Dependent time changed processes with applications to nonlinear ocean waves.

Pierre Ailliot\textsuperscript{1}, Bernard Delyon \textsuperscript{2}, Valérie Monbet \textsuperscript{2,3}, Marc Prevosto \textsuperscript{4}

October 9, 2015

Abstract

Many records in environmental sciences exhibit asymmetric trajectories and there is a need for simple and tractable models which can reproduce such features. In this paper we explore an approach based on applying both a time change and a marginal transformation on Gaussian processes. The main originality of the proposed model is that the time change depends on the observed trajectory. We first show that the proposed model is stationary and ergodic and provide an explicit characterization of the stationary distribution. This result is then used to build both parametric and non-parametric estimates of the time change function whereas the estimation of the marginal transformation is based on up-crossings. Simulation results are provided to assess the quality of the estimates. The model is applied to shallow water wave data and it is shown that the fitted model is able to reproduce important statistics of the data such as its spectrum and marginal distribution which are important quantities for practical applications. An important benefit of the proposed model is its ability to reproduce the observed asymmetries between the crest and the troughs and between the front and the back of the waves by accelerating the chronometer in the crests and in the front of the waves.

Keywords: time change, Gaussian processes, nonlinear processes, ocean waves

1 Introduction

Marine coastal systems are subject to loadings induced by ocean waves and long time series of wave conditions are often needed to study the performances of such systems. It is very difficult to measure wave conditions on long periods of time and thus there is a need for stochastic models which can simulate quickly long time series of wave conditions with realistic characteristics. The systems are often located near the shore where the waves are known to be non-linear. Figure 1 (bottom plot) shows a typical time series of sea-surface elevation in shallow water. It exhibits both different shapes for crests and troughs, with sharp crests and flat troughs, and asymmetries
between the steeper front of waves and their backs. Many other situations in environmental
science or in econometric lead to study asymmetric records and there is a need for simple and
tractable alternatives to Gaussian processes. An approach based on applying a time change to
a Gaussian process is investigated in this paper.

![Time-changed process](image)

**Figure 1:** Bottom left: sequence of measured sea-surface elevation (location 7). Middle left and top
left: estimated processes \( \hat{Z} \) and \( \hat{Y} \) for the model defined by (4-5). The right panel shows the estimated
functions \( \hat{h} \) (bottom), \( \hat{f} \) (middle) and the empirical spectral density of \( Y \) (top).

A time-changed process \( Z = (Z_t)_{t \geq 0} \) is defined as

\[
Z_t = Y_{\varphi(t)}
\]

where \( \{\varphi(t)\}_{t \geq 0} \) denotes a non-decreasing stochastic process referred to as the time change here-
after. \( Y = (Y_t)_{t \geq 0} \) is a stochastic process which represents the evolution of the phenomenon
of interest in the original time scale. Bold letters are used to stress the fact that \( Y \) and \( Z \) are
possibly multivariate processes. \( Y \) is generally referred to as the base process in the literature.

Models with time change have become popular for financial applications since they were intro-
duced in [5] (see e.g. [16] for a recent review). In this application field the modified time \( \varphi(t) \) is
generally interpreted as a ‘business time’ and \( Y \) and \( \varphi \) are generally assumed to be independent
Levy processes. Deformations have also been extensively used for modeling non-stationary fields
in environmental applications since the seminal paper [13]. It generally consists in seeking a
deterministic transformation \( \varphi \) to map a stationary field \( Y \) into a non-stationary field \( Z \) through
(1). These methods can deal with multidimensional (e.g. spatial) fields but they generally re-
quire replicates of the random field as input to fit the model. More recently methods that can
deal with a unique realization of a densely observed field were proposed (see [1] and references
therein). Bayesian approaches where the deformation is modeled using a Gaussian prior were
also proposed in [15] and then extended to include covariates in [14].

Specific models which include a time change have already been proposed for modeling wave
data. Of particular interest is the two-dimensional stochastic Gauss-Lagrange ocean wave model
discussed in [9]. The “space wave” which describes the one-dimensional sea-surface elevation \( Z \)
at a fixed time is defined implicitly as

\[
Z_{\psi(t)} = Y_t.
\]

where \( \psi \) and \( Y \) denote correlated Gaussian processes which describe respectively the horizontal
and vertical motions. Note that the observed process is the sea surface elevation which is a
univariate process. In this model the trajectories are defined uniquely and explicitly as \( \psi \) only
if \( \psi \) is one to one which does not always hold true. Another model based on applying a moving
average on a Laplace motion (or variance Gamma process) is used in [11] and validated on the
same dataset than considered in this paper. A Laplace motion can be defined as \( \psi \) with \( Y \) a
Brownian motion and \( \phi \) an independent Gamma process.

The originality of the model considered in this paper consists in assuming that the time change
\( \varphi \) is a stochastic process which depends on the base process \( Y \). More precisely, the first model
considered in this paper assumes that

\[
Z_t = Y_{\varphi(t)}, \quad \varphi(t) = \int_0^t f(Y_{\varphi(s)})ds
\]

with \( Y \) a stationary (ergodic) process and \( f \) a positive function referred to as the time change
function hereafter. In this model the time change is thus a function of the base process and \( f \)
links the speed of the chronometer to the observed process. This model is quite general since \( Y \)
can be multivariate and include for example covariates or latent variables. Note that (2) can be
rewritten equivalently as

\[
Z_{\psi(t)} = Y_t, \quad \psi(t) = \int_0^t \frac{1}{f(Y_s)}ds
\]

where \( \psi \) denotes the reciprocal function of \( \varphi \).

Then we consider a more specific model for sea-surface wave elevation where the time change
\( \varphi \) is a deterministic function of the path of the process and its derivative. More precisely we
assume that

\[
Z_t = Y_{\varphi(t)}, \quad \varphi(t) = \int_0^t f(Y_{\varphi(s)}, \dot{Y}_{\varphi(s)})ds
\]

with \( Y \) a univariate differentiable stationary ergodic Gaussian process, \( \dot{Y} \) the derivative of \( Y \)
and \( f \) a positive function. A Gaussian process \( Y \) has symmetric paths and (4) will permit to
create both crests-troughs or vertical asymmetries (i.e. different shapes for the crests and the
troughs) and front-back or horizontal asymmetries (i.e. different shapes for the front and the
back of the waves). An example of sequence simulated with this model is shown on Figure 1.
The chosen \( f \) function is increasing in both variables. It implies that the modified time \( \varphi(t) \)
increases quicker when the process is at high levels and thus that the crests of the time modified
process \( \{Z_t\} \) are narrower than the ones of \( \{Y_t\} \) with the opposite holding true for the troughs.
The chronometers also accelerates with the derivatives of the process and this leads to steeper
wave fronts compared to the backs.

The time change impacts the marginal distribution and a process defined by (1) may have non-
Gaussian marginal distribution. However the time change does not modify the level of the process
and as a consequence the up-crossings of \( Z \) and \( Y \) share similar characteristics. A marginal
transformation is then added to add flexibility in the modeling of the up-crossing intensity which
is an important quantity for many applications. More precisely we assume that

\[
h(X_t) = Z_t
\]

with \( Z \) defined by (1), \( X \) the observed process and \( h \) an increasing function referred to as the
marginal link function hereafter.

If \( f = 1 \) then the process \( X \) defined by (4) satisfies \( h(X_t) = Y_t \) with \( \{Y_t\} \) a Gaussian process.
This model, which is sometimes referred to as transformed Gaussian model in the literature,
is very classical for modeling processes or time series with non-Gaussian marginal distribution
and has been extensively used in particular for wave data (see e.g. [12]). Gaussian processes have symmetric paths and this has implications on the paths of a transformed Gaussian process which can handle some crests-troughs asymmetries but no front-back asymmetries. This is a clear limitation of transformed Gaussian models which cannot reproduce wave sequences such as the one shown on Figure 1 which exhibits both kind of asymmetries.

The modeling principle is summarized on Figure 1. Starting from an observed trajectory with asymmetric path (bottom left), we first apply a marginal transformation (5) where the estimated function $h$ is shown on the bottom right panel. The concavity of this estimated function leads to decreasing the steepness of the crests compared to the troughs (middle left). Then a time deformation $f$ is applied. The estimated $f$ function is increasing in both variables (see middle right panel) and thus the chronometer accelerates with the level and the process and its derivatives. It permits to act on both vertical and horizontal asymmetries and obtain a transformed trajectory which has approximatively symmetric path (top left) and can reasonably be modeled using a Gaussian process. The practical motivation for this work was to develop methods to generate quickly long time series of sea-surface elevation with similar characteristics than the observed one. Using one of the methods that have been proposed in the literature to simulate Gaussian processes, realizations of $Y$ can be quickly generated. It is then easy to deduce realizations of the transformed process $X$ once the transformation $f$ and $h$ have been estimated.

The distribution of the process $X$ defined by (4-5) is characterized by the functions $f$ and $h$ and by the second order structure of the Gaussian process $Y$. This paper proposes methods to estimate these functions. For simplicity reasons we first focus on the time change model with no marginal transformation. In Section 2 we consider the general model defined by (3) and give conditions which imply stationarity and ergodicity. This result is then used in Sections 4 and 5 to build respectively parametric and nonparametric estimates of $f$ in the model (1). Section 6 discusses the estimation of $f$ and $h$ and the second order structure of the Gaussian process $Y$ in the model defined by (1). Section 7 discusses the results obtained when fitting this model to a specific wave dataset and conclusions are given in Section 7.

2 Stationarity and ergodicity

In this section we consider the model defined by (3). $Z$ and $Y$, as well as $Z_t$ and $Y_t$, will denote the processes $(Z_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$. The stationarity of $Y$ means that for any measurable function
\( g, g(Y_s) \) has same distribution as \( g(Y_{s+s}) \) for any \( s > 0 \). The ergodicity means that for any bounded measurable function \( g \)

\[
\lim_{H \to \infty} \frac{1}{H} \int_0^H g(Y_{s+s}) = E[g(Y_s)]
\]

In order to check stationarity and ergodicity, it suffices to consider functions \( g \) of the form

\[
g(Y) = g_0(Y_{t_1}, \ldots, Y_{t_n})
\]

where \( g_0 \) is a bounded Borel measurable function; we may even assume that \( g_0 \) is continuous. We will however keep the notation \( g(Y) \) since it will appear to be handier.

**Proposition 1.** Let \((Y_t)_{t \geq 0}\) be a stationary process with values on \( \mathbb{R}^d \) and continuous sample paths and \( f(.) \) be a measurable real positive function defined on \( \mathbb{R}^d \) such that \( q_0 = E \left[ \frac{1}{f(Y_0)} \right] < +\infty \). Define the process \((Z_t)_{t \geq 0}\) by \( (8) \). Then \( Z \) is stationary for the measure \( P' \) defined as

\[
E'[g(Z_s)] = q_0^{-1} E \left[ g(Z_0) \frac{1}{f(Z_0)} \right].
\]

If \( Y \) is ergodic then \( Z \) is ergodic under \( P' \). In particular for any measurable real function \( g \) defined on \( \mathbb{R}^d \) such that \( E \left[ \frac{g(Y_0)}{f(Y_0)} \right] < \infty \)

\[
E'[g(Z_s)] = \lim_{H \to \infty} \frac{1}{H} \int_0^H g(Z_s) ds = q_0^{-1} E \left[ \frac{g(Y_0)}{f(Y_0)} \right].
\]

The proof of this proposition is given in the appendix. \((7)\) is the key result for the rest of this paper. It shows that empirical means computed on a realization of the time changed process \( Z \) converges when the length of the realization tends to infinity. The limit depends both on the distribution of the base process \( Y \) at time 0 and on the time change function \( f \).

### 3 Method of moments

In this section we consider the model defined by \((4)\). It is a particular case of the model defined by \((2)\) with \( Y_t = (Y_t, \dot{Y}_t) \) and Proposition \(1\) gives stationarity and ergodicity conditions for \( Z \). We further assume that \( Y \) is a Gaussian process and thus for any time \( t \) the random variables \( Y_t \) and \( \dot{Y}_t \) are independent Gaussian variables. We have \( E[\dot{Y}_t] = 0 \) and we can assume without loss of generality that \( var(\dot{Y}_t) = 1 \). We also assume that \( E[Y_t] = 0 \) and \( var(Y_t) = 1 \). It is not restrictive in the practical application discussed in Section \(6\) since this condition will be taken into account when estimating the marginal transformation \( h \) (see Section \(5\)).

#### 3.1 Moments in the general case

The following proposition gives a general expression for the joint moments of \((Z_t, \dot{Z}_t)\).

**Proposition 2.** Let \( Y \) be a univariate differentiable ergodic stationary Gaussian process such that \( E[Y_t] = 0 \) and \( var(Y_t) = var(\dot{Y}_t) = 1 \). Let \( U = (U_1, U_2) \) denote a two dimensional standard normal vector. Assume that \((4)\) holds true with \( f \) a positive measurable real function such that

\[
q_0 = E \left[ \frac{1}{f(U)} \right] < +\infty.
\]
Then $Z$ is a differentiable ergodic stationary process. Let $g$ be a measurable real function defined on $\mathbb{R}^2$ such that $E \left[ \frac{g(U)}{f(U)} \right] < \infty$. The stationary measure $P'$ is such that

$$E'[g(Z_s, \dot{Z}_s)] = \lim_{H \to \infty} \frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s) ds = \frac{1}{q_0} E \left[ \frac{g(U_1, U_2 f(U))}{f(U)} \right].$$ \hspace{1cm} (9)

**Proof.** We have

$$\frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s) ds = \frac{1}{H} \int_0^H g \left( Y_{\varphi(s)}, f \left( Y_{\varphi(s)}, \dot{Y}_{\varphi(s)} \right) \dot{Y}_{\varphi(s)} \right) ds.$$  

Now by applying Proposition 3 with $Y_t = (Y_t, \dot{Y}_t)$ we get that $Z_t = (Y_{\varphi(t)}, \dot{Y}_{\varphi(t)})$ is a stationary ergodic process such that

$$\lim_{H \to \infty} \frac{1}{H} \int_0^H g \left( Y_{\varphi(s)}, f \left( Y_{\varphi(s)}, \dot{Y}_{\varphi(s)} \right) \dot{Y}_{\varphi(s)} \right) ds = q_0^{-1} E \left[ \frac{g(Y_0, f(Y_0, \dot{Y}_0)) \dot{Y}_0}{f(Y_0, \dot{Y}_0)} \right]$$

with $q_0 = E \left[ \frac{1}{f(Y_0, \dot{Y}_0)} \right]$. Thus (9) holds true since $Y_0$ and $\dot{Y}_0$ are independent $\mathcal{N}(0, 1)$ variables. \qed

The previous proposition shows that the time change function $f$ is related to the joint distribution of $(Z_t, \dot{Z}_t)$ and thus suggests estimating $f$ to match the empirical joint distribution of $(Z_t, \dot{Z}_t)$. Note however that the relation is not straightforward and that it seems difficult to deduce an estimate of $f$ from an estimate of this joint distribution in the general case. Two different simplifications are considered hereafter. In the first one, which is discussed in the rest of this section, we assume a parametric shape for $f$. In the second one, which is discussed in the Section 4 we assume that $f$ is product of two functions. This leads to an explicit expression for the joint probability density function (pdf) of $(Z_t, \dot{Z}_t)$ and the derivation of nonparametric estimates for $f$.

### 3.2 Moments in the quadratic exponential model

In the rest of this section we assume that the time change function $f$ has a quadratic exponential shape defined as

$$f(y, \dot{y}) = \exp(\beta_0 + \beta_1 y + \beta_2 \dot{y} + \frac{\beta_{1,1}}{2} y^2 + \beta_{1,2} y \dot{y} + \frac{\beta_{2,2}}{2} \dot{y}^2).$$ \hspace{1cm} (10)

with unknown parameter $\beta = (\beta_0, \beta_1, \beta_2 \beta_{1,1}, \beta_{1,2}, \beta_{2,2}) \in \mathbb{R}^6$. $\beta_1$ and $\beta_{1,1}$ describe the effect of the level of the base process on the time change and thus the vertical asymmetry whereas $\beta_2$ and $\beta_{1,2}$ describe the effect of the derivative and the horizontal asymmetry. $\beta_{2,2}$ may be interpreted as an interaction coefficient and $\beta_0$ as a scale factor. We discuss the realism of this condition for wave data in Section 6. Other parametric time change function may be dealt with in a similar way.

**Proposition 3.** Let $Y$ be a univariate differentiable ergodic stationary Gaussian process such that $E[Y_t] = 0$ and $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$. Assume that (9) holds true with $f$ given by (10) and that the matrix $Q = \left( \begin{array}{cc} \beta_{1,1} + 1 & \beta_{1,2} \\ \beta_{1,2} & \beta_{2,2} + 1 \end{array} \right)$ is definite positive. Then $Z$ is a differentiable ergodic
stationary process. The stationary marginal distribution of $Z$ is Gaussian with mean $E'[Z_t]$ and variance $\text{var}'(Z_t)$ such that

\[
\text{var}'(Z_t) = \frac{1}{1 + \beta_{1,1} - \frac{\beta_{1,2}^2}{1 + \beta_{2,2}}}.
\]

(11)

\[
\frac{E'[Z_t]}{\text{var}'(Z_t)} = -\beta_1 + \frac{\beta_2 \beta_{1,2}}{1 + \beta_{2,2}}.
\]

(12)

Denote $m_1 = E[|U_1| \log |U_1|] \simeq 0.046$, $m_2 = E[|U_1| (\log |U_1|)^2] \simeq 0.33$, $m_3 = E[|U_1|^3 \log |U_1|] = 2m_1 + \sqrt{\frac{2}{\pi}}$. Then

\[
q_0 E'[\tilde{Z}_t] = \sqrt{\frac{2}{\pi}}.
\]

(13)

\[
q_0 E'[\tilde{Z}_t \log(|\tilde{Z}_t|)] = \beta_2.
\]

(14)

\[
q_0 E'[Z_t \tilde{Z}_t \log(|\tilde{Z}_t|)] = \beta_{1,2}.
\]

(15)

\[
q_0 E'[\tilde{Z}_t \log(|\tilde{Z}_t|)] = m_1 + (\beta_0 + \frac{1}{2} \beta_{1,1} + \beta_{2,2}) \sqrt{\frac{2}{\pi}}.
\]

(16)

\[
q_0 E'[Z_t \tilde{Z}_t \log(|\tilde{Z}_t|)] = \beta_1 \sqrt{\frac{2}{\pi}}.
\]

(17)

\[
q_0 E'[|Z_t| \tilde{Z}_t \log(|\tilde{Z}_t|)] = m_1 \sqrt{\frac{2}{\pi}} + \frac{\beta_0}{\beta} (\beta_0 + \beta_{1,1} + \beta_{2,2}).
\]

(18)

\[
q_0 E'\left[|Z_t| \left(\log(|\tilde{Z}_t|) - \beta_1 Z_t - \frac{\beta_{1,1}}{2} \tilde{Z}_t^2\right)^2\right] = m_2 + \beta_0^2 \sqrt{\frac{2}{\pi}} + 2 \beta_{2,2} \sqrt{\frac{2}{\pi}} + 2 \beta_0 m_1 + \beta_{2,2} \theta_3 + 2 \beta_0 \beta_{2,2} \sqrt{\frac{2}{\pi}} + 2 \sqrt{\frac{2}{\pi}} (\beta_0^2 + \beta_{1,2}^2).
\]

(19)

Proof. If $Q$ is definite positive then condition (8) is satisfied and thus (9) holds true. Applying (9) with $g(u_1, u_2) = g(u_1)$ a measurable bounded real function we get

\[
E'[g(Z_s)] = \frac{1}{2 \pi q_0} \int_{\mathbb{R}^2} \frac{g(u_1)}{f(u_1, u_2)} \exp \left(-\frac{u_1^2 + u_2^2}{2}\right) du_1 du_2.
\]

Then replacing $f$ by its expression (10) and integrating of $u_2$ we obtain

\[
E'[g(Z_s)] = \int_{\mathbb{R}} g(u_1) \phi(u_1) du_1
\]

where $\phi$ denotes the pdf of Gaussian distribution with mean and variances given by (11) and (12).

Now applying (9) with $g(u_1, u_2) = |u_2|$ we get

\[
E'[|\tilde{Z}_s|] = \frac{1}{q_0} \int \frac{|U_2 f(U)|}{f(U)} du_1 = E[|U_2|] = \sqrt{\frac{2}{\pi}}
\]

and thus (13). Similar calculations lead to (14) and (19).

The model which defines $Z$ in Proposition 3 is referred to as the quadratic exponential model hereafter.
3.3 Some simulation results in the quadratic exponential model

Various estimates of $\beta$ can be built based on the relations given in Proposition 3. Two possibilities which we have tried are introduced in this section and compared using simulations.

A first estimate, denoted $\hat{\beta}$ hereafter, can be obtained by solving (14-19) with the theoretical moments replaced by their empirical estimates and $q_0$ estimated using (13). In practice $\beta_2$, $\beta_{1,2}$ and $\beta_1$ are obtained using respectively (14), (15) and (17). Then an expression for $\beta_{1,1}$ is derived from (16) and (18) (note that the term $\beta_0 + \beta_{2,2}$ appears in these two formulas). Finally $\beta_0$ and $\beta_{2,2}$ are estimated using (17) and (19) with the first equation used to express $\beta_{2,2}$ as a linear function of $\beta_0$ and the second one to give a second order equation for $\beta_0$.

Simulation results indicate that $\hat{\beta}$ provides poor estimations for the parameters $\beta_1$ and $\beta_{1,1}$ (see Figure 3) which describe the effect of the level of the process on the time change. It suggests using (11,12) to reestimate these parameters using

$$
\beta_1 = \frac{\beta_0 \beta_{1,2}}{1 + \beta_{2,2}} - \frac{E[Z_t]}{\text{var}(Z_t)}
$$

(20)

$$
\beta_{1,1} = \frac{1}{\text{var}(Z_t)} - 1 + \frac{\beta_{2,2}^2}{1 + \beta_{1,2}}.
$$

with the theoretical moments being replaced by their empirical estimates and $\beta$ by $\hat{\beta}$ in the right hand terms. We denote $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_{1,1}, \hat{\beta}_{1,2}, \hat{\beta}_{2,2})$ the corresponding estimates.

Let us now introduce the numerical setup used to perform the simulations. We have chosen to consider parameters values obtained when fitting the model to the wave data shown on Figure 1. The parameters $\hat{\beta}_{1,2}$ was found to be close to 0 (see Section 6.3 for a more detailed discussion) and it was decided to force this parameter to be equal to 0. It permits to obtain a model with no interaction which is compatible with the assumptions made in Section 4 and allows comparison between the parametric and the nonparametric approaches. More precisely the chosen parameter values is $\beta^* = (1.253, 0.110, 0.096, 0.068, 0, 0.088)$. The corresponding $f$ function is shown on Figure 3 and the spectral density which has been used to simulate the Gaussian process $Y$ is very close to the one shown on Figure 1. The peak period for $Z$ is about 2.2s, the sampling time step is 0.027s and the length of the simulated sequence is $10^5$ time points. It corresponds to approximatively 1200 waves. This is equivalent to a few hours of wave measurements in the ocean and to the length of the measurements described more precisely in Section 6. The simulated sequences have similar characteristics than the sequence $Z$ shown on Figure 1 (left middle plot) with both crests-troughs and front-back asymmetries.

Figure 3 shows an important improvement in the estimation of $\beta_1$ and $\beta_{1,1}$ when using $\hat{\beta}$ and $\hat{\beta}$, the other estimates being unchanged by construction. Simulations with other parameter values confirmed these results and led us to keep the estimate $\hat{\beta}$ hereafter.

4 Non-parametric estimation

In this section we still consider the model defined by (4) with $Y$ a stationary differentiable ergodic Gaussian process. In order to get an explicit and tractable expression for the joint pdf of $(Z_t, Z_{t+1})$ and deduce non-parametric estimate of $f$ we further assume that there is no interaction between the effect of the level and the derivative, i.e. for any $(y, \dot{y}) \in \mathbb{R}^2$

$$
f(y, \dot{y}) = f_1(y)f_2(\dot{y}).
$$

(22)
Figure 3: Distribution of $\tilde{\beta} - \beta^*$ (left) and $\tilde{\beta} - \beta^*$ (right) in the quadratic exponential model. Boxplots based on 1000 simulated sequences. The simulation experiment is described in Section 3.3.

We discuss the realism of this condition for wave data in Section 6. In this model $f_1$ is related to the vertical asymmetries whereas $f_2$ describes horizontal asymmetries. An increasing $f_1$ function leads to a process $Z$ with crests narrower than the troughs and if $f_2$ is increasing then the front of the waves are steeper than the back.

The joint pdf of $(Z_t, \dot{Z}_t)$ is given in Section 4.1. This result is used in Section 4.2 to build non-parametric estimates of $f_1$ and $f_2$. A simulation experiment is performed in Section 4.3 to assess the performance of the estimates.

4.1 Joint distribution of $(Z_t, \dot{Z}_t)$ in the model with no interaction

Proposition 4. Let $Y$ be a univariate differentiable ergodic stationary Gaussian process such that $E[Y_t] = 0$ and $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$. Assume that (4) holds true with $f$ a positive continuously differentiable real function such that (5) and (22) hold true. Assume further that $f_1$ and $f_2$ are continuously differentiable functions and that the function $\dot{y} \mapsto \int f_2(\dot{y})$ is one to one and denote $k$ its inverse function. The stationary measure $P'$ is such that for any bounded measurable function $g: \mathbb{R}^2 \to \mathbb{R}$

$$E'[g(Z_s, \dot{Z}_s)] = \lim_{H \to \infty} \frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s)ds = \iint_{\mathbb{R}^2} g(z, \dot{z})p_{Z,\dot{Z}}(z, \dot{z})dzd\dot{z}$$

with joint pdf

$$p_{Z,\dot{Z}}(z, \dot{z}) = \frac{1}{2\pi q_0} \frac{k'(f_1(z))}{f_1(z)^2 f_2 \left(k(\dot{z}/f_1(z))\right)} \exp \left(-\frac{z^2 + k^2(f_1(z))}{2}\right).$$ (23)

Proof. According to Proposition 2 we have

$$\lim_{H \to \infty} \frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s)ds = \frac{1}{2\pi q_0} \int\int_{\mathbb{R}^2} g(u_1, u_2 f(u_1, u_2)) \exp \left(-\frac{u_1^2 + u_2^2}{2}\right) du_1 du_2.$$

The change of variable $z = u_1$ and $\dot{z} = u_2 f(u_1, u_2)$ gives (23).
The model which defines \( Z \) in Proposition 4 is referred to as the model with no interaction hereafter.

4.2 Non-parametric estimation in the model with no interaction

**Identifiability constraints.** In this section we assume that the conditions of Proposition 4 hold true and discuss the estimation of the functions \( f_1 \) and \( f_2 \). These functions are defined up to a multiplicative constant and hereafter we assume, without loss of generality, that

\[
\int_{-\infty}^{+\infty} \frac{\exp\left(-u^2/2\right)}{f_2(u)} du = 1
\]

and thus that \( q_0 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\exp(-u^2/2)}{f_1(u)} du \).

**Marginal and conditional pdf.** We then deduce from (23) that the stationary marginal pdf of \( Z \) and conditional pdf \( \dot{Z}_t \) given \( Z_t = z \) are given respectively by

\[
p_Z(z) = \int_{\mathbb{R}} p_{Z,Z}(z, \dot{z}) d\dot{z} = \frac{1}{2\pi q_0} \frac{\exp\left(-\frac{\dot{z}^2}{2}\right)}{f_1(z)}
\]

and

\[
p_{Z|Z=z}(\dot{z}) = \frac{p_{Z,Z}(z, \dot{z})}{p_Z(z)} = \frac{k'}{f_1(z)} \frac{\dot{z}}{\hat{f}_1(z)^2} \exp\left(-\frac{k'}{2\hat{f}_1(z)^2}\right).
\]

Estimates of \( f_1 \) and \( f_2 \) can be derived from (25) as described below.

**Estimation of \( f_1 \).**

The estimate of \( f_1 \) considered hereafter is based on (26) where \( f_1(z) \) appears as a scale factor of the conditional distribution of \( \dot{Z}_t \) given \( Z_t = z \). More precisely we have

\[
E'[\dot{Z}_t|Z_t = z] = 0 \quad \text{and} \quad E'[|\dot{Z}_t||Z_t = z] = 2f_1(z).
\]

\( f_1 \) can be estimated by plugging in a non-parametric estimate of the conditional expectation \( E'[\dot{Z}_t|Z_t = z] \) in this relation. In practice we used the Nadaraya-Watson kernel estimate originally proposed in [10, 17].

An alternative estimate of \( f_1 \) is suggested by (25) which shows that the marginal distribution of \( Z \) depends only on \( f_1 \) and not on \( f_2 \). More precisely we have

\[
f_1(z) = \frac{1}{2\pi q_0} \frac{\exp(-\frac{z^2}{2})}{p_Z(z)}
\]

and an estimate of \( f_1 \) may be obtained by replacing the true pdf \( p_Z \) by a parametric or non-parametric estimate in this relation. Note however that the unknown constant \( q_0 \) appears in this relation and thus need to be estimated from the data. A possible estimate of \( q_0 \) may be obtained using the up-crossing intensity of the zeros level \( \nu^+(Z, 0) \) which satisfies (see Section 5 for more details)

\[
\nu^+(Z, 0) = \frac{1}{2\pi q_0}.
\]

This estimation strategy is not further discussed hereafter for simplicity reasons but may be useful in applications where the marginal distribution of \( Z_t \) (and its tails for example) is of
particular importance. We found that both approaches give similar results on simulations and the link between both estimates is further discussed in Section \(5\) (see (33)).

**Estimation of** \(f_2\). Let us denote \(V_t = \dot{Z}_t / f_1(Z_t)\). According to (26) \(V_t\) is independent of \(Z_t\) and the pdf \(p_v\) of \(V_t\) is

\[
p_v(v) = k'(v)k(v) \exp\left(-\frac{k(v)^2}{2}\right).
\]

It means in particular that for the model with no interaction the conditional distribution of \(\dot{Z}_t\) given \(Z_t = z\) depends on the level \(z\) only through a scale factor \(f_1(z)\). We then deduce that

\[
k'(v)k(v) \exp\left(-\frac{k(v)^2}{2}\right) = vp_v(v)
\]

and thus, after integration between 0 and \(v\) (note that \(k\) is increasing and \(k(0) = 0\)), we get

\[
1 - \exp(-k(v)^2/2) = \int_0^v up_v(u)du
\]

and finally

\[
k(v) = \text{sign}(v)\sqrt{-2 \ln\left(1 - \int_0^v zp_v(z)dz\right)}.
\] (29)

This expression is used to derive the following non-parametric estimate for \(f_2\) as described below.

1. Compute \(\dot{V}_t = \dot{Z}_t / f_1(Z_t)\) where \(\hat{f}_1\) is the estimate of \(f_1\).

2. Estimate \(\int_0^v zp_v(z)dz\) using a simple Monte Carlo approach, e.g. for \(v \geq 0\) by

\[
\frac{1}{N} \sum_{t:0 \leq V_t \leq v} \dot{V}_t
\]

where \(N\) denotes the length of the observed sequence. Plug this estimate in (29) to deduce an estimate \(\hat{k}\) of \(k\).

3. By definition, \(k\) is the inverse function of \(\dot{z} \mapsto \dot{z}f_2(\dot{z})\) and thus

\[
f_2(k(\dot{z})) = \frac{\dot{z}}{k(\dot{z})}.
\]

The next step consists in inverting numerically this relation after replacing \(k\) by \(\hat{k}\) to deduce an estimate \(\hat{f}_2\) of \(f_2\).

4. In practice the estimate is noisy close to the origin (the computation of \(\dot{z} / k(\dot{z})\) leads to a division by 0 at the origin) and we found it useful to smooth the estimate close to the origin. In this work the LOESS (locally weighted polynomial regression) smoothing method was used.
4.3 Some simulation results

The performance of the estimates discussed in Section 4.2 was assessed using simulations of the quadratic exponential model with no interaction considered in Section 3.3. Note that for this model we have

\[ \dot{y} f_2(\dot{y}) \propto \dot{y} \exp \left( \beta_2 \dot{y} + \frac{\beta_{2,2}}{2} \dot{y}^2 \right) \]

and that this function is strictly increasing on \( \mathbb{R} \) if and only if \( \beta_2^2 - 4\beta_{2,2} < 0 \). This condition is satisfied for the parameters values used for the simulations (see Section 3.3). It implies that the function \( \dot{y} \mapsto \dot{y} f_2(\dot{y}) \) is one to one and thus Proposition 1 applies.

Figure 4 shows typical parametric and non-parametric estimates of \( f \) obtained on a simulated sequence. Both estimates seem to be able to reproduce the global shape of \( f \) with the non-parametric estimate being more noisy. Figure 5 shows fluctuation intervals based on 1000 simulated sequences and permits a more systematic comparison of both estimates. Again both approaches provide reasonable estimates of \( f_1 \) and \( f_2 \) and there is no clear superiority of one of them. The chosen estimation method may thus depend on the application. The parametric estimate is specific to the quadratic exponential model. If this model is not appropriate for a particular application then other parametric model may be developed or the non-parametric approach may be considered. Note however that the non-parametric approach is valid only for the model with no interaction and that this assumption may also be restrictive for some applications. We will see in Section 6 that the quadratic exponential model seems realistic for the particular wave dataset considered in this paper. In such situation the parametric estimate has the additional advantage of providing a description of the observed sequence with a small number of parameters.

![Figure 4: Comparison of the true (left) function \( f \) with the parametric (middle) and non-parametric estimate (right) for a particular simulated sequence. The simulation experiment is described in Section 3.3.](image)

5 Model with marginal link function

In this section we discuss the statistical inference in the model defined by (4-5) where a marginal transformation is added to the model discussed in the previous sections. Section 5.1 is devoted to the estimation of the marginal link function \( h \). The estimation of the time change function \( f \) and
5.1 Up-crossing intensity and estimation of \( h \)

As already mentioned in the introduction, if \( f = 1 \) then the model reduces to \( h(X_t) = Y_t \) with \( \{Y_t\} \) a Gaussian process. This transformed Gaussian model is usual for analyzing wave data and at least two different strategies have been proposed in the literature to estimate \( h \). The first one is based on the marginal distribution of \( X \) which is directly related to \( h \), and the second one uses the up-crossing intensity (see [12, 2] and references therein).

Both approaches may be extended to the model defined by (4-5). However the marginal distribution of \( X \) depends on both \( f \) and \( h \) and it leads to complicated estimation procedures when trying to isolate the effect of \( h \) on this distribution. We have thus chosen to focus on methods based on up-crossings. Indeed the time deformation does not modify the levels of the observed sequence and thus the up-crossing intensity will mainly depend on \( h \) (up to a normalizing constant). This is stated more formally in the proposition below.

**Proposition 5.** Let \( Y \) be a univariate differentiable ergodic stationary Gaussian process such that \( E[Y_t] = 0 \) and \( \text{var}(Y_t) = \text{var}(Y_t) = 1 \). Assume that (4-5) hold true with \( f \) a positive measurable real function such that \( \mathfrak{H} \) holds true and \( h \) an increasing function. Let \( \nu_T^+(X, u) \) denotes the number of up-crossings of the level \( u \) of the process \( X \) on the time interval \([0, T]\) and \( \nu^+(X, u) := \lim_{T \to \infty} \frac{1}{T} \nu_T^+(X, u) \) the up-crossing intensity. Then

\[
\nu^+(X, u) = \nu^+(Z, h(u)) = \frac{1}{q_0} \nu^+(Y, h(u)) = \frac{1}{2\pi q_0} \exp \left( -\frac{h(u)^2}{2} \right).
\] (30)

**Proof.** Since \( h \) is increasing and the time change does not modify the observed level we have

\[
\nu_T^+(X, u) = \nu_T^+(Z, h(u)) = \nu_T^+(Y, h(u)).
\]

We deduce that
\[
\lim_{T \to \infty} \frac{1}{T} \nu_T^+(X, u) = \lim_{T \to \infty} \frac{1}{T} \nu_T^+(Z, h(u)) = \lim_{T \to \infty} \frac{\varphi(T)}{T} \frac{1}{\varphi(T)} \nu_T^+(Y, h(u))
\]

where \( \lim_{T \to \infty} \frac{\varphi(T)}{T} = q_0^{-1} \) (see proof of Proposition 1). The result can then be deduced from the following classical result which holds true for differentiable Gaussian processes (Rice formula, see e.g. [8])

\[
\lim_{T \to \infty} \frac{1}{T} \nu_T^+(Y, u) = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right).
\]

The previous proposition has been used to build non-parametric estimates of \( h \) as described below. According to (30) we get

\[
-\frac{h(x)^2}{2} = \log(\nu^+(X, x)) - \log\left(\frac{1}{2\pi q_0}\right).
\]

\( q_0 \) needs then to be estimated from the data in order to be able to use this relation to estimate \( h^2 \). For that we used that

\[
\sup_x \{\log(\nu^+(X, x))\} = \log\left(\frac{1}{2\pi q_0}\right).
\]

which shows that \( q_0 \) can be estimated from the mode of the up-crossing intensity (ie the intensity of the most often crossed level). Once \( q_0 \) is estimated we are faced to a similar estimation problem than discussed in [12] for transformed Gaussian models. It consists in taking a square root of (31) using the monotony of \( h \) and eventually smoothing the obtained estimates (the loess method was used in this work). An isotonic regression method may also be applied to force the obtained estimates to be increasing but this was not necessary for the practical cases considered in this paper.

Figure 6 shows simulation results obtained with the setting described in Section 3.3 and \( h \) estimated from the data (the true \( h \) function used for the simulations is shown as a full line on the left plot). The bias and variance of the estimate are generally low with higher variances in the tails where the number of up-crossings is lower. Note also that the concavity of \( h \) implies that the marginal distribution of \( X \) has a fatter upper tail compared to the lower tail and this has consequences on the estimate of \( h \) which has a higher variance in the upper tail.

5.2 Estimation of \( f \)

The estimation of \( f \) when the process \( Z \) is observed has been discussed in the previous sections. However only \( X \) is generally observed when considering the model (4-5). Once \( h \) is estimated it is possible to compute an approximation of \( Z \) as

\[
\hat{Z}_t = \hat{h}(X_t)
\]

with \( \hat{h} \) the estimate of \( h \). Estimates of the time change function \( f \) can then be obtained by replacing the latent process \( Z \) by \( \hat{Z} \) in the methods described in Sections 3 and 4. Comparing the right plots of Figures 6 and 3 shows that using \( \hat{Z} \) instead of \( Z \) modifies the statistical
properties of the estimates of \( f \) in the parametric case (similar results were obtained with the non-parametric estimates introduced in Section 4). It is not surprising since \( h \) is estimated such that the up-crossing intensity of \( \hat{Z} \) has a Gaussian shape (see (30)) and this up-crossing intensity is related to the joint distribution of \( (\hat{Z}, \hat{\dot{Z}}) \) via the Rice formula as follows

\[
\nu^+(\hat{Z}, z) = \int_{\mathbb{R}} p_{\hat{Z}, \hat{\dot{Z}}}(z, \hat{\dot{z}}) d\hat{\dot{z}} \approx \frac{1}{2\pi q_0} \exp(-\frac{z^2}{2})
\]

Estimates of \( f \) are based on this joint distribution and thus may be impacted when using \( \hat{Z} \) instead of \( Z \). Note also that this relation implies that

\[
p_{\hat{Z}}(z) E[|\hat{\dot{Z}}_t|^2] \approx \frac{1}{2\pi q_0} \exp(-\frac{z^2}{2}) \tag{33}
\]

and may explain why the non-parametric estimates of \( f_1 \) based on (28) or (27) lead to similar results (it can be shown that a similar relation holds true when replacing \( p_{\hat{Z}}(z) \) and \( E[|\hat{\dot{Z}}_t|^2] \) by non-parametric kernel estimates).

Figure 6: Left plot: fluctuation interval for the estimate of \( h \). The full lines correspond to the true function, the dashed line to the mean estimated values and the dotted line to 95% fluctuation interval. Right plot: distribution of \( \tilde{\beta} - \beta^* \) in the quadratic exponential model when \( h \) is unknown. Results based on 1000 simulated sequences. The simulation experiment is described in Section 3.3.

5.3 Estimation of the second order structure of \( Y \)

\( Y \) is stationary Gaussian process with zero mean and is thus completely characterized by its second order structure. Again the process \( Y \) is not directly observed but can be approximated when the function \( h \) and \( f \) have been estimated. In order to simplify the presentation we assume in this section that the process \( Z \) is observed. In the practical applications it is not observed but replaced by the approximation \( \hat{Z} \) introduced in the previous section. Recall that the time change is defined through the following differential equation

\[
\psi(t) = \int_0^t f(Y_{\psi(s)}, \dot{Y}_{\psi(s)}) ds \tag{34}
\]

with \( Y_{\psi(t)} = Z_t \) and \( \dot{Y}_{\psi(s)} \) related to \( (Z_t, \dot{Z}_t) \) implicitly as follows

\[
\dot{Z}_t = f \left( Z_t, \dot{Y}_{\psi(t)} \right) \tag{35}
\]
A Euler scheme has then been used to compute an approximation of $\varphi$ defined by (34) where the implicit equation (35) is solved at each time step. In the separable model defined in Section 4, it can be easily done using

$$\dot{\varphi}(t) = k \left( \frac{\dot{Z}_t}{J_1(Z_t)} \right)$$

and the function $k$ is directly estimated from the data. A numerical optimization procedure was used to solve (35) in the quadratic exponential model.

We can then deduce an approximation $\hat{Y}$ of $Y$ and compute a standard estimate of the second order structure (autocorrelation or spectral density) on the sample $\hat{Y}$.

6 Application to wave data

This section starts with a brief overview of physically based wave models which lead to time or space deformations. We then introduce the wave dataset and discuss the results obtained when fitting the model (4-5) to this dataset.

6.1 Physical motivations

In the gravity wave theory, the non-linear solution of the free surface evolution has been proven to be well approximated by the first order linear Lagrangian solution of the Euler equations (Gerstner wave in harmonic case, Pierson [7] for multi-harmonic or irregular waves). Let $(p_0 + \delta(t, p_0), \chi(t, p_0))$ denote the two-dimensional linear trajectory of a water particle of the free surface located in $p_0$ at time $t = 0$, such that the height of the water surface at time $t$ and location $p_0 + \delta(t, p_0)$ is $\chi(t, p_0)$. The non-linear (Eulerian) free surface elevation $\eta$ is thus given by

$$\eta(p_0 + \delta(t, p_0), t) = \chi(t, p_0).$$

It corresponds to a modulation in space of the vertical movements of the particles. In the case where the modulation is strong, the solution of this implicit equation may not be unique. It corresponds in the physical meaning to wave breaking.

More recently a canonical Lie transformation was proposed for gravity waves in order to transform the real physical world in a linear (more exactly close to linear) world or inversely (see [6]). In the infinite water depth case it leads to an Inviscid Burgers’ equation on the free surface elevation

$$\frac{\partial}{\partial \lambda} \tilde{\eta}(p, t, \lambda) + \frac{1}{2} \frac{\partial}{\partial p} \tilde{\eta}^2(p, t, \lambda) = 0$$

with $0 \leq \lambda \leq 1$ and $\tilde{\eta}$ the Hilbert transform in space. $\eta_1(p, t) = \eta(p, t, 1)$ and $\eta_0(p, t) = \eta(p, t, 0)$ correspond to respectively the free surface elevation in the linear and non-linear worlds. The solution of the Burgers’ equation satisfies

$$\tilde{\eta}_0(p - \tilde{\eta}_1(p, t), t) = \tilde{\eta}_1(p, t).$$

Again modulation of space appears as a natural way to transform linear waves to non-linear ones. In the case of finite water depth and of more complex situations with varying sea bottom in shallow water, the solution of the Euler equations are not so simple, but the idea of modulation by a function of the process itself could appear still pertinent.
6.2 Data

The proposed methodology was used to analyze wave data from an experiment carried out in a wave flume. A detailed description of the experiment can be found in [4]. The bathymetric profile of the flume experiment is shown on Figure 7 together with the various locations where the sea-surface elevation is measured (numbered from 1 close to the wave-maker to 16). The wave-maker generates random phase waves which are expected to be close to a realization of a Gaussian process with a JONSWAP-type spectra. Then the waves propagate to the right on a varying bathymetry which induces strong non-linearities. Examples of time series measured at locations 2 and 7 are shown on Figure 7. It shows in particular that both horizontal and vertical asymmetries increase when the waves propagate to the right. Free surface elevation is recorded over a duration of 40 minutes with a sampling time-step of 0.070. The peak period is about 2.4s. It corresponds to about 1000 waves with 34 points per wave.

![Figure 7](image_url)

Figure 7: Top panel (from [4]): lay-out of the experimental set-up for the irregular flume experiment. The numbers indicate the position of the wave probes. Sequences of waves measured at stations 2 (middle panel) and 7 (bottom panel)

6.3 Results

The model (4-5) was fitted at the 16 locations where wave measurements are available but in the discussion below we mainly focus on the central location 7. According to Figure 8 the parametric approach for the quadratic exponential model and the non-parametric approach for the model with no interaction lead to estimates with similar shapes for $f$ at location 7. It is a first indication that the choice of the quadratic exponential model is sensible for the wave dataset considered in this paper. Both the parametric and non-parametric approaches were validated using the criteria discussed below and generally slightly better results were obtained with the parametric approach. We have thus chosen to focus on this approach hereafter. It has the
additional advantage of providing a parsimonious description of the observed sequence with only 6 parameters involved. Figure 9 shows the evolution of the estimates with the measurement location. At locations 1 and 2, where it is expected that the observed sequences may be modeled as a realization of a Gaussian process, low estimated values are obtained for all the parameters except the intercept $\beta_0$. It corresponds to an almost constant time change function as in the Gaussian case. Then from location 3 the parameters values evolve with the varying bathymetry. This evolution is relatively smooth in space. This could be used for example to interpolate the parameter values and simulate synthetic waves at locations with no observation. Note that parametric models such as power transformations and classical wave spectral models may also be considered for $h$ and the spectrum of $Y$ to get a full parametric model.

According to Figure 9 the values of the interaction term $\beta_{1,2}$ is always low and this suggests that a model with no interaction ($\beta_{1,2} = 0$) may be appropriate. Fluctuation intervals for the estimates can be computed using similar simulations as in Section 5.2. The 95% interval obtained for $\beta_{1,2}$ at location 7 is $[-0.021, -0.008]$. It suggests that the interaction term is low but still statistically significant.

Let us now focus on location 7 and validate the fitted quadratic exponential model. Figure 1 shows the estimates of $h$, $f$ and the spectral density of $Y$ at location 7. As already mentioned in the introduction the estimated $f$ function is increasing in both variables and this permits to create both vertical and horizontal asymmetries by accelerating the chronometer in the crests and in the front of the waves. $h$ is concave and thus enhance again the sharpness of the crests by boosting high levels. If the fitted model is appropriate then the marginally and time transformed sequence $\hat{Y}$ should have paths which can be reasonably modeled as a realization of a Gaussian process. No formal test was performed but Figure 1 shows a sequence of this process and that the marginal and time transformations permit to symmetrize the path of the observed sequence. Another way to validate the model consists in generating sequences of the fitted model and compare the statistical properties of the simulated sequences with the ones of the original data. We focus on this validation strategy hereafter since it is directly related to the practical application (we want a model which can generate realistic wave sequences) and it is also probably easier to interpret than validation in the Gaussian domain. Note that the fitted model can be very quickly simulated (about one half second to generate 1000 waves on a basic laptop using Matlab).

An example of simulated sequence is shown on Figure 10. It looks visually coherent with the
Figure 9: Estimates of $\beta$ at the 16 measurement locations.

Figure 10: A sequence of observation (top) and simulated data (bottom) at location 7.

observed sequence shown on the same figure. A more systematic comparison is performed below.

Figure 11 shows the marginal and joint pdf of $(X_t, \dot{X}_t)$. This distribution is relatively complex
with an asymmetric distribution for $X_t$ (the empirical skewness of $X_t$ is 0.75), a leptokurtic
marginal distribution for $\dot{X}_t$ (the empirical skewness of $\dot{X}_t$ is 37.8) and a complex relation between
both variables with for example higher steepness in the crests. The fitted model is able to reproduce the complexity of this distribution. The marginal distribution of $X_t$ is of particular importance for the applications and the quantile-quantile plot shown on Figure 12 confirms that the fit is good except maybe in the upper tail where the fit is slightly less good. An appropriate parametric model for the tails of $h$ may be useful here.

**Figure 11:** Joint distribution of $X_t$ (x-axis) and $\dot{X}_t$ (y-axis) for the observation (left) and the fitted model (right).

**Figure 12:** Quantile-quantile plot, empirical spectra and autocorrelation functions at location 7 for the data (grey curves) and the fitted quadratic exponential model (black curves). The dotted lines are 95% fluctuation intervals computed using simulations.

Figure 12 also shows that the second order structure of the data is well reproduced by the fitted model except the third peak in the observed spectrum. This third peak is induced by non linear effect which create a low amplitude high frequency components which phase is linked to the other frequencies. The fitted model is not able to reproduce this non-linear interaction but still reproduces the associated asymmetries. Indeed Figure 13 shows the distribution of the various slopes (crest front, crest back, trough front, trough back) computed after extracting individual waves using zeros-crossings (see [11] for a more precise definition and [3] for a review of existing measures of asymmetries). The differences in these four distributions are linked to the asymmetries in the sea surface elevation. For example it shows that the fronts of the crests are the steepest whereas the fronts of the troughs are the flattest. These slope distributions are well reproduced by the fitted model.
Figure 13: From left to right: distribution of the crest front steepness, crest back steepness, trough front steepness, trough back steepness. For each criterion, the first boxplots correspond to the observation and the second one to the fitted model.

7 Conclusions and perspectives

This paper proposes a new model for asymmetric time sequences. The originality of the proposed model consists in introducing a time change which is dependent of the observed process. The proposed statistical inference is based on the up-crossings and on the joint distribution of the process and its derivatives. Simplifying assumptions are imposed on the shape of the time change function $f$ in order to be able to derive estimates for $f$. We show that these assumptions seems realistic for the particular wave dataset considered in this paper but it may be useful to consider other cases for other applications. This remains to be further investigated together with other modeling issues such as the extension to a spatial or space-time context.

Appendix: proof of Proposition 1

Proof. Let us denote the shifted processes $Y^\tau = Y_{,+\tau}$ and $(\psi^\tau, Z^\tau)$ the processes obtained by replacing $Y$ is replaced with $Y^\tau$ in (3). We get

\[
\psi^\tau(t) = \int_0^t \frac{1}{f(Y_s)} \, ds = \int_0^t \frac{1}{f(Y_{s+\tau})} \, ds = \psi(\tau + t) - \psi(\tau)
\]

\[
Z^\tau_{(\psi(\tau+t)-\psi(\tau))} = Z^\tau_{\psi^\tau(t)} = Y^\tau_t = Y(t + \tau) = Z_{\psi(\tau+t)}.
\]

It can then be deduced that $Z^\tau = Z_{,+\psi(\tau)}$ which means that the effect of a shift on $Y$ is a random shift on $Z$.

Now, by the stationarity of $Y$, we have for any $\tau > 0$ and any function $g$ of the form (6)

\[
q_0 E'[g(Z_\tau)] = E \left[ \frac{g(Z_\tau)}{f(Z_0)} \right] = E \left[ \frac{g(Z^\tau)}{f(Z_0)} \right] = E \left[ \frac{g(Z_{,+\psi(\tau)})}{f(Z_{,\psi(\tau)})} \right].
\]

Hence, for any $H > 0$

\[
q_0 E'[g(Z_\tau)] = \frac{1}{H} \int_0^H E \left[ \frac{g(Z_{,+\psi(\tau)})}{f(Z_{,\psi(\tau)})} \right] \, d\tau.
\]
\[
\begin{align*}
&= E \left[ \frac{1}{H} \int_0^{\psi(H)} g(Z_{s+s}) \frac{\dot{\varphi}(s)}{f(Z_s)} ds \right] \\
&= E \left[ \frac{1}{H} \int_0^{\psi(H)} g(Z_{s+s}) ds \right]
\end{align*}
\]

and for any \( t > 0 \)
\[
q_0 \left( E'[g(Z_s)] - E'[g(Z_s + t)] \right) = E \left[ \frac{1}{H} \int_0^{\psi(H)} g(Z_{s+s}) ds \right] - E \left[ \frac{1}{H} \int_0^{\psi(H)+t} g(Z_{s+s}) ds \right]
\]
\[
= E \left[ \frac{1}{H} \int_0^{t} g(Z_{s+s}) ds \right] - E \left[ \frac{1}{H} \int_{\psi(H)+t}^{\psi(H)+t+s} g(Z_{s+s}) ds \right]
\]
\[
q_0 |E'[g(Z_s)] - E'[g(Z_s + t)]| \leq \frac{2tH}{H} \|g\|_{\infty}.
\]

Since the right hand side can be made arbitrarily small, we have proven the stationarity. Similar equations can be used for proving the ergodicity. Indeed we have
\[
\frac{1}{T} \int_0^T g(Z_{u+s}) ds = \frac{1}{T} \int_0^{\varphi(T)} \frac{g(Z_{u+\varphi(u)})}{f(Z_{\varphi(u)})} du = \frac{\varphi(T)}{T} \frac{1}{T} \int_0^{\varphi(T)} \frac{g(Z_{u})}{f(Z_{u})} du = \frac{\varphi'(T)}{T} \int_0^{T} \frac{du}{f(Y_u)} (36)
\]

where \( \varphi \) denotes the reciprocal function of \( \psi \). Note that
\[
\lim_{T \to \infty} \frac{\varphi(T)}{T} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{du}{f(Y_u)} = q_0
\]

by the ergodicity of \( Y \) and thus \( \lim_{T \to \infty} \frac{\varphi(T)}{T} = q_0^{-1} \). Using again the ergodicity of \( Y \) and that \( \lim_{T \to \infty} \varphi(T) = +\infty \) we deduce that the second term in the right hand size term of (36) converges to a deterministic limit which is its expected value.

(37) is a particular case of the general convergence result proved above with \( g(Y) = g_0(Y_t) \) and remarking that \( Y_0 = Z_0 \).

\section*{References}

[1] E. B. Anderes and M. L. Stein. Estimating deformations of isotropic gaussian random fields on the plane. \textit{The Annals of Statistics}, pages 719–741, 2008.

[2] J.-M. Azaïs and M. Wschebor. \textit{Level sets and extrema of random processes and fields}. John Wiley & Sons, 2009.

[3] A. Baxevani, K. Podgórski, and J. Wegener. Sample path asymmetries in non-gaussian random processes. \textit{Scandinavian Journal of Statistics}, 41(4):1102–1123, 2014.

[4] F. Becq-Girard, P. Forget, and M. Benoit. Non-linear propagation of unidirectional wave fields over varying topography. \textit{Coastal Engineering}, 38(2):91–113, 1999.

[5] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. \textit{Econometrica: journal of the Econometric Society}, pages 135–155, 1973.

[6] D. B Creamer, F. Henyey, R. Schult, and J. Wright. Improved linear representation of ocean surface waves. \textit{Journal of Fluid Mechanics}, 205:135–161, 1989.
[7] Pierson W.J. Jr. Models of random seas based on the lagrangian equations of motion. Technical report, Technical report, College of Engineering, Research Division, New York University., 1961.

[8] G. Lindgren. Lectures on stationary stochastic processes. Lecture Notes, Lund University, 2006.

[9] G. Lindgren. Exact asymmetric slope distributions in stochastic gauss–lagrange ocean waves. Applied Ocean Research, 31(1):65–73, 2009.

[10] E.A. Nadaraya. On estimating regression. Theory of Probability & Its Applications, 9(1):141–142, 1964.

[11] N. Raillard, M. Prevosto, and P. Ailliot. Modeling process asymmetries with Laplace moving average. Computational Statistics & Data Analysis, 81:24–37, 2015.

[12] I. Rychlik, P. Johannesson, and M.R. Leadbetter. Modelling and statistical analysis of ocean-wave data using transformed gaussian processes. Marine Structures, 10(1):13–47, 1997.

[13] P.D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association, 87(417):108–119, 1992.

[14] A.M. Schmidt, P. Guttorp, and A. O’Hagan. Considering covariates in the covariance structure of spatial processes. Environmetrics, 22(4):487–500, 2011.

[15] A.M. Schmidt and A. O’Hagan. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(3):743–758, 2003.

[16] A.E.D. Veraart and M. Winkel. Time change. Encyclopedia of quantitative finance, 2010.

[17] G.S. Watson. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A, pages 359–372, 1964.