


Time-change models for asymmetric processes

Pierre Ailliot¹  | Bernard Delyon² | Valérie Monbet^{2,3} | Marc Prevosto⁴

¹Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, Brest, France

²Univ Rennes, CNRS, IRMAR - UMR 6625, Rennes, France

³INRIA/ASPI, Rennes, France

⁴IFREMER, Marine Structures Laboratory, Plouzané, France

Correspondence

Pierre Ailliot, Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, Brest, France.
Email: pierre.ailliot@univ-brest.fr

Abstract

Many records in environmental sciences exhibit asymmetric trajectories. The physical mechanisms behind these records may lead for example to sample paths with different characteristics at high and low levels (up–down asymmetries) or in the ascending and descending phases leading to time irreversibility (front–back asymmetries). Such features are important for many applications, and there is a need for simple and tractable models that can reproduce them. In this paper, we explore original time-change models where the clock is a stochastic process that depends on the observed trajectory. The ergodicity of the proposed model is established under general conditions, and this result is used to develop nonparametric estimation procedures based on the joint distribution of the process and its derivative. The methodology is illustrated on meteorological and oceanographic data sets. We show that, combined with a marginal transformation, the proposed methodology is able to reproduce important characteristics of the data set such as marginal distributions, up-crossing intensity, and up–down and front–back asymmetries.

KEYWORDS

asymmetric records, ergodicity, Gaussian processes, time change, time irreversibility

1 | INTRODUCTION

Many situations in environmental science or in econometrics lead to study asymmetric records. For example, the propagation of ocean waves in shallow water leads to waves with sharper crests compared to the troughs (referred to as *up–down asymmetries* hereafter) and steeper fronts

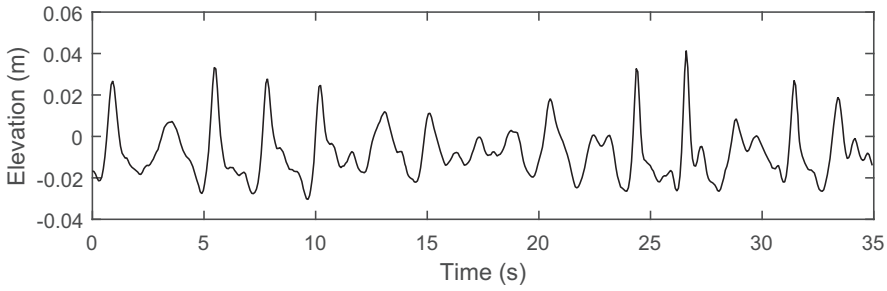


FIGURE 1 Sequence of sea-surface elevation in shallow water (see Section 4 for a description of the data)

compared to the backs (*front-back asymmetries*). This leads to modeling records such as the one shown in Figure 1 where the two asymmetries can easily be seen.

Gaussian processes fail to represent such asymmetric features. A very classical alternative in time series analysis, sometimes referred to as the *transformed Gaussian* (TG) model, consists in applying a marginal transformation to a Gaussian process where the marginal transformation is generally chosen to match non-Gaussian margins (see, e.g., Grigoriu, 1998). TG processes can reproduce some constrained form of up-down asymmetries but no front-back asymmetries. In this paper, we propose to add a time change in the static TG framework. It allows to stretch symmetric Gaussian sequences both vertically and horizontally and gives flexibility in the modeling of up-down asymmetries and front-back asymmetries.

A multivariate *time-changed process* $\mathbf{Z} = (\mathbf{Z}_t)_{t \geq 0}$ is defined as

$$\mathbf{Z}_t = \mathbf{Y}_{\varphi(t)}, \tag{1}$$

where $\{\varphi(t)\}_{t \geq 0}$ denotes a nondecreasing stochastic process referred to as the *time change* or *clock* hereafter and \mathbf{Y} denotes a multivariate process referred to as the *base process*. Hereafter, we assume that only the time-changed process \mathbf{Z} is observable, and we aim at finding a time change, which is such that the nonobservable (or latent) process \mathbf{Y} is a multivariate stationary Gaussian process. An originality of the *dependent time-change model* considered in this paper consists in assuming that the time change φ is a stochastic process that depends on the base process \mathbf{Y} as follows:

$$\varphi(t) = \int_0^t f(\mathbf{Y}_{\varphi(s)}) ds, \tag{2}$$

where f is a positive function that describes the speed of the clock. It is referred to as the *time-change function* hereafter. In Section 2, an ergodicity result is given for this general model. Note that \mathbf{Y} is a multivariate process, and thus, the speed of the clock may depend on several covariates.

For example, in order to generate asymmetric trajectories from a univariate Gaussian process Y with symmetric sample path, we can let the time change depend on $\mathbf{Y} = (Y, \dot{Y})$, where \dot{Y} denotes the derivative of Y and consider the processes Z defined as

$$Z_t = Y_{\varphi(t)}. \tag{3}$$

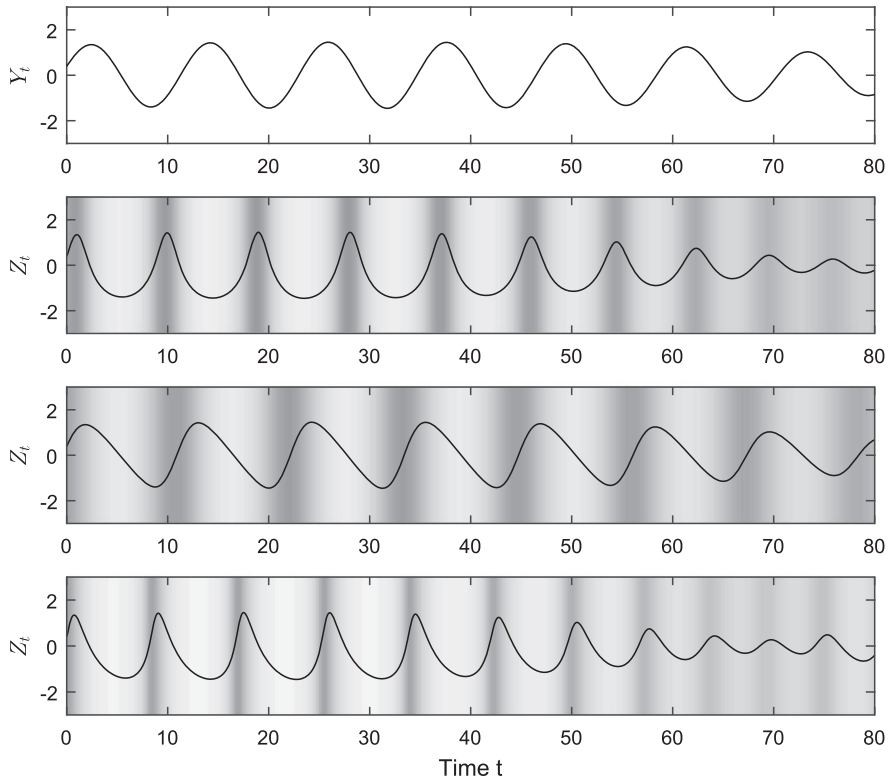


FIGURE 2 Sequences simulated with the time-change model (3), (4). (Top) Gaussian sequence ($f_1(y) \equiv f_2(\dot{y}) \equiv 1$). (Second row) sequence with up–down asymmetries ($f_1(y) = 2s(y)$, $f_2(\dot{y}) \equiv 1$ with $s(y) = \frac{\exp(y)}{1+\exp(y)}$). (Third row) sequence with front–back asymmetries ($f_1(y) \equiv 1$, $f_2(\dot{y}) = 2s(\dot{y})$). (Bottom) sequence with up–down and front–back asymmetries ($f_1(y) = 2s(y)$, $f_2(\dot{y}) = s(\dot{y})$). The background color corresponds to the value of $f_1(Y_{\varphi(t)})f_2(\dot{Y}_{\varphi(t)})$ (speed of the clock)

In the sequel, we focus on the special case on separable time-change functions f , which can be written as the product of two functions f_1 and f_2 such that

$$\varphi(t) = \int_0^t f_1(Y_{\varphi(s)}) f_2(\dot{Y}_{\varphi(s)}) ds. \tag{4}$$

Figure 2 shows that it is possible to generate asymmetries with the model (3), (4). The top panel corresponds to the case where $f_1 \equiv f_2 \equiv 1$, thus to the case where $Z_t = Y_t$ is a Gaussian sequence with symmetric path. The second row of panels corresponds to the case where f_1 is an increasing function and $f_2(\dot{y}) \equiv 1$. In this case, the clock accelerates with the level of the base process leading to a time-changed process Z with narrower crest compared to the ones of Y and, thus, up–down asymmetries. The opposite behavior will be observed if f_1 is a decreasing function. In the third case, we assume that f_2 is an increasing function and $f_1(y) \equiv 1$; thus, the clock accelerates with the derivatives of the process. It leads to front–back asymmetries with steeper wave fronts compared to the backs. In the bottom plot, we assume that f_1 and f_2 are increasing functions; thus, the clock also accelerates with the derivatives of the process, and this leads to both up–down and front–back asymmetries.

In this paper, we propose to combine marginal transformation and time change and consider the model where the observed X satisfies

$$h(X_t) = Z_t,$$

where Z is the time-changed process defined by (3, 4) and the *marginal link function* h is assumed to be one to one. This paper mainly aims at investigating some theoretical properties of this model and proposing methods to fit it on real data, more precisely, to estimate h , f_1 , f_2 from the observation of a trajectory of $(X_t)_{0 \leq t \leq T}$.

Let us shortly discuss how this model compares to previous works. Models with time change have become popular for financial applications because they were introduced in Clark (1973); see also, for example, Veraart and Winkel (2010) for a recent review. In this application field, the modified time $\varphi(t)$ is generally interpreted as a “business time”: periods with higher market activity corresponds to periods when the markets move at a higher speed and higher volatility. In these financial applications, Y and φ are generally assumed to be *independent* Levy processes, whereas in the model considered in this paper, they are strongly dependent of each other.

Deformations have also been extensively used for modeling nonstationary fields in environmental applications since the seminal paper of Sampson and Guttorp (1992). It generally consists in seeking a transformation φ to map a stationary field Y into a nonstationary field Z through (1). These methods can deal with multidimensional (e.g., spatial) fields, but the estimation of φ generally require *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 have been proposed (see Anderes & Stein, 2008, and the references therein). Note that, in this application field, φ is generally assumed to be *deterministic*, whereas in the present paper, it is a stochastic process and we add the additional constraint that Y is not only stationary but also Gaussian.

Another related technique is dynamic time warping where the transformation φ is chosen to “align” the processes Y and Z by minimizing some criteria (see, e.g., Wang & Gasser, 1997). When using dynamic time warping methods, *both Y and Z are observed*, whereas hereafter, Y is not directly observable and the transformation φ is estimated such that it transforms the observed sequence Z into a Gaussian sequence Y .

The paper is organized as follows. The ergodicity of the general model (1, 2) is discussed in Section 2.1. This is the key result that is used in the following sections to fit time-change models. Then, in Section 2.2, the joint density of the time-changed process (Z_t, \dot{Z}_t) is derived (as a function of f_1 and f_2). Section 3 is dedicated to the estimation of the marginal transformation function h and time-change functions f_1 and f_2 . In Section 4, the model is applied to real data. We show that dynamic time-change and static marginal transformations can be combined in order to reproduce asymmetries, which are observed in atmospheric pressure time series and shallow water wave records. Conclusions are given in Section 5. Theoretical results and proofs are given in the Appendix.

2 | TIME-CHANGE MODEL

2.1 | Ergodicity of the dependent time-change model

In this section, we discuss the ergodicity of general time-changed process \mathbf{Z} defined by (1, 2). This is important because it shows that statistical properties of the model can be derived from a single long-enough realization as soon as the marginal distribution of the base process Y is known.

It is useful to consider the multivariate setting for the next sections and bold letters are used to stress that \mathbf{Y} and \mathbf{Z} have values on \mathbb{R}^d . In order to ensure that the model is well defined, we assume that $(\mathbf{Y}_t)_{t \geq 0}$ has continuous sample paths and that $f(\cdot)$ is a measurable real-valued positive function defined on \mathbb{R}^d . Let g be a measurable real-valued function defined on \mathbb{R}^d and let us discuss the convergence of the time average estimate defined as $\frac{1}{H} \int_0^H g(\mathbf{Z}_s) ds$ when H tends to infinity. Using the model definition and a change of variables, we obtain

$$\frac{1}{H} \int_0^H g(\mathbf{Z}_s) ds = \frac{\varphi(H)}{H} \frac{1}{\varphi(H)} \int_0^{\varphi(H)} \frac{g(\mathbf{Y}_s)}{f(\mathbf{Y}_s)} ds. \tag{5}$$

In order to show the convergence of the quantities that appear in the left-hand side of (5), we further assume that \mathbf{Y} is a stationary ergodic process. This implies the convergence almost everywhere (a.e.) of time averages to a deterministic value, which is its expected value and in particular that

$$\lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H \frac{1}{f(\mathbf{Y}_s)} ds = q_0 \quad \text{a.e.,}$$

provided that $q_0 = E[\frac{1}{f(\mathbf{Y}_0)}] < +\infty$.

Now, remark that the function ψ defined as $\psi(t) = \int_0^t \frac{1}{f(\mathbf{Y}_s)} ds$, which appears in the previous expression, is the reciprocal function of φ . We have $\lim_{H \rightarrow \infty} \frac{\psi(H)}{H} = q_0$ a.e., and thus,

$$\lim_{H \rightarrow \infty} \frac{\varphi(H)}{H} = q_0^{-1} \text{ a.e.}$$

This shows the convergence of the first term, which appears in the right-hand term of (5). For the convergence of the second term, we further assume that $E[\frac{|g(\mathbf{Y}_0)|}{f(\mathbf{Y}_0)}] < \infty$. Using again that \mathbf{Y} is ergodic and that $\lim_{H \rightarrow \infty} \varphi(H) = +\infty$ a.e., we obtain that

$$\lim_{H \rightarrow \infty} \frac{1}{\varphi(H)} \int_0^{\varphi(H)} \frac{g(\mathbf{Y}_s)}{f(\mathbf{Y}_s)} ds = E \left[\frac{g(\mathbf{Y}_0)}{f(\mathbf{Y}_0)} \right] \text{ a.e.}$$

Finally, we have shown that

$$\lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(\mathbf{Z}_s) ds = q_0^{-1} E \left[\frac{g(\mathbf{Y}_0)}{f(\mathbf{Y}_0)} \right] \text{ a.e.} \tag{6}$$

This is the key result for the rest of this paper. It shows that empirical means computed on a realization of the time-changed process \mathbf{Z} converge, when the length of the realization tends to infinity, to a limit that depends on the marginal distribution of the base process \mathbf{Y} and the time-change function f in a simple way.

A more formal and general statement of this result is given in Theorem 1, which shows that the process \mathbf{Z} is stationary and ergodic under a measure equivalent to E . This formalism is not necessary to understand the rest of this paper and is given in the Appendix, together with some comments on the theoretical properties of the model.

2.2 | Transformed time-changed process

In the remainder of this paper, we focus on the particular models that are defined below.

Definition 1. A **transformed time-changed Gaussian** (TTCG) process X is defined as

$$h(X_t) = Z_t, \quad Z_t = Y_{\varphi(t)}, \quad \varphi(t) = \int_0^t f_1(Y_{\varphi(s)})f_2(\dot{Y}_{\varphi(s)})ds \tag{7}$$

with h a continuously differentiable increasing function, Y a continuously differentiable ergodic stationary Gaussian process such that $E[Y_t] = 0$ and $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$, and f_1 and f_2 positive continuous real-valued functions defined on \mathbb{R} such that $q_0 = E[\frac{1}{f_1(Y_0)f_2(\dot{Y}_0)}] < +\infty$. When $h(x) = x$ (no marginal transformation), the process $X = Z$ is referred to as the **time-changed Gaussian** (TCG) process.

Note first that only the product f_1f_2 appears in the model definition, and thus, f_1 and f_2 are defined up to a multiplicative constant. A constraint is thus needed to ensure identifiability. Hereafter, we assume, without loss of generality, that

$$\int_{-\infty}^{+\infty} \frac{\exp(-u^2/2)}{f_2(u)} du = \frac{1}{\sqrt{2\pi}}. \tag{8}$$

It implies that $q_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{\exp(-u^2/2)}{f_1(u)} du$, which is useful to simplify various expressions in the sequel.

Note that scaling h, f_1 , and f_2 permits to assume, without restriction, that $E[Y_t] = 0$, $\text{var}(Y_t) = 1$ and $\text{var}(\dot{Y}_t) = 1$ (see the discussion in Appendix A.3).

In the TTCG model, starting from a Gaussian process Y with symmetric trajectories, we combine both a vertical stretching (associated to the marginal transformation h) and horizontal stretching (associated to the time change) to describe asymmetric trajectories. As illustrated in Figure 2, f_1 describes the effect of the level of the process on the time change and is thus related to up–down asymmetries, whereas f_2 models the effect of the slope on the time-change and front–back asymmetries. We expect that an increasing f_1 (respectively f_2) function will lead to a process Z with crests narrower than the troughs (respectively ascending phases steeper than descending phases). The time-change function $f(y, \dot{y}) = f_1(y)f_2(\dot{y})$ has a product form, and thus, there is no interaction between the effect of the level and the derivative (separable model). Some results given in this paper are also valid in the more general case where f is not separable. However, the separability assumption leads to a more tractable expression for the joint stationary probability density function (pdf) and simplifies the statistical inference.

The TCG model of Definition 1 is a particular case of the general time-changed model (1)–(2) with $\mathbf{Y}_t = (Y_t, \dot{Y}_t)$, and the results given in Section 2.1 apply. The conditions on the base process Y and the time-change function f ensure that the process Z is well defined and continuously differentiable. The following proposition gives the joint stationary pdf of (Z_t, \dot{Z}_t) for the TCG model. It is one of the key results used to build nonparametric estimates for f_1 and f_2 in Section 3.

Proposition 1. *Let Z be a TCG process with a separable time-change function. Assume further that f_1 and f_2 are continuously differentiable positive functions and that the function $\dot{y} \mapsto \dot{y}f_2(\dot{y})$ is a bijection from \mathbb{R} to \mathbb{R} and denote by k its inverse function. Then, for any bounded measurable function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$,*

$$\lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s) ds = \iint_{\mathbb{R}^2} g(z, \dot{z}) p_{Z, \dot{Z}}(z, \dot{z}) dz d\dot{z} \text{ a.e.}$$

with joint stationary pdf

$$p_{Z,\dot{Z}}(z, \dot{z}) = \frac{1}{q_0} \frac{\dot{k}\left(\frac{\dot{z}}{f_1(z)}\right)}{f_1(z)^2 f_2\left(k\left(\frac{\dot{z}}{f_1(z)}\right)\right)} p_{Y,\dot{Y}}\left(z, k\left(\frac{\dot{z}}{f_1(z)}\right)\right), \tag{9}$$

where \dot{k} denotes the derivative of k and $p_{Y,\dot{Y}}$ denotes the joint pdf of (Y_t, \dot{Y}_t) .

The proof of this proposition is given in the Appendix. It is based on the application of Equation (6) with $\mathbf{Y}_t = (Y_t, \dot{Y}_t)$ and $\mathbf{Z}_t = (Y_{\varphi(t)}, \dot{Y}_{\varphi(t)})$ (note that $\mathbf{Z}_t \neq (Z_t, \dot{Z}_t)$).

Proposition 1 is not restricted to the case where the base process Y is Gaussian (see the discussion in Section 3.5). If we assume that Y is a Gaussian process, then Y_t and \dot{Y}_t are independent Gaussian variables and we obtain the following expression for the joint stationary pdf of (Z_t, \dot{Z}_t) :

$$p_{Z,\dot{Z}}(z, \dot{z}) = \frac{1}{2\pi q_0} \frac{\dot{k}\left(\frac{\dot{z}}{f_1(z)}\right)}{f_1(z)^2 f_2\left(k\left(\frac{\dot{z}}{f_1(z)}\right)\right)} \exp\left(-\frac{z^2 + k\left(\frac{\dot{z}}{f_1(z)}\right)^2}{2}\right). \tag{10}$$

It is then possible to deduce the stationary marginal pdf of Z and the conditional pdf \dot{Z}_t given $Z_t = z$, which are given respectively by

$$p_Z(z) = \frac{1}{\sqrt{2\pi} q_0} \frac{\exp\left(-\frac{z^2}{2}\right)}{f_1(z)} \tag{11}$$

$$p_{\dot{Z}|Z=z}(\dot{z}) = \frac{1}{\sqrt{2\pi}} \frac{\dot{k}\left(\frac{\dot{z}}{f_1(z)}\right) k\left(\frac{\dot{z}}{f_1(z)}\right)}{\dot{z}} \exp\left(-\frac{k\left(\frac{\dot{z}}{f_1(z)}\right)^2}{2}\right). \tag{12}$$

These expressions are used in the next section to build nonparametric estimates of the functions f_1 and f_2 . Remark in particular that the marginal distribution of Z only depends on f_1 and that if $f_1(y) \equiv 1$, such as on the third example of Figure 2; then, the marginal distribution of Z is Gaussian.

3 | NONPARAMETRIC ESTIMATION

In this section, we first discuss the estimation of the marginal transformation h using up-crossings. Then, we discuss the estimation of the time-change functions f_1 and f_2 using the results obtained in the previous section. Finally, we show that the TTCG model has enough flexibility to describe important characteristics of observed trajectories related to the joint distribution of the process and its derivative in Section 3.5.

3.1 | Marginal transformation function h

A classical way to ‘‘Gaussianize’’ a non-Gaussian sequence consists in applying a marginal transformation and consider a TG process X defined as

$$h(X_t) = Y_t, \tag{13}$$

where Y is a stationary Gaussian process and h is an increasing real-valued function. It is a particular case of the TTCG model with $f_1 \equiv f_2 \equiv 1$.

Different parametric models, such as the power Box–Cox transformation and the Tuckey transformation (see Yan & Genton, 2017, and the references therein), have been proposed for the transformation h . In this work, we focus on nonparametric approaches. In this setting, h is generally chosen such that the transformed process Y has a Gaussian marginal distribution and is estimated using the probability integral transform and the empirical distribution function (see, e.g., Grigoriu, 1998).

Here, we propose an alternative approach that takes advantage of the properties of the up-crossing intensity as in Rychlik, Johannesson, and Leadbetter (1997) and Azaïs and Wschebor (2009) and has the advantage to generalize to the TTCG model. More precisely, let $v_H^+(X, u)$ denote the number of up-crossings of the level u of the process X on the time interval $[0, H]$ and

$$v^+(X, u) := \lim_{H \rightarrow \infty} \frac{1}{H} v_H^+(X, u)$$

denote the up-crossing intensity, where $:=$ holds for “is defined by.” $v^+(X, u)$ can be interpreted as the expected number of up-crossings of the level u per time unit. It is an important quantity for many applications for which only few analytical results are known (see Lindgren, 2012, for a review) with the noticeable exception of the so-called Rice formula

$$v^+(X, u) = p_X(u) E[|\dot{X}| \mid X = u]. \quad (14)$$

It shows that the expected number of $v^+(X, u)$ depends only on the marginal pdf p_X of X , and the so-called *derivative scale function* $E[|\dot{X}_t| \mid X_t = x]$. This formula holds true for stationary differentiable stationary processes, and thus, for the Gaussian base process Y for which we get

$$v^+(Y, u) = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right).$$

The proposition below shows that the Rice formula also holds true for a TTCG process X and that the up-crossing intensity mainly depends on the marginal link function h . An intuitive explanation is that the horizontal stretching associated with the time deformation does not modify the levels of the observed sequence, and thus, the up-crossing intensities of the processes Z and Y are equal up to a scaling factor.

Proposition 2. *If X is a TTCG process, then*

$$v^+(X, u) = v^+(Z, h(u)) = \frac{1}{q_0} v^+(Y, h(u)) = \frac{1}{2\pi q_0} \exp\left(-\frac{h(u)^2}{2}\right). \quad (15)$$

The proof of this proposition is given in the Appendix. Note that the two first equalities of Proposition 2 hold true in the general case where Y is not Gaussian.

According to (15), we get

$$-\frac{h(u)^2}{2} = \log(v^+(X, u)) - \log\left(\frac{1}{2\pi q_0}\right). \quad (16)$$

This relation shows that an estimate of h can be deduced from an estimate of the up-crossing intensity (see Section 3.4 for the details).

3.2 | Up–down asymmetries time-change function f_1

In this section, we assume that the process Z is observed. In practice, if only the process X is observed and an estimate \hat{h} of h is available; then, it is possible to compute an approximation $\hat{Z}_t = \hat{h}(X_t)$ of Z_t and replace Z by \hat{Z} in the estimates defined below.

According to (11), f_1 is related to the marginal pdf of Z_t , but using this relation requires a nonparametric estimate of $\frac{1}{p_Z}$. This approach was found to be numerically unstable. Instead, we propose to use (12) where f_1 appears as a scale factor of the conditional distribution of the derivative. More precisely, f_1 is proportional to the derivative scale function

$$f_1(z) = \sqrt{\frac{\pi}{2}} E[|\dot{Z}_t| \mid Z_t = z]. \tag{17}$$

This expression can be derived from (12) by applying the change of variable $u = k(\frac{z}{f_1(z)})$, which is valid because k is assumed to be a bijection from \mathbb{R} to \mathbb{R} . An estimate of f_1 can be deduced from any nonparametric estimate of the conditional expectation $E[|\dot{Z}_t| \mid Z_t = z]$ (see Section 3.4 for more details).

3.3 | Front-back asymmetries time-change function f_2

In this section, we assume that Z_t is observed and f_1 is known (in practice, these unknown quantities will be replaced by the estimates defined in the previous section). We propose to estimate f_2 using the relation (12), which shows that f_2 is related to the shape of the pdf of the derivative \dot{Z}_t .

Let us first notice that, according to 10, the scaled derivative $V_t = \frac{\dot{Z}_t}{f_1(Z_t)}$ is independent of Z_t and that the pdf p_V of V_t is

$$p_V(v) = \frac{\dot{k}(v)k(v)}{v\sqrt{2\pi}} \exp\left(-\frac{k(v)^2}{2}\right). \tag{18}$$

If f_2 is a constant function, then k is proportional to the identity function and V_t is a Gaussian variable. In the general case, this distribution may be non-Gaussian with, for example, skewed distributions associated to front-back asymmetries in the trajectories.

We deduce from (18) that

$$\dot{k}(v)k(v) \exp\left(-\frac{k(v)^2}{2}\right) = \sqrt{2\pi}vp_V(v),$$

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

$$1 - \exp(-k(v)^2/2) = \sqrt{2\pi} \int_0^v up_V(u)du,$$

and finally,

$$k(v) = \text{sign}(v) \sqrt{-2 \ln \left(1 - \sqrt{2\pi} \int_0^v zp_V(z)dz \right)}. \tag{19}$$

In this expression, the function k appears as a function of the expected value

$$\int_0^v zp_V(z)dz = E[V_t \mathbb{1}_{[0,v]}(V_t)],$$

which can be estimated from the data. Then, using the link between the functions k and f_2 , it is possible to deduce an estimate of f_2 . This is discussed more precisely in Section 3.4.

3.4 | Estimation procedure

Using the results of Sections 3.1, 3.2, and 3.3, we now propose a three-stage procedure for fitting the TTCG model. h is first estimated using up-crossings; then, f_1 is estimated using the derivative

scale function of the marginally transformed sequence and, finally, f_2 is estimated using the scaled derivative.

Estimation of h . The estimation of the marginal function h is based on up-crossings as discussed in Section 3.1. According to (16), h is directly related to the up-crossing intensity function up to the scaling factor $q_0 = E[\frac{1}{f(Y_0)}]$, which depends on the time-change function f . In order to estimate q_0 , we use that this quantity is related to the mode of the up-crossing intensity (i.e., the intensity of the most often crossed level) because

$$\sup_x \{ \log(v^+(X, x)) \} = \log \left(\frac{1}{2\pi q_0} \right). \quad (20)$$

Once q_0 is estimated, we are faced with a similar estimation problem as discussed in Rychlik et al. (1997) for TG models. It consists in taking a square root of (16) using the monotonicity of h . The obtained estimate was found to be noisy, and the LOESS method (locally weighted polynomial regression; see Cleveland, 1979) is used to smooth it. An isotonic regression method may also be applied to force the obtained estimate to be increasing, but this was not necessary for the practical cases considered in this paper.

After estimating h , it is possible to compute an approximation \hat{Z} of the process $Z_t = h(X_t)$. The final estimates of f_1 and f_2 are obtained by replacing Z by \hat{Z} in the expressions below.

Estimation of f_1 . f_1 is estimated by plugging in a nonparametric estimate of the conditional expectation $E[|\dot{Z}_t| \mid Z_t = z]$ in (17). Again, the LOESS approach is used to estimate this conditional expectation. The corresponding estimate is denoted \hat{f}_1 .

Estimation of f_2 . The estimation of f_2 is based on the four following steps.

1. Compute an approximation \hat{V}_t of V_t as $\hat{V}_t = \frac{Z_t}{\hat{f}_1(Z_t)}$.
2. Estimate the expectation $\int_0^v z p_V(z) dz$ using the corresponding sample mean, for example, for $v \geq 0$, by

$$\frac{1}{N} \sum_{t: 0 \leq \hat{V}_t \leq v} \hat{V}_t,$$

where N denotes the length of the observed sequence. Plug this estimate into (19) to deduce an estimate \hat{k} of k .

3. By definition, k is the inverse function of $z \mapsto z f_2(z)$, 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. The obtained estimate is noisy near the origin (the computation of $\frac{\dot{z}}{k(\dot{z})}$ leads to a division by values close to 0 at the origin), and we found it useful to smooth the estimate. The LOESS smoothing method is used again.

We performed a simulation study to assess the quality of the estimates. The model considered for simulation is inspired from the real data of Section 4.2. It is a TTCG model with a separable time-change function. The functions f_1 and f_2 are shown in Figure 3 and the other quantities,

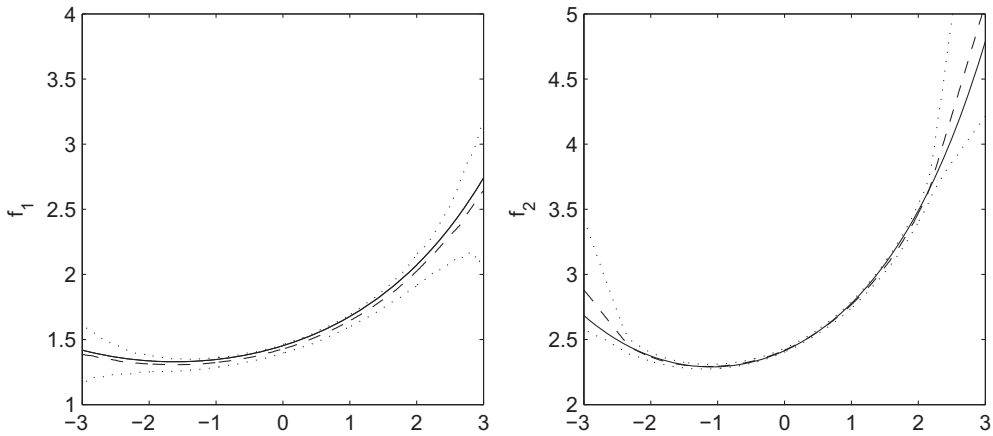


FIGURE 3 Fluctuation intervals for the estimates of f_1 (left) and f_2 (right). The full lines correspond to the true functions used for the simulation, the dashed line to the mean estimated values, and the dotted line to 95% fluctuation intervals. Results based on 1000 sequences simulated with the model fitted on the wave data

which define the model (function h and second-order structure of the base process Y) in Figure 6. 1000 sequences were generated with the model. Each sequence corresponds to a record of 40 min with a sampling time step of 0.070 s comparable to the amount of data available when considering wave observations (about 1000 waves with 34 points per wave). According to Figure 3, the proposed nonparametric estimates seem to be able to retrieve the shape of f_1 and f_2 with a small bias and reasonable fluctuation intervals.

3.5 | Model versatility

In this section, we discuss the versatility of the proposed approach and show that combining a marginal transformation and a time change permits to handle important properties of observed sequences.

Let us first focus on the effect of applying a time change and consider a general transformed time-changed model defined by (7) where Y is a continuously differentiable ergodic stationary process (not necessarily Gaussian as for the TTCCG model). Then, (9) gives the stationary joint distribution of (Z_t, \dot{Z}_t) . If we further assume that $f_2 \equiv 1$, then it can be easily deduced that

$$E[|\dot{Z}_t| \mid Z_t = z] = f_1(z)E[|\dot{Y}_t| \mid Y_t = z]. \tag{21}$$

The derivative scale function $E[|\dot{Z}_t| \mid Z_t = z]$ describes the dispersion of the derivative at the level z . (21) shows that the application of a time change permits to transform a process Z into a process Y such that the derivative scale function does not depend on the level y . Such a process will be referred to as *conditionally homoscedastic* hereafter by analogy to discrete-time models where homoscedasticity implies that the future time variability does not depend on the past (see, e.g., Fan & Yao, 2003). The derivative scale function is also somewhat related to vertical asymmetries. Indeed, if the processes Z and $-Z$ have the same distribution, then the derivative scale function

is even (i.e., $E[|\dot{Z}_t| \mid Z_t = z] = E[|\dot{Z}_t| \mid Z_t = -z]$) and this relation holds true for conditionally homoscedastic processes. On the opposite, if the derivative scale function is a decreasing function of z , as it is the case for the data set considered in Section 4.1, then the trajectories will typically have higher slopes at low levels and look asymmetric.

Time-change models have thus enough versatility to match any prescribed derivative scale function, and Proposition 2 shows that adding a marginal transformation permits to act on the up-crossing intensity. By combining both transformations as in the TTTCG model, we obtain a model that has enough flexibility to generate a process with given up-crossing intensity function and derivative scale function. Then, the Rice formula (14) implies that the marginal distributions also match (see proposition 3). This is a clear improvement compared to the more usual TG model, which can handle only one of these quantities.

This is stated more precisely in the proposition below. It shows that, under general conditions, the application of both a marginal transformation and a time change permits to transform a process X into a process Y , which is conditionally homoscedastic and has Gaussian margin. Reciprocally, TTTCG processes have enough versatility to simulate sequences with a prescribed derivative scale function and marginal distribution.

Proposition 3. *Let X denote a continuously differentiable ergodic stationary process. Assume that there exists a positive function f and an increasing function h such that*

$$v^+(X, u) = \frac{r_0}{2\pi} \exp\left(-\frac{h(u)^2}{2}\right) \quad (22)$$

$$f_1(z) = \sqrt{\frac{\pi}{2}} E[|\dot{Z}_t| \mid Z_t = z] \quad (23)$$

with $Z_t = h(X_t)$ and $r_0 = E[f(Z_0)] < \infty$.

Then, the time-changed process Y defined as

$$h(X_t) = Z_t, \quad Y_t = Z_{\psi(t)}, \quad \psi(t) = \int_0^t \frac{1}{f_1(Y_s)} ds$$

is such that (7) holds true, $v^+(Y, u) = \frac{1}{2\pi} \exp(-\frac{u^2}{2})$, $E[|\dot{Y}_t| \mid Y_t = y] = \sqrt{\frac{2}{\pi}}$ and Y has standard Gaussian margins.

Reciprocally, let Y^o be any continuously differentiable ergodic stationary Gaussian process with standard Gaussian margin such that $\text{var}(\dot{Y}_t^o) = 1$ and assume that $q_0 = E[\frac{1}{f_1(Y_0^o)}] < +\infty$. Then, the TTTCG process X^o defined as

$$h(X_t^o) = Z_t^o, \quad Z_t^o = Y_{\varphi^o(t)}^o, \quad \varphi^o(t) = \int_0^t f_1\left(\frac{Y_{\varphi^o(s)}^o}{\varphi^o(s)}\right) ds$$

is such that, for all x $v^+(X, x) = v^+(X^o, x)$, $E[|\dot{X}_t| \mid X_t = x] = E[|\dot{X}_t^o| \mid X_t = x]$ and $p_X(x) = p_{X^o}(x)$.

The proof of this proposition is given in the Appendix. This result is illustrated in Section 4.1 on a meteorological time series.

4 | NUMERICAL RESULTS

In this section, the proposed TTCG model is applied to real data. Firstly, an atmospheric pressure time series is considered. In this example, we focus on up–down asymmetries. We highlight that the combination of a marginal transformation h and a time-change function that depends on the level ($f_2 \equiv 1$) allows to describe the complex features of the time series much better than a marginal transformation alone. In the second application, we consider a time series of shallow water wave data, which exhibits both up–down and front–back asymmetries.

4.1 | Up–down asymmetries: Atmospheric pressure data

To illustrate the proposed models, we consider a particular time series X_t of hourly sea-level atmospheric pressure data measured at Guipavas (northwest of France, latitude $48^\circ 26' 36''$ N, longitude $4^\circ 24' 42''$ W). We focus on the winter months (January, February, and March) to reduce seasonal variations. One of the 17 winters available in the data set is shown in Figure 4a. It shows an alternation of periods with high-pressure (anticyclonic) and low-pressure (cyclonic) conditions with different characteristics. Cyclonic conditions tend to be less stable with successions of low-pressure systems coming from the Atlantic Ocean and reaching the west coast of France, whereas anticyclonic conditions generally corresponds to large and slowly evolving high-pressure systems. Visually, it leads to up–down asymmetries with different behaviors at low and high levels. It can also be seen on the empirical joint distribution of (X_t, \dot{X}_t) (see Figure 4b). In particular, the marginal distribution of X is asymmetric and the scatterplot shows heteroscedasticity with more variability in the derivative when the pressure is low.

Similar behavior can also be seen on other weather variables such as rainfall, temperature, or wind, but here, we have chosen to focus on pressure to simplify the interpretation of the model (pressure is more directly related to synoptic weather conditions). The model proposed below could be used as an input to a stochastic weather generator (see Ailliot, Allard, Monbet, & Naveau, 2015) for the other weather variables mentioned above. The most classical approach to describe the abovementioned alternation is based on the so-called weather-type models (see, e.g., Ailliot et al., 2015, for a recent review). In weather-type models, a discrete variable is introduced to describe explicitly the meteorological regimes (e.g., cyclonic/anticyclonic conditions), and the temporal evolution of the weather variables is modeled conditionally to this discrete variable. In this section, we explore a method based on transforming a Gaussian process.

Recall that, for a Gaussian process Y , the process Y_t and the derivative \dot{Y}_t at the same time are independent and Gaussian. This is clearly not true for the pressure data considered here. More generally, a stationary Gaussian process Y is symmetric about its mean μ (the processes $Y - \mu$ and $\mu - Y$ have the same distribution) and, thus, cannot describe up–down asymmetries.

Let us first focus on the TG model (TTCG model with $f_1 \equiv f_2 \equiv 1$). Figure 4e shows some properties of the sequence $\hat{Y}_t = \hat{h}(X_t)$ with \hat{h} estimated using (15). The empirical joint distribution shown in Figure 4f indicates that the marginal transformation was able to “Gaussianize” the marginal distribution but not the joint distribution of the process and its derivatives with still higher temporal variability at low levels. This is an indication that the transformed sequence is not Gaussian and that the fitted TG model is not able to reproduce some properties of the data such as up-crossing intensity. This is confirmed when looking at the transformed sequence shown in Figure 4e. The marginal transformation has reduced up–down asymmetries compared to the original sequence, but different behaviors at high and low levels are still visible.

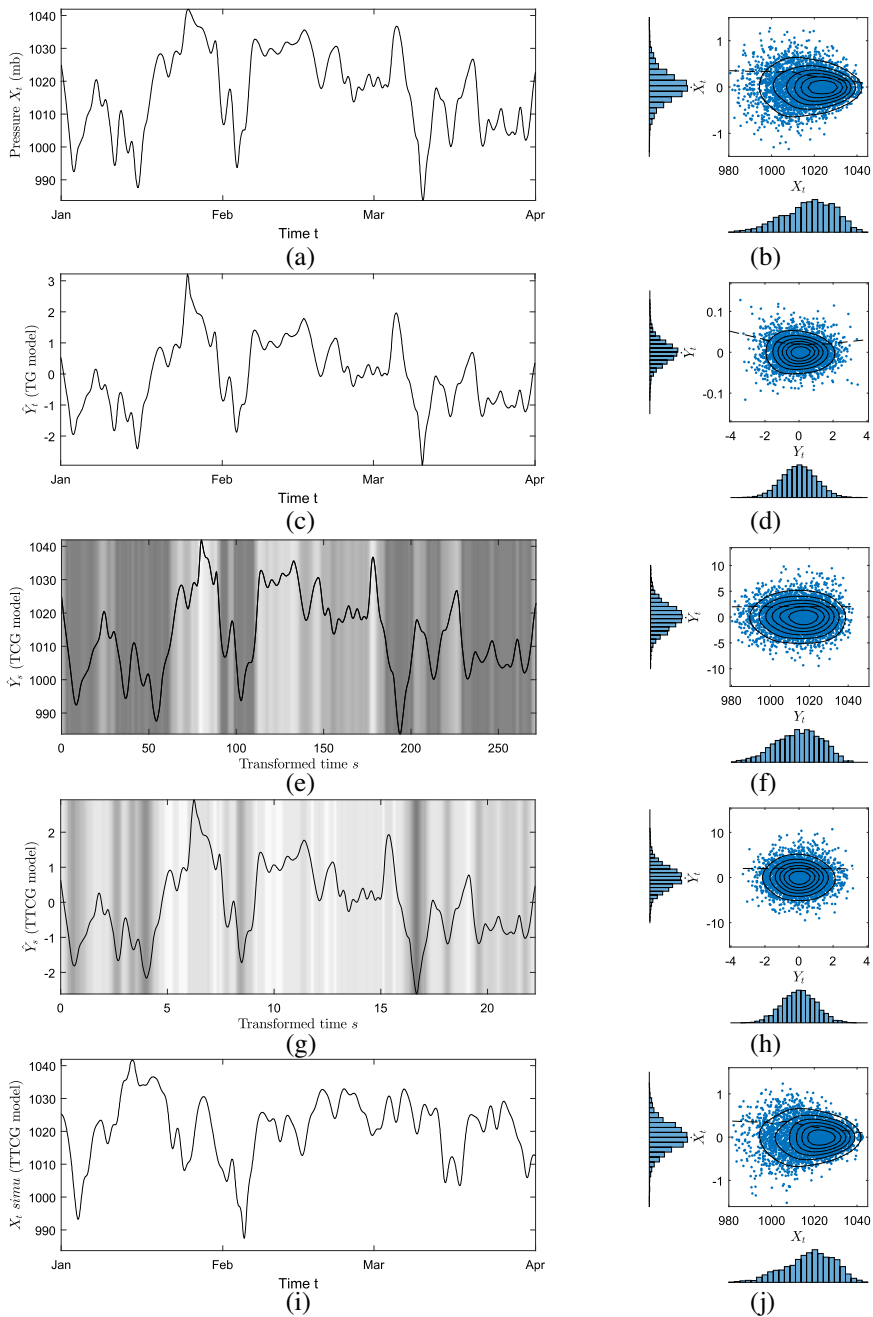


FIGURE 4 Left plots, from top to bottom: (a) sea-level pressure measured at Guipavas during the winter 1999–2000, (c) marginally transformed sequence $\hat{Y}_t = \hat{h}(X_t)$ for the fitted transformed Gaussian (TG) model, (e) estimated base process for the fitted time-changed Gaussian (TCG) model ($h(x) = x, f_2 \equiv 1$), (g) estimated base process for the fitted transformed time-changed Gaussian (TTCG) model with $f_2 \equiv 1$, (i) example of sequence simulated with the fitted TTCG model. The background colors in panels (c) and (g) correspond the value of $f(\hat{Y}_t)$ (speed of the clock). Right plots: corresponding joint distribution of the process (x -axis) and its derivative at the same time (y -axis) with nonparametric estimates of the joint probability density function (full lines) and of the conditional expectation (17) (dashed line) [Colour figure can be viewed at wileyonlinelibrary.com]

Let us now consider the TCG model (TTCG model with $h(x) = x$) without front-back asymmetries ($f_2 \equiv 1$). As discussed in Section 3.4, $f_1(x)$ is estimated using the derivative scale function $E[\dot{X}_t | X_t = x]$. The obtained estimate is superimposed on the scatterplot 4b. It is decreasing because the temporal variability is higher when the pressure is low. Once f_1 is estimated, it is possible to derive an estimate of the time change φ , using (7) with f_1 replaced by its estimate and, then, an approximation \hat{Y} of the base process Y . An example of transformed sequence \hat{Y} is shown in Figure 4c. Because the estimated time-change function is decreasing, the transformed sequence \hat{Y} has paths with shorter excursions at high levels and longer excursions at low levels and, thus, more symmetric paths compared to the original sequence. The value of $f_1(\hat{Y}_t)$ is superimposed on this plot to show the effect of the pressure on the speed of the clock. According to the scatterplot of the estimated transformed process \hat{Y} and its derivative (see Figure 4d), the transformed sequence seems to be approximatively conditionally homoscedastic as expected from the fitting procedure. However, the marginal distribution of \hat{Y} is asymmetric with a negative skewness. It is a clear indication that applying only a time change does not permit to “Gaussianize” the observed sequence.

Let us now focus on the TTCG model with $f_2 \equiv 1$. It is also possible to compute an approximation \hat{Y} of the base process Y , which should be close to Gaussian if the fitted model is valid. No obvious asymmetry can be seen on the sequence shown in Figure 4g. As expected from Proposition 3, the empirical distribution of \hat{Y} seems to be close to Gaussian and the derivative scale function close to constant (see Figure 4h). Other criteria may be considered, but as far as we know, there is no well-established formal test that could be used to check if the obtained sequence is Gaussian. 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. For this, the autocovariance function (ACF) of the base process Y is first estimated using the sample ACF of \hat{Y} . Then, a Gaussian sequence with this ACF is generated before applying the time-change and the marginal transformation. An example of simulated sequence is shown in Figure 4i. Visually, it seems to exhibit a behavior similar to the original pressure sequence (Figure 4a) with some up-down asymmetries. Various statistics of the simulated sequences were compared to the one of the original data. The results were compared with the ones obtained with the classical TG model. We found that both models are able to reproduce the marginal of the process. Compared to the TG model, the TTCG model clearly improves the description of the joint distribution of the process and its derivatives at the same time (see Figure 4j) and the up-crossing intensity (not shown). This is expected given the fitting procedure and Proposition 3.

We also found that the both models are able to reproduce the second-order structure of the data (not shown), but this is not ensured by the fitting procedure. A more elaborate approach for estimating the second-order structure of the latent Gaussian process Y would consist in choosing it in such a way the ACF of the transformed process X match the one of the observed process. For this, we would need an explicit expression that relates the second-order structure of the processes X and Y . Unfortunately, we could not derive such an expression for the time-changed models (see discussion of Theorem 1 in the Appendix). Note that such an expression is available for TG models when a specific marginal transformation such as the Tuckey transformation is used (see Xu & Genton, 2017) but not for general TG models.

4.2 | Up-down and front-back asymmetries: Sea wave data

Marine coastal systems are subject to loadings induced by ocean waves, and long time series of wave conditions are needed to study the performances of such systems. The systems are often

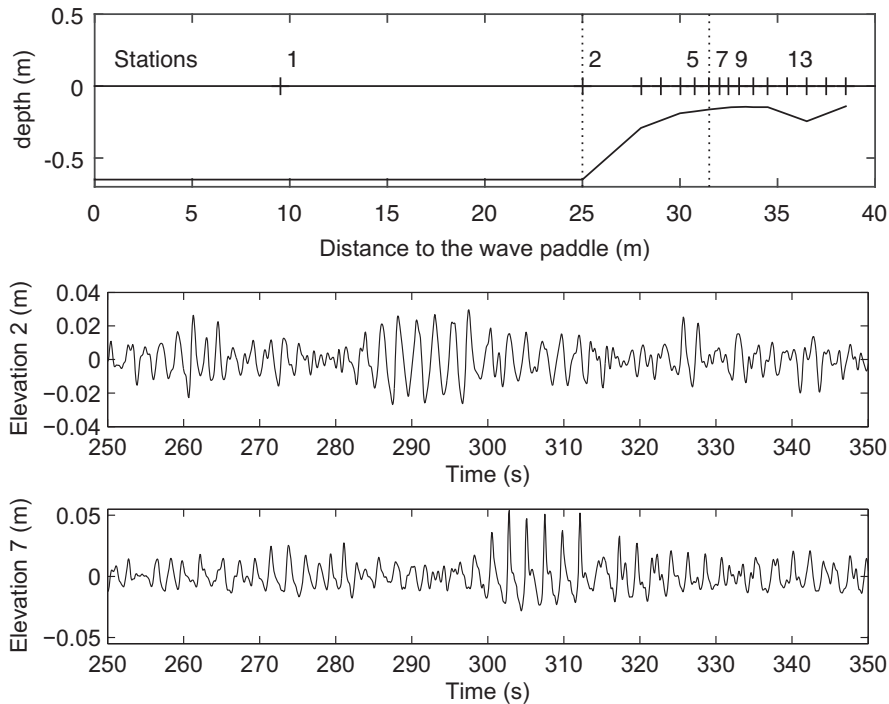


FIGURE 5 (A detailed description of the considered experiment can be found in Becq-Girard et al., 1999) layout of the experimental setup for the irregular flume experiment. The numbers indicate the position of the wave probes. Sequences of waves measured at Stations 2 (middle panel) and 7 (bottom panel)

located near the shore where the waves are asymmetric. It is known that it is very difficult to make in situ measurements of such wave conditions. A classical alternative consists in carrying experiments in a wave flume. A detailed description of the considered experiment can be found in Becq-Girard, Forget, and Benoit (1999). The bathymetric profile of the flume experiment is shown in Figure 5 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 (wave paddle) generates random-phase waves, which are expected to be close to a realization of a Gaussian process with a JONSWAP-type spectrum, which is one of the classical parametric models for wave spectra (see Holthuijsen, 2010). Then, the waves propagate to the right on a varying bathymetry, which induces strong nonlinearities. Examples of time series measured at Locations 2 and 7 are shown in Figure 5. At location 2, before the waves are modified by the varying bottom, we obtain a measurement with no obvious asymmetries as expected for a Gaussian sequence, whereas at Location 7, both up-down and front-back asymmetries are prominent. Free-surface elevation is recorded over a duration of 40 min with a sampling time step of 0.070 s. The mean wave period is around 2.4 s.

In this section, we investigate a stochastic model that could simulate quickly long or repeated time series of wave data with similar characteristics than the ones observed at Location 7. This generator could be used as an alternative to physically based wave models. In this section, X denotes the process that describes water elevation at Location 7. The TTCG modeling principle is summarized in Figure 6. Starting from an observed trajectory of X with asymmetric path (bottom left), we first apply a marginal transformation h whose estimate is shown on the

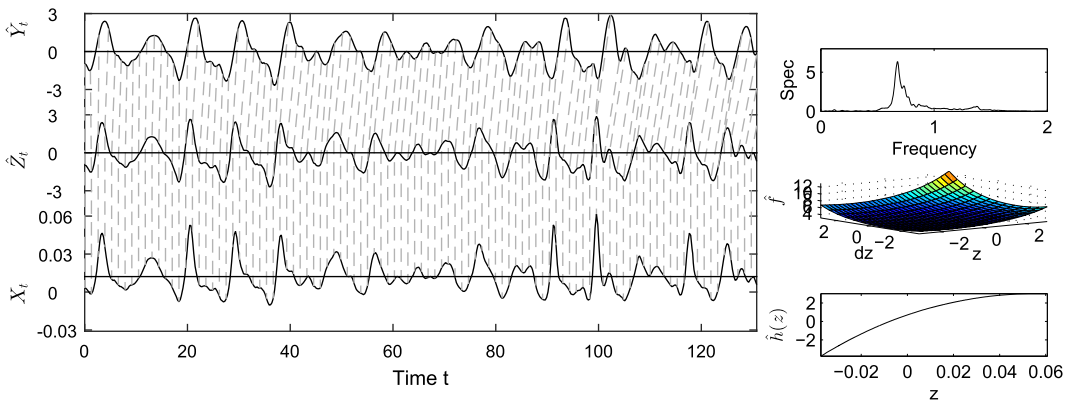


FIGURE 6 (Bottom left) sequence of measured sea-surface elevation (Location 7). (Middle left and top left) estimated processes \hat{Z}_t and \hat{Y}_t for the transformed time-changed Gaussian model. The vertical dotted lines materialize the time change introduced in the model. The right panel shows the estimated functions \hat{h} (bottom) and $\hat{f} = \hat{f}_1 \hat{f}_2$ (middle), and the empirical spectral density of Y (top) [Colour figure can be viewed at wileyonlinelibrary.com]

bottom right panel. After estimating h , it is possible to compute an approximation $\hat{Z}_t = \hat{h}(X_t)$ of X_t . The concavity of h leads to decreasing the steepness of the crests compared to the troughs (middle left). Then, a time deformation is applied. The estimated f_1 and f_2 functions are increasing (see their product $f_1 f_2$ on the middle right panel), and thus, the chronometer accelerates with the level of the process and its derivatives. It permits to act both on up-down and front-back asymmetries and obtain a transformed trajectory \hat{Y}_t (top left), which has an approximately symmetric path and can reasonably be modeled as a realization of a Gaussian process. No obvious asymmetry can be seen on the sequence \hat{Y}_t shown in Figure 7c and the empirical distribution of \hat{Y}_t , and its derivative seems to be close to Gaussian (see Figure 7d). We also compared the distributions of the four slopes defined in Figure 9, which have similar distributions (not shown) as expected for a Gaussian process that has symmetric paths.

Then, we checked that the fitted TTCG model is able to generate realistic wave sequences. Note that the fitted model can be simulated very quickly (about one-half second to generate 1000 waves on a basic laptop using MATLAB). An example of simulated sequence is shown in Figure 7e. It looks visually coherent with the observed sequence shown on the same figure.

It is not straightforward to define quantitative criteria, which measure the asymmetries of a trajectory (see Baxevani, Podgórski, & Wegener (2014, for a discussion). It is partly related to the joint distribution of (X_t, \dot{X}_t) . Indeed, up-down asymmetries, with crests that are more peaked and narrower compared to the troughs, lead to a heteroscedastic joint distribution for (X_t, \dot{X}_t) with higher derivatives at high levels (see Figure 7b). Front-back asymmetries, with steeper fronts compared to the backs, lead to an asymmetric distribution for \dot{X}_t with a positive skewness (the empirical skewness is equal to 0.53). According to Figure 7f, the fitted TTCG model seems to be able to reproduce the shape of this bivariate distribution and clearly outperforms the TG (see Figure 7h). The quantile-quantile plots shown in Figure 8 confirm that both models are able to reproduce the marginal distribution of X but the TG model cannot reproduce the marginal distribution of \dot{X}_t . This can be problematic for applications sensitive to the wave steepness.

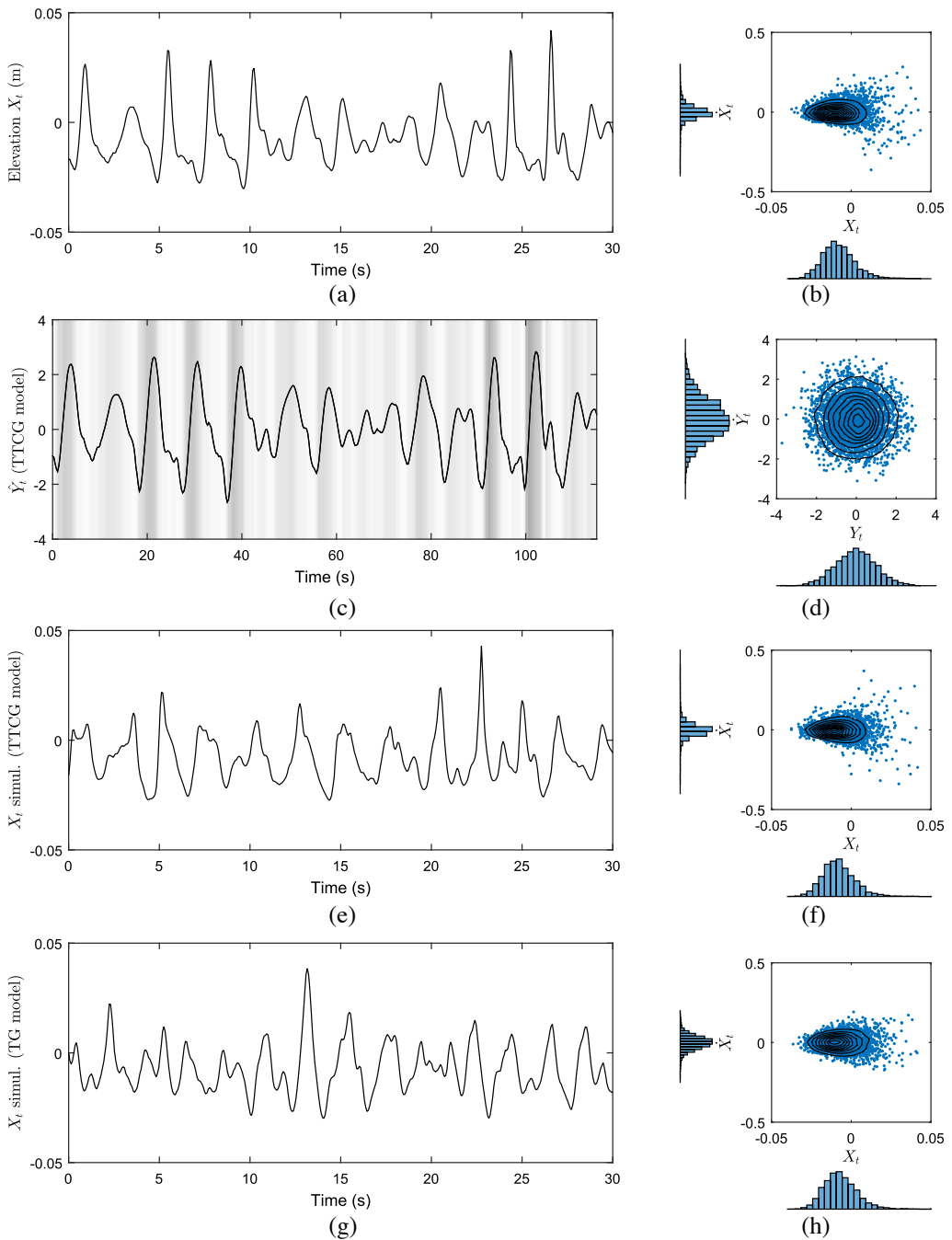


FIGURE 7 Left plots, from top to bottom: (a) elevation measured at Location 7, (c) estimated base process for the transformed time-changed Gaussian (TTCG) model, (e) example of sequence simulated with the fitted TTCG model, (g) example of sequence simulated with the fitted transformed Gaussian (TG) model. The background color on the second plot corresponds to the value of $f_1(Y_t)f_2(\dot{Y}_t)$ (speed of the clock). Right plots: corresponding joint distribution of the process (x -axis) and its derivative at the same time (y -axis) with nonparametric estimates of the joint probability density function (full lines) [Colour figure can be viewed at wileyonlinelibrary.com]

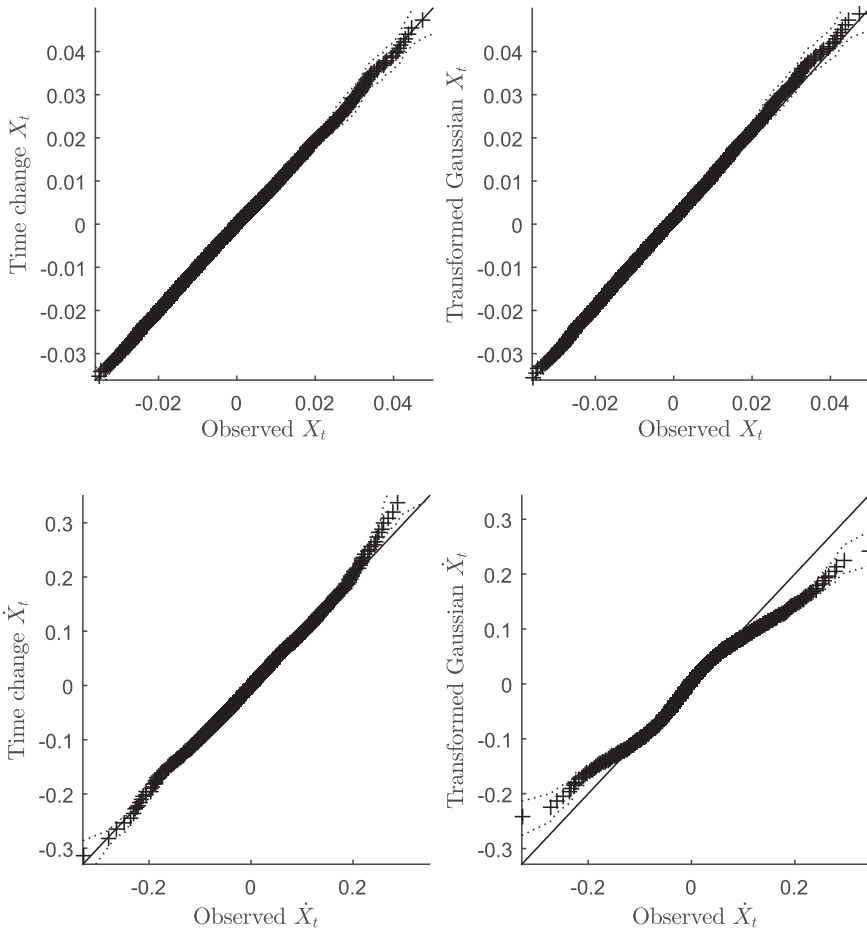


FIGURE 8 Left: quantile–quantile plot of the data versus the fitted TTCG model. Right: quantile–quantile plot of the data versus the fitted TG model. The top plots correspond to the process and the bottom plots to the derivative. Results obtained at Location 7. The dotted lines are 95% prediction intervals computed using simulations

Note that Gaussian processes, and thus, TG processes are time reversible, and Figure 7g confirms visually that the TG model cannot reproduce the observed front–back asymmetries. A time-reversible process X is such that the conditional distribution of the slope \dot{X}_t given $X_t = x$ is symmetric for all level x , and thus, the distribution of \dot{X}_t is symmetric. It can be shown that the conditional distribution of \dot{X}_t is also symmetric for the TTCG model in the particular case when f_2 is constant. A nonconstant f_2 function permits to create asymmetries in this distribution.

Another usual characterization of asymmetries in ocean waves is based on the comparison of the distributions of the various slopes (trough front, trough back, crest front, and crest back) shown on the top panel of Figure 9, where individual waves are extracted using zeros-crossings. The differences in these four distributions (see bottom panel of Figure 9) are linked to waves asymmetries. For example, it shows that the fronts of the crests are the steepest, whereas the fronts of the troughs are the flattest. Figure 9 shows that the fitted TTCG model is able to reproduce

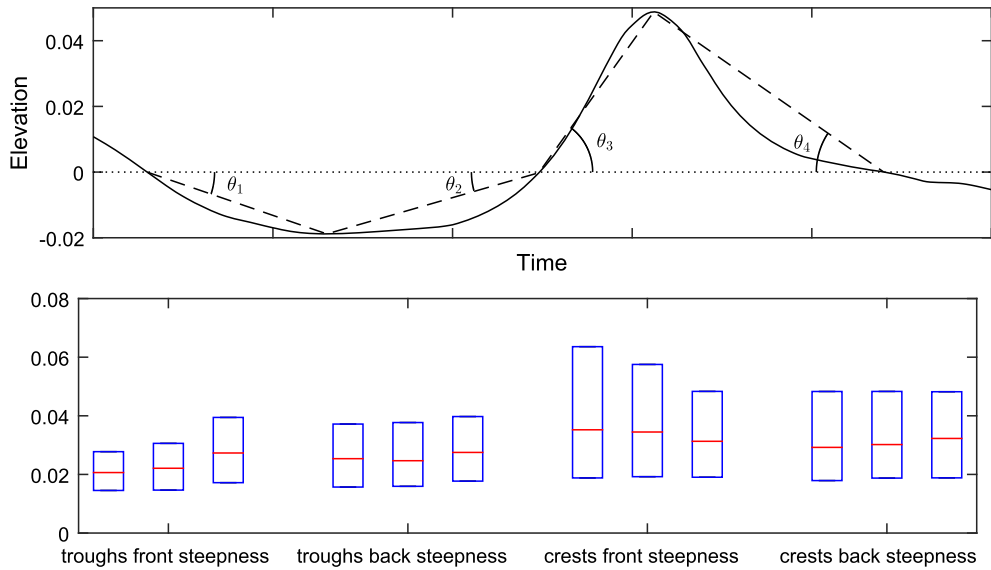


FIGURE 9 Definition of the trough front steepness (θ_1), trough back steepness (θ_2), crest front steepness (θ_3) and crest back steepness (θ_4) (top) and corresponding box plots (bottom). For each criterion, the first box plots correspond to the observation (Location 7), the second ones to the fitted TTGC model and the third ones to the fitted TG process. Results obtained using simulations [Colour figure can be viewed at wileyonlinelibrary.com]

these distributions. Note that the fitted TG model also reproduces some up–down asymmetries (the distributions for the crests and the troughs are different) but no front–back asymmetries (the distributions for the fronts and the backs are the same).

5 | CONCLUSIONS AND PERSPECTIVES

This paper proposes an original framework for modeling asymmetric sequences where a time-change is combined with a marginal transformation. The ergodicity of the proposed time-changed model is established under general conditions. This result is used to build non-parametric estimation procedures. The marginal transformation is estimated beforehand using up-crossings. Then, the estimation of the time-change function is based on the joint distribution of the process and its derivative. The methodology is illustrated on two environmental time series. We show that the fitted models are able to reproduce important characteristics of the data sets such as marginal distributions, conditional heteroscedasticity, up-crossing intensity, and up–down and front–back asymmetries.

The proposed methodology only applies to univariate temporal sequences. Generalizations to multivariate processes and spatial or spatiotemporal processes will be the topic of future work. On the theoretical side, important properties of the process, such as its second-order properties, are still unknown and need to be further investigated. Finally, we plan to develop parametric approaches. They may be useful for applications such as wave modeling where the amount and quality of observations are not always sufficient to use the nonparametric fitting procedures discussed in this paper.

ACKNOWLEDGEMENT

The authors are grateful to the anonymous reviewers for their constructive comments and suggestions, which led to significant improvements in this paper.

ORCID

Pierre Ailliot  <https://orcid.org/0000-0002-7883-6970>

REFERENCES

- Ailliot, P., Allard, D., Monbet, V., & Naveau, P. (2015). Stochastic weather generators: An overview of weather type models. *Journal de la Société Française de Statistique*, 156, 101–113.
- Anderes, E. B., & Stein, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. *The Annals of Statistics*, 719–741.
- Azaïs, J.-M., & Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. Hoboken, NJ: John Wiley & Sons.
- Baxevani, A., Podgórski, K., & Wegener, J. (2014). Sample path asymmetries in non-Gaussian random processes. *Scandinavian Journal of Statistics*, 41, 1102–1123.
- Becq-Girard, F., Forget, P., & Benoit, M. (1999). Non-linear propagation of unidirectional wave fields over varying topography. *Coastal Engineering*, 38, 91–113.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, 135–155.
- Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829–836.
- Fan, J., & Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods*. New York, NY: Springer.
- Grigoriu, M. (1998). Simulation of stationary non-Gaussian translation processes. *Journal of Engineering Mechanics*, 124, 121–126.
- Holthuijsen, L. H. (2010). *Waves in oceanic and coastal waters*. Cambridge, UK: Cambridge University Press.
- Lindgren, G. (2012). *Stationary stochastic processes: Theory and applications*. Boca Raton, FL: CRC Press.
- Rychlik, I., Johannesson, P., & Leadbetter, M. (1997). Modelling and statistical analysis of ocean-wave data using transformed Gaussian processes. *Marine Structures*, 10, 13–47.
- Sampson, P., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87, 108–119.
- Veraart, A. E. D., & Winkel, M. (2010). Time change. In: *Encyclopedia of quantitative finance*. New York, NY: Wiley.
- Wang, K., & Gasser, T. (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25, 1251–1276.
- Xu, G., & Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 1236–1249.
- Yan, Y., & Genton, M. G. (2017). Non-Gaussian autoregressive processes with Tukey g-and-h transformations. arXiv preprint arXiv:1711.07516.

How to cite this article: Ailliot P, Delyon B, Monbet V, Prevosto M. Time-change models for asymmetric processes. *Scand J Statist*. 2019;1–26. <https://doi.org/10.1111/sjos.12382>

APPENDIX A

This appendix provides formal statements and proofs of the various theoretical results used in this paper.

A.1 | Ergodicity of the time-changed process

Theorem 1. Let $(\mathbf{Y}_t)_{t \geq 0}$ be a stationary process with values on \mathbb{R}^d and continuous sample paths and let $f(\cdot)$ be a measurable real-valued positive function defined on \mathbb{R}^d such that $E[\frac{1}{f(\mathbf{Y}_0)}] < +\infty$. Denote $q_0 = E[\frac{1}{f(\mathbf{Y}_0)}]$ and define the process $(\mathbf{Z}_t)_{t \geq 0}$ by (1)–(2). Then, \mathbf{Z} is stationary for the measure P' associated to the conditional expectation E' defined as

$$E'[g(\mathbf{Z}_\cdot)] = q_0^{-1} E \left[\frac{g(\mathbf{Z}_\cdot)}{f(\mathbf{Z}_0)} \right]. \quad (\text{A1})$$

If \mathbf{Y} is ergodic, then \mathbf{Z} is ergodic under P' . In particular, for any measurable real-valued function g defined on \mathbb{R}^d such that $E[\frac{|g(\mathbf{Y}_0)|}{f(\mathbf{Y}_0)}] < \infty$,

$$E'[g(\mathbf{Z}_t)] = \lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(\mathbf{Z}_s) ds = q_0^{-1} E \left[\frac{g(\mathbf{Y}_0)}{f(\mathbf{Y}_0)} \right] \text{ a.e.} \quad (\text{A2})$$

Proof of Theorem 1. Let us first introduce some notations. \mathbf{Z} and \mathbf{Y} , as well as \mathbf{Z}_\cdot and \mathbf{Y}_\cdot , will denote the processes $(\mathbf{Z}_t)_{t \geq 0}$ $(\mathbf{Y}_t)_{t \geq 0}$. The stationarity of \mathbf{Y} means that, for any measurable function g , $g(\mathbf{Y}_\cdot)$ has the same distribution as $g(\mathbf{Y}_{\cdot+s})$ for any $s > 0$. The ergodicity means that, for any bounded measurable function g ,

$$\lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(\mathbf{Y}_{\cdot+s}) ds = E[g(\mathbf{Y}_\cdot)] \text{ a.e.}$$

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

$$g(\mathbf{Y}) = g_0(\mathbf{Y}_{t_1}, \dots, \mathbf{Y}_{t_n}), \quad (\text{A3})$$

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

Let us denote by ψ the reciprocal function of φ . Remark that Equation (2) means that $\dot{\varphi}(t) = f(\mathbf{Y}_{\varphi(t)})$, that is, setting $s = \varphi(t)$, $\dot{\psi}(s)^{-1} = f(\mathbf{Y}_s)$. Finally,

$$\mathbf{Z}_{\psi(t)} = \mathbf{Y}_t, \quad (\text{A4})$$

$$\psi(t) = \int_0^t \frac{1}{f(\mathbf{Y}_s)} ds. \quad (\text{A5})$$

This alternative formulation provides a more direct and tractable definition of the model.

Let us denote the shifted processes $\mathbf{Y}^\tau = \mathbf{Y}_{\cdot+\tau}$ and $(\psi^\tau, \mathbf{Z}^\tau)$ denote the processes obtained by replacing \mathbf{Y} with \mathbf{Y}^τ in (A4–A5). We get

$$\psi^\tau(t) = \int_0^t \frac{1}{f(\mathbf{Y}_s^\tau)} ds = \int_0^t \frac{1}{f(\mathbf{Y}_{s+\tau})} ds = \psi(\tau + t) - \psi(\tau).$$

In addition, (A4) becomes $\mathbf{Z}_{\psi^\tau(t)}^\tau = \mathbf{Y}_t^\tau$; hence,

$$\mathbf{Z}_{\psi^\tau(\tau+t)-\psi(\tau)}^\tau = \mathbf{Z}_{\psi^\tau(t)}^\tau = \mathbf{Y}_t^\tau = \mathbf{Y}(t + \tau) = \mathbf{Z}_{\psi(t+\tau)}.$$

Thus, $\mathbf{Z}^\tau = \mathbf{Z}_{\cdot+\psi(\tau)}$, which means that the effect of a shift on \mathbf{Y} is a random shift on \mathbf{Z} .

Now, by the stationarity of \mathbf{Y} , we have for any $\tau > 0$ and any function g of the form (A3)

$$q_0 E' [g(\mathbf{Z}_.)] = E \left[\frac{g(\mathbf{Z}_.)}{f(\mathbf{Z}_0)} \right] = E \left[\frac{g(\mathbf{Z}_t^.)}{f(\mathbf{Z}_0^t)} \right] = E \left[\frac{g(\mathbf{Z}_{.+\psi(\tau)})}{f(\mathbf{Z}_{\psi(\tau)})} \right].$$

Because this is true for any $t \geq 0$, for any $H > 0$,

$$\begin{aligned} q_0 E' [g(\mathbf{Z}_.)] &= \frac{1}{H} \int_0^H E \left[\frac{g(\mathbf{Z}_{.+\psi(\tau)})}{f(\mathbf{Z}_{\psi(\tau)})} \right] d\tau \\ &= E \left[\frac{1}{H} \int_0^{\psi(H)} \frac{g(\mathbf{Z}_{.+s})}{f(\mathbf{Z}_s)} \dot{\varphi}(s) ds \right], \quad (\tau = \varphi(s)) \\ &= E \left[\frac{1}{H} \int_0^{\psi(H)} g(\mathbf{Z}_{.+s}) ds \right] \end{aligned}$$

and, for any $t > 0$,

$$\begin{aligned} q_0 (E' [g(\mathbf{Z}_.)] - E' [g(\mathbf{Z}_. + t)]) &= E \left[\frac{1}{H} \int_0^{\psi(H)} g(\mathbf{Z}_{.+s}) ds \right] - E \left[\frac{1}{H} \int_0^{\psi(H)} g(\mathbf{Z}_{.+t+s}) ds \right] \\ &= E \left[\frac{1}{H} \int_0^t g(\mathbf{Z}_{.+s}) ds \right] - E \left[\frac{1}{H} \int_{\psi(H)}^{\psi(H)+t} g(\mathbf{Z}_{.+s}) ds \right] \\ q_0 |E' [g(\mathbf{Z}_.)] - E' [g(\mathbf{Z}_. + t)]| &\leq \frac{2t}{H} \|g\|_\infty, \end{aligned}$$

where $\|g\|_\infty < +\infty$ denotes the sup norm of the bounded function g . Because 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

$$\begin{aligned} \frac{1}{H} \int_0^H g(\mathbf{Z}_{.+s}) ds &= \frac{1}{H} \int_0^{\varphi(H)} \frac{g(\mathbf{Z}_{.+\psi(\tau)})}{f(\mathbf{Z}_{\psi(\tau)})} d\tau \quad (s = \psi(\tau)) \\ &= \frac{\varphi(H)}{H} \frac{1}{\varphi(H)} \int_0^{\varphi(H)} \frac{g(\mathbf{Z}_t^.)}{f(\mathbf{Z}_0^t)} d\tau. \end{aligned} \tag{A6}$$

Note that

$$\lim_{H \rightarrow \infty} \frac{\psi(H)}{H} = \lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H \frac{du}{f(\mathbf{Y}_u)} = q_0 \text{ a.e.}$$

by the ergodicity of \mathbf{Y} , and thus, $\lim_{H \rightarrow \infty} \frac{\varphi(H)}{H} = q_0^{-1}$ a.e. Using again the ergodicity of \mathbf{Y} and that $\lim_{H \rightarrow \infty} \varphi(H) = +\infty$ a.e., we deduce that the second term in the right-hand side term of (A6) converges to a deterministic limit that is its expected value.

(A2) is a particular case of the general convergence result proved above with $g(Y) = g_0(Y_t)$ and remarking that $Y_0 = Z_0$. □

Some remarks on the general time-change model.

1. The limit E' is a measure equivalent to E and this equivalence explains why the middle limit in (A2) holds w.p.1 under E' and E . E' is essentially the limit distribution of $(\mathbf{Z}_{t+s})_{s \geq 0}$ when t tends to infinity. One has indeed from (A1) and from the stationarity of \mathbf{Z} under E' that, for any $t \geq 0$ and for any bounded function h ,

$$E' [h(\mathbf{Z}_.)] = E' [h(\mathbf{Z}_{t+.})] = q_0^{-1} E \left[\frac{h(\mathbf{Z}_{t+.})}{f(\mathbf{Z}_0)} \right]$$

(we have applied (A1) to $g(\mathbf{Z}_t) = h(\mathbf{Z}_{t+})$). If for any bounded function h , the correlation between $h(\mathbf{Z}_{t+})$ and $1/f(\mathbf{Z}_0)$ tends to zero as $t \rightarrow \infty$ (this is a reasonable mixing assumption), we get that

$$E'[h(\mathbf{Z}_t)] = \lim_{t \rightarrow \infty} q_0^{-1} E \left[\frac{h(\mathbf{Z}_{t+})}{f(\mathbf{Z}_0)} \right] = \lim_{t \rightarrow \infty} E[h(\mathbf{Z}_{t+})].$$

In other terms, \mathbf{Z} is asymptotically stationary, and unsurprisingly, the empirical means capture this limit stationary measure.

2. Equation (A1) implies actually the more general statement

$$E'[g(\mathbf{Z}_t)] = \lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(\mathbf{Z}_{t+s}) ds = q_0^{-1} E \left[\frac{g(\mathbf{Z}_t)}{f(\mathbf{Z}_0)} \right] \quad \text{a.e.,}$$

which appears to be uneasy to utilize unless $g(x)$ is a function of x_0 . This explains why it is difficult to work with joint distributions at lagged times (e.g., autocovariance functions) in this model.

3. **Unicity of time change.** Notice that $1/q_0 f$ is the density of E' w.r.t. the distribution of \mathbf{Z}_0 (use (A2) and notice that $\mathbf{Y}_0 = \mathbf{Z}_0$). The function f is thus uniquely determined, up to a scalar factor, by the distribution of the observation process \mathbf{Z} . In practice, if only one trajectory is observed, the distribution of \mathbf{Z}_0 is unknown. f is characterized by the fact that the process \mathbf{Y} given by (A4) is stationary, but finding the function f that would make the \mathbf{Y} process stationary is impossible because one cannot measure the stationarity of \mathbf{Y} based on a single trajectory. This makes the estimation of f impossible. However, under additional assumptions on \mathbf{Y} , f may actually be a function of the stationary measure E' , which means that it can be estimated by using statistics calculated over only one trajectory. In Section 3.2, we show that if \mathbf{Y} is a univariate differentiable Gaussian process, then f can be consistently estimated from empirical means $\frac{1}{H} \int_0^H g(\mathbf{Z}_s) ds$. The necessity of the differentiability assumption is an open problem.

A.2 | Proof of Proposition 1, Proposition 2, and Proposition 3

Proof of Proposition 1. Set

$$f(u, v) = f_1(u) f_2(v).$$

We have

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

Now, by applying Theorem 1 with $\mathbf{Y}_t = (Y_t, \dot{Y}_t)$, we get that $\mathbf{Z}_t = (Y_{\varphi(t)}, \dot{Y}_{\varphi(t)})$ is a stationary ergodic process such that

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

with $q_0 = E[\frac{1}{f(Y_0, \dot{Y}_0)}]$. Hence,

$$\begin{aligned} \lim_{H \rightarrow \infty} \frac{1}{H} \int_0^H g(Z_s, \dot{Z}_s) ds &= \frac{1}{q_0} \iint \frac{g(u_1, u_2 f(u_1, u_2))}{f(u_1, u_2)} p_{Y, \dot{Y}}(u_1, u_2) du_1 du_2 \quad \text{a.e.} \\ &= \frac{1}{q_0} \int \left(\int \frac{g(u_1, u_2 f_1(u_1) f_2(u_2))}{f_1(u_1) f_2(u_2)} p_{Y, \dot{Y}}(u_1, u_2) du_2 \right) du_1. \end{aligned}$$

The change of variable $z = u_2 f_1(u_1) f_2(u_2)$ (or equivalently $u_2 = k(z/f_1(u_1))$) in the inner integral gives (9), proving Proposition 1. □

Proof of Proposition 2. Because h is increasing and the time change does not modify the observed level we have $v_H^+(X, u) = v_H^+(Z, h(u)) = v_{\varphi(H)}^+(Y, h(u))$. We deduce that

$$\begin{aligned} \lim_{H \rightarrow \infty} \frac{1}{H} v_H^+(X, u) &= \lim_{H \rightarrow \infty} \frac{1}{H} v_H^+(Z, h(u)) \\ &= \lim_{H \rightarrow \infty} \frac{\varphi(H)}{H} \frac{1}{\varphi(H)} v_{\varphi(H)}^+(Y, h(u)), \end{aligned}$$

where $\lim_{H \rightarrow \infty} \frac{\varphi(H)}{H} = q_0^{-1}$ (see proof of Theorem 1). The result can then be deduced from the following classical result, which holds true for differentiable Gaussian processes (Rice formula; see, e.g., Lindgren, 2012):

$$\lim_{H \rightarrow \infty} \frac{1}{H} v_H^+(Y, u) = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right),$$

where we have used that $E[Y_t] = 0$ and $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$. □

Proof of Proposition 3. Using similar arguments than in the proof of Proposition 2, we get

$$v^+(Y, h(u)) = \frac{1}{r_0} v^+(X, u)$$

and, then, using (22), that $v^+(Y, u) = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right)$. Then, remark that Y is a time-change process with base process Z and time-change function $\frac{1}{f_1}$. Using (21), we get that

$$E[|\dot{Y}_t| \mid Y_t = y] = \frac{1}{f_1(z)} E[|\dot{Z}_t| \mid Z_t = y]$$

and, then, using (23), that $E[|\dot{Y}_t| \mid Y_t = y] = \sqrt{\frac{2}{\pi}}$. Finally, using (14), we obtain that $p_X(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$.

Applying (15), we get $v^+(X, u) = \frac{1}{2\pi q_0} \exp(-\frac{h(u)^2}{2})$. Now, using (9), we obtain that $p_Y = \frac{1}{r_0} f_1 p_Z$ and, thus, $E[\frac{1}{f_1(Y_0)}] = \frac{1}{r_0}$. Because the processes Y and Y^o have both standard Gaussian margins, we deduce that $q_0 = \frac{1}{r_0}$ and, using (22), that $v^+(X, x) = v^+(X^o, x)$. We can also deduce that Z and Z^o and, then, X and X^o have the same marginal distributions. □

A.3 | The identifiability constraint $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$

We consider the TTCG model and show that a scaling of (h, f_1, f_2) permits to assume, without restriction, that $\text{var}(Y_t) = \text{var}(\dot{Y}_t) = 1$. Set

$$\begin{aligned} \sigma &= \text{var}(Y_t)^{1/2}, & \alpha &= \frac{1}{\sigma} \text{var}(\dot{Y}_t)^{1/2} \\ \tilde{Z}_t &= \frac{1}{\sigma} Z_t, & \tilde{Y}_t &= \frac{1}{\sigma} Y_t / \alpha, & \tilde{\varphi}(t) &= \alpha \varphi(t) \\ \tilde{h}(x) &= \frac{1}{\sigma} h(x), & \tilde{f}_1(x) &= \alpha f_1(\sigma x), & \tilde{f}_2(x) &= f_2(\alpha \sigma x). \end{aligned}$$

Then, one has clearly $\tilde{h}(X_t) = \tilde{Z}_t$, $\tilde{Z}_t = \tilde{Y}_{\tilde{\varphi}(t)}$ and

$$\begin{aligned} \frac{d}{dt} \tilde{\varphi}(t) &= \alpha f_1(Y_{\varphi(t)}) f_2(\dot{Y}_{\varphi(t)}) \\ &= \alpha f_1(\sigma \tilde{Y}_{\alpha\varphi(t)}) f_2(\sigma \alpha \dot{\tilde{Y}}_{\alpha\varphi(t)}) \quad \left(\dot{\tilde{Y}}_t = \frac{1}{\sigma \alpha} \dot{Y}_{t/\alpha} \right) \\ &= \tilde{f}_1(\tilde{Y}_{\tilde{\varphi}(t)}) f_2'(\dot{\tilde{Y}}_{\tilde{\varphi}(t)}). \end{aligned}$$

In addition, $\text{var}(\tilde{Y}_t) = \text{var}(\dot{\tilde{Y}}_t) = 1$.