



A novel surrogate for extremes of random functions

Hui Xu ^{a,*}, Mircea D. Grigoriu ^{a,b}, Kurtis R. Gurley ^c

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

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

^c Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL, USA

ARTICLE INFO

MSC:

60

62

Keywords:

Extremes

Finite dimensional (FD) models

Polynomial chaos (PC)

Polynomial chaos translation (PCT)

FD surrogates

ABSTRACT

Numerical solutions of stochastic problems require the representation of random functions in their definitions by finite dimensional (FD) models, i.e., deterministic functions of time and finite sets of random variables. It is common to represent the coefficients of these FD surrogates by polynomial chaos (PC) models. We propose a novel model, referred to as the polynomial chaos translation (PCT) model, which matches exactly the marginal distributions of the FD coefficients and approximately their dependence. PC- and PCT-based FD models are constructed for a set of test cases and a wind pressure time series recorded at the boundary layer wind tunnel facility at the University of Florida. The PCT-based models capture the joint distributions of the FD coefficients and the extremes of target time series accurately while PC-based FD models do not have this capability.

1. Introduction

Most records of time series are insufficient to estimate the distributions of extremes and other functionals of these records over bounded time intervals. Available data need to be supplemented by probabilistic models which deliver these distributions under assumptions that are physically acceptable and can be validated. For example, the extreme value theory may be unsatisfactory in many applications since it assumes that the time series is infinite while the reference time is bounded [1]. In addition, extreme value estimates based on experimental data of typical duration exhibit notable statistical variability when fitted to insufficiently long records [2,3]. Also, the validity of approximations of the distributions of extremes of time series based on the mean rates at which they cross specified levels is questionable since the existence of mean crossing rates is difficult to check for non-Gaussian processes [4] (Theorem 7.2.4). These limitations justify the need for robust numerical methods for characterizing extremes of random functions. Their implementation requires the representation of the target random functions by finite dimensional (FD) models, i.e., deterministic functions of time and finite sets of random variables.

The samples of FD models are elements of linear spaces spanned by specified deterministic basis functions/vectors with random coefficients. The form of the FD models in this study is that of truncated Karhunen-Loève (KL) series for random processes and vectors, i.e., linear forms of eigenfunctions/eigenvectors of the correlation function/matrix of the target random processes/time series. In contrast to standard KL representations whose random coefficients are defined

partially by their first two moments for non-Gaussian time series, the random coefficients of FD models are specified by their samples, which are obtained by projecting target samples on basis functions/vectors. The resulting random coefficients have the same first two moments as those of the KL representations.

It is common to represent the random coefficients of FD models by truncated polynomial chaos (PC) series fitted to the available information, which may consist of data or statistics. PC-based FD models have been used extensively in applications to, e.g., construct efficient kriging-based surrogates [5,6], estimate reliability via subset simulation [6], quantify the uncertainty in the eigenvalues and eigenfunctions of the Boltzman stochastic differential equation [7]. Algorithms based on λ -distributions [8], sequential sampling [9,10] and machine learning [11] have been developed to improve the accuracy and efficiency of PC-based FD models. Other available algorithms for constructing efficient and accurate PC-based FD models use distribution-free PC expansion [12], resampling techniques [13], Maratantola transformation [14], fractional moments [15], and active learning functions [16].

A common feature of the existing PC-based FD models is that their performance is assessed by the mean square error of the discrepancy between target random functions/vectors and these models. This metric is adequate for a broad range of applications which deal with both epistemic and aleatoric uncertainties [17,18]. Moreover, it can be shown that under mild conditions the PC-based FD models of target random functions converge to these functions in mean square

* Corresponding author.

E-mail addresses: hx223@cornell.edu (H. Xu), mdg12@cornell.edu (M.D. Grigoriu), kgurl@ce.ufl.edu (K.R. Gurley).

(m.s.) as their truncation level increases indefinitely [19, Chap. 9]. This convergence implies that the joint distributions of the PC-based FD models converge to those of the target functions. Unfortunately, this asymptotic result is irrelevant in applications since the truncation levels of PC-based FD models must be kept low to limit calculations so that, generally, the joint distributions of target random functions and their PC-based FD models differ. This observation is relevant for large dimensional random vectors, e.g., wind pressure time series recorded in wind tunnels, since their KL representations have large numbers of terms. To minimize calculations, they need to be truncated so that the target and KL statistics differ.

A main contribution of the proposed model is that it matches exactly the marginal distributions of the target random vector for any truncation level of the underlying PC polynomial. The dependence structure of the proposed PC-based FD model is captured approximately and improves with the PC truncation level. These features are essential for estimating extremes of random functions and vectors and are not available in the current FD models.

We propose an alternative model for the coefficients of FD representations of random functions, referred to as the polynomial chaos translation (PCT) model, which is constructed as follows. Let ξ_d^{PC} be a PC model of an $\mathbb{R}^n$ -valued random vector ξ , where d denotes the stochastic dimension of the FD model. The PC model belongs to the linear space spanned by, e.g., Hermite polynomials of the independent standard Gaussian variables $(G_1, \dots, G_d)$ whose degrees are limited by the selected truncation level. The components of the PCT models ξ_d^{PCT} are obtained from those of the PC models by translation [20] (Chap. 3) so that their distributions match exactly the target marginal distributions for any PC coefficients. The optimal PCT coefficients minimize the discrepancy between the dependence of the target and PCT components. Joint distributions and spectral measures are used as metrics for the dependence among ξ and ξ_d^{PCT} components.

The paper is organized as follows. The proposed FD models are defined in Section 2. This section also outlines properties of these models and examines their performance for various basis vectors. PC and PCT models of the random coefficients of FD models are constructed and examined in Section 3. The performance of FD models with PC and PCT coefficients is evaluated in Section 4 for wind pressure records from the UFBLWT facility. It is found that FD models with PCT-based random coefficients are superior in the sense that they best describe extremes of wind pressure time series. Final comments are in Section 5.

2. Bases for finite dimensional (FD) models

Consider a stationary ergodic time series $Y_1, Y_2, \dots, Y_n$ with the time step Δt , which is defined on a probability space $(\Omega, \mathcal{F}, P)$. For example, the wind pressure record in Section 4 can be viewed as a sample of this vector. Generally, the dimension n of the random vector $\xi = (Y_1, Y_2, \dots, Y_n)$ is very large. Our objective is to develop representations of ξ which are accurate and depend on just $d \ll n$ random variables. We refer to these representations as FD models. We recognize that the concept of FD models can be confusing in this context since the target ξ is also finite dimensional. This is not the case when dealing with random functions since their KL representations have infinite numbers of terms. We agree to call FD models the representations of ξ which depend on $d < n$ random variables. The following two subsections present two types of FD models whose relative performance is assessed in the third subsection.

Throughout this paper, we use $\|\cdot\|_2$ to denote the L_2 norm, defined by $\|x\|_2 = (\sum_{i=1}^l x_i^2)^{1/2}$ and $\|A\|_2 = \sup_{u \neq 0} \frac{\|Au\|_2}{\|u\|_2}$, where x is a l dimensional vector and A is a $l \times r$ matrix. Similarly, $\|\cdot\|$ means the L_∞ norm, i.e., $\|x\| = \max_{1 \leq i \leq l} |x_i|$ and $\|A\| = \max_{1 \leq i \leq l, 1 \leq j \leq r} |A_{ij}|$. The two norms are equivalent since we deal with matrices and vectors in finite dimensional linear spaces, see [21], Chap. 3 for technical details.

2.1. Covariance eigenvectors

Let ξ be an n -dimensional (column) real-valued random vector defined on a probability space $(\Omega, \mathcal{F}, P)$ with mean $E[\xi] = \mathbf{0}$ and covariance matrix $\gamma = E[\xi \xi^T]$. The assumption $E[\xi] = \mathbf{0}$ is not restrictive since, if the mean is not zero, it can be added to the samples of ξ . It is assumed that (1) the covariance matrix γ is not singular so that its eigenvalues $\{\lambda_k\}$ are strictly positive and (2) the eigenvalues $\{\lambda_k\}$ are distinct so that the eigenvectors $\{v_k\}$ of γ span $\mathbb{R}^n$. The random vector ξ admits the Karhunen-Loève (KL) representation

$$\xi_{\text{KL}} = \sum_{k=1}^n Z_{\text{KL},k} v_k, \quad (2.1)$$

where the eigenvectors $\{v_k\}$ are orthonormal, i.e., $v_k^T v_l = \delta_{kl}$, and $\{Z_{\text{KL},k}\}$ are uncorrelated random variables with mean $E[Z_{\text{KL},k}] = 0$ and variance $E[Z_{\text{KL},k}^2] = \lambda_k$, $k = 1, \dots, n$. The random vector ξ_{KL} is partially defined by its first two moments which match the corresponding moments of the target vector ξ , unless ξ is Gaussian in which case $\{Z_{\text{KL},k}\}$ are independent Gaussian variables. If ξ is not Gaussian, $\{Z_{\text{KL},k}\}$ are uncorrelated but dependent non-Gaussian variables with unknown distributions.

The KL representation of (2.1) can be generalized to characterize fully ξ if the random coefficients in the expression of ξ_{KL} are defined by the projections of the samples $\xi(\omega)$ of ξ on the eigenvectors of the covariance matrix γ of this vector, i.e., $Z_k(\omega) = \langle \xi(\omega), v_k \rangle = \xi(\omega)^T v_k$. Accordingly, the samples $\xi(\omega)$ of ξ admit the representation

$$\xi(\omega) = \sum_{k=1}^n Z_k(\omega) v_k, \quad \omega \in \Omega, \quad (2.2)$$

which holds for almost all samples of ξ .

If the dimension n of ξ is large, it is convenient to represent ξ by truncated versions of (2.2) which contains the top $d < n$ eigenvectors of γ , i.e., the eigenvectors corresponding to the largest d eigenvalues of the covariance matrix of ξ . The representation has the form

$$\xi_d(\omega) = \sum_{k=1}^d Z_k(\omega) v_k, \quad \omega \in \Omega, \quad (2.3)$$

and is referred to as the finite dimensional (FD) model of ξ . Since ξ_d depends on d random variables, its stochastic dimension is d . The L_2 norm of the error of the FD model is

$$\|\xi - \xi_d\|_2^2 = (\xi - \xi_d)^T (\xi - \xi_d) = \sum_{k=d+1}^n \sum_{l=d+1}^n Z_k Z_l v_k^T v_l = \sum_{k=d+1}^n Z_k^2, \quad a.s.,$$

by the orthonormality of the eigenvectors $\{v_k\}$, where *a.s.* means that the probability of the subset of the sample space Ω on which the random variables $\|\xi - \xi_d\|_2^2$ and $\sum_{k=d+1}^n Z_k^2$ differ is zero. Its mean and variance are

$$\begin{aligned} E[\|\xi - \xi_d\|_2^2] &= \sum_{k=d+1}^n E[Z_k^2] = \sum_{k=d+1}^n E[(\xi^T v_k)^T (\xi^T v_k)] \\ &= \sum_{k=d+1}^n v_k^T \gamma v_k = \sum_{k=d+1}^n \lambda_k \end{aligned} \quad (2.4)$$

and

$$\begin{aligned} \text{Var}[\|\xi - \xi_d\|_2^2] &= \text{Var}\left[\sum_{k=d+1}^n Z_k^2\right] = \sum_{k=d+1}^n \text{Var}[Z_k^2] = \sum_{k=d+1}^n \text{Var}[(\xi^T v_k)^T \xi^T v_k] \\ &= \sum_{k=d+1}^n \left(E[v_k^T \xi \xi^T v_k v_k^T \xi \xi^T v_k] - (E[v_k^T \xi \xi^T v_k])^2 \right) \\ &= \sum_{k=d+1}^n v_k^T \left(E[\xi \xi^T v_k v_k^T \xi \xi^T] - \gamma v_k v_k^T \gamma \right) v_k. \end{aligned} \quad (2.5)$$

Note that the expectation of the error decreases with d since $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ by assumption so that there exists d such that the variance can be made as small as desired.

2.2. Linearly independent vectors

Let $\{\tilde{v}_k\}$ be a set of n orthonormal n -dimensional vectors so that they are linearly independent and span $\mathbb{R}^n$. Since the samples of ξ are elements of $\mathbb{R}^n$ and the vectors $\{\tilde{v}_k\}$ span this space, we have

$$\xi(\omega) = \sum_{k=1}^n \tilde{Z}_k(\omega) \tilde{v}_k, \quad \omega \in \Omega, \quad (2.6)$$

where $\tilde{Z}_k(\omega) = \xi(\omega)^T \tilde{v}_k$ are the projections of $\xi(\omega)$ on these vectors, see (2.2). The corresponding FD model has the form

$$\tilde{\xi}_d(\omega) = \sum_{k=1}^d \tilde{Z}_k(\omega) \tilde{v}_k, \quad \omega \in \Omega, \quad (2.7)$$

where $\tilde{v}_k$, $k = 1, \dots, d$, is an arbitrary subset of the vectors under consideration. The L_2 norm of the error of $\tilde{\xi}_d$ is

$$\|\xi - \tilde{\xi}_d\|_2^2 = (\xi - \tilde{\xi}_d)^T (\xi - \tilde{\xi}_d) = \sum_{k=d+1}^n \sum_{l=d+1}^n \tilde{Z}_k \tilde{Z}_l \tilde{v}_k^T \tilde{v}_l = \sum_{k=d+1}^n \tilde{Z}_k^2, \quad a.s., \quad (2.8)$$

by the orthonormality of the vectors $\{\tilde{v}_k\}$. The definition of $\tilde{Z}_k$ in (2.7) gives

$$\begin{aligned} E[\|\xi - \tilde{\xi}_d\|_2^2] &= \sum_{k=d+1}^n E[\tilde{Z}_k^2] = \sum_{k=d+1}^n E[(\xi^T \tilde{v}_k)^T (\xi^T \tilde{v}_k)] \\ &= \sum_{k=d+1}^n \tilde{v}_k^T \gamma \tilde{v}_k = \sum_{k=d+1}^n \left(\sum_{i=1}^n b_{ki}^2 \lambda_i \right) \end{aligned} \quad (2.9)$$

and

$$\begin{aligned} \text{Var}[\|\xi - \tilde{\xi}_d\|_2^2] &= \text{Var} \left[\sum_{k=d+1}^n \tilde{Z}_k^2 \right] \\ &= \sum_{k=d+1}^n \text{Var}[\tilde{Z}_k^2] = \sum_{k=d+1}^n \text{Var}[(\xi^T \tilde{v}_k)^T \xi^T \tilde{v}_k] \\ &= \sum_{k=d+1}^n \left(E[\tilde{v}_k^T \xi \xi^T \tilde{v}_k \tilde{v}_k^T \xi \xi^T \tilde{v}_k] - (E[\tilde{v}_k^T \xi \xi^T \tilde{v}_k])^2 \right) \\ &= \sum_{k=d+1}^n \tilde{v}_k^T \left(E[\xi \xi^T \tilde{v}_k \tilde{v}_k^T \xi \xi^T] - \gamma \tilde{v}_k \tilde{v}_k^T \gamma \right) \tilde{v}_k, \end{aligned} \quad (2.10)$$

where $\tilde{v}_k = \sum_{i=1}^n b_{ki} v_i$ holds since $\tilde{v}_k \in \mathbb{R}^n$ and $\{v_i\}$ is a basis of this space. If $\tilde{v}_k = v_k$, then $b_{ki} = \delta_{ki}$ and $\tilde{v}_k^T \gamma \tilde{v}_k = \lambda_k$.

2.3. Comparison of the two FD models

This section assesses the accuracy of the FD models ξ_d in (2.3) and $\tilde{\xi}_d$ in (2.7) via the L_2 norms $\|\xi - \xi_d\|_2^2$ and $\|\xi - \tilde{\xi}_d\|_2^2$. The following theorem shows that ξ_d is superior to any other FD models in the sense of the L_2 norm. The proof of this known result can be found in [22] and is given here for convenience.

Theorem 2.1. *If the covariance matrix of ξ is positive definite and has distinct eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$, then its eigenvectors $\{v_k\}$, $k = 1, \dots, n$, are optimal in the sense that $E[\|\xi - \xi_d\|_2^2] \leq E[\|\xi - \tilde{\xi}_d\|_2^2]$ for ξ_d in (2.3) and $\tilde{\xi}_d$ in (2.7).*

Proof. The difference between the errors of the FD models ξ_d and $\tilde{\xi}_d$ can be expressed in the form, see (2.4) and (2.9),

$$E[\|\xi - \tilde{\xi}_d\|_2^2] - E[\|\xi - \xi_d\|_2^2] = \sum_{k=d+1}^n \left(\sum_{i=1}^n b_{ki}^2 \lambda_i - \lambda_k \right),$$

where $\sum_{i=1}^n b_{ki}^2 = 1$, $k = 1, \dots, n$, since the vectors $\tilde{v}_k$ have unit length so that $1 = \|\tilde{v}_k\|_2^2 = \|\sum_{i=1}^n b_{ki} v_i\|_2^2 = \sum_{i,j=1}^n b_{ki} b_{kj} v_i^T v_j = \sum_{i=1}^n b_{ki}^2$. The assumed orthogonality of the vectors $\{\tilde{v}_k\}$ implies $0 = \tilde{v}_k^T \tilde{v}_k = (\sum_{i=1}^n b_{ki} v_i)^T (\sum_{j=1}^n b_{lj} v_j) = \sum_{i=1}^n b_{ki} b_{li}$. Note that

$$\sum_{k=d+1}^n \left(\sum_{i=1}^n b_{ki}^2 \lambda_i - \lambda_k \right) \geq \sum_{k=d+1}^n \left(\sum_{i=1}^{d+1} b_{ki}^2 \lambda_{d+1} + \sum_{i=d+2}^n b_{ki}^2 \lambda_i - \lambda_k \right)$$

Table 1

$E[\|\xi - \xi_d\|_2^2]$ and $E[\|\xi - \tilde{\xi}_d\|_2^2]$ for $d = 10, 30, 50$.

	d=10	d=30	d=50
$E[\ \xi - \xi_d\ _2^2]$	71.9224	37.1370	20.5518
$E[\ \xi - \tilde{\xi}_d\ _2^2]$	90.2117	70.3436	50.0748

$$\begin{aligned} &= \left(\sum_{k=d+1}^n \sum_{i=1}^{d+1} b_{ki}^2 - 1 \right) \lambda_{d+1} + \sum_{i=d+2}^n \sum_{k=d+1}^n b_{ki}^2 \lambda_i - \sum_{k=d+2}^n \lambda_k \\ &= \left(\sum_{k=d+1}^n \sum_{i=1}^{d+1} b_{ki}^2 - 1 \right) \lambda_{d+1} + \sum_{i=d+2}^n \left(\sum_{k=d+1}^n b_{ki}^2 - 1 \right) \lambda_i \\ &\geq \left(\sum_{k=d+1}^n \sum_{i=1}^{d+1} b_{ki}^2 - 1 \right) \lambda_{d+1} + \sum_{i=d+2}^n \left(\sum_{k=d+1}^n b_{ki}^2 - 1 \right) \lambda_{d+1} = 0 \end{aligned}$$

which implies that the eigenvectors $\{v_k\}$ provide a superior FD model in the sense of the error under consideration. $\square$

We conclude this subsection with an example illustrating that the FD models based on eigenvectors are superior to those based on arbitrary linearly independent vectors in the sense of Theorem 2.1.

Example 2.1. Let ξ be a zero-mean n -dimensional Gaussian vector with covariance matrix $\gamma = \{\gamma_{ij} = \rho^{|i-j|}, i, j = 1, \dots, n\}$, $0 < \rho < 1$. We construct the FD models ξ_d in (2.3) and $\tilde{\xi}_d$ in (2.7) with $\tilde{v}_{k,i} = \delta_{ki}$.

The following numerical results are for $n = 100$, $\rho = 0.5$ and 3000 samples of ξ . The left, middle and right panels of Fig. 1 show the histograms of $\|\xi - \xi_d\|_2^2$ for $d = 10, 30$ and 50 . Similar histograms are in Fig. 2 for $\|\xi - \tilde{\xi}_d\|_2^2$ and the same values of d . The plots show, in agreement with our theoretical results, that the discrepancy between the samples of ξ and ξ_d and between the samples of ξ and $\tilde{\xi}_d$ can be made as small as desired by increasing the stochastic dimension d . Note also that histograms of Fig. 1 have a smaller range than those of Fig. 2 and are closer to zero. Table 1 shows that the expected L_2 norm of the error for ξ_d is smaller than the corresponding error for $\tilde{\xi}_d$ for $d = 10, 30, 50$.

3. PC and PCT models

Consider the FD model ξ_d in (2.3) whose samples are elements of the linear space spanned by the top d eigenvectors of the covariance matrix of the target random vector ξ . The vector $Z = (Z_1, \dots, Z_d)$ of its random coefficients is non-Gaussian with dependent components unless ξ is Gaussian. Our objective is to construct models of Z from its samples which are **accurate** in the sense that their joint distributions match the joint distribution of ξ , and **efficient**, i.e., standard Monte Carlo algorithms can be used to generate samples of these models.

The Rosenblatt transformation [23] provides a model with these features. It shows that the components of $Z = (Z_1, \dots, Z_d)$ can be related to the components of, e.g., a vector $G = (G_1, \dots, G_d)$ with independent standard Gaussian variables, by the mapping

$$Z_1 = F_1^{-1} \circ \Phi(G_1)$$

$$Z_k | Z_{k-1}, \dots, Z_1 = F_{k|k-1, \dots, 1}^{-1} \circ \Phi(G_k), \quad 2 \leq k \leq d, \quad (3.1)$$

where F_k is the distribution of Z_k , $F_{k|k-1, \dots, 1}$ is the distribution of $Z_k | Z_{k-1}, \dots, Z_1$. If the mapping in (3.1) is available, samples of Z can be obtained from samples of G , which can be generated by standard algorithms. Since the conditional distributions in mapping $G \mapsto Z$ are available analytically only in special cases, they have to be constructed numerically in most applications. Their construction from the joint distribution of Z is computationally demanding and the resulting conditional distributions are likely to be unsatisfactory, particularly when dealing with heavy tail distributions. The construction of the conditional distributions $F_{k|k-1, \dots, 1}$ from data is not feasible when dealing with high dimensional vectors and relatively small data sets.

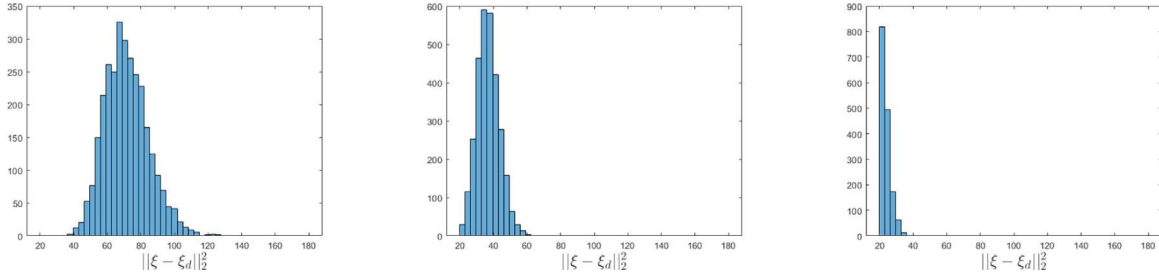


Fig. 1. Histograms of $\|\xi - \xi_d\|_2^2$ for $d = 10, 30, 50$ (left, middle and right panels) based on 3000 samples.

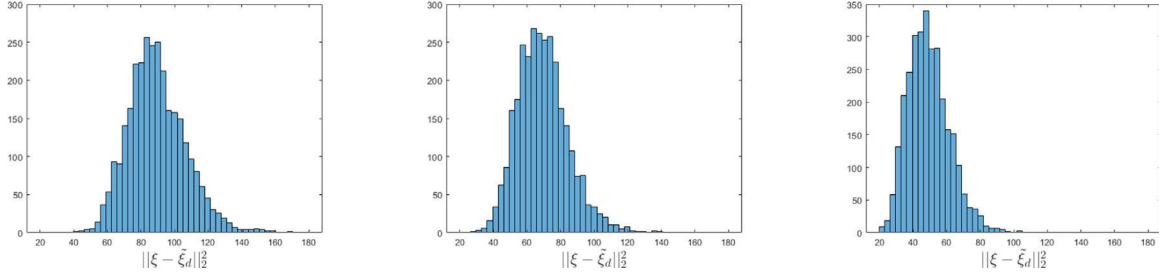


Fig. 2. Histograms of $\|\xi - \xi_d\|_2^2$ for $d = 10, 30, 50$ (left, middle and right panels) based on 3000 samples.

This section develops approximations of the Rosenblatt transformation for the random coefficients $(Z_1, \dots, Z_d)$ of the FD models in (2.3) based on polynomial chaos (PC) and an extension of this representation, referred to as PCT models. These models of $(Z_1, \dots, Z_d)$ are denoted by $Z^{PC} = (Z_1^{PC}, \dots, Z_d^{PC})$ and $Z^{PCT} = (Z_1^{PCT}, \dots, Z_d^{PCT})$.

3.1. Definitions

Unless stated otherwise, the PC models considered here are quadratic forms of independent standard Gaussian variables $G_1, \dots, G_d$ defined by

$$Z_k^{PC} = E[Z_k] + \sum_{j=1}^d a_{k,j} G_j + \sum_{1 \leq j < l \leq d} a_{k,j,l} (G_j G_l - \delta_{jl}), \quad k = 1, \dots, d, \quad (3.2)$$

where $\{a_{k,j}, a_{k,j,l}\}$ are yet undetermined coefficients. This form is used to limit the computational effort related to the determination of the unknown coefficients in the expressions of $\{Z_k^{PC}\}$. Note that $E[Z_k^{PC}] = E[Z_k]$ by construction, since $E[G_j] = 0$ and $E[G_j G_l - \delta_{jl}] = 0$ for any $1 \leq j < l \leq d$. The coefficients $\{a_{k,j}, a_{k,j,l}\}$ in (3.2) are determined by minimizing the objective function

$$\begin{aligned} e_1(a_{k,j}, a_{k,j,l}) = & g_1 E[\|Z - Z^{PC}\|_2^2] \\ & + g_2 \max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PC}(\cdot | a_{k,j}, a_{k,j,l})\|_2 \\ & + g_3 (\|E[Z Z^T] - E[Z^{PC} (Z^{PC})^T]\|) \end{aligned} \quad (3.3)$$

where $h_{i_1, i_2}(\cdot)$ is the histogram of (Z_{i_1}, Z_{i_2}) and $h_{i_1, i_2}^{PC}(\cdot | a_{k,j}, a_{k,j,l})$ is the histogram of $(Z_{i_1}^{PC}, Z_{i_2}^{PC})$ for given coefficients $\{a_{k,j}, a_{k,j,l}\}$. The Matlab function `histcounts2` is used to construct the two dimensional histograms of (Z_{i_1}, Z_{i_2}) and $(Z_{i_1}^{PC}, Z_{i_2}^{PC})$. The error between the two matrices is described by the norm $\|\cdot\|_2$, i.e., the absolute largest eigenvalue of the error matrix. We consider the set of all pairs of components rather than all components to minimize calculations. The weighting coefficients g_1, g_2, g_3 are such that the components $E[\|Z - Z^{PC}\|_2^2]$, $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PC}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\|E[Z Z^T] - E[Z^{PC} (Z^{PC})^T]\|$ contribute equally to the objective function (3.3). We set $g_1 = 0$ if Z and Z^{PC} are not defined on the same probability space since the mean squared error $E[\|Z - Z^{PC}\|_2^2]$ cannot be calculated.

The components of the PCT models are defined by

$$Z_k^{PCT} = F_k^{-1} \circ F_k^{PC}(Z_k^{PC}), \quad k = 1, \dots, d, \quad (3.4)$$

where F_k^{PC} is the distribution of Z_k^{PC} for given coefficients $\{a_{k,j}, a_{k,j,l}\}$ and F_k is the distribution of Z_k . In applications, F_k^{PC} and F_k are empirical distributions estimated from data, see Section 4.4. The coefficients $\{a_{k,j}, a_{k,j,l}\}$ in (3.4) are determined by minimizing the objective function

$$\begin{aligned} e_2(a_{k,j}, a_{k,j,l}) = & w_1 E[\|Z - Z^{PCT}\|_2^2] \\ & + w_2 \max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2 \\ & + w_3 \max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2, \end{aligned} \quad (3.5)$$

where $h_{i_1, i_2}(\cdot)$ is as in (3.3), $s_{i_1, i_2}(\cdot)$ is the spectral measure of (Z_{i_1}, Z_{i_2}) , $s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})$ and $h_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})$ are the spectral measure and the histogram of $(Z_{i_1}^{PCT}, Z_{i_2}^{PCT})$ for given coefficients $\{a_{k,j}, a_{k,j,l}\}$. Spectral measures of (Z_{i_1}, Z_{i_2}) are metrics which quantify the likelihood that (Z_{i_1}, Z_{i_2}) are simultaneously large, see (5.3) and (5.4) in [24] for definitions and [25], Chap.6 for technical details. We sort the samples of the two-dimensional vectors (Z_{i_1}, Z_{i_2}) and $(Z_{i_1}^{PCT}, Z_{i_2}^{PCT})$ according to their lengths such that the first sample is the furthest to the origin and construct the spectral measures from the top 10% of these samples. The Matlab function `histcounts2` is used to construct the two dimensional histograms and spectral measures of (Z_{i_1}, Z_{i_2}) and $(Z_{i_1}^{PCT}, Z_{i_2}^{PCT})$. We consider the set of all pairs of components rather than all components to minimize calculations. The weighting coefficients w_1, w_2, w_3 are such that the components $E[\|Z - Z^{PCT}\|_2^2]$, $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ contribute equally to the objective function (3.5). We set $w_1 = 0$ if Z and Z^{PCT} are not defined on the same probability space since the mean squared error $E[\|Z - Z^{PCT}\|_2^2]$ cannot be calculated. The second and third terms of $e_2(a_{k,j}, a_{k,j,l})$ quantify differences between the dependence structure of Z and Z^{PCT} . The third term is an approximate metric for the differences between the joint distributions of Z and Z^{PCT} while the second term measures the differences between the tail dependence of these random vectors.

3.2. PC versus PCT random coefficients

Consider the FD model ξ_d in (2.3) whose random coefficients $(Z_1, \dots, Z_d)$ are given by the PC and PCT models of the previous

subsection. They have the form

$$\xi_d^{PC} = \sum_{k=1}^d Z_k^{PC} v_k \quad \text{and} \quad \xi_d^{PCT} = \sum_{k=1}^d Z_k^{PCT} v_k \quad (3.6)$$

with Z_k^{PC} and Z_k^{PCT} in (3.2) and (3.4), respectively.

The PC-based FD models ξ_d^{PC} have been studied extensively [26–30]. They provide accurate approximations for the means and correlations of target random vectors, processes and fields. Since the PC vector $\{Z_k^{PC}\}$ converges in mean square to the target vector $\{Z_k\}$ as the truncation level increases indefinitely, the joint distribution of $\{Z_k^{PC}\}$ matches that of $\{Z_k\}$ for a sufficiently large truncation level [19, Chap. 9]. However, this argument does not hold in applications since, generally, the truncation levels for PC expansions have to be kept low to limit calculations, which increase exponentially with the truncation level [31].

The PCT-based FD models ξ_d^{PCT} are novel. Their defining mapping $\{G_k\} \mapsto \{Z_k^{PCT}\}$ is a nonlinear transformation of the defining mapping $\{G_k\} \mapsto \{Z_k^{PC}\}$ of PC models. The transformation guarantees that the marginal distributions of $\{Z_k^{PCT}\}$ match exactly the target marginal distributions for any values of the coefficients of the PC representation. The unknown coefficients of the mapping $\{G_k\} \mapsto \{Z_k^{PCT}\}$ are found by minimizing differences between joint properties of the components of $\{Z_k^{PCT}\}$ and $\{Z_k\}$, as discussed in the previous subsection. Accordingly, the resulting PCT models capture approximately the dependence between the components of this vector, in addition to matching exactly the marginal distributions of $\{Z_k\}$.

The practical implication of the differences between PC- and PCT-based FD models is illustrated in the subsequent sections. They show that samples of $\{Z_k^{PCT}\}$ and $\{Z_k\}$ are similar while those of $\{Z_k^{PC}\}$ and $\{Z_k\}$ differ significantly and that extremes of PCT-based FD models capture accurately extremes of target processes while PC-based FD models do not have this capability.

3.3. Numerical illustrations

The following three examples are presented to illustrate the implementation and accuracy of the PC and PCT models. The Rosenblatt map $G \rightarrow Z$ can be obtained analytically in the first two examples so that samples of Z can be paired with samples of Z^{PC} and Z^{PCT} since they are calculated from the same samples $\{G(\omega)\}$ of G . It is not possible to pair samples of Z with samples of Z^{PC} and Z^{PCT} in the last example since the Rosenblatt map is not available analytically.

Example 3.1. Let $M = (M_1, M_2) \in \mathbb{R}^2$ be a Gaussian vector with zero-mean, unit variance and correlation $E[M_1 M_2] = \rho$, $0 < \rho < 1$. Let $Z = (Z_1, Z_2)$ be a random vector, where $Z_k = F_k^{-1} \circ \Phi(M_k)$ with (i) F_k is uniform in $(0, 1)$, $k = 1, 2$ and (ii) F_k is exponential with parameters $\mu_k > 0$, $k = 1, 2$.

Note that the joint distribution of (Z_1, Z_2) and the conditional distribution of $Z_2|Z_1$ are

$$P(Z_1 \leq z_1, Z_2 \leq z_2) = P(M_1 \leq \Phi^{-1} \circ F_1(z_1), M_2 \leq \Phi^{-1} \circ F_2(z_2))$$

and

$$\begin{aligned} & F_{Z_2|Z_1=z_1}(z_2) \\ &= \int_{-\infty}^{\Phi^{-1} \circ F_2(z_2)} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp \\ & \quad \times \left\{ -\frac{1}{2(1-\rho^2)} \left((\rho \Phi^{-1} \circ F_1(z_1))^2 - 2\rho \Phi^{-1} \circ F_1(z_1)u + u^2 \right) \right\} du \\ &= \int_{-\infty}^{\Phi^{-1} \circ F_2(z_2)} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(u - \rho \Phi^{-1} \circ F_1(z_1) \right)^2 \right\} du \\ &= \Phi \left(\frac{\Phi^{-1} \circ F_2(z_2) - \rho \Phi^{-1} \circ F_1(z_1)}{\sqrt{1-\rho^2}} \right). \end{aligned}$$

The definition of $\{Z_k\}$ and the above distributions give the Rosenblatt map

$$\begin{aligned} G_1 &= \Phi^{-1} \circ F_1(Z_1) = M_1 \\ G_2 &= \Phi^{-1} \circ F_{2|1}(Z_2|Z_1) = \frac{M_2 - \rho M_1}{\sqrt{1-\rho^2}}. \end{aligned} \quad (3.7)$$

where G_1 and G_2 are independent standard Gaussian variables.

The PC representation $Z^{PC} = (Z_1^{PC}, Z_2^{PC})$ is given by (3.2) with $d = 2$, where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PC model are determined by minimizing the objective function in (3.3). The PCT model $Z^{PCT} = (Z_1^{PCT}, Z_2^{PCT})$ is defined by (3.4) with $d = 2$, where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PCT model are determined by minimizing the objective function in (3.5).

Since Z , Z^{PC} and Z^{PCT} are defined on the same probability space, the terms $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ of the objective functions (3.3) and (3.5) can be calculated. We set $g_2 = g_3 = 0$, since the resulting objective function (3.3) is adequate for constructing PC models [9]. We set $w_3 = 0$, since we are interested in the tail dependence, see comments following (3.5). The values of w_1 and w_2 in (3.5) are given in the subsequent two cases.

Case (i): F_k is uniform in $(0, 1)$ for $k = 1, 2$. Generally, it is not possible to obtain analytical expressions of coefficients $\{a_{k,j}, a_{k,j,l}\}$ of the PC model. We show that the expressions of $\{a_{k,j}, a_{k,j,l}\}$ can be found in this special case. Note that the following two equations hold for any $k = 1, 2$,

$$\begin{aligned} E[M_k \Phi(M_k)] &= \int_{-\infty}^{\infty} \frac{x \Phi(x)}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} -\frac{\Phi(x)}{\sqrt{2\pi}} d e^{-x^2/2} \\ &= \int_{-\infty}^{\infty} \frac{e^{-x^2}}{2\pi} dx = \frac{1}{2\sqrt{\pi}} \end{aligned}$$

and

$$\begin{aligned} E[M_k^2 \Phi(M_k)] &= \int_{-\infty}^{\infty} \frac{x^2 \Phi(x)}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} -\frac{x \Phi(x)}{\sqrt{2\pi}} d e^{-x^2/2} \\ &= \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \Phi(x) = \int_{-\infty}^{\infty} \frac{\Phi(x)}{\sqrt{2\pi}} e^{-x^2/2} dx \\ & \quad + \int_{-\infty}^{\infty} \frac{x e^{-x^2}}{2\pi} dx = \frac{1}{2} \Phi^2(x) \Big|_{-\infty}^{\infty} = \frac{1}{2}. \end{aligned}$$

Since $M_2 - \rho M_1$ is independent of $\Phi(M_1)$ and $\rho M_2 - M_1$ is independent of $\Phi(M_2)$, then

$$\begin{aligned} E[Z_1 Z_1^{PC}] &= \frac{1}{2} E[\Phi(M_1)] + a_{1,1} E[M_1 \Phi(M_1)] \\ & \quad + a_{1,2} E \left[\frac{M_2 - \rho M_1}{\sqrt{1-\rho^2}} \Phi(M_1) \right] \\ & \quad + a_{1,1,1} E[M_1^2 \Phi(M_1) - \Phi(M_1)] \\ & \quad + a_{1,1,2} E \left[\frac{M_1(M_2 - \rho M_1)}{1-\rho^2} \Phi(M_1) \right] \\ & \quad + a_{1,2,2} E \left[\frac{(M_2 - \rho M_1)^2}{1-\rho^2} \Phi(M_1) - \Phi(M_1) \right] \\ &= \frac{1}{4} + \frac{a_{1,1}}{2\sqrt{\pi}} \end{aligned}$$

and

$$\begin{aligned} E[Z_2 Z_2^{PC}] &= \frac{1}{2} E[\Phi(M_2)] + a_{2,1} E[M_1 \Phi(M_2)] \\ & \quad + a_{2,2} E \left[\frac{M_2 - \rho M_1}{\sqrt{1-\rho^2}} \Phi(M_2) \right] \\ & \quad + a_{2,1,1} E[M_1^2 \Phi(M_2) - \Phi(M_2)] \\ & \quad + a_{2,1,2} E \left[\frac{M_1(M_2 - \rho M_1)}{\sqrt{1-\rho^2}} \Phi(M_2) \right] \\ & \quad + a_{2,2,2} E \left[\frac{(M_2 - \rho M_1)^2}{1-\rho^2} \Phi(M_2) - \Phi(M_2) \right] \end{aligned}$$

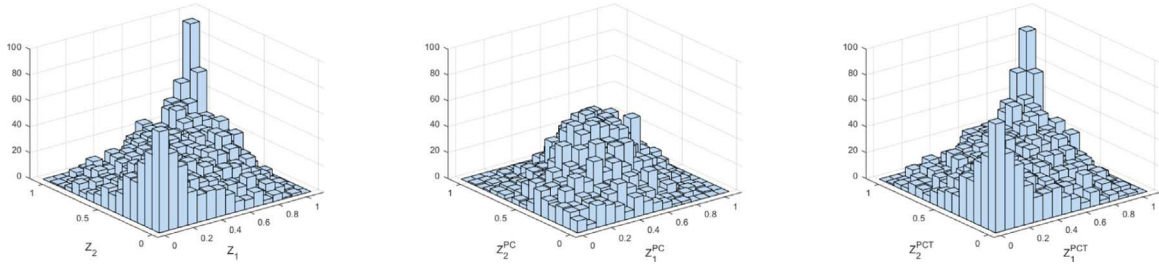


Fig. 3. Histograms of (Z_1, Z_2) , (Z_1^{PC}, Z_2^{PC}) obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and (Z_1^{PCT}, Z_2^{PCT}) obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (228, 500, 0)$ based on 5000 samples (left, middle and right panels).

$$= \frac{1}{4} + \frac{\rho a_{2,1}}{2\sqrt{\pi}} + \frac{a_{2,2}}{2} \sqrt{\frac{1-\rho^2}{\pi}}.$$

Therefore from

$$\begin{aligned} & E[(Z_k^{PC})^2] \\ &= E\left[\left(\frac{1}{2} + a_{k,1}G_1 + a_{k,2}G_2 + a_{k,1,1}(G_1^2 - 1) + a_{k,1,2}G_1G_2 + a_{k,2,2}(G_2^2 - 1)\right)^2\right] \\ &= \frac{1}{4} + a_{k,1}^2 + a_{k,2}^2 + 2a_{k,1,1}^2 + a_{k,1,2}^2 + 2a_{k,2,2}^2, \quad k = 1, 2, \end{aligned}$$

we have

$$\begin{aligned} & e_1(a_{k,j}, a_{k,j,l}) \\ &= E[Z_1^2] - 2E[Z_1 Z_1^{PC}] + E[(Z_1^{PC})^2] + E[Z_2^2] - 2E[Z_2 Z_2^{PC}] + E[(Z_2^{PC})^2] \\ &= \frac{2}{3} - 1 - \frac{a_{1,1}}{\sqrt{\pi}} - \frac{\rho a_{2,1}}{\sqrt{\pi}} - a_{2,2} \sqrt{\frac{1-\rho^2}{\pi}} + \frac{1}{4} + a_{1,1}^2 + a_{1,2}^2 + 2a_{1,1,1}^2 + a_{1,1,2}^2 + 2a_{1,2,2}^2 \\ &\quad + \frac{1}{4} + a_{2,1}^2 + a_{2,2}^2 + 2a_{2,1,1}^2 + a_{2,1,2}^2 + 2a_{2,2,2}^2 \\ &= \frac{1}{6} - \frac{1}{2\pi} + \left(a_{1,1} - \frac{1}{2\sqrt{\pi}}\right)^2 + \left(a_{2,1} - \frac{\rho}{2\sqrt{\pi}}\right)^2 + a_{2,2}^2 + \left(a_{2,2} - \frac{1}{2} \sqrt{\frac{1-\rho^2}{\pi}}\right)^2 \\ &\quad + 2a_{1,1,1}^2 + a_{1,1,2}^2 + 2a_{1,2,2}^2 + 2a_{2,1,1}^2 + a_{2,1,2}^2 + 2a_{2,2,2}^2, \end{aligned}$$

which implies the optimal values are $a_{1,1} = 1/(2\sqrt{\pi})$, $a_{2,1} = \rho/(2\sqrt{\pi})$, $a_{2,2} = \sqrt{(1-\rho^2)/\pi}/2$ and $a_{1,2} = a_{1,1,1} = a_{1,1,2} = a_{1,2,2} = a_{2,1,1} = a_{2,1,2} = a_{2,2,2} = 0$.

The following numerical results are for $\rho = 0.5$ and are based on 5000 samples. The left, middle and right panels of Fig. 3 show 5000 samples of (Z_1, Z_2) and the corresponding samples of (Z_1^{PC}, Z_2^{PC}) obtained under (3.3) and (Z_1^{PCT}, Z_2^{PCT}) obtained under (3.5) with $(w_1, w_2) = (228, 500)$. For these values of (w_1, w_2) the components $E[\|Z - Z^{PCT}\|_2^2]$ and $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ of the objective function (3.5) have similar magnitudes. The two dimensional histograms of the middle panel suggest that the joint densities of (Z_1, Z_2) and (Z_1^{PC}, Z_2^{PC}) differ significantly. The left, middle and right panels of Fig. 4 show the histograms of $\|Z - Z^{PC}\|_2^2$ obtained under (3.3) and of $\|Z - Z^{PCT}\|_2^2$ obtained under (3.5) with weighting coefficients $(w_1, w_2) = (1, 0)$ and $(w_1, w_2) = (228, 500)$. They suggest that the PCT model is superior to the PC model. We take two different sets of (w_1, w_2) to assess the usefulness of including the spectral measures in the objective function (3.5). The plots suggest that objective functions with the spectral measures delivered improved PCT models.

Case (ii): F_k is exponential with parameters $\mu_k > 0$, $k = 1, 2$. Then we have $Z_k = -\log(1 - \Phi(M_k))/\mu_k$ and the Rosenblatt mapping $G \rightarrow Z$ is given by (3.7).

The following numerical results are for $\rho = 0.5$, $\mu_1 = 1$, $\mu_2 = 2$, and are based on 5000 samples. The left, middle and right panels of Fig. 5 show 5000 samples of (Z_1, Z_2) and the corresponding samples of (Z_1^{PC}, Z_2^{PC}) obtained under (3.3) and (Z_1^{PCT}, Z_2^{PCT}) obtained under (3.5) with $(w_1, w_2) = (3000, 500)$. For these values of (w_1, w_2) the components $E[\|Z - Z^{PCT}\|_2^2]$ and $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ of the objective function (3.5) have similar magnitudes. The two dimensional histograms of the middle panel suggest that the joint densities of (Z_1, Z_2) and (Z_1^{PC}, Z_2^{PC}) differ significantly. The left, middle and

right panels of Fig. 6 show the histograms of $\|Z - Z^{PC}\|_2^2$ obtained under (3.3) and of $\|Z - Z^{PCT}\|_2^2$ obtained under (3.5) with weighting coefficients $(w_1, w_2) = (1, 0)$ and $(w_1, w_2) = (3000, 500)$. They suggest that the PCT model is superior to the PC model. We take two different sets of values of (w_1, w_2) to assess the usefulness of including the spectral measures in the objective function (3.5). The plots suggest that objective functions with the spectral measures delivered improved PCT models.

Example 3.2. Let $Z_1 = R_1$ and $Z_2 = R_1 + R_2$, where R_1, R_2 are independent gamma random variables with distribution functions F_k , $k = 1, 2$, and density functions

$$f_{R_k}(r_k) = \frac{1}{\Gamma(\theta_k)} e^{-r_k} r_k^{\theta_k-1}, \quad r_k, \theta_k > 0, \quad k = 1, 2.$$

The CDF of the conditional random variable $Z_2|Z_1 = z_1$ is

$$F_{Z_2|Z_1=z_1}(z_2) = P(R_1 + R_2 \leq z_2 | R_1 = z_1) = P(R_2 \leq z_2 - z_1) = F_{R_2}(z_2 - z_1).$$

so that the Rosenblatt map has the form

$$\begin{aligned} G_1 &= \Phi^{-1} \circ F_1(Z_1) \\ G_2 &= \Phi^{-1} \circ F_{2|1}(Z_2|Z_1) = \Phi^{-1} \circ F_{R_2}(Z_2 - Z_1) = \Phi^{-1} \circ F_{R_2}(R_2). \end{aligned} \quad (3.8)$$

where G_1 and G_2 are independent standard Gaussian variables.

The PC representation $Z^{PC} = (Z_1^{PC}, Z_2^{PC})$ is given by (3.2) with $d = 2$, where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PC model are determined by minimizing the objective function in (3.3). The PCT model $Z^{PCT} = (Z_1^{PCT}, Z_2^{PCT})$ is defined by (3.4) with $d = 2$, where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PCT model are determined by minimizing the objective function in (3.5).

Since Z, Z^{PC} and Z^{PCT} are defined on the same probability space, the terms $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ of the objective functions (3.3) and (3.5) can be calculated. We set $g_2 = g_3 = 0$, since the resulting objective function (3.3) is adequate for constructing PC models [9]. We set $w_3 = 0$, since we are interested in the tail dependence, see comments following (3.5). The values of w_1 and w_2 in (3.5) are given in the subsequent numerical experiment.

The following numerical results are for $\theta_1 = 2$, $\theta_2 = 3$ and are based on 5000 samples. The left, middle and right panels of Fig. 7 show 5000 samples of (Z_1, Z_2) and the corresponding samples of (Z_1^{PC}, Z_2^{PC}) obtained under (3.3) and (Z_1^{PCT}, Z_2^{PCT}) obtained under (3.5) with $(w_1, w_2) = (3000, 500)$. For these values of (w_1, w_2) the components $E[\|Z - Z^{PCT}\|_2^2]$ and $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ of the objective function (3.5) have similar magnitudes. The two dimensional histograms of the middle panel suggest that the joint densities of (Z_1, Z_2) and (Z_1^{PC}, Z_2^{PC}) differ significantly. The left, middle and right panels of Fig. 8 show the histograms of $\|Z - Z^{PC}\|_2^2$ obtained under (3.3) and of $\|Z - Z^{PCT}\|_2^2$ obtained under (3.5) with weighting coefficients $(w_1, w_2) = (1, 0)$ and $(w_1, w_2) = (3000, 500)$. They suggest that the PCT model is superior to the PC model. We take two different sets of (w_1, w_2) to assess the contribution of the spectral measure to the objective function (3.5). The plots suggest that the incorporation of the

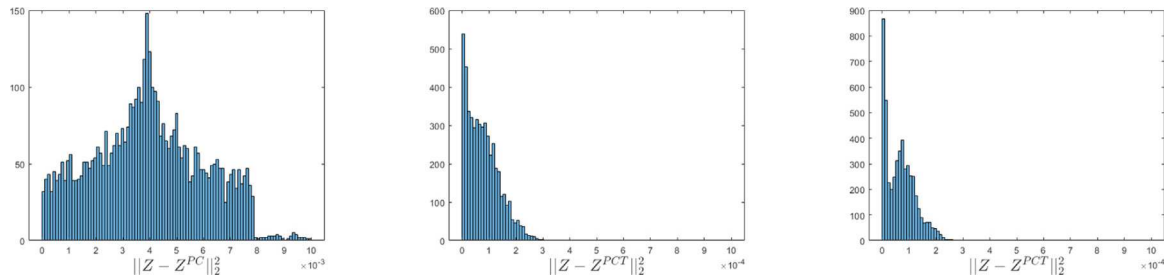


Fig. 4. Histograms of $\|Z - Z^{PC}\|_2^2$ obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and of $\|Z - Z^{PCT}\|_2^2$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (1, 0, 0)$ and $(w_1, w_2, w_3) = (228, 500, 0)$ based on 5000 samples (left, middle and right panels).

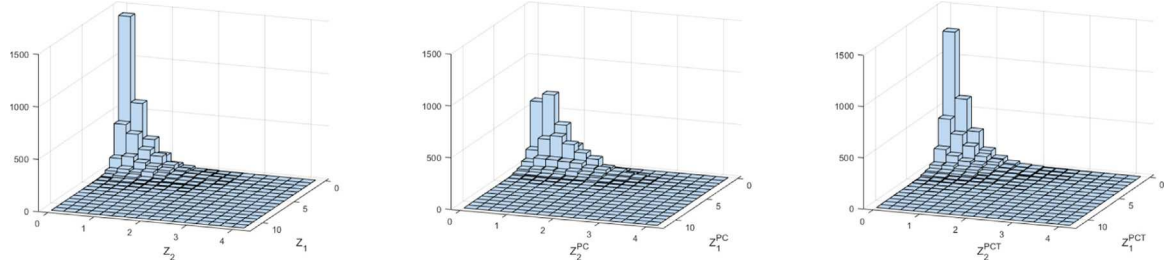


Fig. 5. Histograms of (Z_1, Z_2) , (Z_1^{PC}, Z_2^{PC}) obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and (Z_1^{PCT}, Z_2^{PCT}) obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (3000, 500, 0)$ based on 5000 samples (left, middle and right panels).

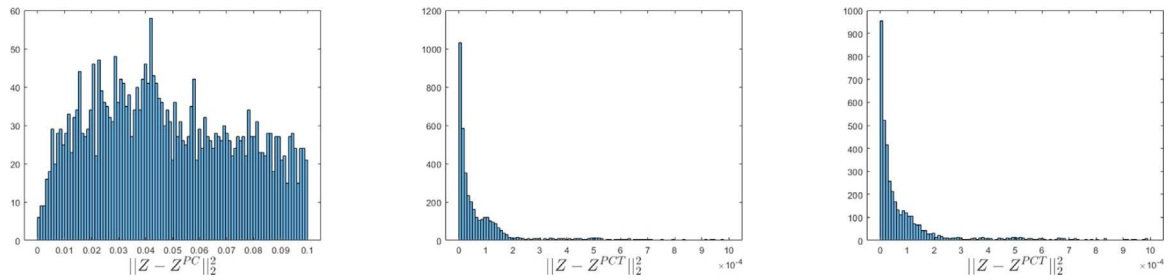


Fig. 6. Histograms of $\|Z - Z^{PC}\|_2^2$ obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and of $\|Z - Z^{PCT}\|_2^2$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (1, 0, 0)$ and $(w_1, w_2, w_3) = (3000, 500, 0)$ based on 5000 samples (left, middle and right panels).

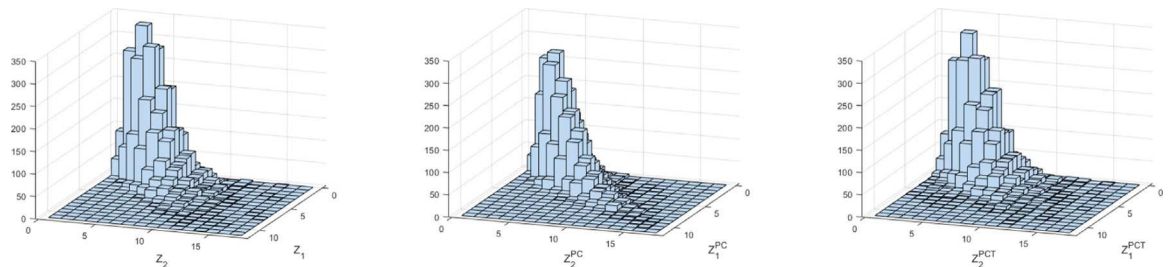


Fig. 7. Histograms of (Z_1, Z_2) , (Z_1^{PC}, Z_2^{PC}) obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and (Z_1^{PCT}, Z_2^{PCT}) obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (3000, 500, 0)$ based on 5000 samples (left, middle and right panels).

spectral measures in the objective function improves the quality of PCT models.

Example 3.3. Let $X_1(t), X_2(t)$, $0 \leq t \leq \tau$, be real-valued processes defined by the differential equations

$$\begin{aligned} \ddot{X}_1(t) + \alpha_1 \dot{X}_1(t) + \beta_1 X_1(t) &= \kappa_1 V(t), \\ \ddot{X}_2(t) + \alpha_2 \dot{X}_2(t) + \beta_2 X_2(t) &= \kappa_2 V(t), \quad 0 \leq t \leq \tau \end{aligned} \quad (3.9)$$

with the initial conditions $X_i(0) = 0$ and $\dot{X}_i(0) = 0$, $i = 1, 2$, where $\alpha_i, \beta_i, \kappa_i > 0$, $i = 1, 2$ are constants. The input is the translation process $V(t) = F^{-1} \circ \Phi(W(t))$, where F is the Gamma distribution function with

the shape parameter ν and scale parameter 1, $W(t)$ is the stationary solution of $dW(t) = -\vartheta W(t)dt + \sqrt{2\vartheta}dB(t)$, $\vartheta > 0$, and B denotes the standard Brownian motion.

From [32] (Chap.2), the solution of (3.9) is

$$X_i(t) = \int_0^t \frac{\kappa_i}{\psi_i} e^{-\alpha_i(t-u)/2} \sin(\psi_i(t-u)) V(u) du, \quad i = 1, 2, \quad 0 \leq t \leq \tau, \quad (3.10)$$

where $\psi_i = (\beta_i - \alpha_i^2/4)^{1/2}$, $i = 1, 2$.

Our objective is to construct FD models for the vector-valued process $(X_1(t), X_2(t))$. Since (3.9) has to be solved numerically, $V(t)$ and $(X_1(t), X_2(t))$ are defined and calculated at a finite set of times, e.g., the

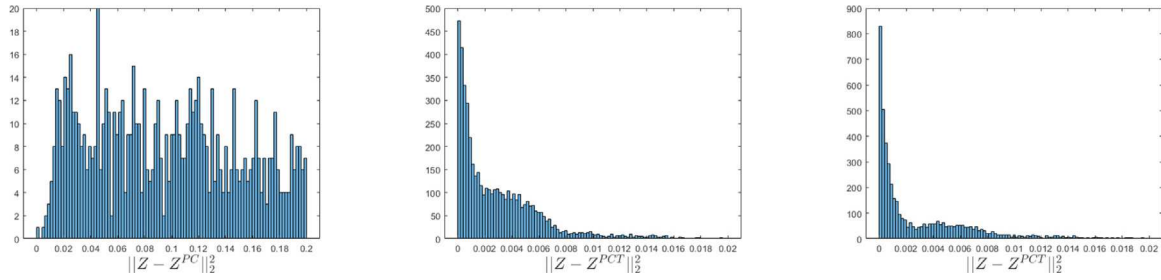


Fig. 8. Histograms of $\|Z - Z^{PC}\|_2^2$ obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (1, 0, 0)$ and of $\|Z - Z^{PCT}\|_2^2$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (1, 0, 0)$ and $(w_1, w_2, w_3) = (3000, 500, 0)$ based on 5000 samples (left, middle and right panels).

equally spaced times $t_i = i \Delta t$, $i = 1, \dots, n$, where $\Delta t = \tau/n$ denotes the integration time step. Denote by $\eta = (V(t_1), \dots, V(t_n))$ and $\zeta_i = (X_i(t_1), \dots, X_i(t_n))$, $i = 1, 2$, the discrete versions of the input $V(t)$ and of the processes $X_i(t)$, $i = 1, 2$. The random vector η admits the representation $\eta = \sum_{k=1}^n Z_k v_k$, where $\{v_k\}$ are the eigenvectors of the covariance matrix $E[\eta\eta^T]$ and the random coefficients $\{Z_k\}$ are defined sample by projection, i.e., $Z_k(\omega) = \eta^T(\omega) v_k$, $\omega \in \Omega$. The corresponding FD model is $\eta_d = \sum_{k=1}^d Z_k v_k$. Since the differential Eqs. (3.9) are linear, their solutions to η and η_d are linear forms of $\{Z_k\}$ denoted by $\zeta_i = \{\zeta_{i,j}\}$ and $\zeta_{d:i} = \{\zeta_{d:i,j}\}$, $i = 1, 2$, $j = 1, \dots, n$.

We construct PC and PCT models of the random vector $(Z_1, \dots, Z_d)$. The PC representation $Z^{PC} = (Z_1^{PC}, \dots, Z_d^{PC})$ is given by (3.2), where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PC model are determined by minimizing the objective function in (3.3). The PCT model $Z^{PCT} = (Z_1^{PCT}, \dots, Z_d^{PCT})$ is defined by (3.4), where the coefficients $\{a_{k,j}, a_{k,j,l}\}$ for the PCT model are determined by minimizing the objective function in (3.5). In this example, which mimics the common situation in applications, target and FD samples cannot be paired as in the previous two examples, so that we can only compare global properties, e.g., distributions. As a result, the mean squared errors $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ are not available and are removed from the objective functions (3.3) and (3.5) by setting $g_1 = w_1 = 0$. The discrete versions of the FD models of $X_i(t)$, $i = 1, 2$, corresponding to the PC and PCT models Z^{PC} and Z^{PCT} of Z are denoted by $\zeta_{d:i}^{PC}$ and $\zeta_{d:i}^{PCT}$. They are elements of the same n -dimensional Euclidean space. An additional FD model denoted by $\zeta_{d:i}$ is constructed. It has the same functional form as $\zeta_{d:i}^{PC}$ and $\zeta_{d:i}^{PCT}$ but the samples of its random coefficients are given by $Z_k(\omega) = \eta^T(\omega) v_k$, see discussion following (3.10).

The following numerical results are for $\nu = 2$, $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, $\beta_1 = 25$, $\beta_2 = 9$, $\kappa_1 = 1$, $\kappa_2 = 10$, $\vartheta = 5$, $\tau = 10$, $n = 1000$, $d = 20$ and the time step $\Delta t = 0.01$. We set $g_1 = w_1 = 0$ since the mean squared errors $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ are not available, see comments following (3.3) and (3.5). The other weighting coefficients are $(g_2, g_3) = (1, 300)$ and $(w_2, w_3) = (1, 1.1)$. For these values, the components $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PC}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\|E[Z Z^T] - E[Z^{PC} (Z^{PC})^T]\|$ of the objective function (3.3) and the components $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1, i_2}(\cdot) - s_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1, i_2}(\cdot) - h_{i_1, i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ of the objective function (3.5) have similar magnitudes so that they contribute equally to the objective functions. All the plots are based on 10000 samples.

The left, middle and right panels of Figs. 9 and 10 show the two dimensional histograms of (Z_i, Z_j) , (Z_i^{PC}, Z_j^{PC}) obtained under (3.3) and (Z_i^{PCT}, Z_j^{PCT}) obtained under (3.5). Visual inspection of the plots in Figs. 9 and 10 suggests that the PCT histograms are closer to the target histograms than PC histograms. This qualitative observation is confirmed by the plots of Fig. 11 which show with estimates of extremes for PC- and PCT-based FD models. The PCT-based estimates of extremes follow closely the target estimates in contrast to the PC-based estimates which are unsatisfactory.

The thin solid lines of the left and right panels of Fig. 11 are estimates of $P(\|\zeta_i\| > x)$ for $i = 1$ and $i = 2$ which are obtained directly from data, where $\|\zeta_i\| = \max_{1 \leq j \leq n} |\zeta_{i,j}|$. These probabilities are viewed

as truth. The other lines of the figure are calculated from samples of $\zeta_{d:i}$ (heavy solid lines), $\zeta_{d:i}^{PC}$ (dotted lines) and $\zeta_{d:i}^{PCT}$ (dashed lines) for the first and second components (left and right panels). The heavy solid lines are the closest to the truth. The next best model is $\zeta_{d:i}^{PCT}$ while $\zeta_{d:i}^{PC}$ differs significantly from the truth. We prefer $\zeta_{d:i}^{PCT}$ to $\zeta_{d:i}$ since the set of samples of $\zeta_{d:i}$ is defined by the available data so that it cannot be extended. In contrast, samples of any size can be generated from $\zeta_{d:i}^{PCT}$ since its probability law is known.

4. Wind data analysis

PC- and PCT-based FD models are developed for the vector-valued wind pressure time series recorded in the University of Florida boundary layer wind tunnel facility (UFBLWT) on the surface of a bluff body at $m = 6$ pressure taps. The experimental setup and recorded data are discussed in the following subsection. Sections 4.2 and 4.3 construct PC- and PCT-based FD models and assess their performance. The focus is on the capability of these models to predict extremes of the recorded wind pressure time series. The optimization algorithm used to construct the FD models for the wind time series is presented in Section 4.4.

4.1. Experimental setting and data

The pressure data set used in this section was generated in October of 2021 as a part of an investigation into the influence of raised planter bed shapes on their vulnerability to high winds. The study was inspired by Hurricane Irma (2017) which damaged the plastic coating over the tops of raised planter beds in agricultural fields in Florida. These coatings are necessary to optimize the retention of moisture and nutrients, and their damage due to high winds represents an economic loss.

The premise of the experimental study was to evaluate the wind pressures over the surfaces of a series of adjacent agricultural planter bed rows using 1:10 scale models of two different planter bed shapes, see the top panel of Fig. 12. The complete model consisted of 5 adjacent rows of raised planter beds, see the bottom panel of Fig. 12. Three of these rows were fitted with 84 pressure taps each over half the length of the row. The left panel of Fig. 13 shows the five adjacent bed model, the location of the 252 taps, and the approach wind direction employed in this section. The box on the center row in the left panel shows the location of the six taps used in this section, which are numbered 79–84 from left to right. The model was mounted on a turntable to change the approach wind direction. The right panel of Fig. 13 shows the model in the wind tunnel in the zero degree approach wind angle.

For both planter bed shapes, the pressure data were sampled at 625 Hz for 180 s for 36 directions and two different boundary layer roughness regimes (open and suburban exposure).

The subset of data accessed for the study presented in the following two subsections corresponds to the wide planter bed shape in the right panel of Fig. 12, zero degree wind approach in the right panel of Fig. 13, the open terrain exposure, and the six taps identified in the

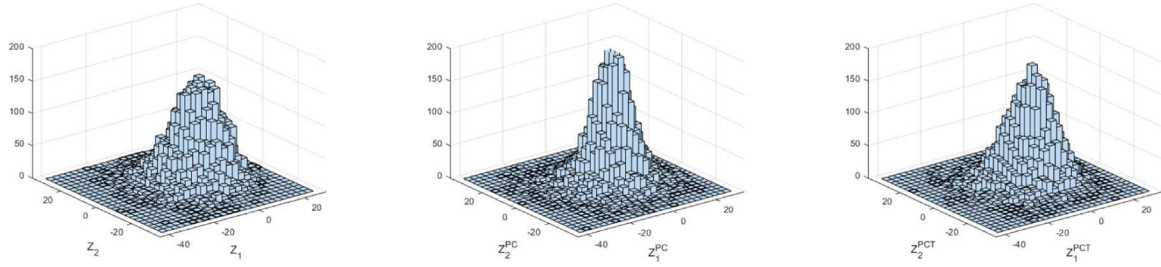


Fig. 9. Histograms of (Z_1, Z_2) , (Z_1^{PC}, Z_2^{PC}) obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (0, 1, 300)$ and (Z_1^{PCT}, Z_2^{PCT}) obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (0, 1, 1.1)$ based on 10000 samples (left, middle and right panels).

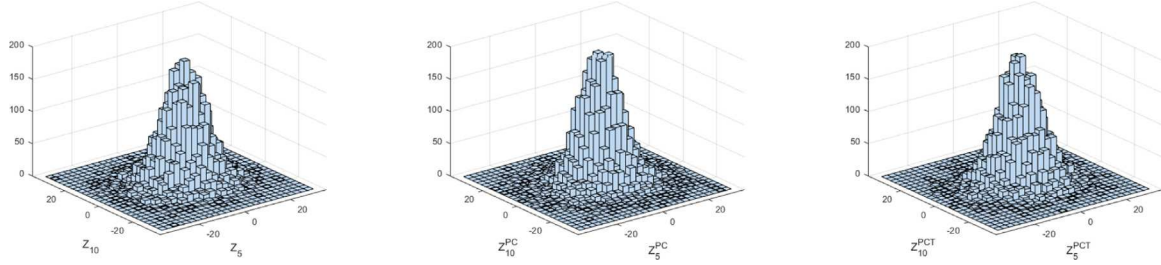


Fig. 10. Histograms of (Z_5, Z_{10}) , (Z_5^{PC}, Z_{10}^{PC}) obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (0, 1, 300)$ and $(Z_5^{PCT}, Z_{10}^{PCT})$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (0, 1, 1.1)$ based on 10000 samples (left, middle and right panels).

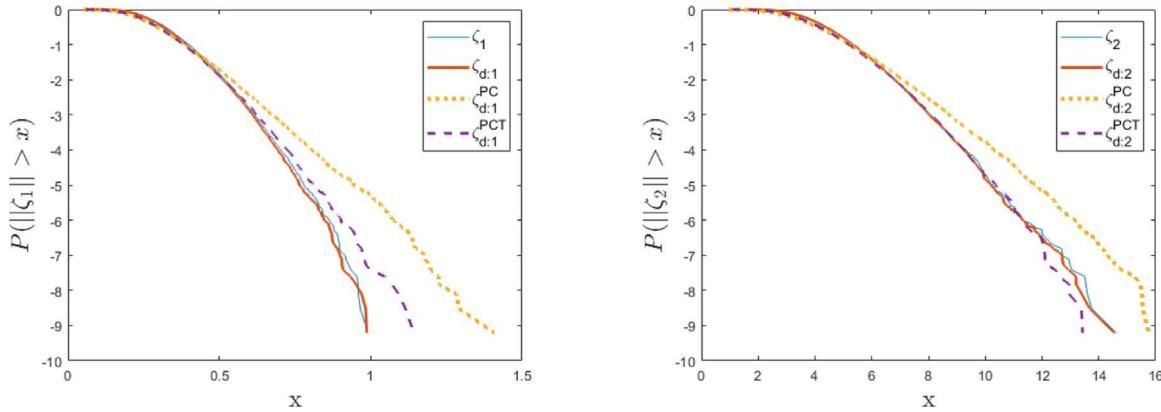


Fig. 11. Estimates of the target probability $P(\|\zeta_i\| > x)$ (thin solid line), estimates based on FD model (heavy solid line), estimates based on finite dimensional PC model (dotted line) and estimates based on finite dimensional PCT model in logarithmic scale (dashed line) for ζ_1 and ζ_2 (left and right panels).

left panel of Fig. 13. All experiments were conducted with a steady fan speed of 1200 RPM, resulting in stationary ergodic datasets.

The left and right panels of Fig. 14 show segments of length 1000 of the wind pressure records at taps 79 and 80. The lack of symmetry of these records suggests that the wind pressure process is non-Gaussian. This qualitative observation is quantified in the following two subsections.

4.2. Construction of FD models

Let $(y_1, y_2, \dots, y_n)$ with $y_j = (y_{1,j}, \dots, y_{m,j})^T$, $j = 1, \dots, n$ be the wind pressure record at m pressure taps. The analysis in this section considers the pressure taps 79 to 84 so that $m = 6$, (Fig. 13, left). The taps 79 – 84 are renamed 1 – 6 in the remainder of this section. The proximity of these six taps is such that strong spatial correlation exists over adjacent pairs. The record has length $n = 110000$ and time step $\Delta t = 0.0016$ seconds. It is assumed that the record is a sample of an $m = 6$ dimensional time series which is stationary and ergodic. Our objective is to construct FD models of this series whose random coefficients are represented by PC and PCT models, see (3.2) and (3.4).

The construction involves the following three steps.

- **Step 1:** The available record is partitioned in q segments of length $N = \lfloor n/q \rfloor$ each, i.e., $(y_1, y_2, \dots, y_N)$ is the first sample, $(y_{N+1}, y_{N+2}, \dots, y_{2N})$ is the second sample and so on. These segments are assumed to be q independent realizations of an N -dimensional vector $(Y_1, Y_2, \dots, Y_N)$. This heuristic assumption is supported by the estimates of the correlation functions of the time series based on the entire record and on its q segments shown in Figs. 15 and 16 by heavy solid lines and heavy dotted lines which nearly coincide. The above partition of the record violates our assumption of independence. To satisfy this assumption, we should skip every other segment of length N . Yet, numerical studies suggest that such precaution is unnecessary and is not implemented.
- **Step 2:** Let $\mathcal{X}_i = (Y_{i,1}, Y_{i,2}, \dots, Y_{i,N})^T$, $i = 1, \dots, m$, denote the m -dimensional time series of length N describing the wind model. We construct the FD models $\mathcal{X}_{d:i} = (\mathcal{X}_{d:i,1}, \dots, \mathcal{X}_{d:i,N})^T$ of the type

$$\mathcal{X}_{d:i} = \sum_{k=1}^d Z_{i,k} v_{i,k}, \quad i = 1, \dots, m, \quad (4.1)$$

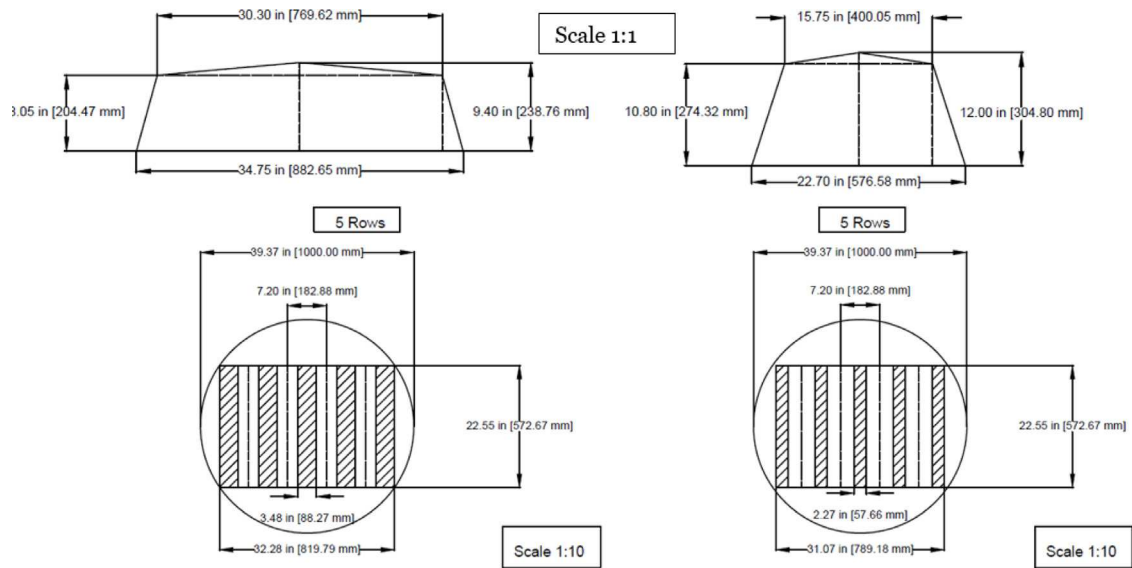


Fig. 12. Top Row: Full-scale dimensions of the cross section of a single row of the wide (left panel) and compact (right panel) planter bed model. Bottom Row: The 1:10 scale full 5-row model of the wide (left panel) and compact (right panel) planter beds as installed on the 1 m diameter turntable.

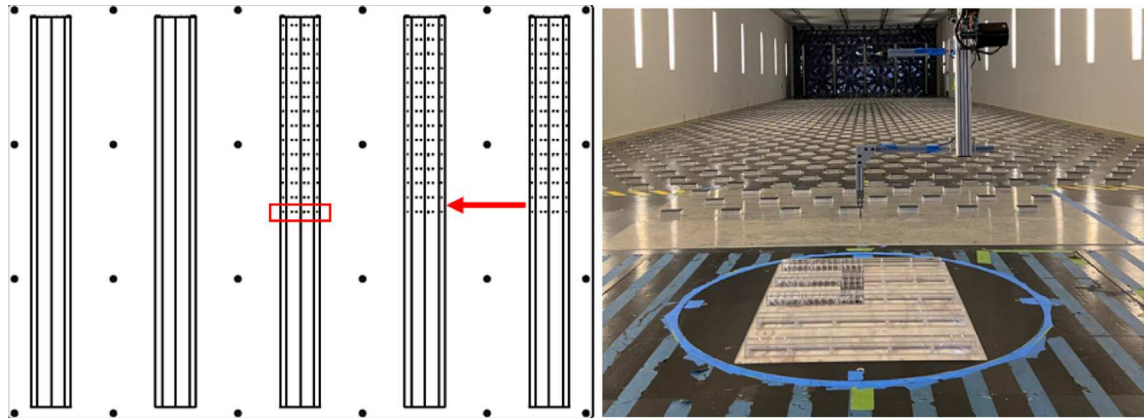


Fig. 13. Left panel: Locations of 252 pressure taps on the upper half of three out of five planter bed rows. The box on the center row shows the location of taps 79–84 (left to right) used in this study. The arrow indicates the zero degree wind approach direction used in this study. Right panel: Model in the wind tunnel at zero degree wind approach angle.

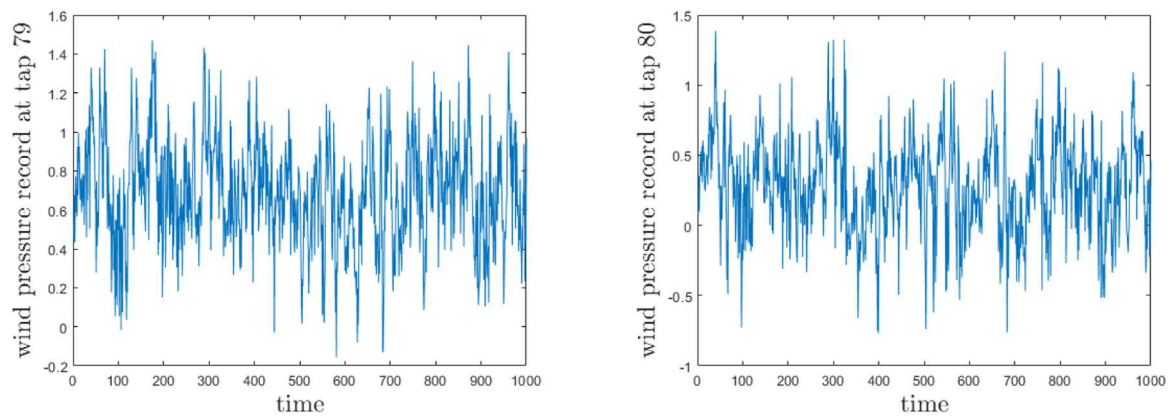


Fig. 14. Segments of length 1000 of wind pressure records at taps 79 and 80 (left and right panels).

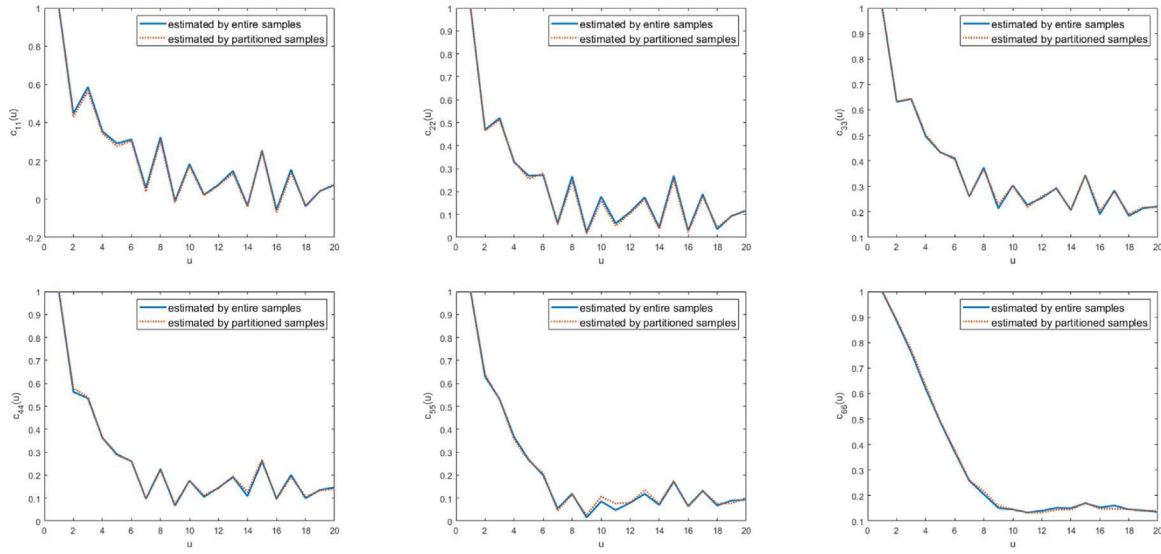


Fig. 15. Correlation function $c_{ii}(u) = E[Y_{i,1}Y_{i,u}]$, $i = 1, \dots, 6$ for the time lag $u = 1, \dots, 20$ estimated by the entire records and the partitioned samples (heavy solid lines and heavy dotted lines).

where $\{v_{i,k}\}$ are the eigenvectors of the covariance matrix $E[\mathcal{X}_i \mathcal{X}_i^T]$ and the random coefficients $\{Z_{i,k}\}$ are defined sample by projection, i.e., $Z_{i,k}(\omega) = \mathcal{X}_i(\omega)^T v_{i,k}$, $\omega \in \Omega$. These samples are used to estimate the marginal distributions $F_{i,k}$ of $Z_{i,k}$.

- Step 3: We construct PC and PCT models of the random vectors $Z = (Z_{1,1}, \dots, Z_{1,d}, \dots, Z_{m,d})^T$. The set of FD models in (4.1) depends on md random coefficients which are related to md independent standard Gaussian variables $G_1, G_2, \dots, G_{md}$ by

$$\begin{aligned} Z_{i,k}^{PC} = & E[Z_{i,k}] + \sum_{j=1}^{md} a_{i,k,j} G_j + \sum_{1 \leq j_1 \leq l_1 \leq md} a_{i,k,j_1,l_1} (G_{j_1} G_{l_1} - \delta_{j_1,l_1}) \\ & + \sum_{1 \leq j_1 \leq l_1 \leq l_2 \leq md} a_{i,k,j_1,l_1,l_2} G_{j_1} G_{l_1} G_{l_2} \\ & + \sum_{1 \leq j_1 \leq l_1 \leq l_2 \leq l_3 \leq md} a_{i,k,j_1,l_1,l_2,l_3} (G_{j_1} G_{l_1} G_{l_2} G_{l_3} - E[G_{j_1} G_{l_1} G_{l_2} G_{l_3}]), \end{aligned}$$

$$i = 1, \dots, m, \quad k = 1, \dots, d, \quad (4.2)$$

where the coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$ are determined by minimizing the objective function (3.3). The above equation gives us the PC model $Z^{PC} = (Z_{1,1}^{PC}, \dots, Z_{1,d}^{PC}, \dots, Z_{m,d}^{PC})^T \in \mathbb{R}^{md}$ and extends (3.2), by increasing the degree of the polynomial chaos from two to four. The PCT model $Z^{PCT} = (Z_{1,1}^{PCT}, \dots, Z_{1,d}^{PCT}, \dots, Z_{m,d}^{PCT})^T \in \mathbb{R}^{md}$ is defined by

$$Z_{i,k}^{PCT} = F_{i,k}^{-1} \circ F_{i,k}^{PC}(Z_{i,k}^{PC}), \quad i = 1, \dots, m, \quad k = 1, \dots, d, \quad (4.3)$$

where $F_{i,k}^{PC}$ is the distribution of $Z_{i,k}^{PC}$ for given coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$. These coefficients are determined by minimizing the objective function (3.5). Since target and FD samples cannot be paired as in Examples 3.1 and 3.2, we can only compare global properties, e.g., distributions. As a result, the mean squared errors $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ are not available and are removed from the objective functions (3.3) and (3.5) by setting $g_1 = w_1 = 0$.

The functional form of the FD models under consideration is given by (4.1). The models, denoted by $\mathcal{X}_{d:i}^{IND}$, $\mathcal{X}_{d:i}^{PC}$ and $\mathcal{X}_{d:i}^{PCT}$, are elements of the space spanned by the same vectors $\{v_{i,k}\}$, but their coefficients differ. The random coefficients of $\mathcal{X}_{d:i}^{PC}$ are given by (4.2). The random coefficients of $\mathcal{X}_{d:i}^{IND}$ and $\mathcal{X}_{d:i}^{PCT}$ have the same marginal distributions but they are independent for $\mathcal{X}_{d:i}^{IND}$ and dependent given by (4.3) for $\mathcal{X}_{d:i}^{PCT}$.

4.3. Extremes of wind pressures by FD models

The following numerical results are for $q = 5500$, $N = 20$ and $d = 15$. As mentioned in Section 4.2, we set $g_1 = w_1 = 0$ since the mean squared errors $E[\|Z - Z^{PC}\|_2^2]$ and $E[\|Z - Z^{PCT}\|_2^2]$ are not available, see comments following (3.3) and (3.5). The other weighting coefficients are $(g_2, g_3) = (1, 100)$ and $(w_2, w_3) = (1, 1.1)$. For these values, the components $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1,i_2}(\cdot) - h_{i_1,i_2}^{PC}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\|E[Z Z^T] - E[Z^{PC} (Z^{PC})^T]\|$ of the objective function (3.3) and the components $\max_{1 \leq i_1 < i_2 \leq d} \|s_{i_1,i_2}(\cdot) - s_{i_1,i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ and $\max_{1 \leq i_1 < i_2 \leq d} \|h_{i_1,i_2}(\cdot) - h_{i_1,i_2}^{PCT}(\cdot | a_{k,j}, a_{k,j,l})\|_2$ of the objective function (3.5) have similar magnitudes so that they contribute equally to the objective functions. All the plots are based on 5500 samples.

The left, middle and right panels of Figs. 17 and 18 show the two dimensional histograms of $(Z_{i_1,k_1}, Z_{i_2,k_2})$, $(Z_{i_1,k_1}^{PC}, Z_{i_2,k_2}^{PC})$ obtained under (3.3) and $(Z_{i_1,k_1}^{PCT}, Z_{i_2,k_2}^{PCT})$ obtained under (3.5). Visual inspection of the plots in Figs. 17 and 18 suggests that the PCT histograms are closer to the target histograms than PC histograms. This qualitative observation is confirmed by the plots of Fig. 19 which show with estimates of extremes for PC- and PCT-based FD models. The PCT-based estimates of extremes follow closely the target estimates in contrast to the PC-based estimates which are unsatisfactory. The thin solid lines of Fig. 19 are estimates of $P(\|\mathcal{X}_i\| > x)$ for $i = 1, \dots, 6$ which are obtained directly from data, where $\|\mathcal{X}_i\| = \max_{1 \leq j \leq N} |Y_{i,j}|$. These probabilities are viewed as truth. The other lines of the figure are calculated from samples of $\mathcal{X}_{d:i}$ in (4.1) (heavy solid lines), $\mathcal{X}_{d:i}^{PC}$ (dotted lines) and $\mathcal{X}_{d:i}^{PCT}$ (dashed lines) for taps 79 to 84. The heavy solid lines are the closest to the truth. The next best model is $\mathcal{X}_{d:i}^{PCT}$ while $\mathcal{X}_{d:i}^{PC}$ differs significantly from the truth. The PC and PCT models are based on polynomial chaos of degree two, i.e., without the last two terms in (4.2).

The plots of Fig. 20 explore the effects of the dependence between the random coefficients of FD models on extremes. The solid lines are estimates of the probability $P(\|\mathcal{X}_5\| > x)$ obtained from the wind record at tap 83 for which the performance of the PCT model was less satisfactory. The other lines are extremes of $\mathcal{X}_{d:5}$ in (4.1) with random coefficients $\{Z_{5,k}\}$ which have different dependencies: independent (dash-dotted lines), Z^{PC} based on a second degree polynomial chaos (dotted line, left panel), Z^{PC} based on a fourth degree polynomial chaos (dotted line, right panel), Z^{PCT} based on a second degree polynomial chaos (dashed line, left panel) and Z^{PCT} based on a fourth degree polynomial chaos (dashed line, right panel) for tap 83. The estimates are unsatisfactory if the components of Z^{PCT} are assumed

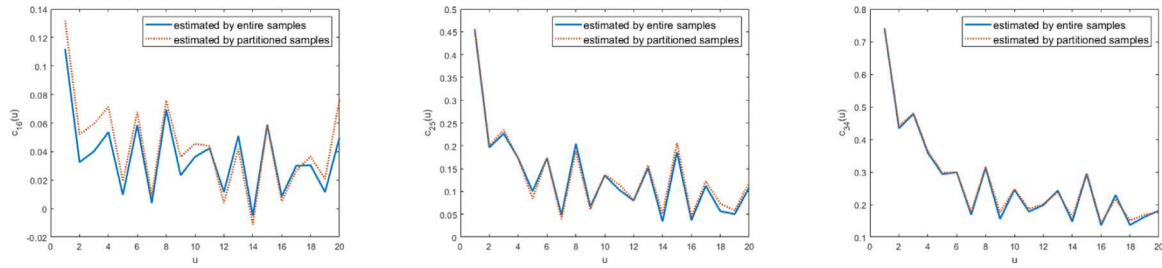


Fig. 16. Correlation function $c_{ij}(u) = E[Y_{i,1}Y_{j,u}]$, $(i,j) = (1,6), (2,5), (3,4)$ for the time lag $u = 1, \dots, 20$ estimated by the entire records and the partitioned samples (heavy solid lines and heavy dotted lines).

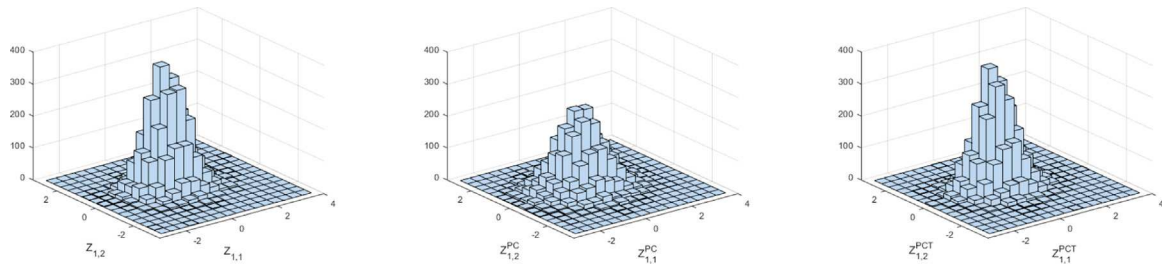


Fig. 17. Histograms of $(Z_{1,1}, Z_{1,2})$, $(Z_{1,1}^{PC}, Z_{1,2}^{PC})$ obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (0, 1, 100)$ and $(Z_{1,1}^{PCT}, Z_{1,2}^{PCT})$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (0, 1, 1.1)$ based on 5500 samples (left, middle and right panels).

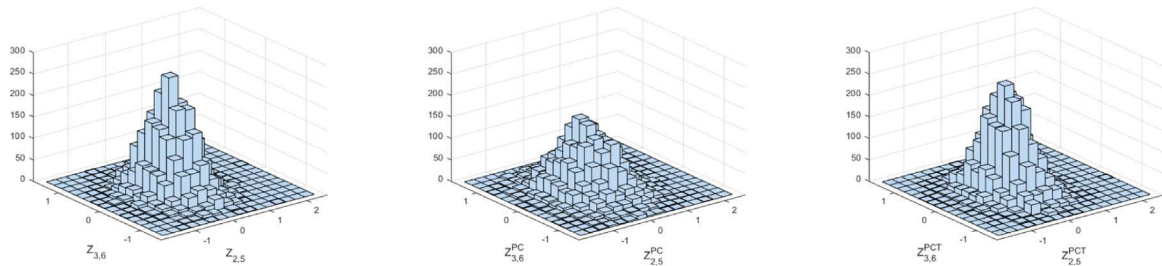


Fig. 18. Histograms of $(Z_{2,5}, Z_{3,6})$, $(Z_{2,5}^{PC}, Z_{3,6}^{PC})$ obtained under the objective function (3.3) with $(g_1, g_2, g_3) = (0, 1, 100)$ and $(Z_{2,5}^{PCT}, Z_{3,6}^{PCT})$ obtained under the objective function (3.5) with $(w_1, w_2, w_3) = (0, 1, 1.1)$ based on 5500 samples (left, middle and right panels).

independent, an expected result since the resulting FD wind model is approximately Gaussian. They approach the target probability as the degree of the polynomial chaos is increased from two to four since this increases results in a superior representation of the dependence between the random variables $Z_{i,k}$. However, increasing the degree of the polynomial chaos does not improve the estimates based on PC models, since Z and Z^{PC} have different marginal distributions.

4.4. Optimization algorithm

The following five-step method is used to identify the optimal coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$ for the PC and PCT models of the random variables $Z_{i,k}$.

- **Step 1:** The marginal distributions $F_{i,k}$ are estimated from the set of q independent samples of the random variables $Z_{i,k}$.
- **Step 2:** Samples of $G_1, \dots, G_{md}$ are generated by the MATLAB function randn. These samples are mapped into samples of $Z_{i,k}^{PC}$ via (4.2) for given coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$ and are used to estimate the marginal distributions $F_{i,k}^{PC}$ of $Z_{i,k}^{PC}$.

- **Step 3:** Samples of $Z_{i,k}^{PCT}$ are calculated from (4.3) based on estimates of $F_{i,k}$ in step 1 and estimates of $F_{i,k}^{PC}$ corresponding to given coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$.
- **Step 4:** The objective functions for given coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$ are calculated from (3.3) and (3.5), where the histograms and spectral measures in the expressions of these functions are constructed by the MATLAB function histcounts2.
- **Step 5:** The MATLAB genetic algorithm is used to identify the optimal coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$ in the expressions of the objective functions.

The PC and PCT models are based on polynomial chaos of degree two for the numerical experiments of Figs. 17–19 and the left panel of Fig. 20. It takes approximately 12 hours for the PC model and approximately 15 hours for the PCT model to obtain the optimal coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}\}$ using MATLAB genetic algorithms. The numerical experiments of the right panel of Fig. 20 are based on the PCT model of tap 83 with polynomial chaos of degree four only. It takes approximately 3 hours for the PCT model of tap 83 to obtain the optimal coefficients $\{a_{i,k,j}, a_{i,k,j,l_1}, a_{i,k,j,l_1,l_2}, a_{i,k,j,l_1,l_2,l_3}\}$. We use the

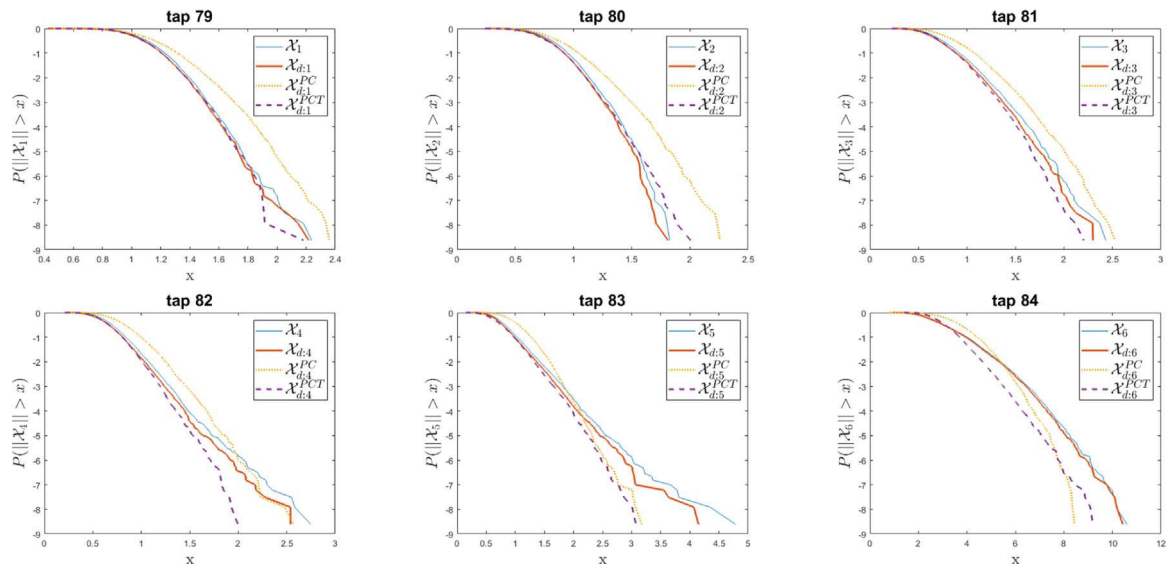


Fig. 19. Estimates of the target probability $P(\|X_i\| > x)$ (solid line), estimates based on FD model (heavy solid lines), estimates based on finite dimensional PC models (dotted lines) and estimates based on finite dimensional PCT models in logarithmic scale (dashed lines), $i = 1, \dots, 6$.

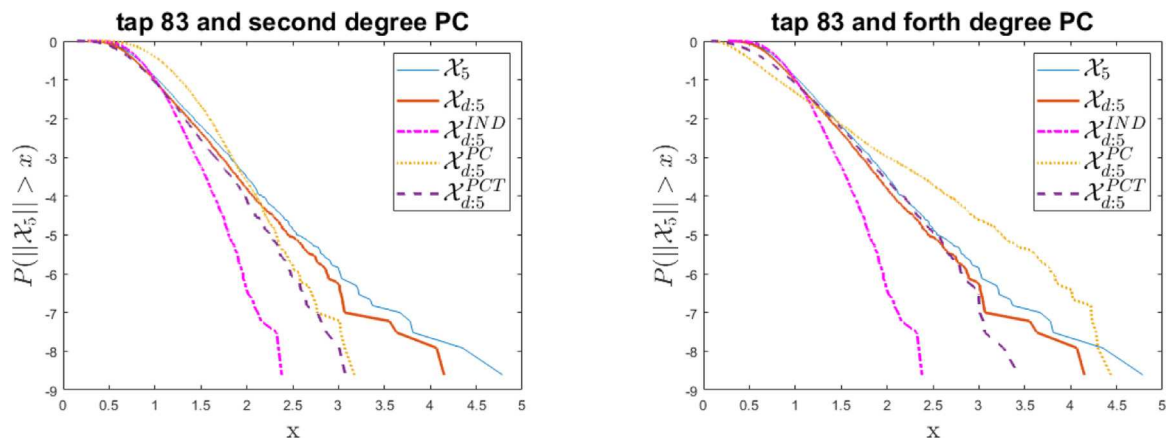


Fig. 20. Estimates of the target probability $P(\|X_5\| > x)$ (thin solid line), estimates based on FD model (heavy solid line), estimates based on finite dimensional independent model (dash-dotted line), estimates based on finite dimensional PC models (dotted line) and estimates based on finite dimensional PCT model in logarithmic scale (dashed line) based on second and fourth degree polynomial chaos (left and right panels).

MATLAB 2022a version for calculations. We perform all the numerical tests on a personal computer with a 3.6 GHz CPU and 16 GB RAM.

5. Conclusions

Finite dimensional (FD) models, i.e., deterministic functions of time and finite sets of random variables, have been constructed for a set of test cases and a wind pressure time series recorded at the UFBLWT facility in Gainesville by using polynomial chaos (PC) and polynomial chaos translation (PCT) models to represent their random coefficients. The components of PCT models are obtained from those of PC models by translation, so that they match exactly the target marginal distributions irrespective of the coefficients in their definition. The optimal values of the PCT coefficients minimize the discrepancy between the PCT and target joint properties, which are quantified by joint distributions and spectral measures. In summary, the PCT models match exactly the marginal distributions of the random coefficients of FD models by construction and capture their dependence with an accuracy that increases with the truncation level of the underlining PC models.

FD models with random coefficients represented by PC and PCT models have been constructed for a set of test cases and a 6-dimensional wind pressure time series recorded in the UFBLWT facility. The FD models with PCT random coefficients are superior to those with PC coefficients in the following sense. First, the PCT models provide a more accurate representation of the joint distributions of the random coefficients of FD models than the PC models. Second, the distributions of extremes of PCT-based FD models are similar to those of target time series while PC-based FD models do not have this capability. It is also shown that the performance of PCT-based FD models can be further improved by increasing their stochastic dimension and/or the order of their underlining PC models.

CRediT authorship contribution statement

Hui Xu: Writing – review & editing, Software, Formal analysis, Data curation. **Mircea D. Grigoriu:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Funding acquisition, Formal analysis. **Kurtis R. Gurley:** Writing – review & editing, Resources, Project administration, Investigation, Data curation.

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

The datasets generated during the current study are not publicly available, since they have been generated for particular applications, but are available from the corresponding author on reasonable 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. The wind tunnel datasets were generated through a grant from the USDA Specialty Crop Block Grant Program and the Florida Department of Agriculture and Consumer Services, United States. Special thanks to Professor Sanjay Shukla at the University of Florida for generously sharing these datasets.

References

- [1] Karpa O, Naess A. Extreme value statistics of wind speed data by the ACER method. *J Wind Eng Ind Aerodyn* 2013;112:1–10.
- [2] Gavanski E, Cook NJ. Evaluation of XIMIS for assessing extreme pressure coefficients. *Front Built Environ* 2019;5(48). <http://dx.doi.org/10.3389/fbuil.2019.00048>.
- [3] Peng X, Yang L, Gavanski E, Gurley K, Prevatt D. A comparison of methods to estimate peak wind loads on buildings. *J Wind Eng Ind Aerodyn* 2014;126:11–23. <http://dx.doi.org/10.1016/j.jweia.2013.12.013>.
- [4] Leadbetter MR, Lindgren G, Rootzén H. *Extremes and related properties of random sequences and processes*. Springer; 1983.
- [5] Che Y, Guo Z, Cheng C. Generalized polynomial Chaos-informed efficient stochastic Kriging. *J Comput Phys* 2021;445:110598.
- [6] Rossat D, Baroth J, Briffaut M, Dufour F. Bayesian inversion using adaptive polynomial Chaos Kriging within subset simulation. *J Comput Phys* 2022;455:110986.
- [7] Poëtte G, Brun E. Efficient uncertain k_{eff} computations with the Monte Carlo resolution of generalised polynomial Chaos based reduced models. *J Comput Phys* 2022;456:111007.
- [8] Cao L, Liu J, Jiang C, Liu G. Optimal sparse polynomial Chaos expansion for arbitrary probability distribution and its application on global sensitivity analysis. *Comput Methods Appl Mech Engrg* 2022;399:115368.
- [9] Novák L, Vořechovský M, Sadílek V, Shields MD. Variance-based adaptive sequential sampling for polynomial chaos expansion. *Comput Methods Appl Mech Engrg* 2021;386:114105.
- [10] Zhang B, Ni Y. A hybrid sequential sampling strategy for sparse polynomial Chaos expansion based on compressive sampling and Bayesian experimental design. *Comput Methods Appl Mech Engrg* 2021;386:114130.
- [11] Zhou Y, Lu Z, Cheng K. Adaboost-based ensemble of polynomial Chaos expansion with adaptive sampling. *Comput Methods Appl Mech Engrg* 2022;388:114238.
- [12] Lim H, Manuel L. Distribution-free polynomial Chaos expansion surrogate models for efficient structural reliability analysis. *Reliab Eng Syst Saf* 2021;205:107256.
- [13] Liu Z, Lesselier D, Sudret B, Wiart J. Surrogate modeling based on resampled polynomial Chaos expansions. *Reliab Eng Syst Saf* 2020;202:107008.
- [14] Mara TA, Becker WE. Polynomial chaos expansion for sensitivity analysis of model output with dependent inputs. *Reliab Eng Syst Saf* 2021;214:107795.
- [15] Zhang R, Dai H. A non-Gaussian stochastic model from limited observations using polynomial Chaos and fractional moments. *Reliab Eng Syst Saf* 2022;221:108323.
- [16] Zhou Y, Lu Z, Yun W. Active sparse polynomial Chaos expansion for system reliability analysis. *Reliab Eng Syst Saf* 2020;202:107025.
- [17] Wang Z, Ghanem R. An extended polynomial Chaos expansion for PDF characterization and variation with aleatory and epistemic uncertainties. *Comput Methods Appl Mech Engrg* 2021;382:113854.
- [18] Wang Z, Ghanem R. A functional global sensitivity measure and efficient reliability sensitivity analysis with respect to statistical parameters. *Comput Methods Appl Mech Engrg* 2022. <http://dx.doi.org/10.1016/j.cma.2022.115175>.
- [19] Kuo Hui-Hsiung. *Introduction to stochastic integration*. New York: Springer; 2006.
- [20] Grigoriu M. *Applied non-gaussian processes: examples, theory, simulation, linear random vibration, and MATLAB solutions*. Englewood Cliffs, NJ: Prentice Hall, Inc.; 1995.
- [21] Rudin W. *Real and complex analysis*. 3rd Ed.. New York: McGraw-Hill, Inc.; 1987.
- [22] Dür A. On the optimality of the discrete Karhunen–Loève expansion. *SIAM J Control Optim* 1998;36(6):1937–9.
- [23] Rosenblatt M. Remarks on a multivariate transformation. *Ann Math Statist* 1952;23:470–2.
- [24] Grigoriu M. PC translation models for random vectors and multivariate extremes. *SIAM J Sci Comput* 2019;41(2):A1228–51.
- [25] Resnick S. *Heavy-tail phenomena: probabilistic and statistical modeling*. New York: Springer; 2007.
- [26] Abraham S, Tsirikoglou P, Miranda J, Lacor C, Contino F, Ghorbaniasl G. Spectral representation of stochastic field data using sparse polynomial Chaos expansions. *J Comput Phys* 2018;367:109–20.
- [27] Alexanderian A, Gremaud PA, Smith RC. Variance-based sensitivity analysis for time-dependent processes. *Reliab Eng Syst Saf* 2019;196:106722.
- [28] Jacquelin E, Baldanzini N, Bhattacharyya B, Brizard D, Pierini M. Random dynamical system in time domain: A POD-PC model. *Mech Syst Signal Pr* 2019;133:106251.
- [29] Raisee M, Kumar D, Lacor C. A non-intrusive model reduction approach for polynomial Chaos expansion using proper orthogonal decomposition. *Int J Numer Meth Eng* 2015;103(4):293–312.
- [30] Sun X, Pan X, Choi J-Il. Non-intrusive framework of reduced-order modeling based on proper orthogonal decomposition and polynomial Chaos expansion. *J Comput Appl Math* 2021;390:113372.
- [31] Xiu D, Karniadakis GE. The Wiener–Askey polynomial Chaos for stochastic differential equations. *SIAM J Sci Comput* 2002;24(2):619–44.
- [32] Grigoriu M. *Linear dynamical systems*. 1st Ed.. Cham: Springer; 2021.