



Finite dimensional models for wind diffusion process

Hui Xu ^{a,*}, Mircea D. Grigoriu ^{a,b}

^a Center for Applied Mathematics, Cornell University, Ithaca, NY, USA

^b Department of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA

ARTICLE INFO

MSC:

60

62

Keywords:

Diffusion process

Extremes

Finite dimensional model

Karhunen–Loève (KL) representation

Monte Carlo simulation algorithms

ABSTRACT

Diffusion processes with linear drift and translations of Gaussian/non-Gaussian diffusion processes are fitted to wind pressure time series recorded at the University of Florida boundary layer wind tunnel facility (UFBWLT). The processes match exactly and approximately the marginal distributions and the correlation functions of these records. It is shown that these simple processes and their finite dimensional models characterize accurately the extremes of the wind record under consideration provided their correlation functions minimize objective functions which account for extremes.

1. Introduction

The information on physical quantities, such as wind pressures on buildings, sea wave heights and material microstructure features, consists of values recorded at finite sets of times and/or locations. For example, suppose that n_s independent samples of a real-valued stationary process $X(t)$ are recorded in a bounded time interval $[0, \tau]$ at a time step $\Delta\tau = \tau/n$, so that the data set $\{x_i(j\Delta\tau), j = 0, 1, \dots, n\}$, $i = 1, \dots, n_s$, provides n_s independent samples of the $(n+1)$ -dimensional random vector $V = (X(t_0), X(t_1), \dots, X(t_n))$. Generally, the data set is sufficient to estimate the mean vector, the correlation matrix and the marginal distribution of V but is insufficient to characterize the extreme random variable $\max_{0 \leq j \leq n} |V_j|$. Hence, even if the random variables $\sup_{t \in [0, \tau]} |X(t)|$ and $\max_{0 \leq j \leq n} |V_j|$ have similar distributions, the distribution of $\sup_{t \in [0, \tau]} |X(t)|$ cannot be approximated solely from data. The information provided by the data set has to be augmented to characterize extremes of $X(t)$.

There are at least two approaches which can be employed to extend the available information on $X(t)$ and use it to characterize the extreme random variable $\sup_{t \in [0, \tau]} |X(t)|$. A common step of these approaches is the estimation of the mean vector and correlation matrix of the random vector V , which can be viewed as approximations of the mean and correlation functions of $X(t)$ provided that the time step $\Delta\tau$ is sufficiently small. The first approach constructs reduced order models $\{V_d\}$, $d = 1, 2, \dots$, of V , see [1]. The models $\{V_d\}$ are truncated Karhunen–Loève (KL) representations of V whose random entries are described by polynomial chaoses (PCs) and translation polynomial chaoses (PCTs) with coefficients which minimize objective functions quantifying the

discrepancy between data and PC/PCT histograms and other properties. It was shown that extremes of V can be approximated by extremes of V_d with PCT coefficients provided that the stochastic dimension d is sufficiently large. We note that V is the target random element in this approach i.e., the time series derived from $X(t)$, rather than the process $X(t)$ itself.

The second approach, which is presented in this study, assumes that the available records are samples of a stationary and even ergodic process $X(t)$ with unknown probability law and that the data set can only provide reliable estimates of the mean, correlation and marginal distributions functions of this process. Accordingly, there exists an infinite family of random processes which match these statistics. Of this family, we consider exponentially correlated diffusion processes, i.e., diffusion processes with linear drift and multiplicative noise, and translations of exponentially correlated Gaussian and non-Gaussian diffusion processes. These processes are required to match exactly the record marginal distribution and minimize the discrepancy between the target and model correlation functions and/or extremes.

For each process $X(t)$ under consideration, we construct finite dimensional (FD) models $\{X_d(t)\}$, $d = 1, 2, \dots$, for $t \in [0, \tau]$, i.e., deterministic functions of time and d random variables $\{Z_k\}$, $k = 1, \dots, d$. The samples of the FD processes $\{X_d(t)\}$ are elements of the linear space spanned by the top eigenfunctions of the estimated correlation function of $X(t)$, $0 \leq t \leq \tau$, i.e., the eigenfunctions corresponding to the largest d eigenvalues, referred to as basis functions. The samples of $\{Z_k\}$ result by projecting samples of $X(t)$ on the basis functions. Conditions are established under which extremes of, e.g., the distribution

* Corresponding author.

E-mail addresses: hx223@cornell.edu (H. Xu), mdg12@cornell.edu (M.D. Grigoriu).

of $\sup_{t \in [0, \tau]} |X_d(t)|$ converges to that of $\sup_{t \in [0, \tau]} |X(t)|$ as $d \rightarrow \infty$ so that the distribution of the target quantity of interest $\sup_{t \in [0, \tau]} |X_d(t)|$ can be estimated from samples of $X_d(t)$ for a sufficiently large d . This property is essential in applications since samples of $X_d(t)$ can be generated by standard Monte Carlo algorithms while samples of $X(t)$ are not available since $X(t)$ has infinite stochastic dimension as an uncountable family of random variables indexed by $t \in [0, \tau]$.

Our objective is to construct simple probabilistic models for uni- and multi-variate wind pressure wind tunnel records which are capable to characterize accurately extremes of these records. It is assumed that the wind records are sufficiently long such that their mean, correlation and marginal distribution functions can be estimated accurately. Of the infinite family of processes which can be fitted to these estimates or only to some of them, we have selected one- and two-dimensional diffusion processes with linear drift coefficients and translations of Gaussian/non-Gaussian diffusion processes. The selected models fit exactly the marginal distribution of the target wind records. Their correlation function is selected to minimize an objective function which depends on the discrepancies between target and model correlations and extremes, so that it may or may not fit accurately the target correlation function. It is shown that even such simple models can provide accurate estimates of the distribution of extremes of wind pressure records.

The paper is organized as follows. We introduce the one- and two-dimensional diffusion processes with linear drifts and construct their FD models in Section 2. Translation diffusion processes and their FD models are defined and examined in Section 3. The processes of the previous two sections are fitted to wind data in Section 4. The section also assess the performance of the resulting processes and of their FD models. Final comments are in Section 5.

2. Diffusion processes with linear drift

We review briefly two of the diffusion processes constructed in [2,3] and give conditions under which the stochastic differential equations defining these processes admit unique strong solutions. These conditions illustrate some of the limitations of two-dimensional diffusion processes under consideration.

2.1. One-dimensional processes

Let $X(t)$, $t \in [0, \tau]$, be a real-valued diffusion process defined by the stochastic differential equation

$$dX(t) = -\rho X(t)dt + D(X(t))dB(t), \quad t \geq 0, \quad (2.1)$$

where $B(t)$ is the standard Brownian motion process and $\rho > 0$. The stationary solution of this equation is a zero-mean non-Gaussian process with correlation function $c(s, t) = E[X(s)X(t)] = \exp(-\rho|s - t|)$. Our objective is to estimate the distribution of the extreme $\sup_{t \in [0, \tau]} |X(t)|$ of the stationary process $X(t)$ during a time interval of length $\tau > 0$.

The characterization of extremes of $X(t)$ requires detailed knowledge of the samples of this process, which is provided by the strong solution of (2.1), see [4] (Sect. 4.7). The following theorem gives conditions under which this solution exists and is unique.

Theorem 2.1. *The stochastic differential Eq. (2.1) admits a unique strong solution $X(t)$ with stationary marginal density function f , if f is continuously differentiable on $\mathbb{R}$, $\lim_{x \rightarrow \infty} -2\rho x f(x)/f'(x) < \infty$ and*

$$D(x) = \left(-\frac{2\rho}{f(x)} \int_{-\infty}^x u f(u) du \right)^{1/2}. \quad (2.2)$$

Proof. It has been shown in [2] that the stationary process $X(t)$ of (2.1) has a specified marginal density f if its diffusion coefficient has the expression in (2.2).

The strong solutions of (2.1) exists and are unique if $D(x)$ is locally Lipschitz continuous and satisfies the growth condition [4, Chap. 4], i.e., if for each $\xi > 0$, there exists $C_\xi > 0$ such that

$$|D(x_1) - D(x_2)| \leq C_\xi |x_1 - x_2|, \quad |x_1|, |x_2| < \xi, \quad (2.3)$$

and there exists $K > 0$ such that

$$D(x)^2 \leq K(1 + x^2), \quad x \in \mathbb{R}. \quad (2.4)$$

Since $f(x) > 0$ and $\int_{-\infty}^x u f(u) du$ are continuous functions on $\mathbb{R}$, then from (2.2), $D(x)^2$ is also continuous on $\mathbb{R}$. Therefore, for any $|x| \leq \xi$ with $0 < \xi < \infty$, $f(x) > 0$ so that $D(x)^2 < \infty$. Moreover, $D(x)^2$ is bounded on $\mathbb{R}$ since $\int_{-\infty}^{\infty} u f(u) du = E[X(t)] = 0$ so that

$$\lim_{x \rightarrow \infty} D(x)^2 = -\lim_{x \rightarrow \infty} \frac{2\rho}{f(x)} \int_{-\infty}^x u f(u) du = -\lim_{x \rightarrow \infty} \frac{2\rho x f(x)}{f'(x)} < \infty.$$

by using l'Hopital rule. Since $D(x)^2$ is bounded on $\mathbb{R}$, then there exists a constant $K > 0$ such that $D(x)^2 \leq K(1 + x^2)$, which means that $D(x)$ satisfies condition (2.4). The differentiation of (2.2) gives

$$D'(x) = -\frac{\rho x}{D(x)} - \frac{f'(x)D(x)}{2f(x)}.$$

so that this function is continuous since $f(x)$, $f'(x)$ and $D(x)$ are continuous. Therefore, $D(x)$ is continuously differentiable so that it is locally Lipschitz continuous. This means that $D(x)$ satisfies condition (2.3). $\square$

2.2. Two-dimensional processes

Let $X(t) = (X_1(t), X_2(t))^T$, $t \in [0, \tau]$ be a diffusion process defined by the stochastic differential equations

$$\begin{aligned} dX_1(t) &= (a_{11}X_1(t) + a_{12}X_2(t))dt + D_1(X_1(t), X_2(t))dB_1(t) \\ dX_2(t) &= (a_{21}X_1(t) + a_{22}X_2(t))dt + D_2(X_1(t), X_2(t))dB_2(t), \quad t \geq 0 \end{aligned} \quad (2.5)$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motion processes. The stationary solution $X(t)$ of (2.5) has zero mean and correlation function $c(s, t) = E[X(s)X(t)^T] = \exp(A(t-s))c(s, s)$ [4, Chap. 7], provided that the real part of the eigenvalues of $A = [a_{11} \ a_{12}; a_{21} \ a_{22}]$ are negative, where $\exp(\cdot)$ denotes the matrix exponential function [5, Sect. 1.5].

It was found in [3] that the stationary bivariate process $X(t)$ has the marginal density f if the diffusion coefficients have the expressions

$$\begin{aligned} D_1(x_1, x_2) &= \left(\frac{2a_{11}}{f(x_1, x_2)} \int_{-\infty}^{x_1} u f(u, x_2) du \right)^{1/2} \\ D_2(x_1, x_2) &= \left(\frac{2a_{22}}{f(x_1, x_2)} \int_{-\infty}^{x_2} u f(x_1, u) du \right)^{1/2}, \quad -\infty < x_1, x_2 < \infty. \end{aligned} \quad (2.6)$$

These results have been obtained by solving the time-invariant Fokker-Planck equation

$$\begin{aligned} \frac{\partial}{\partial x_1} \left((a_{11}x_1 + a_{12}x_2)f(x_1, x_2) - \frac{1}{2} \frac{\partial}{\partial x_1} (D_1(x_1, x_2)^2 f(x_1, x_2)) \right) \\ + \frac{\partial}{\partial x_2} \left((a_{21}x_1 + a_{22}x_2)f(x_1, x_2) - \frac{1}{2} \frac{\partial}{\partial x_2} (D_2(x_1, x_2)^2 f(x_1, x_2)) \right) = 0. \end{aligned} \quad (2.7)$$

under the conditions

$$a_{12}x_2 \frac{\partial}{\partial x_1} f(x_1, x_2) + a_{21}x_1 \frac{\partial}{\partial x_2} f(x_1, x_2) = 0 \quad (2.8)$$

and

$$\begin{aligned} a_{11}x_1 f(x_1, x_2) - \frac{1}{2} \frac{\partial}{\partial x_1} [D_1(x_1, x_2)^2 f(x_1, x_2)] &= 0 \\ a_{22}x_2 f(x_1, x_2) - \frac{1}{2} \frac{\partial}{\partial x_2} [D_2(x_1, x_2)^2 f(x_1, x_2)] &= 0. \end{aligned} \quad (2.9)$$

The condition of (2.8) is satisfied if, e.g., f is a function of $a_{21}x_1^2 - a_{12}x_2^2$ or $a_{12} = a_{21} = 0$. These assumptions impose rather severe restrictions on the form of the density f and/or the correlation function.

As previously noted, our objective is to characterize extremes of the diffusion process $X(t)$ whose characterization requires detailed information on the sample properties of this process, which are provided by the strong solution of (2.5). The following theorem provides conditions under which (2.5) admits a unique strong solution.

Theorem 2.2. *The stochastic differential Eq. (2.5) admits a unique strong solution $X(t)$ with stationary marginal density f , if the drift coefficients $D_1(x)$ and $D_2(x)$ are local Lipschitz continuous (2.10), i.e., for each $\xi > 0$, there exists $C_\xi > 0$ such that*

$$\|D(x) - D(z)\|_2 \leq C_\xi \|x - z\|_2, \quad \|x\|_2, \|z\|_2 < \xi, \quad x = (x_1, x_2)^T, \quad z = (z_1, z_2)^T, \quad (2.10)$$

where $D(x_1, x_2) = (D_1(x_1, x_2), D_2(x_1, x_2))^T$, $\|\cdot\|_2$ denotes the L_2 norm, and satisfy the growth condition, i.e., there exists $K > 0$ such that

$$D_1(x_1, x_2)^2 + D_2(x_1, x_2)^2 \leq K(1 + x_1^2 + x_2^2), \quad x_1, x_2 \in \mathbb{R}. \quad (2.11)$$

Proof. The above conditions are those in (4.74) and (4.75) of [4, Chap. 4]. We note that there are processes $X(t)$ of the type in (2.5) which satisfy (2.10) and (2.11), see Section 3.2.2. Yet, these conditions are rather restrictive. For example, if the joint density of $(X_1(t), X_2(t))$ is a zero-mean Gaussian distributions with covariance matrix $\begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}$, $0 < |c| < 1$, then

$$D_1(x_1, x_2)^2 = \frac{2a_{11}}{f(x_1|x_2)} \int_{-\infty}^{x_1} u f(u|x_2) du, \quad -\infty < x_1, x_2 < \infty,$$

where $f(x_1|x_2)$ denotes the conditional density of $X_1(t)|X_2(t)$. For any fixed $x_2 \in \mathbb{R}$,

$$\int_{-\infty}^{x_1} u f(u|x_2) du \rightarrow \int_{-\infty}^{\infty} u f(u|x_2) du = E[X_1(t)|X_2(t) = x_2] = cx_2, \quad x_1 \rightarrow \infty.$$

and, since $1/f(x_1|x_2)$ increases much faster than x_1^2 as $x_1 \rightarrow \infty$, the growth condition (2.11) is not satisfied. $\square$

We conclude with the observation that the condition of (2.8) limits the use of $X(t)$ in (2.5) since the set of marginal distributions and correlation functions of $X(t)$ under this condition is small so that few data sets can be described by this process.

2.3. Finite dimensional (FD) models

Let $X(t)$, $0 \leq t \leq \tau$, be a zero-mean, finite-variance $\mathbb{R}^n$ -valued stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$, which may or may not be stationary. The FD models $\{X_d(t)\}$ of $X(t)$ are defined by

$$X_{i,d_i}(t) = \sum_{k=1}^{d_i} Z_{i,k} \varphi_{i,k}(t), \quad d_i = 1, 2, \dots, \quad t \in [0, \tau], \quad i = 1, \dots, n, \quad (2.12)$$

where $\{\varphi_{i,k}(t)\}$, $k = 1, \dots, d_i$, are the top d_i eigenfunctions of the correlation function $\{c_{ii}(s, t) = E[X_i(s)X_i(t)]\}$, i.e., the eigenfunctions corresponding to the largest d_i eigenvalues of the correlation functions $\{c_{ii}(s, t)\}$, and $\{Z_{i,k}\}$ are random coefficients whose samples are obtained by projecting samples $X_i(t, \omega)$, $\omega \in \Omega$, of $X_i(t)$ on the basis functions $\{\varphi_{i,k}(t)\}$, i.e.,

$$Z_{i,k}(\omega) = \int_0^\tau X_i(t, \omega) \varphi_{i,k}(t) dt, \quad k = 1, 2, \dots, \quad \omega \in \Omega, \quad i = 1, \dots, n.$$

The following theorem from [6] states that the extremes of $X_{i,d_i}(t)$ converges to those of $X_i(t)$ weakly as d_i increases indefinitely under some conditions. This means that the distribution of the target quantity of interest $\sup_{t \in [0, \tau]} |X_i(t)|$ can be estimated from samples of $X_{i,d_i}(t)$ for sufficiently large stochastic dimensions $\{d_i\}$.

Theorem 2.3. *If $X(t)$ has continuous samples, the correlation functions of the components of $X(t)$ are continuous and $\sum_{k=1}^\infty \lambda_{i,k} \sup_{t \in [0, \tau]} \varphi_{i,k}(t)^2 < \infty$*

for any $i = 1, \dots, n$, then

$$\sup_{t \in [0, \tau]} |X_{i,d_i}(t)| \xrightarrow{w} \sup_{t \in [0, \tau]} |X_i(t)|, \quad d_i \rightarrow \infty, \quad i = 1, \dots, n \quad (2.13)$$

and

$$\sup_{t \in [0, \tau]} \|X_d(t)\| \xrightarrow{w} \sup_{t \in [0, \tau]} \|X(t)\|, \quad \min_{1 \leq i \leq n} d_i \rightarrow \infty, \quad (2.14)$$

where w denotes weak convergence, $\|\cdot\|$ is the L_∞ norm and $X_d(t) = (X_{1,d_1}(t), \dots, X_{n,d_n}(t))$.

This theorem is essential in applications since it shows that the distributions of the target extremes $\sup_{t \in [0, \tau]} |X_i(t)|$ and $\sup_{t \in [0, \tau]} \|X(t)\|$ can be estimated from samples of FD models of $X(t)$ provided that the stochastic dimensions $\{d_i\}$ are sufficiently large, which can be generated by standard Monte Carlo algorithms.

The following Corollary specializes the statement of the previous theorem to the case of real-valued stationary processes with exponential correlation function.

Corollary 2.1. *If a real-valued, zero-mean, weakly stationary stochastic process $X(t)$ has exponential correlation function, then*

$$\sup_{t \in [0, \tau]} |X_d(t)| \xrightarrow{w} \sup_{t \in [0, \tau]} |X(t)|, \quad d \rightarrow \infty,$$

where $d = d_1$ and $X_d(t) = X_{1,d_1}(t)$.

Proof. The eigenfunctions of the exponential correlation function $c(s, t) = e^{-\rho|s-t|}$ of $X(t)$ are given by (2.51) and (2.52) of [7]. They have the expressions

$$\varphi_{2k-1}(t) = \frac{\sin(\xi_k(t + \tau)/2)}{\sqrt{\tau - \sin(2\tau\xi_k)/(2\xi_k)}} \quad \text{and} \quad \varphi_{2k}(t) = \frac{\cos(\eta_k(t + \tau)/2)}{\sqrt{\tau + \sin(2\tau\eta_k)/(2\eta_k)}},$$

where ξ_k and η_k are solutions of $\xi_k + \rho \tan(\tau\xi_k) = 0$ and $\eta_k \tan(\tau\eta_k) = \rho$, $k = 1, 2, \dots$. Since

$$|\varphi_{2k-1}(t)| = \frac{|\sin(\xi_k(t + \tau)/2)|}{\sqrt{\tau + \cos^2(\tau\xi_k)/(2\rho)}} \leq \frac{1}{\sqrt{\tau}}, \quad k = 1, 2, \dots,$$

and

$$|\varphi_{2k}(t)| = \frac{|\cos(\eta_k(t + \tau)/2)|}{\sqrt{\tau + \sin^2(\tau\eta_k)/(2\rho)}} \leq \frac{1}{\sqrt{\tau}}, \quad k = 1, 2, \dots,$$

we have $|\varphi_k(t)| \leq 1/\sqrt{\tau}$ for all $k \geq 1$ so that

$$\sum_{k=1}^\infty \lambda_k \sup_{t \in [0, \tau]} \varphi_k(t)^2 \leq \frac{1}{\sqrt{\tau}} \sum_{k=1}^\infty \lambda_k < \infty,$$

since $\sum_{k=1}^\infty \lambda_k < \infty$ [8, Chap. 4]. Therefore,

$$\sup_{t \in [0, \tau]} |X_d(t)| \xrightarrow{w} \sup_{t \in [0, \tau]} |X(t)|, \quad d_i \rightarrow \infty,$$

by Theorem 2.3. $\square$

This corollary holds also for vector-valued processes provided that the components of these processes are exponentially correlated.

The stochastic process with exponential correlation function is not differentiable. As noted in [9,10], replacing the exponential correlation function by a smoother function can have certain advantages without necessarily compromising the modeling accuracy. The second derivative of these modified correlation functions exist and are continuous. The FD models constructed under these modified correlation functions converge to the target processes weakly under the sup-norm, see Theorem 3.2 in [6].

3. Translation diffusion processes

Let $X(t) = (X_1(t), \dots, X_n(t))^T$ be an n -variate zero-mean stationary process with correlation function $c(u) = E[X(t)X(t+u)^T]$ and marginal distributions F_i , $i = 1, \dots, n$ of the components of $X(t)$. We construct

translation models for $X(t)$ by mapping stationary Gaussian and non-Gaussian processes into processes which match exactly the marginal distributions of the components of $X(t)$ and approximately the correlation functions of this process. We also examine the extremes of the resulting translation processes which are equivalent in the sense that they have the same marginal distributions and similar correlation functions.

3.1. Gaussian diffusion processes

The translation version $X_T(t)$ of $X(t)$ has the components $X_{T,i}(t) = F_i^{-1} \circ \Phi(G_i(t))$, $i = 1, \dots, n$, where $\{G_i(t)\}$ are zero-mean, unit-variance stationary Gaussian processes with correlation functions $c_{G,i,j}(u) = E[G_i(t)G_j(t+u)]$. Note that the processes $X_{T,i}(t)$ and $X_i(t)$ have the same marginal distributions. Generally, the correlation functions of the components of $X_T(t) = (X_{T,1}(t), \dots, X_{T,n}(t))$ and $G(t) = (G_1(t), \dots, G_n(t))$ are similar particularly when dealing with positive and mildly negative correlations [11, Sect. 3]. For simplicity, we approximate the correlation function of $X_T(t)$ by that of $G(t)$.

The Gaussian process $G(t)$ can be defined as the stationary solution of the stochastic differential equation

$$dG(t) = \alpha G(t)dt + \beta dB(t), \quad t \in [0, \tau], \quad (3.1)$$

where α and β are (n, n) and (n, m) real-valued matrices which need to be determined and $B(t)$ denotes an m -dimensional process whose components are independent standard Brownian motions. The above equation admits stationary solution if the real parts of the eigenvalues of α are negative. Under this condition, the covariance matrix $\gamma = E[G(t)G(t)^T]$ and the correlation function of the stationary process $G(t)$ are such that

$$\alpha\gamma + \gamma\alpha^T + \beta\beta^T = 0, \quad (3.2)$$

and $E[G(t)G(t+u)^T] = \exp(\alpha u)\gamma$, where $\exp(\cdot)$ denotes the matrix exponential function [4, Chap. 7].

The matrices α and β can be determined in two steps. First, the matrix α is selected to minimize the objective function

$$e_1(\alpha) = w_1 \int_0^\tau \|c(u) - \exp(\alpha u)\gamma\|_2^2 + w_2 \sum_{i=1}^n \|H_i(\cdot) - H_{T,i}(\cdot)\|_1^2 \quad (3.3)$$

where $\|\cdot\|_i$ denotes the L_i norm, $i = 1, 2$, $c(u) = \{c_{ij}(u)\}$, $H_i(\cdot)$ is the histogram of $\sup_{t \in [0, \tau]} |X_i(t)|$ and $H_{T,i}(\cdot)$ is the histogram of $\sup_{t \in [0, \tau]} |X_{T,i}(t)|$ obtained from the translation model, where the histograms have been calculated by the Matlab histograms function. The weighting coefficients w_1, w_2 are such that the two components of $e_1(\alpha)$ in (3.3) contribute equally to the objective function.

We conclude with the observation that the translation process $X_T(t) = (X_{T,1}(t), \dots, X_{T,n}(t))$ exists for any α and β under the stated conditions provided that the target distribution functions $\{F_i\}$ are strictly monotonically increasing and admit densities. The components of $X_T(t)$ also satisfy the stochastic differential equations, see [4, Chap. 4],

$$\begin{aligned} dX_{T,i}(t) &= (\nabla h_i)^T dG(t) + \frac{1}{2}(dG(t))^T [\nabla^2 h_i] dG(t) \\ &= \left((\nabla h_i)^T \alpha + \frac{1}{2} \text{Tr}(\beta^T [\nabla^2 h_i] \beta) \right) dt + (\nabla h_i)^T \beta dB(t), \quad i = 1, \dots, n, \end{aligned} \quad (3.4)$$

where ∇h_i and $\nabla^2 h_i$ are the gradient and Hessian matrix of h_i and Tr is the trace operator.

3.2. Non-Gaussian diffusion processes

The construction of the previous subsection is extended to translation models which are obtained by memoryless transformations of non-Gaussian diffusion processes. The resulting translation models are equivalent in the sense that they have the same marginal distributions and similar correlation functions. Yet, they may have very different extremes.

3.2.1. One dimensional case

Let $X(t)$ be the target process with zero-mean, correlation function $c(u) = E[X(t+u)X(t)] = e^{-\rho u}$, $\rho > 0$, and marginal distribution F . We construct translation models $\{X_T^{(i)}(t)\}$ of $X(t)$ from two (non-Gaussian) diffusion processes $\{Y^{(i)}(t)\}$ defined by the stochastic differential equations

$$dY^{(i)}(t) = -\rho Y^{(i)}(t)dt + D_i(Y^{(i)}(t))dB(t), \quad t \in [0, \tau], \quad i = 1, 2, \quad (3.5)$$

where $D_1(x) = \sqrt{2\rho}$ and $D_2(x) = (2\rho(x+2/\theta)/\theta)^{1/2}$ with $\rho, \theta > 0$. The first two moments of these processes are $E[Y^{(i)}(t)] = 0$ and $E[Y^{(i)}(t+u)Y^{(i)}(t)] = e^{-\rho u}$. Their marginal distributions are Gauss and Gamma with the densities $f_1(x) = \phi(x)$, $x \in \mathbb{R}$ and $f_2(x) = \theta(\theta x + 2)e^{-\theta x + 2}$, $x > -2/\theta$, where ϕ is the standard Gaussian density [12].

The stochastic differential Eq. (3.5) admits a unique strong solution since the diffusion coefficients $D_i(x)$, $i = 1, 2$, are locally Lipschitz, a property which results directly from their expressions, and satisfy the growth condition in [4, Chap. 4] since there exist constants $K_1 = 2\rho$ and $K_2 = (2\rho\theta + 4\rho)/\theta^2$ such that

$$\begin{aligned} D_1(x)^2 &= 2\rho \leq K_1(1+x^2), \quad x \in \mathbb{R} \\ D_2(x)^2 &= \frac{2\rho}{\theta} \left(x + \frac{2}{\theta} \right) \leq K_2(1+x^2), \quad x > -2/\theta. \end{aligned} \quad (3.6)$$

The translation processes defined by the mapping $X_T^{(i)}(t, \omega) = F^{-1} \circ F_{Y^{(i)}}(Y^{(i)}(t, \omega))$, $i = 1, 2$, have the marginal distribution F and their correlation functions are similar to the exponential correlation function of the processes $Y^{(i)}(t)$. Note also that the extremes of $\{X_T^{(i)}(t)\}$ occur at the same time as those of $\{Y^{(i)}(t)\}$.

3.2.2. Two dimensional case

Suppose that the target is a real-valued, zero-mean, stationary process $X(t) = (X_1(t), X_2(t))^T$ with correlation function $c(u) = E[X(t)X(t+u)^T]$ and component marginal distributions F_j , $j = 1, 2$. We construct translation models of $X(t)$ from two diffusion processes $Y^{(i)}(t) = (Y_1^{(i)}(t), Y_2^{(i)}(t))^T$, $i = 1, 2$, which have the same correlation function as $X(t)$ by translating them such that their components match the marginal distributions $\{F_j\}$ of $\{X_j(t)\}$, $j = 1, 2$.

The diffusion processes $\{Y^{(i)}(t)\}$, $i = 1, 2$ are defined by the stochastic differential equations

$$\begin{aligned} dY_1^{(i)}(t) &= (a_{11}Y_1^{(i)}(t) + a_{12}Y_2^{(i)}(t))dt + D_{11}^{(i)}(Y_1^{(i)}(t), Y_2^{(i)}(t))dB_1(t) \\ &\quad + D_{12}^{(i)}(Y_1^{(i)}(t), Y_2^{(i)}(t))dB_2(t) \\ dY_2^{(i)}(t) &= (a_{21}Y_1^{(i)}(t) + a_{22}Y_2^{(i)}(t))dt + D_{21}^{(i)}(Y_1^{(i)}(t), Y_2^{(i)}(t))dB_1(t) \\ &\quad + D_{22}^{(i)}(Y_1^{(i)}(t), Y_2^{(i)}(t))dB_2(t), \end{aligned} \quad (3.7)$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motion processes, $a_{11} < 0$, $a_{22} < 0$, $a_{12}a_{21} < 0$ and the diffusion coefficients

$$\begin{aligned} D_{ij}^{(1)}(x_1, x_2) &= b_{ij}, \quad x_1, x_2 \in \mathbb{R}, \quad i, j = 1, 2, \\ D_{ij}^{(2)}(x_1, x_2) &= \frac{\kappa_{ij}}{\sqrt{2(\theta-1)}} \left(\frac{(x_1 - x_2)^2}{4(1-r)} + \frac{(x_1 + x_2)^2}{4(1+r)} + \zeta \right)^{1/2}, \\ x_1, x_2 &\in \mathbb{R}, \quad \theta > 1, \zeta > 0, \end{aligned} \quad (3.8)$$

with $\kappa = \{\kappa_{ij}\}$, $\kappa\kappa^T = -(AQQ^T + QQ^TA^T)$, $A = [a_{11} \ a_{12}; a_{21} \ a_{22}]$, $Q = [-\sqrt{1-r} \ \sqrt{1+r}; \sqrt{1-r} \ \sqrt{1+r}]$ and $r = E[Y_1^{(2)}(0)Y_2^{(2)}(0)]$.

The stationary solutions $\{Y^{(i)}(t)\}$ of (3.7) are zero-mean processes with correlation function $c(s, t) = \exp(A|t-s|)\gamma$, where $A = [a_{11} \ a_{12}; a_{21} \ a_{22}]$, $\exp(\cdot)$ denotes the matrix exponential function [5, Sect. 1.5] and γ is the stationary covariance matrix of $\{Y^{(i)}(t)\}$, which exists if the real part of the eigenvalues of the matrix A are negative. The process $\{Y^{(1)}(t)\}$ is Gaussian while $\{Y^{(2)}(t)\}$ is not. The joint densities of the two-dimensional vector $(Y_1^{(2)}(t), Y_2^{(2)}(t))$ is

$$f_2(x_1, x_2) = \frac{C_1}{2\sqrt{(1-r)(1+r)}} \left(\frac{(x_1 - x_2)^2}{4(1-r)} + \frac{(x_1 + x_2)^2}{4(1+r)} + \zeta \right)^{-\theta},$$

where C_1 is the normalizing constant.

For the diffusion coefficients under consideration, the stochastic differential Eq. (3.7) admits strong unique solution since the diffusion coefficients $D_{ij}^{(k)}(x_1, x_2)$ are locally Lipschitz continuous and satisfy the growth condition (2.11). For example, the drift and diffusion coefficients of $Y^{(2)}(t)$ are locally Lipschitz continuous since they are continuously differentiable by definition. The growth condition is satisfied since there exists $K_2 = \max\{1, \zeta\}(\kappa_{11}^2 + \kappa_{12}^2 + \kappa_{21}^2 + \kappa_{22}^2)/(4(\theta - 1))$ such that for any $x_i \in \mathbb{R}$, $i = 1, 2$,

$$\begin{aligned} & D_{11}^{(2)}(x_1, x_2)^2 + D_{12}^{(2)}(x_1, x_2)^2 + D_{21}^{(2)}(x_1, x_2)^2 + D_{22}^{(2)}(x_1, x_2)^2 \\ &= \frac{\kappa_{11}^2 + \kappa_{12}^2 + \kappa_{21}^2 + \kappa_{22}^2}{4(\theta - 1)}(x_1^2 + x_2^2 + \zeta) \leq K_3(1 + x_1^2 + x_2^2). \end{aligned}$$

As for the translation processes of the previous section, we note that the translation processes $X_{T,j}^{(i)}(t) = F_j^{-1} \circ F_{Y_j^{(i)}}(Y_j^{(i)}(t))$, $j = 1, 2$, $i = 1, 2$, have the marginal distributions $\{F_j\}$ and correlation functions similar to the exponential correlation function of the processes $Y^{(i)}(t)$, see [11, Sect. 3]. Note also that the extremes of $\{X_{T,j}^{(i)}(t)\}$ and $\{Y_j^{(i)}(t)\}$ occur at the same times.

3.3. Finite dimensional (FD) models

The FD models $\{X_{T,d}(t)\}$ of the translation model $X_T(t)$ have the form

$$X_{T,d,i}(t) = F_i^{-1} \circ F_{Y_i}(Y_{i,d,i}(t)), \quad i = 1, \dots, n, \quad (3.9)$$

where $Y_{i,d,i}(t)$ are finite dimensional models of the stationary stochastic process $Y_i(t)$ of the type in (2.12) with marginal distribution F_{Y_i} , $i = 1, \dots, n$. As previously, we are interested in extremes $\sup_{t \in [0, \tau]} |X_{T,i}(t)|$ of $X_{T,i}(t)$ in the stationary regime during a time interval of length $\tau > 0$.

The following Corollary from [6] shows that the extremes of $X_{T,d,i}(t)$ converges to those of $X_{T,i}(t)$ in probability as d_i increases indefinitely if the input Y_i is Gaussian. This means that the distribution of the target quantity of interest $\sup_{t \in [0, \tau]} |X_{T,i}(t)|$ can be estimated from samples of $X_{T,d,i}(t)$ for a sufficiently large d_i since convergence in probability implies convergence in distribution.

Corollary 3.1. *If $Y_i(t)$ is Gaussian and has continuous samples and continuous correlation function, and F_i is continuous and strictly monotonically increasing for each $i = 1, \dots, n$, then*

$$\sup_{t \in [0, \tau]} \|X_{T,d}(t) - X_T(t)\| \xrightarrow{p} 0, \quad \min_{1 \leq i \leq n} d_i \rightarrow \infty,$$

where p denotes convergence in probability.

4. Wind tunnel records and probabilistic models

The random processes discussed in the previous sections are fitted to wind pressure time series recorded at the University of Florida boundary layer wind tunnel facility (UFBLWT). The data set is discussed extensively in [1]. It consists of wind pressure time series at 252 taps recorded under a steady fan speed of 1200 RPM.

The following subsection presents the data set and estimates the first two moments and the marginal distributions of the wind pressure time series. The subsequent sections fit the processes discussed in the first part of this study to these estimates, construct FD models for the resulting processes, referred to as target processes, and assess performance by comparing target and FD extremes.

4.1. Estimates of mean, correlation and marginal distributions

The wind pressure data was recorded for $T = 180$ seconds at a time step $\Delta T = 0.00164$ second, so that it consists of $n = T/\Delta T = 110,000$ readings. We consider records at two pressure taps and denote them by $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$, where $\tilde{y}_j = (\tilde{y}_{1,j}, \tilde{y}_{2,j})^T$, $j = 1, \dots, n$. The taps, labeled 120 and 126 in the experiment, have been selected since their skewness

and kurtosis coefficients are 1.2138 and 5.8509 for tap 120 and 1.1926 and 5.9813 for tap 126. Segments of the record $\{\tilde{y}_j\}$ of length $1000\Delta T$ are shown in the left and right panels of Fig. 1.

We assume that the wind records are samples of an ergodic process $\tilde{X}(t) = (\tilde{X}_1(t), \tilde{X}_2(t))^T$ and that these samples are sufficiently long to estimate the marginal distribution and the correlation function of this process. Let $\hat{\mu}_i = \sum_{j=1}^n \tilde{y}_{i,j}/n$ and $\hat{\sigma}_i^2 = \sum_{j=1}^n (\tilde{y}_{i,j} - \hat{\mu}_i)^2/n$, $i = 1, 2$, denote the estimated mean and variance of the recorded time series $\{\tilde{y}_j\}$ and set $y_{i,j} = (\tilde{y}_{i,j} - \hat{\mu}_i)/\sqrt{\hat{\sigma}_i}$, $i = 1, 2$.

The estimates of the correlation function and the marginal densities of the normalized components $X_i(t) = (\tilde{X}_i(t) - E[\tilde{X}_i(t)])/\text{Std}[\tilde{X}_i(t)]$ of $\tilde{X}(t)$ are

$$\hat{c}_{ij}(k) = \frac{1}{n-k} \sum_{l=1}^{n-k} y_{i,l} y_{j,l+k}, \quad k = 0, 1, \dots, \quad i, j = 1, 2. \quad (4.1)$$

and

$$\hat{f}_i(x) = \frac{1}{n\kappa} \sum_{j=1}^n \phi\left(\frac{x - y_{i,j}}{\kappa}\right), \quad i = 1, 2, \quad (4.2)$$

where ϕ denotes the density of the standard Gaussian variable and κ is a bandwidth parameter set equal to $\kappa = (4/3n)^{1/5}$ [13, Chap.3]. These estimates are viewed as the actual correlation functions and marginal distributions of $X(t)$. The left and right panels of Fig. 2 show the histograms and the estimates $\hat{f}_i(x)$ of the normalized wind pressure records at taps 120 and 126. The plots show that the estimates of (4.2) are consistent with data. The estimates of the marginal distributions $\hat{F}_i$ of $X_i(t)$ result from $\hat{f}_i$ by integration, $i = 1, 2$. The left and right panels of Fig. 3 show the marginal distributions obtained from data and the estimates $\hat{F}_i(x)$ of the normalized wind pressure records at taps 120 and 126 in logarithmic scale.

We also estimate the distributions of the extremes $\sup_{t \in [0, \tau]} |X_i(t)|$ of the normalized components $X_i(t) = (\tilde{X}_i(t) - E[\tilde{X}_i(t)])/\text{Std}[\tilde{X}_i(t)]$ of $\tilde{X}(t)$ by the following procedure. First, the wind record is partitioned in non-overlapping segments of duration $1000\Delta T = 1.62$ sec, so that there are $q = 1100$ such segments. Second, the extremes of the individual segments are recorded and used to estimate the distribution of $\sup_{t \in [0, \tau]} |X_i(t)|$ at each pressure tap by using, e.g., the MATLAB ecdf function.

4.2. One-dimensional exponential diffusion processes

We fit the diffusion process $X(t)$ in (2.1) to the normalized wind pressure records at tap 120 by using the properties of this process summarized in Theorem 2.1. We construct FD models $\{X_d(t)\}$ for fitted diffusion process $X(t)$ and assess their performance by comparing the distributions of $\sup_{t \in [0, \tau]} |X(t)|$ and $\sup_{t \in [0, \tau]} |X_d(t)|$ for $\tau = 0.162$ sec., based on 5000 samples of $X(t)$ with the time step $\Delta\tau = 3.27 \times 10^{-5}$ sec. Note that this time step has no relation to that of the recorded wind pressure series. The diffusion process in (2.5) cannot be fitted to wind pressure records at two taps because of restrictions on its distributions, see Section 2.2.

4.2.1. Process fitted to wind data

The parameter ρ of the exponential diffusion process in (2.1) is selected to minimize the objective function (3.3). Its diffusion coefficient $D(x)$ is calculated from (2.2) with f set equal to the estimate $\hat{f}$ of the marginal distribution f given by (4.2). We note that the defining equation of $X(t)$ admits a unique strong solutions by Theorem 2.1 since $f(x)$ defined by (4.2) is continuously differentiable and $\lim_{x \rightarrow \infty} -2\rho x f(x)/f'(x) = 2\rho\kappa^2 < \infty$.

The left panel of Fig. 4 shows the target correlation function $\hat{c}(k)$ for $k = 1, \dots, 100$ estimated from the wind record at tap 120 (solid line) and the corresponding correlation functions of $X(t)$ obtained under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted line) and with $(w_1 = 1, w_2 = 10)$ (dashed line).

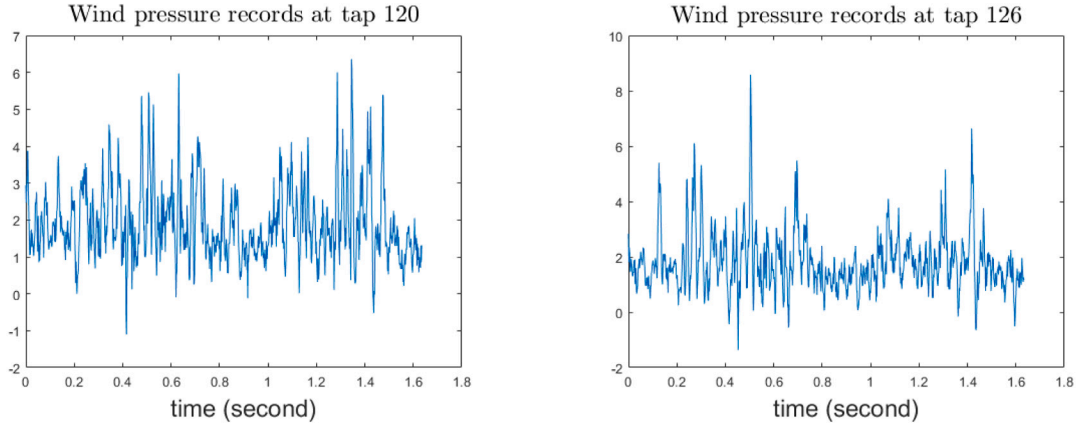


Fig. 1. Segments of length $1000\Delta t = 1.62$ of wind pressure records at taps 120 and 126 (left and right panels).

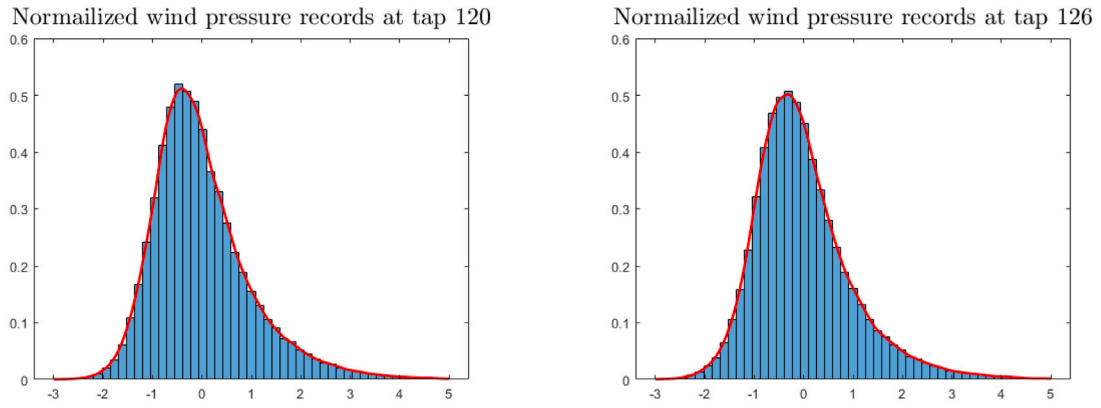


Fig. 2. Histograms of normalized wind pressure records at taps 120 and 126 with the marginal density $\hat{f}_i$ (left and right panels).

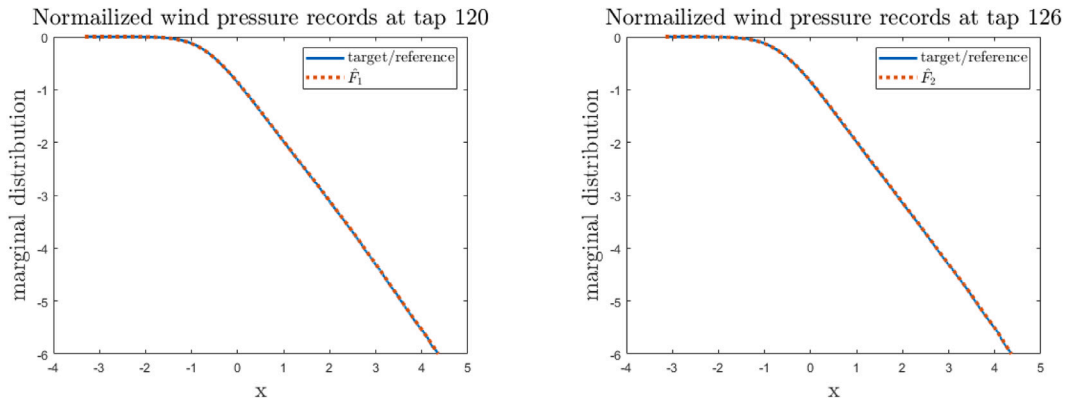


Fig. 3. Marginal distributions obtained from data and estimates $\hat{F}_i$ of normalized wind pressure records at taps 120 and 126 in logarithmic scale (left and right panels).

The right panel of Fig. 4 shows with solid line in semi logarithmic scale the estimate of the target probability $P(\sup_{t \in [0, \tau]} |X(t)| > x)$ at tap 120 obtained from wind data. The dotted and dashed lines are the corresponding probabilities of the diffusion process $X(t)$ defined under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ and $(w_1 = 1, w_2 = 10)$. The extremes of $X(t)$ corresponding to the objective function with $(w_1 = 1, w_2 = 10)$ are superior although its correlation function differs notable from the target correlation.

We conclude with the observation that the diffusion process $X(t)$ in (2.1) can only fit arbitrary marginal distributions. Its extremes

may or may not match target extremes depending on the objective function used to select the correlation parameter ρ of $X(t)$. Correlation functions rather different from the target correlation function may yield accurate estimates of extremes. This is an example of a simple model which, although captures exactly a single feature of the target process $X(t)$, the marginal distribution of $X(t)$ in this example, characterizes the extremes of this process accurately. A possible explanation for this behavior is that (1) the fitted exponential correlation (objective function with $(w_1 = 1, w_2 = 0)$) is too simplistic to capture the complex dependence of the wind time series and (2) the correlation

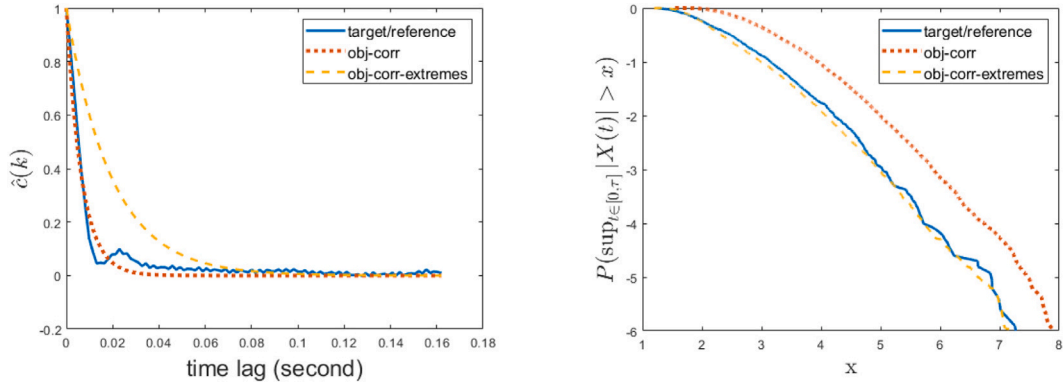


Fig. 4. Left panel: Estimated (solid line) and calculated correlation functions under (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted line), and $(w_1 = 1, w_2 = 10)$ (dashed line). Right panel: Estimated (solid line) and calculated probability $P(\sup_{t \in [0, \tau]} |X(t)| > x)$ under (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted line), and $(w_1 = 1, w_2 = 10)$ (dashed line).

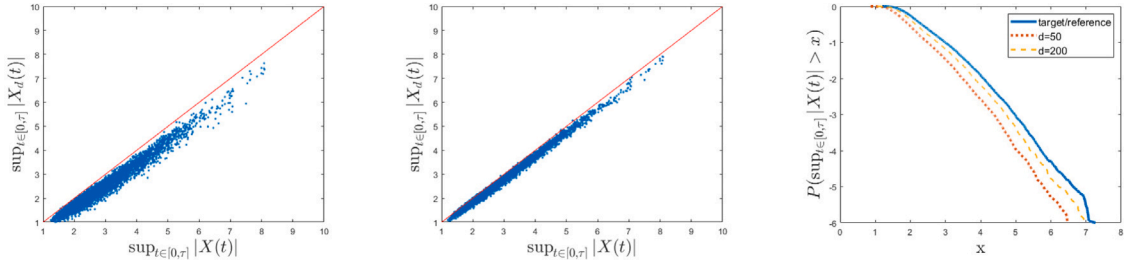


Fig. 5. Scatter plots of $\sup_{t \in [0, \tau]} |X(t)|$ and $\sup_{t \in [0, \tau]} |X_d(t)|$ for $d = 50, 200$ (left and middle panels). Estimate of the target probability $P(\sup_{t \in [0, \tau]} |X(t)| > x)$ (solid line) and estimates of $P(\sup_{t \in [0, \tau]} |X_d(t)| > x)$ for $d = 50$ (dotted line) and $d = 200$ (dashed line) in logarithmic scale (right panel).

function corresponding to the objective function with $(w_1 = 1, w_2 = 10)$ constitutes an “equivalent” correlation function designed to capture extremes.

4.2.2. Finite dimensional (FD) models

We develop FD models $X_d(t)$ for the process $X(t)$ constructed for the tap 120 under the objective function (3.3) with $(w_1 = 1, w_2 = 10)$. The FD models $X_d(t)$ are defined by (2.12) for $n = 1$. Since the correlation function is exponential, $\sup_{t \in [0, \tau]} |X_d(t)|$ converges weakly to $\sup_{t \in [0, \tau]} |X(t)|$ as $d \rightarrow \infty$ by Corollary 2.1. This means that the distribution of $\sup_{t \in [0, \tau]} |X(t)|$ can be estimated from samples of $X_d(t)$ for a sufficiently large d .

The left and middle panels of Fig. 5 show scatter plots of $(\sup_{t \in [0, \tau]} |X(t)|, \sup_{t \in [0, \tau]} |X_d(t)|)$ for $d = 50, 200$. The solid line in the right panel of Fig. 5 is an estimate of $P(\sup_{t \in [0, \tau]} |X(t)| > x)$ which is obtained directly from data. It is viewed as reference. The other lines of the figure are calculated from FD models for $d = 50$ (dotted line), $d = 200$ (dashed line). The dashed line is the closest to the reference. These plots show, in agreement with our theoretical results, that the discrepancy between the distributions of extremes of $X(t)$ and $X_d(t)$ decrease with the stochastic dimension d .

4.2.3. Synthetic exponential data

We examine the potential of the FD models constructed in this study and in [6] for the special case in which data set consists of samples of a zero-mean stationary diffusion process $Y(t)$ with exponential correlation function $E[Y(s)Y(t)] = e^{-\kappa|s-t|}$ and shifted Gamma marginal distribution with shape and scale parameters 2 and θ . The process is defined by the stationary solution of the stochastic differential equation

$$dY(t) = -\kappa Y(t)dt + \left(\frac{2\kappa}{\theta} Y(t) + \frac{4\kappa}{\theta^2}\right)^{1/2} dB(t), \quad t \geq 0,$$

with $\kappa = 1$ and $\theta = 1$.

The following numerical results are based on 5000 independent samples of $Y(t)$ with 5000 time steps $\Delta\tau = 0.0002$. The correlation

function and the marginal density of $Y(t)$ are estimated from $c_Y(s, t) = \sum_{i=1}^{5000} \xi_i(s)\xi_i(t)/5000$ and (4.2) for an arbitrary time t , where $\xi_i(t)$, $i = 1, \dots, 5000$ denote the samples of $Y(t)$. The marginal distribution of the diffusion process $X(t)$ is that of $Y(t)$ and its correlation function is selected to minimize the objective function (3.3) with $(w_1 = 1, w_2 = 0)$.

Two types of FD models have been constructed. They have the same basis functions, the top d eigenfunctions $\{\varphi_k(t)\}$ of the estimated correlation function of $X(t)$. The first FD model $X_d(t)$ is that of (2.12) with $n = 1$, i.e., $X_d(t) = \sum_{k=1}^d Z_k \varphi_k(t)$ with $Z_k(\omega) = \int_0^t X(t, \omega) \varphi_k(t) dt$. The second FD model $X_d^{\text{PCT}, m}(t) = \sum_{k=1}^d Z_k^{\text{PCT}} \varphi_k(t)$ represents the random coefficients $\{Z_k^{\text{PCT}}\}$ by PCT models, where m denotes the degree of the underlying PC model, see [1] for details.

The left panel of Fig. 6 shows the correlation function of the target process $Y(t)$ (solid line) and the estimated correlation function of $X(t)$ (dotted line). The right panel of Fig. 6 shows with solid line in semi-logarithmic scale an estimate of the target probability $P(\sup_{t \in [0, \tau]} |Y(t)| > x)$. The dotted, thin dashed, heavy dashed and dotted-dashed lines are the corresponding probabilities of the diffusion processes $X(t)$, its FD models $X_d(t)$, the PCT models $X_d^{\text{PCT}, 2}(t)$ and $X_d^{\text{PCT}, 4}(t)$ for $d = 15$.

The probability $P(\sup_{t \in [0, \tau]} |Y(t)| > x)$ has an abrupt change of slope at $x = 2$ since

$$P\left(\sup_{t \in [0, \tau]} |Y(t)| > x\right) = P\left(\sup_{t \in [0, \tau]} Y(t) < -x\right) + P\left(\sup_{t \in [0, \tau]} Y(t) > x\right)$$

and $Y(t)$ is supported on $[-2, \infty]$, the above two probabilities have different tails and $P(\sup_{t \in [0, \tau]} Y(t) < -x) = 0$ when $x \geq 2$.

4.3. Translation Gaussian diffusion processes

We fit the diffusion process $X_T(t)$ of (3.4) to the wind pressure records at the taps 120 and 126, construct FD models for this process and compare extremes of $X_T(t)$ with those of their FD models $X_{T,d}(t)$

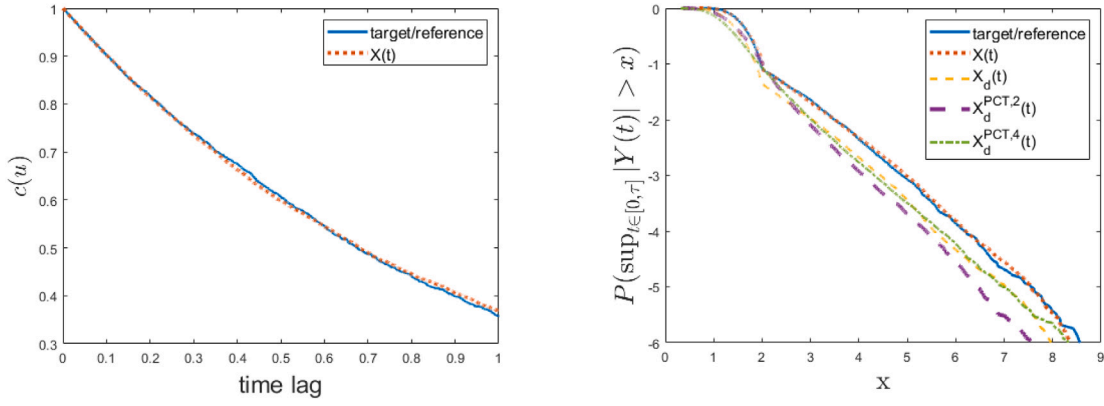


Fig. 6. Left panel: Estimated correlation functions of the target process $Y(t)$ (solid line) and $X(t)$ (dotted line). Right panel: Estimate of the target probability $P(\sup_{t \in [0, \tau]} |Y(t)| > x)$ (solid line) and the corresponding probabilities of $X(t)$ (dotted line), $X_d(t)$ for $d = 15$ (thin dashed line), $X_d^{\text{PCT},2}(t)$ (heavy dashed line) and $X_d^{\text{PCT},4}(t)$ (dotted-dashed line) for $d = 15$ in logarithmic scale.

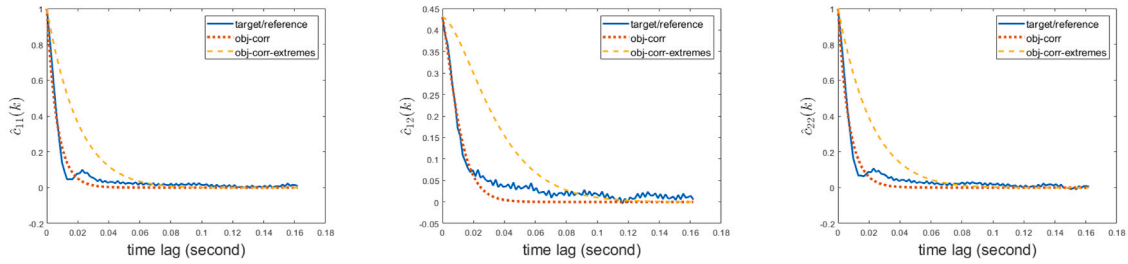


Fig. 7. Correlation functions $\hat{c}_{11}(k)$, $\hat{c}_{12}(k)$ and $\hat{c}_{22}(k)$ for $k = 1, \dots, 100$ (left, middle and right panels) (solid lines) and the corresponding correlation functions of $X_T(t)$ obtained under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted lines), and $(w_1 = 1, w_2 = 50)$ (dashed lines).

4.3.1. Process fitted to wind data

The Gaussian image of the translation processes of (3.4) for the wind records at taps 120 and 126 is constructed by using the objective function in (3.3) with $(w_1 = 1, w_2 = 0)$ and $(w_1 = 1, w_2 = 50)$.

The left, middle and right panels of Fig. 7 show the target correlation functions $\hat{c}_{11}(k)$, $\hat{c}_{12}(k)$ and $\hat{c}_{22}(k)$ for $k = 1, \dots, 100$ estimated from wind records at taps 120 and 126 (solid lines) and the corresponding correlation functions of $X_T(t)$ obtained under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted lines) and with $(w_1 = 1, w_2 = 50)$ (dashed lines).

The left and right panels of Fig. 8 show with solid lines in semi logarithmic scale the estimates of the target probability $P(\sup_{t \in [0, \tau]} |X_i(t)| > x)$ for taps 120 and 126 based on the wind records. The dotted and dashed lines are the corresponding probabilities of the translation diffusion process $X_T(t)$ defined under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ and $(w_1 = 1, w_2 = 50)$. The extremes of $X_T(t)$ corresponding to the objective function with $(w_1 = 1, w_2 = 50)$ are superior although its correlation functions differ notable from the target correlations.

We conclude with the observation that the translation diffusion process $X_T(t)$ in (3.4) can fit exactly the marginal distribution of the target process $X_i(t)$ and approximately its correlation function. However, its extremes may or may not match target extremes depending on the objective function used to select the correlation parameter α of $X_T(t)$. Correlation functions rather different from the target correlation may yield accurate estimates of extremes. This is another example of model which provides accurate estimates of a particular quantity of interest although it does not capture some target properties.

4.3.2. Finite dimensional (FD) models

We develop FD models $X_{T,d}(t)$ for the process $X_T(t)$ constructed for the taps 120 and 126 under the objective function (3.3) with $(w_1 = 1, w_2 = 50)$. The FD models $X_{T,d}(t)$ are defined by (3.9) with $Y_{i,d_i}(t) =$

$G_{i,d_i}(t)$, where $G_{i,d_i}(t)$ defined by (2.12) for $n = 1$ is the FD model of the Gaussian process $G_i(t)$. Since the input $G_i(t)$ is Gaussian has continuous correlation function and continuous samples, $\sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)|$ converges to $\sup_{t \in [0, \tau]} |X_{T,i}(t)|$ in probability as $d_i \rightarrow \infty$ by Corollary 3.1. This means that the distribution of $\sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)|$ can be estimated from samples of $X_{T,i}(t)$ for a sufficiently large d_i .

The left and right panels of Fig. 9 show scatter plots of $(\sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)|, \sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)|)$, $i = 1, 2$ for $d_i = 50$ and 200. The solid lines in the left and right panels of Fig. 10 are estimates of $P(\sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)| > x)$ which are obtained directly from samples of $X_{T,i,d_i}(t)$. These probabilities are viewed as reference. The other lines of the figure are calculated from FD models for $d = 50$ (dotted lines), $d = 200$ (dashed lines). The dashed lines are the closest to the reference. These plots show, in agreement with our theoretical results, that the discrepancy between the distributions of extremes of $X_{T,i,d_i}(t)$ and $X_{T,i}(t)$ decreases with stochastic dimensions d_i , $i = 1, 2$.

4.4. Translation non-Gaussian diffusion processes

Translation models in Section 4.3 can be extended to map non-Gaussian diffusion processes rather the Gaussian diffusion processes. The following two subsections show results for one- and two-dimensional translation non-Gaussian diffusion processes. We note that FD models for these processes can be constructed by the approach of Section 4.3 and are reported here.

4.4.1. One dimensional processes

Consider the one-dimensional translation diffusion processes

$$X_T^{(i)}(t) = \hat{F}^{-1} \circ F_{Y^{(i)}}(Y^{(i)}(t)), \quad i = 1, 2, \quad (4.3)$$

where $\hat{F}$ is the distribution of the estimated marginal density in (4.2) for the wind record at the tap 120, $Y^{(i)}(t)$ is the non-Gaussian diffusion

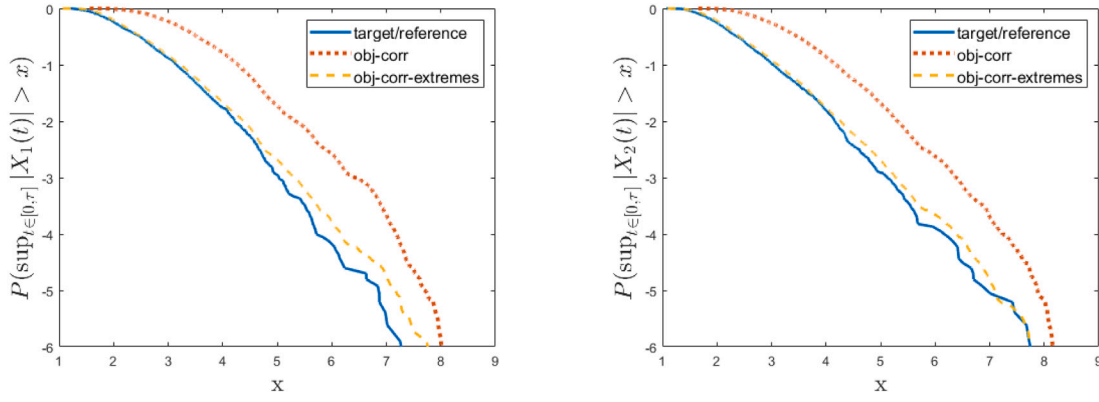


Fig. 8. Estimates of the target probability $P(\sup_{t \in [0, \tau]} |X_i(t)| > x)$ for taps 120 (left panel) and 126 (right panel) based on the wind records (solid lines), the corresponding probabilities of the translation diffusion process $X_T(t)$ defined under the objective function (3.3) with $(w_1 = 1, w_2 = 0)$ (dotted lines), and $(w_1 = 1, w_2 = 50)$ (dashed lines) in logarithmic scale.

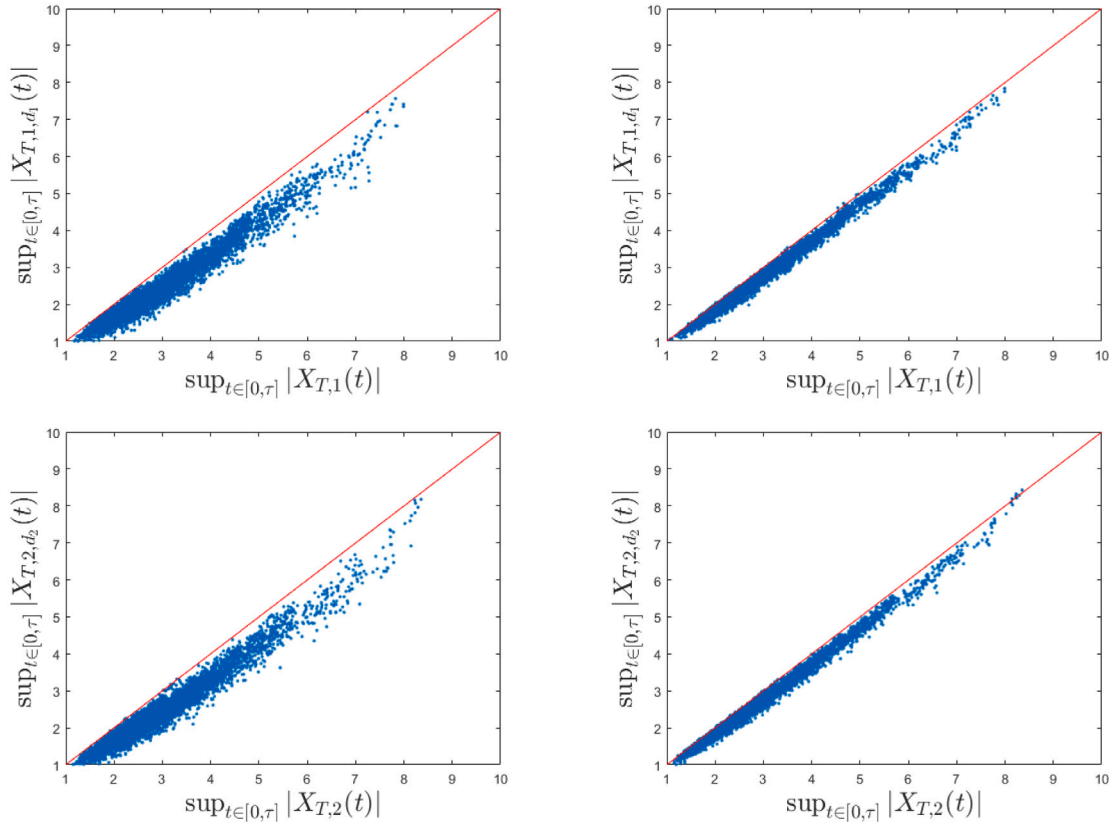


Fig. 9. Scatter plots of $\sup_{t \in [0, \tau]} |X_{T,i}(t)|$ and $\sup_{t \in [0, \tau]} |X_{T,i,d_i}(t)|$ for $d_i = 50, 200$ (left and right panels) and $i = 1, 2$ (upper and lower panels).

processes of (3.5) with diffusion coefficients in (3.6), $\theta = 1$ and $F_{Y^{(i)}}(t)$ is the empirical distribution estimated from samples of $Y^{(i)}(t)$, $i = 1, 2$.

We fit the one-dimensional translation diffusion processes $\{X_T^{(i)}(t)\}$ in (4.3) to the wind pressure record at the tap 120. The estimates of the correlation function and the marginal distribution of the wind record at the tap 120 are given by (4.1) and (4.2). The parameter ρ is selected to minimize the objective function (3.3) with $(w_1 = 1, w_2 = 0)$.

The left panel of Fig. 11 shows the scatter plot of $(\sup_{t \in [0, \tau]} |X_T^{(1)}(t)|, \sup_{t \in [0, \tau]} |X_T^{(2)}(t)|)$. The middle panel of Fig. 11 shows the correlation function $\hat{c}(k)$ for $k = 1, \dots, 100$ estimated from the wind record at tap 120 (solid line) and the correlation functions of the translation diffusion processes $\{X_T^{(i)}(t)\}$ (dotted and dashed lines).

The right panel of Fig. 11 shows with solid line in semi logarithmic scale an estimate of the target probability $P(\sup_{t \in [0, \tau]} |X(t)| > x)$ obtained from wind records at the tap 120. The dotted and dashed

lines are the corresponding probabilities of the translation diffusion processes $\{X_T^{(i)}(t)\}$.

We conclude with the observation that translation diffusion processes $\{X_T^{(i)}(t)\}$ based on the non-Gaussian, exponentially correlated diffusion processes of (3.5), which are equivalent in the sense that they have the marginal of the wind pressure time series recorded at a pressure tap and similar correlation functions, can have very different extremes. This is another illustration of the fact that extremes require information beyond the first two moments and marginal distributions.

4.4.2. Two dimensional processes

Consider the two-dimensional translation diffusion processes

$$X_{T,j}^{(i)}(t) = \hat{F}_j^{-1} \circ F_{Y_j^{(i)}}(Y_j^{(i)}(t)), \quad i, j = 1, 2 \quad (4.4)$$

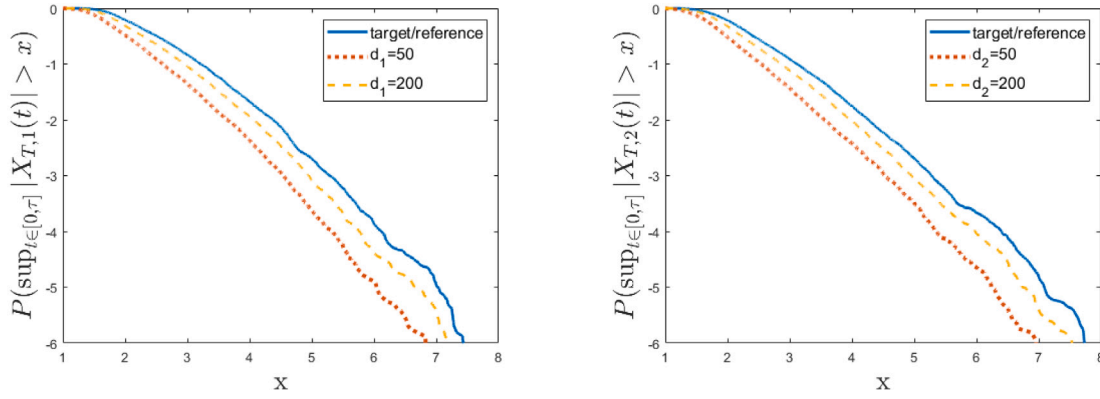


Fig. 10. Estimates of the target probability $P(\sup_{t \in [0, \tau]} |X_{T,j}(t)| > x)$ (solid lines) and estimates of $P(\sup_{t \in [0, \tau]} |X_{T,j,d_i}(t)| > x)$ for $d_i = 50$ (dotted lines) and $d_i = 200$ (dashed lines) in logarithmic scale for $X_{T,1}(t)$ and $X_{T,2}(t)$ (left and right panels).

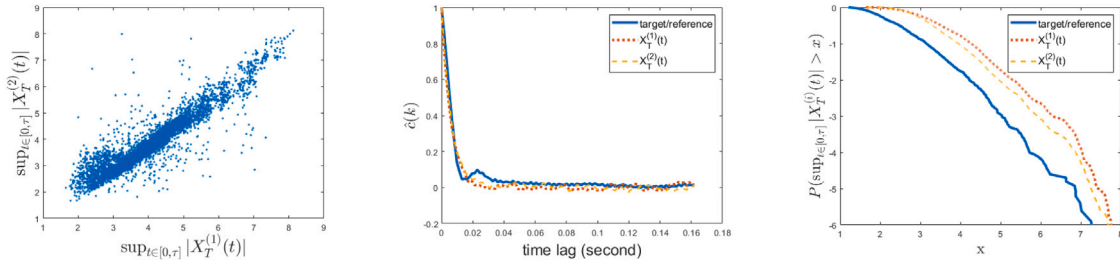


Fig. 11. Left panel: Scatter plots of $\sup_{t \in [0, \tau]} |X_T^{(1)}(t)|$ and $\sup_{t \in [0, \tau]} |X_T^{(2)}(t)|$. Middle panel: Correlation function $\hat{c}(k)$ for $k = 1, \dots, 100$ estimated from the wind record at tap 120 (solid line) and the correlation functions of $\{X_T^{(i)}(t)\}$ (dotted and dashed lines). Right panel: Estimate of the target probability $P(\sup_{t \in [0, \tau]} |X_T^{(j)}(t)| > x)$ at the tap 120 obtained from wind data (solid line) and the corresponding probabilities of $\{X_T^{(i)}(t)\}$ (dotted and dashed lines) in logarithmic scale.

where $\hat{F}_j$, $j = 1, 2$ are the distribution of the estimated marginal density in (4.2) for the wind records at the pressure taps 120 and 126, $Y_j^{(i)}(t)$ is the non-Gaussian diffusion processes of (3.7) with diffusion coefficients in (3.8) with $\theta = 3$ and $\zeta = 1$ and $F_{Y_j^{(i)}}$ is the empirical distribution estimated from samples of $Y_j^{(i)}(t)$, $i, j = 1, 2$. The process $\{Y_j^{(1)}(t)\}$ is Gaussian while $\{Y_j^{(2)}(t)\}$ is not.

We fit the two dimensional translation diffusion processes $\{X_{T,j}^{(i)}(t)\}$ in (4.4) to the wind records at the pressure taps 120 and 126. The estimates of the correlation function and the marginal distribution of the wind records at the pressure taps 120 and 126 are given by (4.1) and (4.2). Note that the correlation functions of $G(t)$ in (3.1) and $\{Y_j^{(i)}(t)\}$ in (3.7) are completely defined by the matrices $\alpha = [a_{11} \ a_{12}; a_{21} \ a_{22}]$ up to a scale parameters. The optimal values of these matrices minimize the objective function in (3.3) with the weighting coefficients ($w_1 = 1, w_2 = 50$).

The left and right panels of Figs. 12 show scatter plots of $(\sup_{t \in [0, \tau]} |X_{T,j}^{(1)}(t)|, \sup_{t \in [0, \tau]} |X_{T,j}^{(2)}(t)|)$ for $j = 1, 2$. The left, middle and right panels of Fig. 13 show the target correlation functions $\hat{c}_{11}(k)$, $\hat{c}_{12}(k)$ and $\hat{c}_{22}(k)$ for $k = 1, \dots, 100$ estimated from wind records at taps 120 and 126 (solid lines) and the corresponding correlation functions of the translation diffusion processes $\{X_{T,j}^{(i)}(t)\}$ (dotted and dashed lines).

The left and right panels of Fig. 14 show with solid lines in semi logarithmic scale the estimates of the target probability $P(\sup_{t \in [0, \tau]} |X_i(t)| > x)$ obtained from wind records at taps 120 and 126. The dotted and dashed lines are the corresponding probabilities of the translation diffusion processes $\{X_{T,j}^{(i)}(t)\}$.

We conclude with the observation that translation diffusion processes $\{X_{T,j}^{(i)}(t)\}$ based on the non-Gaussian diffusion processes of (3.7), which are equivalent in the sense that they have the marginal of the wind pressure time series recorded at two pressure taps and similar correlation functions, can have very different extremes. This is another

illustration of the fact that extremes require information beyond the first two moments and marginal distributions.

5. Conclusions

Two classes of random processes are constructed to describe wind pressures recorded at the University of Florida boundary layer wind tunnel facility (UFBWLWT). The first consists of diffusion processes with linear drift. The functional form of the correlation functions of these processes is fixed, e.g., exponential correlation for one-dimensional processes. The second class consists of translations of Gaussian/non-Gaussian diffusion processes. These processes can characterize a broader set of wind data than those of the first class when dealing with multivariate time series.

It is shown that the processes of both classes can describe accurately the extremes of wind pressure records if their correlation functions are selected to minimize objective functions quantifying the discrepancy between correlations and extremes of these processes and of wind pressure records. It is also shown that the finite dimensional (FD) models developed for these processes can be used to estimate extreme wind pressures. The FD models are deterministic functions of time and finite sets of random variables.

Diffusion processes with linear drift and translations of Gaussian/non-Gaussian diffusion processes are fitted to wind pressure time series recorded at the University of Florida boundary layer wind tunnel facility (UFBWLWT). The processes match exactly and approximately the marginal distributions and the correlation functions of these records. It is shown that (1) the extreme of these simple processes characterize accurately the extremes of the wind record if their correlation functions minimize objective functions which account for extremes and (2) the finite dimensional (FD) models of these processes can be used to estimate extremes of the wind records.

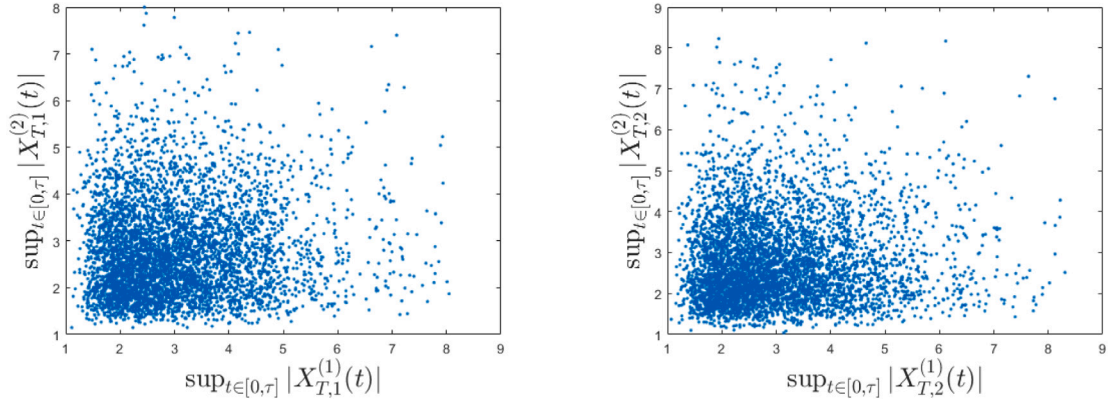


Fig. 12. Scatter plots of $\sup_{t \in [0, \tau]} |X_{T,j}^{(1)}(t)|$ and $\sup_{t \in [0, \tau]} |X_{T,j}^{(2)}(t)|$ for $j = 1, 2$ (left and right panels).

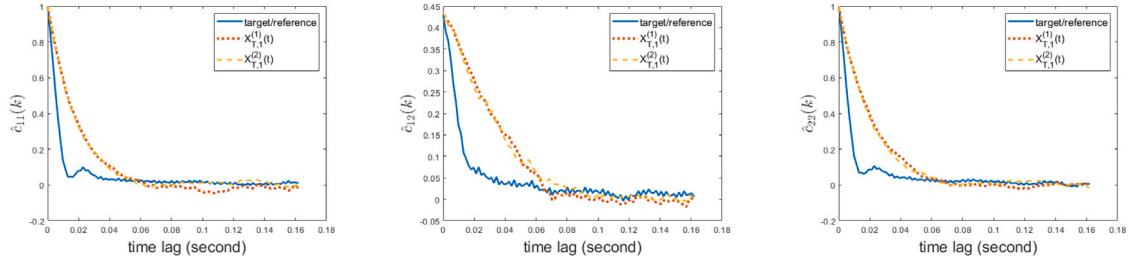


Fig. 13. Correlation functions $\hat{c}_{11}(k)$, $\hat{c}_{12}(k)$ and $\hat{c}_{22}(k)$ for $k = 1, \dots, 100$ (left, middle and right panels) (solid lines) and the correlation functions of $\{X_{T,j}^{(i)}(t)\}$ (dotted and dashed lines).

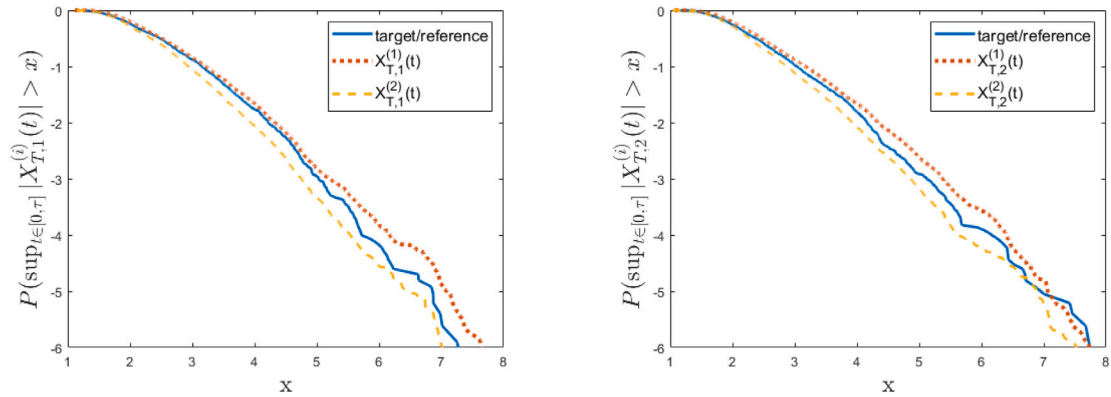


Fig. 14. Estimates of the target probability $P(\sup_{t \in [0, \tau]} |X_i(t)| > x)$ for taps 120 (left panel) and 126 (right panel) based on the wind records (solid lines) and the corresponding probabilities of $\{X_{T,j}^{(i)}(t)\}$ (dotted and dashed lines) in logarithmic scale.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work reported in this paper has been partially supported by the National Science Foundation, United States under the grant CMMI-2013697. This support is gratefully acknowledged.

References

- [1] H. Xu, M. Grigoriu, A novel surrogate for extremes of random functions, *Reliab. Eng. Syst. Saf.* (2023) <http://dx.doi.org/10.1016/j.ress.2023.109493>.
- [2] G.Q. Cai, Y.K. Lin, Generation of non-Gaussian stationary stochastic processes, *Phys. Rev. E* 54 (1) (1996) 299.
- [3] G.Q. Cai, C. Wu, Modeling of bounded stochastic processes, *Probab. Eng. Mech.* 19 (3) (2004) 197–203.
- [4] M. Grigoriu, *Stochastic calculus, Applications in Science and Engineering*, Birkhäuser, Boston, 2002.
- [5] R.W. Brockett, *Finite Dimensional Linear Systems*, John Wiley & Sons Inc., New York, 1970.
- [6] H. Xu, M. Grigoriu, Extremes of vector-valued processes by finite dimensional models, *Methodol. Comput. Appl. Probab.* (2023) (submitted for publication).
- [7] R.G. Ghanem, P.D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Courier Corporation, 2003.
- [8] I. Gohberg, S. Goldberg, *Basic Operator Theory*, Birkhäuser, 2013.
- [9] M.G.R. Faes, M. Broggi, P.D. Spanos, M. Beer, Elucidating appealing features of differentiable auto-correlation functions: A study on the modified exponential kernel, *Probab. Eng. Mech.* 69 (2022) 103269.
- [10] P.D. Spanos, M. Beer, J. Red-Horse, Karhunen–Loève expansion of stochastic processes with a modified exponential covariance kernel, *J. Eng. Mech.* 133 (7) (2007) 773–779.
- [11] M. Grigoriu, *Applied Non-Gaussian Processes: Examples, Theory, Simulation, Linear Random Vibration, and MATLAB Solutions*, Prentice Hall, Inc, Englewood Cliffs, NJ, 1995.
- [12] M. Grigoriu, Translation models revisited, *Prob. Eng. Mech.* (48) (2017) 68–75.
- [13] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, London, ISBN: 978-0-412-24620-3, 1986, p. 45.