducibility of the time series and calls for stochastic modelling, (ii) a statistical dependence between observations separated by a relative small time period, (iii) a pronounced seasonal variability within the annual cycle, (iv) a yearly statistical periodicity induced by the meteorological annual cycle, and possibly (v) an overyear trend, which is difficult to identify.

For this purpose, many-yearlong time series of significant wave height have been analyzed and modelled. It has been found that these time series can be modelled as periodically correlated (cyclostationary) stochastic processes of the form

$$X(\tau) = \overline{X}_{\rm tr}(\tau) + \mu(\tau) + \sigma(\tau)W(\tau), \tag{8.1}$$

where $W(\tau)$ is a zero-mean stationary (in general non-Gaussian) stochastic process, $\mu(\tau)$ and $\sigma(\tau)$ are deterministic, yearly periodic functions representing the mean seasonal effects, and $\overline{X}_{\rm tr}(\tau)$ is a deterministic function representing the overyear climatic trend. Note that, the proposed modelling is designed so that to describe the exhibited features by the raw data. No attempt has been made to use additional theoretical assumptions or any other extra characteristic not depicted in the existing time series of wave data.

In Chapter 4, application of the modelling (8.1) to hindcast data of significant wave height for five locations in the north Atlantic Ocean led to the following conclusions:

1. It is very likely that an increasing overyear trend (1-4 cm yr^{-1}) is present in the examined period of time (1956-1975). This is in accordance with the findings of other authors (Carter and Draper, 1988; Neu, 1984; Bacon and Carter, 1991; Watanabe et al., 1992),

2. The interannual variability of the studied H_S time series exhibits a remarkable site independence. It has been found that the variation coefficient of annual mean values is approximately 0.11 (0.106–0.118, dependent on the site),

3. The seasonal variability is well characterized by a 10-yearlong record or longer (see Figure 4.5),

4. The time-lag autocorrelation coefficient function $\rho^{(1)}_{WW}(r)$ and the spectral density $S_{WW}(f)$ of the seasonally standardized time series also show a remarkable site independence (see Figures 4.8 and 4.9).

Modelling (8.1) is further extended to take into account the effect of missing values. In Chapter 5, a new methodology for the missing-value completion of an incomplete nonstationary time series is presented and applied to simulated and to measured wave data. The key idea is to perform the completion after having removed not only the deterministic components of the series, but also the correlation structure of the stationary series. After a detrending and seasonal standardization, a low-order ARMA model is fitted to the (incomplete) residual stationary series using appropriate estimation techniques. The raw spectrum, calculated as the Fourier transform of a consistent estimate of the corresponding autocovariance function [Step 3 of the missing-value procedure; see Section 5.5], is used for the estimation of the ARMA coefficients and the variance of the residuals. The incomplete time series of uncorrelated residuals is then completed by means of simulated data with the same first-order probability structure, and used, along with the ARMA model and the estimated deterministic components, to construct a new time series of the same structure without missing values.

The above procedure has been applied to two simulated nonstationary time series with different percentage of missing values (16.5% and 33%, respectively), and to two measured time series of wave parameters with different percentage of missing values (16.5% and 22.3%, respectively). Comparisons of various statistical characteristics of the initial (incomplete) and reconstructed (completed) time series are satisfactory in both cases. Some general conclusions, concerning these results, are the following:

1. The estimates of the seasonal characteristics calculated from the incomplete and the reconstructed time series are in a very good agreement with the initial one. Even in the case of the 33% missing-value pattern (see Figures 5.3, 5.4, 5.11). Also, estimates from the incomplete and the reconstructed measured time series are in a very good agreement (see Figures 5.15, 5.20).

2. Spectral density estimate $\widehat{S}_{WW}(f)$ is more sensitive to an increase of missing values. However, it manages to describe satisfactorily the analytic density $S_{\mathsf{ARMA}}(f)$ (see Figures 5.13, 5.16,5.21).

3. The residuals of the measured time series are found to be in both cases (after appropriate transformation) Gaussian (or nearly ones) (see Figures 5.17, 5.22). This fact advocates for the ARMA modelling of the stationary residual part $W_{\mathrm{e.v.}}(\tau)$.

4. If $W_{\mathrm{e.v.}}(\tau)$ is found to be non-Gaussian, it should be then (see also Section 5.4) transformed to a Gaussian one before a linear ARMA model is fitted. The most popular method (and the one used in the present work) is the logarithmic transformation (Salas, 1993), especially useful when the original time series follows a lognormal distribution. Salas (Salas et al., 1980; Salas, 1993) give an excellent survey of the various possible transformations of a hydrologic time series, which is very similar to a wave data time series, in order to obtain a stationary (nearly) Gaussian time series before fitting an ARMA model. It remains to each user to decide which trasformation to chose.

The first- and second-order probability density functions of nonstationary model (8.1) are produced in Chapter 6. Two datasets are used for this study: one with hindcast data, and one with *in situ* measurements.

First, the first- and second-order densities of $W(\tau)$ are calculated. The first-order density is modelled by fitting the three-parameter lognormal to the first-order empirical density. Further, an analytic (non-Gaussian) probability model, the so-called Plackett model, is used for the representation of the second-order density. This analytic form is a versatile tool, which can incorporate any form of univariate marginals and the correlation structure is controlled by a reasonable number of parameters. For comparison purposes, the standard bivariate lognormal density is also examined. The Plackett model seems to offer a better performance regarding the description of the initial data, and an increased flexibility regarding the shape of the distribution, at the same computational cost.

Then, by considering modelling (8.1) as a linear time-varying transformation, the first- and second-order densities of $X(\tau)$ are derived from the corresponding densities of $W(\tau)$. In Figures 6.13 and 6.14, the first-order, and in Figures 6.15 and 6.16, the second-order densities of $X(\tau)$ are depicted.

In Chapter 7, a new method for calculating return periods of various level values from nonstationary time series is used, along with the nonstationary modelling (8.1). This method is applied for the first time to non-Gaussian data, and return-period calculations for various level values are produced by means of real data.

Results are compared with results obtained from a) Gumbel approach, and b) POT method; see Figures 7.7 and 7.8. Further, the method is also tested against an earlier variant of its own, the so-called *a priori* Gaussian approach (Athanassoulis et al., 1995b); see Figure 7.13. All comparison results show that predictions based on MENU method are in agreement with predictions from traditional methods.

A sensitivity analysis of the method is also performed using an artificially produced very long (50 years) time series. Return-period calculations are performed by means of this series, as well as of five 10-year long subseries of it. Results are again compared with the ones produced by traditional methods (Gumbel approach and POT method). One may conclude that MENU method is much more stable than classical methods. Cf. Figures 7.14 and 7.15 with Figures 7.16, 7.17, 7.18 and 7.20.

Apart from the above results, the proposed model can also be used for various applications in the context of probabilistic wave climate prediction and engineering design, as, for example

1. Long-term stochastic simulation of wave climate. Such kind of results have been produced in the present work;
2. Construction of complete versions of measured time series of wave data, which are needed as input in numerical models concerning large-scale flow computations (Scheffner and Borgman, 1992);
3. Enhanced calculation of sea state duration;
4. Extreme-value predictions for marine structure responses; and
5. Input to coastal morphodynamics models. Especially, models used for the sediment transport, beach corrosion etc.

The present model can potentially be extended to various directions. The more interesting are the following:

1. Inclusion of multiple periods (apart from the yearly one) in the periodically varying components describing other environmental periodicities, provided that the data exhibits this feature;
2. Application of the model (8.1) to responses of systems excited by nonstationary forcing;
3. Use of other, more general, multivariate probability models, for the representation of the second-order probability struture of the model (8.1); and
4. Extension of the model to bivariate and, eventually, multivariate time series. This is the more ambitious and, at the same time, the more promising direction, which is under study.

It should be also stressed that the modelling of the second-order probability structure (i) is fairly general, (ii) is non-Gaussian, (iii) can incorporate information acquired in previous stages (e.g., univariate marginals), and (iv) is considered in ocean engineering applications for the first time.

Finally, let it be noted that model (8.1) can also be applied to many other ocean and/or atmospheric parameters, such as, e.g., wave period, wave and wind direction, wind speed, sea temperature, air temperature, humidity, sea level, etc., which also exhibit seasonally varying randomness.

Bibliography

Abazaliev, A.K.J., 1968, Characteristic coefficients of two-dimensional distributions and their applications, *Soviet Math. Doklady*, vol. 9, pp. 52–56.

Abrahams, J., 1986, A survey of recent progress on level crossing problems for random processes, in *Communications and Networks. A survey of recent advances*, I.F. Blake and H.V. Poor, eds., Springer-Verlag, pp. 6–25.

Alekseev, V.G., 1991, On spectral density estimates of a Gaussian periodically correlated random field, *Probability and Mathematical Statistics*, vol. 11, no. 2, pp. 157–167.

Ali, M.M., Mikhail, N.N., and Haq, M.S., 1978, A class of bivariate distributions including the bivariate logistic, *Journal of Multivariate Analysis*, vol. 8, pp. 405–412.

Anderson, T.W., 1971, *The statitical analysis of time series*, John Wiley & Sons, New York.

Athanassoulis, G.A. and Soukissian, T.H., 1991, Return periods of extreme sea states in the greek seas, in *Proceedings of the Second Conference on Enironmental Science and Technology*, Th. Lekkas, ed., University of Aegean, Environmental Studies Dept., Mytilini, Greece, pp. 770–778, (in Greek).

Athanassoulis, G.A. and Soukissian, T.H., 1993a, A new model for long-term stochastic prediction of cumulative quantities, in *Twelfth International Conference on Offshore Mechanics and Arctic Engineering. Vol. 2*, Glasgow, Scotland UK, pp. 417–424.

Athanassoulis, G.A. and Soukissian, T.H., 1993b, The effect of nonstationarity and dependence in long-term extreme value prediction of significant wave height, Tech. Rep. DNAME-93-11, National Technical University of Athens, Athens, Greece.

Athanassoulis, G.A. and Soukissian, T.H., 1995, Return periods of extreme sea-states from long-term time series of H_S, Tech. Rep. TEC-3.3-01(0), WAVEMOD Project, MAST2 CT920025, National Technical University of Athens, Athens, Greece.

Athanassoulis, G.A. and Stefanakos, Ch.N., 1995, A nonstationary stochastic model for long-term time series of significant wave height, *Journal of Geophysical Research*, vol. 100, no. C8, pp. 16149–16162.

Athanassoulis, G.A. and Stefanakos, Ch.N., 1996a, Statistical analysis of wave data. Ia: Univariate statistics/Mediterranean Sea, Tech. Rep. WERATLAS Project, EU DGXII Contract No. JOU2-CT93-0390, National Technical University of Athens.

Athanassoulis, G.A. and Stefanakos, Ch.N., 1996b, Statistical analysis of wave data. IIa: Bivariate statistics/Mediterranean Sea, Tech. Rep. WERATLAS Project, EU DGXII Contract No. JOU2-CT93-0390, National Technical University of Athens.

Athanassoulis, G.A. and Stefanakos, Ch.N., 1998, Missing-value completion of nonstationary time series of wave data, in *17th International Conference on Offshore Mechanics and Arctic Engineering*, Lisbon, Portugal.

Athanassoulis, G.A., Vranas, P.B., and Soukissian, T.H., 1992, A new model for long-term stochastic analysis and prediction, Part-I: theoretical background, *Journal of Ship Research*, vol. 36, no. 1, pp. 1–16.

Athanassoulis, G.A., Skarsoulis, E.K., and Belibassakis, K.A., 1994, Bivariate distributions with given marginals with an application to wave climate description, *Applied Ocean Research*, vol. 16, pp. 1–17.

Athanassoulis, G.A., Skarsoulis, E.K., and Stefanakos, Ch.N., 1995a, Application of bivariate probability models to measured wave data, Tech. Rep. TEC-2.4-02(0), WAVEMOD Project, MAS2-CT92-0025, National Technical University of Athens, Athens, Greece.

Athanassoulis, G.A., Soukissian, T.H., and Stefanakos, Ch.N., 1995b, Long-term variability and its impact to the extreme-value prediction from time series of significant wave height, in *Fourth International Workshop on Wave Hindcasting and Forecasting: Climate Variability and Extremes*, V.R. Swail, ed., Atmospheric Environment Service, Banff, Alberta, Canada, pp. 343–358.

Athanassoulis, G.A., Stefanakos, Ch.N., and Barstow, S.F., 1999, A methodology for integration of long-term data from different sources for the description of wave climate variability, *To appear*.

Bacon, S. and Carter, D.J.T., 1991, Wave climate changes in the north Atlantic and North Sea, *International Journal of Climatology*, vol. 11, pp. 545–558.

Bales, S.L., Cummins, W.E., and Comstock, E.N., 1982, Potential impact of twenty year hindcast wind and wave climatology on ship design, *Marine Technology*, vol. 19, no. 2, pp. 111–139.

Ballerini, R. and McCormick, W.P., 1989, Extreme value theory for processes with periodic variances, *Communications in Statistics - Stochastic Models*, vol. 5, no. 1, pp. 45–61.

Bartolini, P. and Salas, J.D., 1993, Modeling of streamflow processes at different time scales, *Water Resources Research*, vol. 29, no. 8, pp. 2573–2587.

Bendat, J.S. and Piersol, A.G., 1971, *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York.

Beneš, V. and Štěpán, J. (eds.), 1997, *Distributions with given marginals and moment problems*, Kluwer Academic Press, Dordrecht, The Netherlands.

Berger, A., 1991, Basic concepts of climate modelling, in *Climate change and impacts: A general introduction*, R. Fantechi, G. Maracchi, and M.E. Almeida-Teixeira, eds., Report EUR 11943 EN, Luxembourg, pp. 49–76.

Berman, S.M., 1992, *Sojourns and Extremes of Stochastic Processes*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove.

Bettencourt, J., 1993, *Characterisation and prediction of wave power resources*, Master's thesis, FCUL, Lisbon Classical University, Lisbon.

Beveridge, S., 1992, Least squares estimation of missing values in time series, *Communications in Statistics - Theory and Methods*, vol. 21, no. 12, pp. 3479–3496.

Bhattacharyya, G.K., 1984, Tests of randomness against trend or serial correlations, in *Handbook of Statistics, Vol. 4: Nonparametric Methods*, P.R. Krishnaiah and P.K. Sen, eds., North-Holland, Amsterdam, pp. 89–111.

Bishop, J.M, 1984, *Applied Oceanography*, John Willey & Sons, New York.

Blake, I.F. and Lindsey, W.C., 1973, Level-crossing problems for random processes, *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 295–315.

Borgman, L.E., 1973, Probabilities for highest wave in hurricane, *Journal of the Waterways, Harbors and Coastal Engineering Division*, vol. 99, no. WW2, pp. 185–207.

Borgman, L.E, 1975, Extremal statistics in ocean engineering, *Civil Engineering in the Oceans*, vol. 3, pp. 117–133.

Borgman, L.E. and Scheffner, N.W., 1991, Simulation of time sequences of wave height, period, and direction, Tech. Rep. DRP-91-2, U.S. Army Corps of Engineers, Washington, D.C.

Bowman, K.O. and Shenton, L.R., 1988, *Properties of estimators for the Gamma distribution*, Marcel Dekker, Inc., New York.

Box, G.E.P. and Jenkins, G.M., 1976, *Time series analysis and its applications*, Second edn., Holden-Day, San Francisco.

Brelsford, W.M., 1967, *Probability predictions and time series with periodic structure*, Ph.D. thesis, Johns Hopkins University, Baltimore, Maryland.

Brockwell, P.J. and Davis, R.A., 1991, *Time series: Theory and methods*, Second edn., Springer-Verlag, New York.

Brown, W.A., 1987, *On the theory of cyclostationary signals*, Ph.D. thesis, University of california, Davis.

Burroughs, W.J., 1992, *Weather cycles. Real or imaginary?*, Cambridge University Press, Cambridge.

Carter, D.J.T. and Challenor, P.G., 1981, Estimating return values of environmental parameters, *Quart. J. R. Met. Soc.*, vol. 107, pp. 259–266.

Carter, D.J.T. and Challenor, P.G., 1983, Methods of fitting the Fisher-Tippett type I extreme value distribution, *Ocean Engineering*, vol. 10, no. 3, pp. 191–199.

Carter, D.J.T. and Draper, L., 1988, Has the north-east Atlantic become rougher?, *Nature*, vol. 332, pp. 494.

Carter, D.J.T., Aage, C., Allan, T., Boonstra, H., Lindgren, G., and Olagnon, M., 1998, *Oceans from Space. A textbook for offshore engineers and naval architects*, Final Report of EU Project SUCCESS.

Castillo, E., 1988, *Extreme Value Theory in Engineering*, Academic Press, Boston.

Castillo, E. and Sarabia, J.M., 1992, Engineering analysis of extreme value data: Selection of models, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 118, no. 2, pp. 129–146.

Challenor, P.G., Guymer, T.H., and Srokosz, M.A., 1986, The influence of spatial and temporal scales on calibration/validation, in *ERS-I Wind and Wave Calibration Workshop*, European Space Agency, Noordwijk, Netherlands.

Cohen, A.C., 1988, Three-parameter estimation of lognormal distribution, in *Lognormal distributions. Theory and applications*, E.L. Crow and K. Shimizu, eds., Marcel Dekker, Inc., New York, pp. 113–137.

Crow, E.L. and Shimizu, K. (eds.), 1988, *Lognormal distributions. Theory and applications*, Marcel Dekker, Inc., New York.

Cuadras, C.M., 1992, Probability distributions with given multivariate marginals and given dependence structure, *Journal of Multivariate Analysis*, vol. 42, pp. 51–66.

Cummins, W.E. and Bales, S.L., 1980, Extreme and rare occurrence wave statistics for northern hemispheric shipping lanes, in *Spring Meeting/STAR Symposium*, Society of Naval Architects and Marine Engineers, Colorado, California.

Dall' Aglio, G., Kotz, S., and Salinetti, G. (eds.), 1991, *Advances in probability distributions with given marginals*, Kluwer Academic Press, Dordrecht, The Netherlands.

Daniell, P.J., 1946, Discussion on "Symposium on autocorrelation in time series", *Journal of the Royal Statistical Society*, vol. 8, no. Supplement, pp. 88–90.

Dehay, D. and Hurd, H., 1994, Representation and estimation for periodically and almost periodically correlated random processes, in *Cyclostationarity in communications and signal processing*, W.A. Gardner, ed., IEEE Press, pp. 295–326.

Dehay, D. and Hurd, H., 1996, Spectral estimations for a strongly periodically correlated random field defined on $\mathbf{Z}^2$, in *Proceedings of the Workshop on cyclostationary prossesses*, P. Loubaton and H. Hurd, eds., Université de Marne-la-Valée, Noisy le Grand, France, July, 1-3, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Technical Report No. 485.

Dragan, Ya.P. and Yavorskii, I.N., 1982, *Rhythmics of Sea Waves and Underwater Acoustic Signals*, Naukova Dumka, Kiev, Ukraine.

Dragan, Ya.P., Rozhkov, V.A., and Yavorskii, I.N., 1984, Applications of the theory of periodically correlated random processes to the probabilistic analysis of oceanological time series, in *Probabilistic Analysis and Modelling of Oceanological Processes*, V.A. Rozhkov, ed., Gidrometeoizdat, Leningrad, Russia, pp. 4–23.

Dunsmuir, W. and Robinson, P.M., 1981a, Asymptotic theory for time series containing missing and amplitude modulated observations, *Sankhyã*, vol. 43, pp. 260–281.

Dunsmuir, W. and Robinson, P.M., 1981b, Parametric estimators for stationary time series with missing observations, *Advances of Applied Probability*, vol. 13, pp. 129–146.

Dunsmuir, W. and Robinson, P.M., 1981c, Estimation of time series models in the presence of missing data, *Journal of the American Statistical Association*, vol. 76, pp. 560–568.

Edgeworth, F.Y., 1898, On the representation of statistics by mathematical formulae, *Journal of the Royal Statistical Society*, vol. 61, pp. 670–700.

Embrechts, P., Klüppelberg, C., and Mikosch, Th., 1997, *Modelling extremal events*, Springer-Verlag, Berlin.

Fisher, R.A. and Tippet, L.H., 1928, Limiting forms of the frequency distributions of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190.

Forsyth, G.E., Malcolm, M.A., and Moler, C.B., 1977, *Computer methods for mathematical computations*, Prentice-Hall, New York.

Fréchet, M., 1951, Sur les tableaux de correlation dont les marges sont donées, *Ann. Univ. Lyon, Sect. A, Ser. 3*, vol. 14, pp. 53–77.

Fréchet, M., 1956, Sur les tableaux de correlation dont les marges sont donées, *Comptes Rendus Acad. Sci., Paris*, vol. 242, pp. 2426.

Fréchet, M., 1958, Remarques au sujet de la note precedente: Distributions à plusieurs variables dont les marges sont donées, *Comptes Rendus Acad. Sci., Paris*, vol. 246, pp. 2719–2720.

Frutuoso, R., Bettencourt, J., and Pontes, M.T., 1993, Stochastic representation of time-dependent spectral parameters, in *Proceedings of the European Wave Energy Symposium*, Edinburgh, U.K.

Furnes, G. and Reistad, M., 1993, Mean wave height variations in the north-east Atlantic (jan. 1881-dec. 1989), in *Proceedings of the 3rd International Conference on Offshore and Polar Engineering*, J.S. Chung et al., ed., vol. 3, International Society of Offshore and Polar Engineers, Singapore, pp. 71–75.

Galambos, J., 1978, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York.

Gardner, W.A., 1972, *Representation and estimation of cyclostationary processes*, Ph.D. thesis, University of Massachusets, Amherst.

Gardner, W.A., 1990, *Introduction to random processes*, Second edn., McGraw-Hill Publishing Company, New York.

Geçkinli, N. and Yavuz, D., 1983, *Discrete Fourier transformation and its applications to power spectra estimation*, Elsevier Science, New York.

Giannakis, G.B. and Zhou, G., 1994, Parameter estimation of cyclostationary AM time series with application to missing observations, *IEEE, Transactions on Signal Processing*, vol. 42, no. 9, pp. 2408–2419.

Gladyshev, E.G., 1961, Periodically correlated random sequences, *Soviet Mathematics*, vol. 2, pp. 385–388.

Gladyshev, E.G., 1962, *Theory of periodically correlated sequences and processes*, Ph.D. thesis, Institute of Atmospheric Physics, Academy of Sciences, USSR, Moscow.

Gladyshev, E.G., 1963, Periodically and almost-periodically correlated random processes with continuous time parameter, *SIAM, Theory of Probability and its Applications*, vol. 8, pp. 173–177.

Gnanadesikan, R., 1977, *Methods for statistical data analysis of multivariate observations*, John Wiley & Sons, New York.

Gnedenko, B.V., 1943, Sur la distribution limite du terme maximum d' une série aléatoire, *Ann. Math.*, vol. 44, pp. 423–453.

Goda, Y., 1988, On the methodology of selecting design wave height, in *Proceedings of the 21st International Conference on Coastal Engineering*, American Society of Civil Engineers, Malaga, Spain, pp. 899–913.

Gómez, V. and Maravall, A., 1994, Estimation, prediction, and interpolation for nonstationary series with Kalman filter, *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 611–624.

Gradshteyn, I.S. and Ryzhik, I.M., 1980, *Table of Integrals, Series, and Products*, Academic Press, New York.

Grimshaw, S.D., 1993, Computing maximum likelihood estimates for the Generalized Pareto Distribution, *Technometrics*, vol. 35, no. 2, pp. 185–191.

Gringorten, I.I., 1963, A plotting rule for extreme probability paper, *Journal of Geophysical Research*, vol. 68, pp. 813–814.

Gudzenko, L.I., 1959, On periodically nonstationary processes, *Radiotekhnika i Elektronika*, vol. 4, no. 6, pp. 1062–1064.

Guedes Soares, C. and Ferreira, A.M., 1995, Analysis of the seasonality in non-stationary time series of significant wave height, in *Computational Stochastic Mechanics*, P.D. Spanos, ed., A.A. Balkema, Rotterdam, pp. 501–521.

Guedes Soares, C., Ferreira, A.M., and Cunha, C., 1995, Autoregressive model for longterm series of significant wave height in the Portuguese coast, in *Modelling of Coastal and Estuarine Processes*, F.J. Seabra Santos and A. Temperville, eds., Coimbra, pp. 59–70.

Hamon, B.V. and Middleton, J.F., 1989, Return periods of extreme sea levels: The exceedance probability method, *Int. Hydrogr. Rev.*, vol. 66, no. 2, pp. 165–177.

Hannan, E.J., 1955, A test for singularities in Sydney rainfall, *Aust. J. Phys.*, vol. 8, no. 2, pp. 289–297.

Haring, R.E. and Heideman, J.C., 1978, Gulf of Mexico rare wave return periods, in *Proceedings of the 10th Offshore Technology Conference*, pp. 1537–1550.

Hasselmann, K. and Barnett, T.P., 1981, Techniques of linear prediction for systems with periodic statistics, *Journal of the Atmospheric Sciences*, vol. 38, pp. 2275–2283.

Hidalgo, O.S., 1995, Interpolation of time series of significant wave height using bivariate models, Tech. Rep. 58, Puertos del Estado, Departamento de Clima Maritimo, Madrid, Spain.

Hidalgo, O.S., Nieto Borge, J.C., Cunha, C.C., and Guedes Soares, C., 1995, Filling missing observations in time series of significant wave height, in *Fourteenth International Conference*

on Offshore Mechanics and Arctic Engineering. Vol. II, C. Guedes Soares, ed., Copenhagen, pp. 9–17.

Hinich, M.V., 1982, Testing for Gaussianity and linearity of a stationary time series, *Journal of Time Series Analysis*, vol. 13, no. 3, pp. 169–176.

Hogben, N., 1995, Increase in wave heights over the north Atlantic: A review of the evidence and some implications for the naval architect, *Transactions of The Royal Institution of Naval Architects*, vol. 137, pp. 93–101.

Hollander, M. and Wolfe, D.A., 1973, *Nonparametric statistical methods*, John Wiley, New York.

Hosking, J.R.M., 1985, Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution, *Applied Statistics*, vol. 34, pp. 301–310.

Hosking, J.R.M. and Wallis, J.R., 1987, Parameter and quantile estimation for the Generalized Pareto Distribution, *Technometrics*, vol. 29, no. 3, pp. 339–349.

Huang, J.S. and Kotz, S., 1984, Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions, *Biometrika*, vol. 71, pp. 633–636.

Hurd, H.L., 1969, *An investigation of periodically correlated stochastic processes*, Ph.D. thesis, Duke University, Durham, N.C.

Hurd, H., Kallianpur, G., and Farshidi, J., 1997, Correlation and spectral theory for periodically correlated random fields indexed on $\mathbf{Z}^2$, Tech. Rep. No. 448, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Hutchinson, T.P. and Lai, C.D., 1990, *Continuous bivariate distributions, emphasising applications*, Rumsby Scientific Publishing, Adelaide, Australia.

Jenkins, G.M. and Watts, D.G., 1968, *Spectral analysis and its applications*, Holden-Day, San Francisco.

Joe, H., 1997, *Multivariate models and dependence concepts*, Chapman & Hall, London.

Johnson, M.E., 1987, *Multivariate statistical simulation*, John Wiley & Sons, New York.

Johnson, N.L., 1949, Bivariate distributions based on simple translation systems, *Biometrika*, vol. 36, pp. 297–304.

Johnson, N.L. and Kotz, S., 1970, *Distributions in statistics. Vol. 2: Continuous univariate distributions. Part 1*, Houghton Mifflin, New York.

Johnson, N.L. and Kotz, S., 1972, *Distributions in statistics. Vol. 4: Continuous multivariate distributions*, John Wiley, New York.

Johnson, N.L. and Kotz, S. (eds.), 1985, *Encyclopedia of statistical sciences, Vol. 7*, John Wiley & Sons, New York.

Jones, R.H., 1962, Spectral analysis with regularly missed observations, *Annals of Mathematical Statistics*, vol. 32, pp. 455–461.

Jones, R.H., 1980, Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, vol. 22, no. 3, pp. 389–395.

Jones, R.H., 1985, Time series analysis with unequally spaced data, in *Handbook of Statistics*,

Vol. 5: Time series in the frequency domain, E.J. Hannan, P.R. Krishnaiah, and M.M. Rao, eds., North-Holland, Amsterdam, pp. 157–177.

Konstant, D.G. and Piterbarg, V.I., 1993, Extreme values of the cyclostationary Gaussian random process, *Journal of Applied Probability*, vol. 30, pp. 82–97.

Kottegoda, N.T., 1980, *Stochastic water resources technology*, Macmillan Press, London.

Kushnir, Y., Cardone, V.J., Greenwood, J.G., and Cane, M.A., 1995, Link between North Atlantic climate/variability of surface wave height and sea level pressure, in *Fourth International Workshop on Wave Hindcasting and Forecasting*, V.R. Swail, ed., Atmospheric Environment Service, Banff, Alberta, Canada, pp. 59–64.

Labeyrie, J., 1990, Stationary and transient states of random seas, *Marine Structure*, vol. 3, pp. 43–58.

Labeyrie, J., 1991, Time scales and statistical uncertainties in the prediction of extreme environmental conditions, *Reliability Engineering and System Safety*, vol. 32, pp. 243–266.

Lamb, H., 1991, *Historic Storms of the North Sea, British Isles and Northwest Europe*, Cambridge University Press, New York.

Lancaster, H.O., 1958, The structure of bivariate distributions, *Annals of Mathematical Statistics*, vol. 29, pp. 719–736.

Lancaster, H.O., 1963, Correlations and canonical forms of bivariate distributions, *Annals of Mathematical Statistics*, vol. 34, pp. 532–538.

Lancaster, H.O., 1969, *The chi-squared distribution*, John Wiley & Sons, New York.

Larson, H.J. and Shubert, B.O., 1979, *Probabilistic models in engineering sciences.Vol I: Random variables and stochastic processes*, John Wiley & Sons, New York.

Lavielle, M. and Rio, E., 1987, Identification d' états de mer stationnaires ou de transitions par détection de ruptures d' un modèle, Tech. Rep. Rapport de Recherche No. 01, Laboratoire d' Études Mathématiques des Phénomènes Aléatoires de Brest, Brest, France.

Lazanoff, S.M. and Stevenson, N.M., 1975, An evaluation of a hemispheric operational wave spectral model, Tech. Rep. 75-3, Fleet Numerical Weather Center, Monterey, California.

Lazanoff, S.M. and Stevenson, N.M., 1978, A twenty-year northern Hemisphere wave spectral climatology, in *Turbulent Fluxes Through the Sea Surface Wave Dynamics and Prediction*, A. Favre and K. Hasselman, eds., Plenum Press, New York.

Leadbetter, M.R., 1974, On extreme values in stationary sequences, *Zeitschrift fur Wahrsch. Verw.*, vol. 28, pp. 289–303.

Leadbetter, M.R., Lindgren, G., and Rootzen, H., 1983, *Extremes and Related Properties of Random Sequences and Series*, Springer-Verlag, New York.

Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, vol. 2, pp. 164–168.

Ljung, G.M., 1989, A note on the estimation of missing values in time series, *Communications in Statistics - Simulation and Computation*, vol. 18, no. 2, pp. 459–465.

Loubaton, P. and Hurd, H. (eds.), 1996, *Workshop on cyclostationary prossesses*, Université de

Marne-la-Valée, Noisy le Grand, France, July, 1-3, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Technical Report No. 485.

Loynes, R.M., 1965, Extreme values in uniformly mixing stationary stochastic processes, *Annals of Mathematical Statistics*, vol. 36, pp. 993–999.

Luceño, A., 1997, Estimation of missing values in possibly partially nonstationary vector time series, *Biometrika*, vol. 84, no. 2, pp. 495–499.

Macleod, A.J., 1989, A remark on algorithm as215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution, *Applied Statistics*, vol. 38, pp. 198–199.

Marco, J.M. and Ruiz-Rivas, C., 1992, On the construction of multivariate distributions with given nonoverlapping multivariate marginals, *Statistics & Probability Letters*, vol. 15, pp. 259–265.

Mardia, K.V., 1970, *Families of bivariate distributions*, Griffin, London.

Marquardt, D., 1963, An algorithm for least-squares estimation of non-linear parameters, *SIAM Journal of Applied Mathematics*, vol. 11, no. 2, pp. 431–441.

Mathiesen, M., 1994, Estimation of wave height duration statistics, *Coastal Engineering*, vol. 23, pp. 167–181.

Medina, J.R., Gimenez, M.H., and Hudspeth, R.T., 1991, A wave climate simulator, in *Proceedings of the Twenty-fourth International Assoc. Hydr. Research Congress*, pp. B.521–528.

Middleton, J.F. and Thompson, K.R., 1986, Return periods of extreme sea levels from short records, *Journal of Geophysical Research*, vol. 91, no. C10, pp. 11707–11716.

Mirsaidi, S., Fleury, G.A., and Oksman, J., 1997, LMS-like AR modeling in case of missing observations, *IEEE, Transactions on Signal Processing*, vol. 45, no. 6, pp. 1574–1583.

Monaldo, F., 1988, Expected differences between buoy and radar altimeter estimates of wind speed and significant wave height and their implications on buoy-altimeter comparisons, *Journal of Geophysical Research*, vol. 93, no. C3, pp. 2285–2302.

Monaldo, F., 1990, Corrected spectra of wind speed and significant wave height, *Journal of Geophysical Research*, vol. 95, no. C3, pp. 3399–3402.

Morgenstern, D., 1956, Einfache beispiele zweidimensionaler verteilungen, *Mitteilungsblatt fur Math. Statistik*, vol. 8, pp. 234–235.

Muir, L.R. and El-Shaarawi, A.H., 1986, On the calculation of extreme wave heights: A review, *Ocean Engineering*, vol. 13, no. 1, pp. 93–118.

NERC, 1975, *Flood studies report*, Natural Environment Research Council, London.

Neu, H.J.A., 1984, Interannual variations and longer-term changes in the sea state of the north Atlantic from 1970 to 1982, *Journal of Geophysical Research*, vol. 89, no. C4, pp. 6397–6402.

Nieto, F.H. and Martínez, J., 1996, A recursive approach for estimating missing observations in an univariate time series, *Communications in Statistics - Theory and Methods*, vol. 25, no. 9, pp. 2101–2116.

Ochi, M.K., 1982, Stochastic analysis and probabilistic prediction of random seas, *Advances in Hydroscience*, vol. 13, pp. 217–375.

Ochi, M., 1990, *Applied probability and stochastic processes*, John Wiley & Sons, New York.

Ochi, M.K., 1993, Marine environment and its impact on the design of ships and marine structures, *SNAME Transactions*, vol. 101, pp. 673–703.

Oliveira Pires, H.N. and Vasconcelos Pessanha, L.E., 1986, Wave power climate of Portugal, in *Hydrodynamics of Ocean Wave-Energy Utilization, IUTAM Symposium 1985*, D.V. Evans and A.F. de O. Falcão, eds., Springer-Verlag, New York, pp. 157–167.

Ortiz, J. and Ruiz de Elvira, A., 1985, A cyclo-stationary model of sea surface temperatures in the pacific ocean, *Tellus*, vol. Series A, 37, no. 1, pp. 14–23.

Papoulis, A., 1991, *Probability, random variables, and stochastic processes*, Third edn., McGraw-Hill, Inc., New York.

Parzen, E., 1957, On consistent estimates of the spectrum of stationary time series, *Annals of Mathematical Statistics*, vol. 28, pp. 329–348.

Parzen, E., 1961, Mathematical considerations in the estimation of spectra, *Technometrics*, vol. 3, no. 2, pp. 167–190.

Parzen, E., 1962, Spectral analysis of asymptotically stationary time series, *Bulletin de l' Institut International de Statistique*, vol. 39, pp. 87–103.

Parzen, E., 1963, On spectral analysis with missing observations and amplitude modulation, *Sankhyã*, vol. 25, pp. 383–392.

Parzen, E. (ed.), 1984, *Time series analysis of irregularly observed data*, New York, Symposium held at Texas A & M University, College Station, Texas, February 10-13, 1983, Lecture Notes in Statistics, Vol. 25, Springer-Verlag.

Parzen, E. and Pagano, M., 1979, An approach to modeling seasonally stationary time series, *Journal of Econometrics*, vol. 9, pp. 137–153.

Pickands, III, J., 1975, Statistical inference using extreme order statistics, *Annals of Statistics*, vol. 3, pp. 119–131.

Pierson, W.J., Tick, L.J., and Baer, L., 1966, Computer based procedures for preparing global wave forecasts and wind field analyses capable of using wave data obtained by a spacecraft, in *Proceedings of the 6th Naval Hydrodynamics Symposium*, Office of Naval Research, Washington, D.C.

Pierson, Jr, W.J., 1982, The spectral ocean wave model (SOWM), a Northern Hemisphere computer model for specifying and forecasting ocean wave spectra, Tech. Rep. DTNSR DC-82/011, David W. Taylor Naval Ship Research and Development Center, Bethesda, Maryland.

Plackett, R.L., 1965, A class of bivariate distributions, *Journal of the American Statistical Association*, vol. 60, pp. 516–522.

Porat, B. and Friedlander, B., 1984, ARMA spectral estimation of time series with missing observations, *IEEE, Transactions on Information Theory*, vol. 30, no. 6, pp. 823–831.

Prescott, P. and Walden, A.T., 1980, Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution, *Biometrika*, vol. 67, pp. 723–724.

Prescott, P. and Walden, A.T., 1983, Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples, *Journal of Statistical Computation and Simulation*, vol. 16, pp. 241–250.

Priestley, M.B., 1981, *Spectral analysis and time series*, Academic Press, London.

Pugh, D.T. and Vassie, J.M., 1978, Extreme sea levels from tide and surge probability, in *Proceedings of the 16th Coastal Engineering Conference*, vol. 1, pp. 911–930.

Pugh, D.T. and Vassie, J.M., 1980, Applications of the joint probability method for extreme sea level computations, *Proc. Instn. Civ. Engrs, Part 2*, vol. 69, pp. 959–975.

Queffeulou, P., 1983, Seasat wave height measurement: A comparison with sea-truth data and a wave forecasting model: Application to the geographic distribution of strong sea states in storms, *Journal of Geophysical Research*, vol. 88, no. C3, pp. 1779–1788.

Resnick, S.I., 1971, Tail equivalence and its applications, *Journal of Applied Probability*, vol. 8, pp. 136–156.

Rice, S.O., 1944/1945, Mathematical analysis of random noise, *Bell System Techical Journal*, vol. 23/24, pp. 282–332/46–156, (Selected Papers on Noise and Stochastic Processes, N. Wax (ed.), Dover, New York, 1954).

Rosen, Y. and Porat, B., 1989b, Optimal ARMA parameter estimation based on the sample covariances for data with missing observations, *IEEE, Transactions on Information Theory*, vol. 35, no. 2, pp. 342–349.

Rosjberg, D. and Knudsen, J., 1984, POT-estimation of extreme sea-states and the benefit of using wind data, in *Statistical Extremes and Applications*, J. Tiago de Oliveira, ed., NATO Advance Science Institutes, D. Reidel Publishing Co., Dordrecht, pp. 611–620.

Rüschendorf, L., 1985, Construction of multivariate distributions with given marginals, *Annals of the Institute of Statistical Mathematics*, vol. 37, no. Part A, pp. 225–233.

Rüschendorf, L., Schweizer, B., and Taylor, M.D. (eds.), 1996, *Distributions with fixed marginals and related topics*, Institute of Mathematical Statistics.

Rytov, S.M., Kravtsov, Yu.A., and Tatarskii, V.I., 1988, *Principles of statistical radiophysics. Vol 2: Correlation theory of random processes*, Springer-Verlag, London.

Salas, J.D., 1993, Analysis and modeling of hydrologic time series, in *Handbook of hydrology*, D.R. Maidment, ed., chap. 19, McGraw-Hill, New York, pp. 19.1–19.72.

Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L., 1980, *Applied modeling of hydrologic time series*, Water Resources Publications, Colorado.

Sarmanov, O.V., 1966, Generalized normal correlation and two-dimensional Fréchet classes, *Soviet Math. Doklady*, vol. 7, pp. 596–599.

Sarmanov, O.V. and Bratoeva, Z.N., 1967, Probabilistic properties of bilinear expansions of Hermite polynomials, *Theory of probability and its applications*, vol. 12, no. 3, pp. 470–481.

Scheffner, N.W. and Borgman, L.E., 1992, Stochastic time-series representation of wave data, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 118, no. 4, pp. 337–351.

Schweizer, B., 1991, Thirty years of copulas, in *Advances in probability distributions with given marginals*, G. Dall'Aglio, S. Kotz, and G. Salinetti, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 13–50.

Shih, S.F. and Cheng, K.S., 1989, Generation of synthetic and missing climatic data for Puerto Rico, *Water Resources Bulletin*, vol. 25, no. 4, pp. 829–836.

Sklar, A., 1959, Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231.

Sklar, A., 1973, Random variables, joint distribution functions and copulas, *Kybernetika*, vol. 9, pp. 449–460.

Smith, R.L., 1984, Threshold methods for sample extremes, in *Statistical Extremes and Applications*, J. Tiago de Oliveira, ed., NATO Advance Science Institutes, D. Reidel Publishing Co., Dordrecht, pp. 621–638.

Smith, R.L., 1986, Extreme value theory based on the r largest annual events, *Journal of Hydrology*, vol. 86, pp. 27–43.

Snedecor, G.W. and Cochran, W.G., 1980, *Statistical methods*, The Iowa State University Press, Ames, Iowa.

Soukissian, T.H., 1995, *Stochastic methods for analysis and prediction of wave climate*, Ph.D. thesis, National Technical University of Athens, Athens, Greece, (in Greek).

Spanos, P-T.D., 1983, ARMA algorithms for ocean wave modelling, *Trans. of the ASME, Journal of Energy Resources Technology*, vol. 105, pp. 300–309.

Stark, H. and Woods, J.W., 1986, *Probability, random processes, and estimation theory for engineers*, Prentice-Hall, Englewood Cliffs, New Jersey.

Stedinger, J.R., 1980, Fitting log normal distributions to hydrologic data, *Water Resources Research*, vol. 16, no. 3, pp. 481–490.

Stefanakos, Ch.N. and Athanassoulis, G.A., 1997, A new methodology for missing-value completion of nonstationary time series, *To appear*.

von Storch, H., Guddal, J., Iden, K., Jonsson, T., Perlwitz, J., Reistad, M., de Ronde, J., Schmidt, H., and Zorita, E., 1994, Changing statistics of storms in the North Atlantic?, Tech. Rep. 116, Max-Planck Institut für Meteorologie.

Tawn, J.A., 1988, An extreme value theory model for dependent observations, *Journal of Hydrology*, vol. 101, pp. 227–250.

Tawn, J.A. and Vassie, J.M., 1989, Extreme sea levels: The joint probabilities method revisited and revised, *Proc. Instn. Civ. Engrs., Part 2*, vol. 78, pp. 429–442.

Teng, C.-C., Timpe, G.L., Palao, I.M., and Brown, D.A., 1994, Design waves and wave spectra for engineering applications, in *Ocean Wave Measurement and Analysis*, O.T. Magoon and J.M. Hemsley, eds., American Society of Civil Engineers & National Oceanic and Atmospheric Administration, New Orleans, Louisiana, pp. 992–1006.

Theil, H., 1950a, A rank-invariant method of linear and polynomial regression analysis. Part 1, *Proc. K. Ned. Akad. Wet.*, vol. Ser. A, 53, pp. 386–392.

Theil, H., 1950b, A rank-invariant method of linear and polynomial regression analysis. Part 2, *Proc. K. Ned. Akad. Wet.*, vol. Ser. A, 53, pp. 521–525.

Theil, H., 1950c, A rank-invariant method of linear and polynomial regression analysis. Part 3, *Proc. K. Ned. Akad. Wet.*, vol. Ser. A, 53, pp. 1397–1412.

Thomas, H.A. and Fiering, M.B., 1962, Mathematical synthesis of streamflow sequences for the analysis or river basins by simulation, in *Design of Water Resource Systems*, A. Mass et al., ed., Harvard University Press, Cambridge, Massachusetts.

Tournadre, J., 1993, Time and space scales of significant wave heights, *Journal of Geophysical Research*, vol. 98, no. C3, pp. 4727–4738.

U.S. Navy, 1983, *Hindcast Spectral Ocean Wave Model Climatic Atlas: North Atlantic Ocean*, Naval Oceanography Command Detachment, Asheville, N.C.

Vecchia, A.V., 1985, Periodic autoregressive-moving average (PARMA) modeling with applications to water resources, *Water Resources Bulletin*, vol. 21, no. 5, pp. 721–730.

van Vledder, G. and Zitman, T.J., 1992, Design waves: Statistics and engineering practice, in *Proceedings of the Second International Offshore and Polar Engineering Conference*, vol. III, pp. 170–178.

van Vledder, G., Goda, Y., Hawkes, P., Mansard, E., Martin, M.J., Mathiesen, M., Peltier, E., and Thompson, E.F., 1994, Case studies of extreme wave analysis - a comparative analysis, in *Ocean Wave Measurement and Analysis*, O.T. Magoon and J.M. Hemsley, eds., American Society of Civil Engineers & National Oceanic and Atmospheric Administration, New Orleans, Louisiana, pp. 977–991.

Walton, Jr., T.L. and Borgman, L.E., 1990, Simulation of nonstationary, non-Gaussian water levels on great lakes, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 116, no. 6, pp. 664–685.

The WAMDI Group, 1988, The WAM model-A third generation ocean wave prediction model, *Journal of Physical Oceanography*, vol. 18, pp. 1775–1810.

WASA group, 1995a, Comment on “Increase in wave heights over the north Atlantic: A review of the evidence and some implications for the naval architect” by N. Hogben, *Transactions of The Royal Institution of Naval Architects*, vol. 137, pp. 107–110, (WASA group: K. Iden, H. Reichardt, M. Reistad, W. Rosenthal, A. Ruiz de Elvira, and H. von Storch).

WASA group, 1995b, The WASA project: Changing storm and wave climate in the northeast Atlantic and adjacent seas?, in *Fourth International Workshop on Wave Hindcasting and Forecasting*, V.R. Swail, ed., Atmospheric Environment Service, Banff, Alberta, Canada, pp. 31–44, (WASA group: H. Alexandersson, E. Bouws, J.C. Carretero, J. Guddal, I. González, H. Günther, D. Jannink, V. Kharin, G. Komen, H. Reichardt, W. Rosenthal, A. Ruiz de Elvira, T. Schmith, M. Stawarz, and H. von Storch).

WASA group, 1998, Results of the WASA wave hindcast, in *Fifth International Workshop*

on Wave Hindcasting and Forecasting, D.T. Resio, ed., Atmospheric Environment Service, Melbourne, Florida, pp. 17–29, (WASA group: H. Günther, W. Rosenthal, M. Stawarz, J.C. Carretero, M. Gomez, I. Lozano, O. Serrano and M. Reistad).

Watanabe, I., Tomita, H., and Tanizawa, K., 1992, *Wind and waves of the North Pacific Ocean (1974-1988)*, Ship Research Institute, Supplement 14, Yokohama, Japan, (in Japanese).

Wei, W.W.S., 1990, *Time series analysis. Univariate and multivariate methods*, Addison-Wesley Publishing Company, Inc., California.

Weissman, I., 1978, Estimation of parameters and large quantiles based on the kth largest observations, *Journal of the American Statistical Association*, vol. 63, pp. 812–815.

Wilk, M.B. and Gnanadesikan, R., 1968, Probability plotting methods for the analysis of data, *Biometrika*, vol. 55, no. 1, pp. 1–17.

Wyland, R.M. and Thornton, E.B., 1991, Extremal wave statistics using three hindcasts, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 117, no. 1, pp. 60–74.

Yaglom, A.M., 1987a, *Correlation theory of stationary and related random functions*, vol. I: Basic results, Springer-Verlag, New York.

Yaglom, A.M., 1987b, *Correlation theory of stationary and related random functions*, vol. II: Supplementary notes and references, Springer-Verlag, New York.

Appendix A

Data used

In the present work, wave data from three different sources are used: a) data generated by means of ocean wave models (***hindcast data***), b) *in situ* measurements (***measured data***), c) data produced by means of appropriate stochastic simulation (***synthetic*** or ***simulated data***). Although the presented nonstationary modelling can be applied to time series of any wave parameter exhibiting seasonally varying randomness, in the present work only long-term time series of significant wave height are considered.

A.1 Hindcast data

First, for the needs of the analysis in Chapter 4, five 20-year long time series of hindcast significant wave height for an equal number of locations in the North Atlantic Ocean (see Table A.1 and Figure A.1). Such time series are available from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). The generation of these data was part of a U.S. Navy project aimed at producing hindcasts of various wind and wave parameters for an extended grid of sea locations in the northern hemisphere (Lazanoff and Stevenson, 1975; Lazanoff and Stevenson, 1978). In the framework of this project, the spectral ocean wave model (SOWM) (Pierson et al., 1966) was implemented, and historical meteorological data, covering the period from January 1956 to December 1975, were used as input. Although SOWM is a first-generation spectral model that neglects wave-wave and wave-current interactions, it gives competent hindcasts for the deep ocean, provided that a strong current is not present (U.S. Navy, 1983). The data sets have been acquired by the Ship and Marine Hydrodynamics Laboratory (SMHL) in 1991.

Table A.1: Examined sites in the North Atlantic Ocean

Grid point	Latitude (degrees)	Longitude (degrees)
2	66.6 N	6.8 E
10	58.6 N	18.2 W
21	50.0 N	11.7 W
27	44.6 N	12.9 W
31	39.3 N	13.9 W

Further, results of the well-known third generatrion wave model WAM are also used (The WAMDI Group, 1988). The model is operational at European Centre for Medium-range Weather Forecasts (ECMWF) in Reading, England in two versions. The first one is running in a global grid and is appropriate for the wave-climate description of open sea areas, such as, e.g., the Atlantic Ocean. The second one is running in various local grids taking into account the particularities of each region separately. This second version can be used for the wave-climate description of closed basins, such as, e.g., the Mediterranean Sea and the Baltic Sea. For the needs of Chapters 6 and 7, an eight-year long data set is analyzed from a site located off the Spanish coasts in the Atlantic Ocean (Charcot, 45^oN, 12^oW); see Figure A.2. The data set has been obtained by SMHL in 1994, in the framework of EU project WERATLAS (Contract No. JOU2-CT93-0390).

A.2 Measured data

Two data sets from *in situ* measurements are used: (i) one from a site off Portuguese coasts (Figueira da Foz, 40.19^oN, 9.15^oW), and (ii) one from a site off Mediterranean Spanish coasts (Palamós, 41.82^oN, 3.18^oE). See Figure A.2 in the end of the present chapter.

The first data set (Figueira da Foz) is a six-year long (1983, 1985–1989) time series of significant wave height H_S. Note that, the in-between time series of 1984 was excluded, because 56% of its values is missing. In the remaining data set, the 16.5% of the values is missing. The second data set (Palamós) is again a six-year long (1990–1995) time series of significant wave height H_S, with a larger amount of missing values, namely 22.3%. It should be noted that, inevitably, measured data sets are always incomplete due to a number of reasons, such as, e.g., instrument failure or extreme weather conditions.

Finally, the two data sets have been kindly offered: the first one by M.T. Pontes, Instituto

Nacional de Engenharia e Tecnologia Industrial, Portugal, and the second one by A. Ruiz de Elvira, Programa Clima Maritimo, Spain. We are grateful to both of them.

A.3 Synthetic data

In various stages of the present work, the need of simulated data with similar stochastic structure as the existing data sets (hindcast or measured) has been arisen. Synthetic data are very useful in order to validate various analysis procedures or hypotheses. For this, a simulation procedure of nonstationary time series has been set up; see Chapter 5, and has been used for the generation of synthetic long-term time series.

In Chapter 5, two such time series has become incompete in order to check the missing-value completion procedure. In Chapter 7, a 50-yearlong time series of the same structure has been produced for the sensitivity analysis of the various extreme-value prediction methods.

A.4 Figures

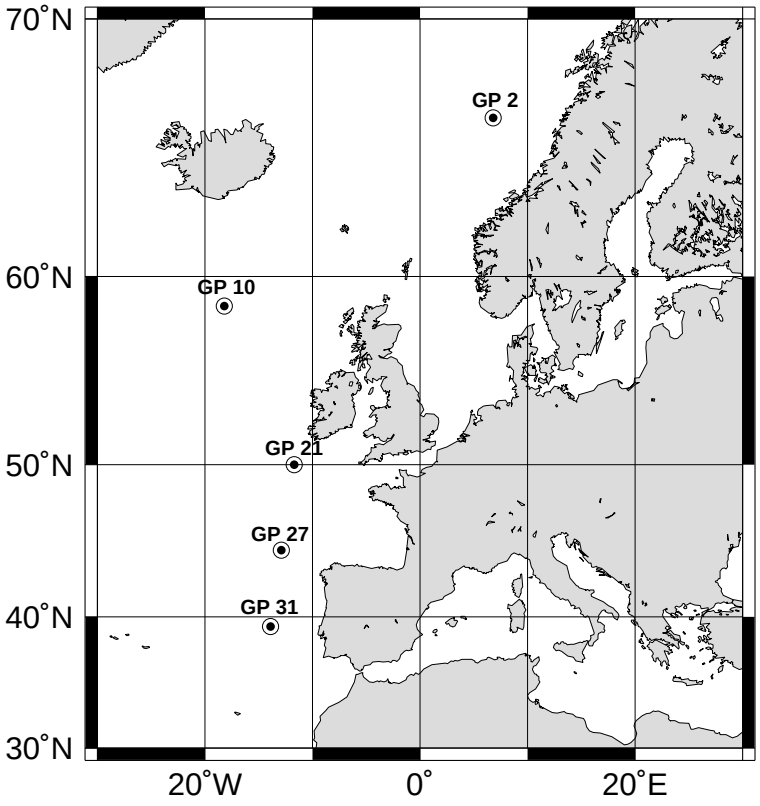


Figure A.1: Examined sites in the North Atlantic Ocean. Grid points (GPs) 2, 10, 21, 27, 31.

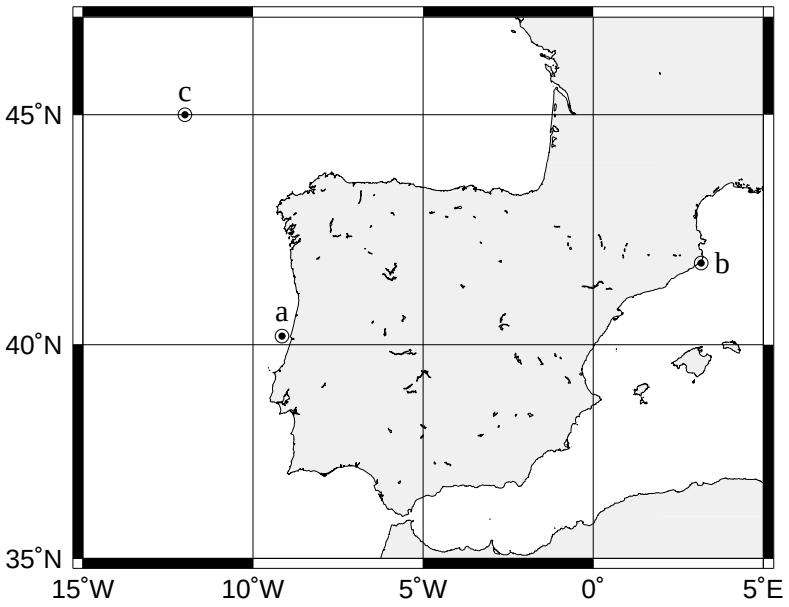


Figure A.2: Examined sites: a) Figueira da Foz, b) Palamós, c) Charcot.