Metamodeling of structural systems with parametric uncertainty subject to stochastic excitation

M. Spiridonakos & E. Chatzi
Institute of Structural Engineering, D-BAUG, ETH Zurich,
Wolfgang-Pauli Str. 15, 8093 Zurich, Switzerland

ABSTRACT:
The problem of estimating reduced order metamodels for the accurate representation of computationally costly numerical models is presently addressed. The introduced metamodeling method is based on Nonlinear AutoRegressive models with eXogenous input (NARX), able to describe nonlinear dynamics, which moreover are characterized by random parameters utilized for the description of the uncertainty propagation. These random parameters are expanded onto a suitably defined finite-dimensional Polynomial Chaos (PC) basis and thus the resulting representation is fully described through a small number of deterministic coefficients of projection. The effectiveness of the proposed method is illustrated through its application on the metamodeling of a five-storey shear frame model paradigm. The results demonstrate the efficiency of the proposed methodology for accurate prediction and simulation of the numerical model dynamics.

1 INTRODUCTION

The assessment of the dynamic behavior of structural systems subjected to extreme loading conditions, such as earthquakes, is of particular importance for the safe operation of civil structures. Nevertheless, for the vast majority of such structures, it is hardly ever possible to perform dynamic response testing on the actual structure or even realistic simulation tests on an appropriately scaled structural model.

The advancements achieved in the field of Finite Element (FE) modeling methods have played a key role in overcoming this hurdle by enabling the realization of sophisticated simulation experiments emulating structural response (Bathe 2009). However, despite the rapidly growing computational power and the continuous development of increasingly efficient algorithms, the also growing complexity of FE models and the necessity for more detailed descriptions of both structural geometry and mechanical properties renders the use of highly detailed FE models almost prohibitive for complex, large structures (Gholizadeh & Salajegheh 2009). The problem is even more pronounced when taking into account that the structural systems are commonly characterized by parameter uncertainty to what concerns for instance the mechanical properties of the structure. Thus, the analyst is faced with the task of performing a number of simulations in order to obtain an accurate numerical model of an existing structure. Moreover, when the linear-

ity assumption is relaxed in favor of increased modeling accuracy the numerical model must be tested for a number of different excitation scenarios. Thus, for the extensive experimentation required for design optimization or model updating procedures based on time history loading, a simpler representation of the FE model should be considered which should also be able to describe the behavior of the structure for a wide range of excitations.

In this work, both structural properties and excitation uncertainties are treated by means of a time-series model with random parameters utilized for the description of uncertainty propagation through the numerical model. More specifically, the random responses of a large-scale numerical model are approximated by a suitably defined Polynomial Chaos Nonlinear AutoRegressive with eXogenous input (PC-NARX) metamodel which is able to describe both types of uncertainties by the expansion of its random model parameters onto a finite-dimensional PC basis (Spiridonakos & Chatzi 2012, Blatman & Sudret 2010). Thus, the PC-NARX metamodel is described by a small number of deterministic coefficients of projection, which may be estimated through the ordinary least squares method.

The method's effectiveness is demonstrated through the identification of a metamodel for a five-storey building with nonlinear material properties. Toward this end, a limited number of simulation experiments is conducted with the numerical model

being subjected to a number of different realizations of synthetic earthquakes. Synthetic earthquakes are produced by filtering a white noise process through a non-stationary impulse response filter and a non-stationary modulating function in order to simulate the time-varying characteristics for both the temporal and spectral content of a real earthquake (Rezaeian & Kiureghian 2010). The non-stationary filter and the modulating functions are described by a small number of uncertain parameters, with sample observations of the latter being estimated through fitting of the model to real recorded earthquake ground motion signals.

2 POLYNOMIAL CHAOS NARX MODELS

Metamodeling refers to the process of identifying a reduced order and computational complexity representation of a large scale numerical model. In the present work, a metamodel is sought for the accurate representation of the dynamics of the numerical model and the accurate simulation of its time history loading response.

Let us consider a structural system represented by a numerical model $\mathcal{M}$ that is characterized by a number of input parameters related with the properties of the modeled structure (mechanical and/or geometric). It is assumed that M_1 of these parameters are subject to uncertainty and that they may be described by independent random variables gathered in a random vector $\boldsymbol{\xi}_S = [\xi_1, \xi_2, \ldots, \xi_{M_1}]^\mathsf{T}$. Superscript T denotes transpose of a vector or a matrix while it is noted that for simplicity of notation no distinction is made between a random variable and its value(s).

Since the dynamic behavior of a nonlinear numerical model is also a function of the input excitation characteristics, it is considered that the set of excitation signals used for the dynamic loading of the numerical model may be described by a small number of random variables $\boldsymbol{\xi}_X = [\xi_{M_1+1}, \xi_{M_1+2}, \ldots, \xi_{M_1+M_2}]^\mathsf{T}$.

As a result of the uncertainty propagation, the dynamic response of the numerical model to a given input excitation will also be a random variable which in addition depends on time, that is

$$y[t, \boldsymbol{\xi}] = \mathcal{M}(x[1, \boldsymbol{\xi}_X], x[2, \boldsymbol{\xi}_X], \ldots, x[t, \boldsymbol{\xi}_X], \boldsymbol{\xi}_S) \quad (1)$$

with $t = 1, 2, \ldots, T$ designating normalized by the sampling period discrete time, $x[t, \boldsymbol{\xi}_X]$ the excitation input signal, $y[t, \boldsymbol{\xi}]$ the corresponding numerical model response signal, and $\boldsymbol{\xi} = [\boldsymbol{\xi}_S^\mathsf{T}, \ \boldsymbol{\xi}_X^\mathsf{T}]^\mathsf{T}$ the M-dimensional ($M = M_1 + M_2$) complete vector of input random variables with known joint probability density function (pdf) $f(\boldsymbol{\xi})$.

In order to approximate the numerical model dynamic response $y[t, \boldsymbol{\xi}]$ for every realization of $\boldsymbol{\xi}$ in an efficient way, a novel metamodeling method based on Polynomial Chaos Nonlinear AutoRegressive with eXogenous input (PC-NARX) models is introduced

in the present study. The general PC-NARX model, in the linear-in-the-parameters form, is given by the following relationship:

$$y[t] = \sum_{i=1}^{n_\theta} \theta_i(\boldsymbol{\xi}) \cdot g_i(\boldsymbol{z}[t]) + e[t] \quad (2)$$

where $g_i(\boldsymbol{z}[t])$ are the nonlinear model terms generated from the regression vector $\boldsymbol{z}[t] = [y[t-1], \ldots, y[t-n_a], x[t], \ldots, x[t-n_b]]^\mathsf{T}$ with n_a, n_b designating the maximum output and input time lags, respectively, and $e[t] \sim \mathrm{NID}(0, \sigma_e^2)$ the model's residual sequence with $\mathrm{NID}(\cdot, \cdot)$ denoting Normally Independently Distributed process with the indicated mean and variance. It should be mentioned that the model terms $g_i(\boldsymbol{z}[t])$ may be constructed from a variety of local or global basis functions including polynomials, splines, neural networks, wavelets and others (Wei & Billings 2009).

The important feature of PC-NARX models, in comparison with the conventional NARX models (Chen & Billings 1989), is that they are characterized by parameters $\theta_i(\boldsymbol{\xi})$ which are random variables themselves. The latter are actually represented by a deterministic mapping which describes their relation with the input random variables. More specifically, assuming that the PC-NARX model parameters $\theta_i(\boldsymbol{\xi})$ have finite variance, they admit the following polynomial chaos representation (Soize & Ghanem 2004):

$$\theta_i(\boldsymbol{\xi}) = \sum_{j=1}^{\infty} \theta_{i,j} \cdot \phi_{\boldsymbol{d}(j)}(\boldsymbol{\xi}) \quad (3)$$

where $\theta_{i,j}$ are unknown deterministic coefficients of projection, $\boldsymbol{d}(j)$ is the multi-indices of the multivariate polynomial basis, and $\phi_{\boldsymbol{d}(j)}$ are multivariate basis functions that are orthonormal with respect to the joint pdf of $\boldsymbol{\xi}$, that is:

$$E[\phi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \phi_{\boldsymbol{\beta}}(\boldsymbol{\xi})] = \delta_{\boldsymbol{\alpha}, \boldsymbol{\beta}} = \begin{cases} 1 & \text{for } \boldsymbol{\alpha} = \boldsymbol{\beta} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Each probability density function may be associated with a well known family of orthogonal polynomials. For instance, normal distribution is associated with Hermite polynomials while uniform distribution with Legendre. A list of the most common probability density functions along with the corresponding orthogonal polynomials and the relations for their construction may be found in (Soize & Ghanem 2004).

For purposes of practicality, the infinite series of expansion of equation (3) must be truncated by selecting an appropriate functional subspace consisting of a finite number of terms p. In this way, the resulting PC-NARX model is fully parametrized in terms of a finite number of deterministic coefficients of projection $\theta_{i,j}$, while the complete PC-NARX identification problem consists of the subproblems of model param-

eter estimation and model structure selection, which are discussed in the following sections.

At this point it should be noted that PC-ARIMA models with a-priori known deterministic coefficients of projection have been considered before for the characterization of terrain topology by Wagner and Ferris (Wagner & Ferris 2007). Moreover, linear PC-ARX models have used in the metamodeling context in (Spiridonakos & Chatzi 2012). Similarly linear ARX models with functionally dependent parameters have also been used in a number of studies for the purposes of structural identification and damage detection (for instance see Kopsaftopoulos & Fassois 2012 and the references therein). However, in these studies the input parameters are not related with a known probability density function and thus the basis is constructed from an arbitrarily selected family of orthogonal basis functions.

2.1 Parameter estimation

As already mentioned, the estimation of a PC-NARX metamodel refers to the determination of the coefficients of projection parameter vector $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = [\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{n_\theta,p}]^\mathsf{T} \tag{5}$$

The calculation is based on the availability of time history data for the input excitation and output response of the numerical model. These may be acquired for a series of simulations, conducted for different realizations of the input random vector using the full scale numerical model.

Let us consider a series of K simulations conducted for a corresponding number of input random vector realizations $\boldsymbol{\xi}_k = [\xi_{k,1}, \xi_{k,2}, \ldots, \xi_{k,M}]^\mathsf{T}$ (for $k = 1, 2, \ldots, K$), and the resulting set of set of excitation signals $x_k^T = \{x_k[1], x_k[2], \ldots, x_k[T]\}$. It is noted that for the purposes of notational simplicity the dependency of the input and output signals on $\boldsymbol{\xi}_k$ is not indicated explicitly in the following relationships.

The corresponding dynamic response of the full scale numerical model is indicated as $y_k^T = \{y_k[1], y_k[2], \ldots, y_k[T]\}$ and as already mentioned it is assumed to follow the general PC-NARX model of equation (2):

$$y_k[t] = \sum_{i=1}^{n_\theta} \theta_i(\boldsymbol{\xi}_k) \cdot g_i(z_k[t]) + e_k[t], \tag{6a}$$

$$e_k[t] \sim \text{NID}(0, \sigma_{e_k}), \tag{6b}$$

$$E\{e_i[t], e_j[t-\tau]\} = \begin{cases} \sigma_{e_i}^2 \cdot \delta_{t,\tau} & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{6c}$$

In the relationships above, the uncorrelatedness of the residual series between different simulation experiments is also assumed. For these simulations, the input vector $\boldsymbol{\xi}_k$ is generated from the input parameter space either randomly or by using a structured sampling technique, such as the Latin Hypercube Sampling (LHS; Helton & Davis 2003).

According to the intended use of the metamodel the estimation of the parameter vector $\boldsymbol{\theta}$ may be based upon the minimization of either the Prediction Error (PE) criterion or Simulation Error (SE) criterion.

2.1.1 Prediction error method

When the structural system corresponding to the numerical model is appropriately instrumented, the PC-NARX PE-based estimation method may provide metamodels for the on-line prediction of its dynamic response, which in turn may be used for a number of purposes such as vibration control, structural health monitoring, model updating and others.

The PE criterion consists of the sum of squares of the model's one-step-ahead prediction errors for the complete set of simulation experiments. As it may be easily shown the one-step-ahead prediction errors coincide with the model's residual sequence and thus $\boldsymbol{\theta}$ has to be estimated by the following minimization problem:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{k=1}^{K} \sum_{t=1}^{T} (y_k[t] - \hat{y}_k[t|t-1])^2 \right\}$$

$$= \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{k=1}^{K} \sum_{t=1}^{T} e_k^2[t] \right\} \tag{7}$$

with $\hat{y}_k[t|t-1]$ designating the model's one-step-ahead prediction and $\arg\min$ minimizing argument. Toward this end, and by using equation (3), equation (6a) may be rewritten as:

$$y_k[t] = \sum_{i=1}^{n_\theta} \sum_{j=1}^{p} \theta_{i,j} \cdot \phi_{\boldsymbol{d}(j)}(\boldsymbol{\xi}_k) \cdot g_i(z_k[t]) + e_k[t] \Rightarrow$$

$$y_k[t] = \underbrace{\begin{bmatrix} \phi_{\boldsymbol{d}(1)}(\boldsymbol{\xi}_k) g_1(z_k[t]) \\ \vdots \\ \phi_{\boldsymbol{d}(p)}(\boldsymbol{\xi}_k) g_{n_\theta}(z_k[t]) \end{bmatrix}}_{\phi[\boldsymbol{\xi}_k, t]}^\mathsf{T} \begin{bmatrix} \theta_{1,1} \\ \vdots \\ \theta_{n_\theta,p} \end{bmatrix} + e_k[t] \tag{8}$$

or by stacking all time-instants in a single vector

$$\underbrace{\begin{bmatrix} y_k[1] \\ y_k[2] \\ \vdots \\ y_k[T] \end{bmatrix}}_{\boldsymbol{y}_{k[T \times 1]}} = \underbrace{\begin{bmatrix} \phi[\boldsymbol{\xi}_k, 1]^\mathsf{T} \\ \phi[\boldsymbol{\xi}_k, 2]^\mathsf{T} \\ \vdots \\ \phi[\boldsymbol{\xi}_k, T]^\mathsf{T} \end{bmatrix}}_{\boldsymbol{\Phi}(\boldsymbol{\xi}_k)_{[T \times n_\theta p]}} \cdot \underbrace{\begin{bmatrix} \theta_{1,1} \\ \vdots \\ \theta_{n_\theta,p} \end{bmatrix}}_{\boldsymbol{\theta}_{[n_\theta p \times 1]}} + \underbrace{\begin{bmatrix} e_k[1] \\ e_k[2] \\ \vdots \\ e_k[T] \end{bmatrix}}_{\boldsymbol{e}_{k[T \times 1]}}$$

where subscripts in brackets indicate the respective matrix/vector dimensions. Finally, by pooling all the available simulation experiments the following linear

regression model is obtained:

$$
\underbrace{\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix}}_{\boldsymbol{Y}_{[KT \times 1]}} = \underbrace{\begin{bmatrix} \boldsymbol{\Phi}(\boldsymbol{\xi}_1) \\ \boldsymbol{\Phi}(\boldsymbol{\xi}_2) \\ \vdots \\ \boldsymbol{\Phi}(\boldsymbol{\xi}_K) \end{bmatrix}}_{\boldsymbol{\Phi}(\boldsymbol{\xi})_{[KT \times n_\theta p]}} \cdot \underbrace{\begin{bmatrix} \theta_{1,1} \\ \vdots \\ \theta_{n_\theta,p} \end{bmatrix}}_{\boldsymbol{\theta}_{[n_\theta p \times 1]}} + \underbrace{\begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_K \end{bmatrix}}_{\boldsymbol{E}_{[KT \times 1]}} \quad (9)
$$

where $\boldsymbol{\Phi}(\boldsymbol{\xi})$ is the regression matrix. Thus, the minimization problem of equation (7) may be rewritten as follows:

$$
\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \boldsymbol{E}^\mathsf{T} \cdot \boldsymbol{E} \right\} \quad (10)
$$

which due to the linear dependence of the residual sequence $\boldsymbol{E}$ on the parameter vector $\boldsymbol{\theta}$ leads to a linear ordinary least squares estimator for the latter, that is:

$$
\widehat{\boldsymbol{\theta}} = \left(\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\xi}) \cdot \boldsymbol{\Phi}(\boldsymbol{\xi}) \right)^{-1} \cdot \left(\boldsymbol{\Phi}(\boldsymbol{\xi})^\mathsf{T} \cdot \boldsymbol{Y} \right) \quad (11)
$$

2.1.2 *Simulation error method*

The SE estimation method must be employed when a PC-NARX metamodel is intended to be used for simulation purposes. In this case, a metamodel that may replace the numerical model for additional simulations is sought.

Let us consider again the same set of K input excitation signals $x_k^T (k = 1, 2, \ldots, K)$. For each x_k^T the simulated response $\bar{y}_k[t]$ of a PC-NARX metamodel may be obtained by using the input excitation signals and the following relationship applied recursively with respect to time t:

$$
\bar{y}_k[t] = \sum_{i=1}^{n_\theta} \theta_i(\boldsymbol{\xi}_k) \cdot g_i(\bar{\boldsymbol{z}}[t]) \quad (12)
$$

with given initial conditions $\{\bar{y}_k[1 - n_a], \ldots, \bar{y}_k[0]\}$ and $\{x_k[1 - n_b, \boldsymbol{\xi}_X], \ldots, x_k[0, \boldsymbol{\xi}_X]\}$. The responses $\bar{y}_k[t]$ of the metamodel should be as close as possible to that of the numerical model $y_k[t]$, thus the estimation of a PC-NARX metamodel may be based on the minimization of the following SE criterion:

$$
\widehat{\boldsymbol{\theta}}_s = \arg\min_{\boldsymbol{\theta}_s} \left\{ \sum_{k=1}^{K} \sum_{t=1}^{T} (y_k[t] - \bar{y}_k[t])^2 \right\}
$$

$$
= \arg\min_{\boldsymbol{\theta}_s} \left\{ \sum_{k=1}^{K} \sum_{t=1}^{T} \varepsilon_k^2[t] \right\} \quad (13)
$$

This optimization problem may be solved by iterative nonlinear optimization methods initialized by the parameter vector $\boldsymbol{\theta}$ obtained through the PE method of section 2.1.1.

2.2 *Model structure selection*

Model structure selection may be considered as the optimization procedure during which models corresponding to various candidate "structures" are estimated (through the methods of the previous section), and the one providing the *"best fitness"* to the simulation data is selected.

The PC-NARX model structure selection procedure concerns the determination of: *i)* the nonlinear model terms $g_i(\boldsymbol{z}[t])$ and *ii)* the PC basis functional subspace.

2.2.1 *NARX terms*

Given the specific family of functions (e.g. polynomials, splines, neural networks, or wavelets), and the maximum output and input lags n_a and n_b, a finite number of appropriate nonlinear terms must be selected. This selection process cannot be based on a trial-and-error approach, since the dimension of the initial search space is usually excessive.

For this reason a more automated approach must be employed. The orthogonal least squares scheme introduced by Chen & Billings (1989) along with the PE estimation method may be used towards this end. This scheme transforms the linear regression model of equation (9) into an auxiliary one given as

$$
\mathcal{Y} = \boldsymbol{\Psi}(\boldsymbol{\xi}) \cdot \boldsymbol{\eta} + \mathcal{E} \quad (14)
$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi})$ is a transformed regressor matrix consisting of orthogonal columns such as $\boldsymbol{\Psi}_i(\boldsymbol{\xi}) \cdot \boldsymbol{\Psi}_j(\boldsymbol{\xi}) = 0$ for $i \neq j$ and $i, j = 1, \ldots, n_\theta p$, while $\boldsymbol{\eta}$ is the corresponding transformed coefficients of projection parameter vector. The orthogonalization of the initial regression model may be realized by applying the Gram-Schmidt procedure and the Orthogonal Forward Regression (OFR) algorithm as described in Chen & Billings (1989). The orthogonality of the transformed regressors permits estimating the novel parameters η_i one at a time. As a consequence, the significance of each regressor term towards the reduction in the total mean square error may be quantified. Hence, the candidate terms are arranged in the order of significance and the most relevant model terms are selected.

2.2.2 *PC basis*

PC functional subspace selection may be formulated as a discrete variable selection problem that may be tackled through a Genetic Algorithm (GA) that minimizes the the Bayesian Information Criterion ("fitness" function) (BIC; Ljung 1999):

$$
\mathrm{BIC} = -\ln(\hat{\sigma}_{\boldsymbol{E}}^2) + \frac{\ln(KT)}{KT} \cdot \dim \boldsymbol{\theta} \quad (15)
$$

where $\hat{\sigma}_{\boldsymbol{E}}^2$ designates the estimation of the residual sequence $\boldsymbol{E}$ variance, and $\dim \boldsymbol{\theta}$ the number of indepen-
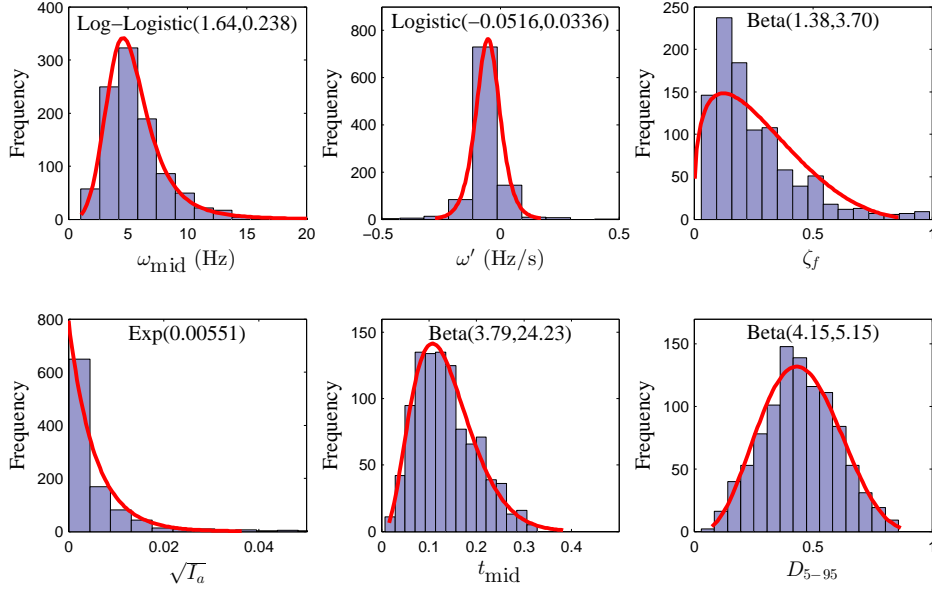
Figure 1: Histograms of the stochastic ground motion model parameters along with the fitted probability density function.

dently adjusted model parameters, that is the number of the deterministic coefficients of projection. Finally, KT is the total number of data samples used for the estimation of the PC-NARX model.

In order to illustrate this procedure consider an initial PC basis with total maximum degree $P = 2$ and $M = 3$ input random variables. The dimensionality of this basis will be equal to (Blatman & Sudret 2010):

$$p = \frac{(M+P)!}{M!P!} = \frac{5!}{3!2!} = 10$$

In order to represent all the functional subspaces included in this initial search space through a compact 10-bit binary vector, each multi-index vector $\boldsymbol{d}(j)$ $(j = 1, \ldots, 10)$ may be represented by a single bit indicating the existence (1) or not (0) of the corresponding PC basis function in the specific model structure.

3 PARAMETRIC REPRESENTATION OF SYNTHETIC EARTHQUAKES

In the present study, the stochastic ground motion model proposed in Rezaeian & Kiureghian 2010 is employed for the parametric representation of synthetic earthquake ground motion acceleration signals. According to this model the ground motion is produced by time-modulating a normalized filtered white noise process. In order to simulate the time-varying characteristics of both the temporal and spectral content of a real earthquake, the modulating function and the Impulse Response Filter (IRF) of the aforementioned model have non-stationary properties.

More specifically, the non-stationary modulating function is defined as a gamma function of the following form:

$$q(t, \boldsymbol{\alpha}) = \alpha_1 t^{\alpha_2 - 1} \exp(-\alpha_3 t) \qquad (16)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^{\mathsf{T}}$, with $\alpha_1 > 0$ the parameter of the process intensity, $\alpha_2 > 1$ the parameter controlling the shape of the modulating function, and $\alpha_3 > 0$ the duration of the motion. The parameters of the modulating function, may be directly estimated from three time-domain characteristics of a ground motion accelerogram: *i)* the expected Arias intensity I_a, that is a measure of the total energy contained in the motion, *ii)* the effective duration of the motion D_{5-95} (the time interval between the instants at which the 5 and 95% of the expected Arias intensities are reached), and finally *iii)* the time t_{mid} at which 45% level of the expected Arias intensity is reached.

On the other hand, the non-stationary IRF is given by the following expression:

$$h[t - \tau] = \frac{\omega_f(\tau) e^{-\zeta_f \omega_f(\tau)(t-\tau)} \cdot \sin\left[\omega_f(\tau)(t-\tau)\sqrt{1-\zeta_f^2}\right]}{\sqrt{1-\zeta_f^2}} \quad (17)$$

for $\tau \leq t$, with ζ_f designating the damping ratio of the filter and $\omega_f(\tau)$ the filter's frequency. The latter is defined as $\omega_f(\tau) = \omega_{\text{mid}} + \omega'(\tau - t_{\text{mid}})$ with ω_{mid} representing the filter frequency at t_{mid}, and ω' the rate of change of the filter frequency with time.

Therefore, given a target accelerogram, the parameters $\omega_{\text{mid}}, \omega', \zeta_f, I_a, D_{5-95}$, and t_{mid} may be identified by matching the properties of the recorded motion with the corresponding statistical measures of the stochastic ground motion model (Rezaeian & Kiureghian 2010). The identified model may be subsequently used for the generation of synthetic ground motion accelerograms with similar time-frequency characteristics with that of the real recorded motion.

In a similar way, a pdf may be identified for each of the stochastic ground motion model parameters in order to represent not a single but a set of real earthquake accelerograms. This procedure is presently applied for the modeling of the 3551 ground motion signals of the PEER database (horizontal components
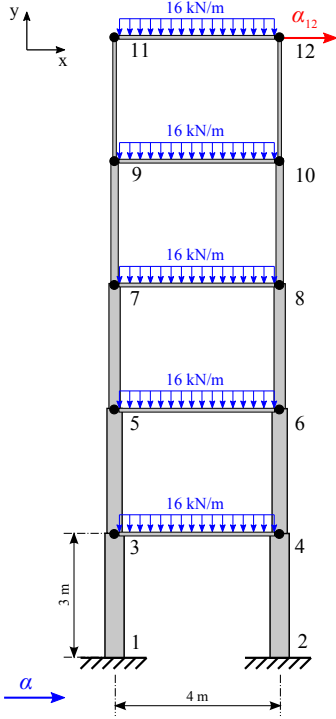
Figure 2: The five-storey shear frame model.

Table 1: Geometric and mechanical properties of the five-storey shear frame model.

Geometric		Mechanical	
Cross-sectional area	cm^2		
1st storey columns	900	Poisson ratio	0.29
2nd storey columns	625	Density (kg/m^3)	7850
3rd storey columns	400	Yield stress (MPa)	200
4th storey columns	225	Tangent modulus (GPa)	10
5th storey columns	100		
Horizontal beams	100		

only; PEER 2012). From the identified sets of parameters, only the 1000 with the best model fitting are kept for the subsequent analysis. The resulting histograms for the identified parameters and the fitted pdfs are shown in figure 1 (D_{5-95} and t_{mid} are shown in normalized discrete time t/T).

4 NUMERICAL EXAMPLE

A FE model of a building frame structure is presently considered for the validation of the introduced method (figure 2). The identification of the frame's metamodel is based on recordings of the horizontal acceleration of the top floor of the building measured at node 12 ($y[t] = \alpha_{12}[t]$; figure 2) obtained by the time history loading of the FE model with various synthetic earthquake ground motion signals.

The elements of the shear frame are considered to have square cross section and being made of steel with isotropic behavior described by a bilinear stress-strain curve. The constant, characterized by negligible uncertainty, mechanical and geometric properties of the structure are summarized in table 1. On the other hand, the steel's Young modulus E is considered to be uncertain, modeled by a random variable following a
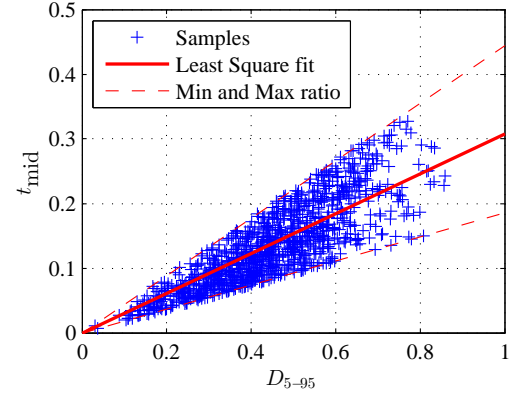

Figure 3: Identified values of D_{5-95} versus t_{mid} for the accelerograms of the PEER ground motion database.

Table 2: Random input variables.

Variable	Distribution	pdf parameters
E (GPa)	Uniform	$\min = 180, \max = 220$
ω_{mid} (Hz)	Log-Logistic	$\alpha = 1.64, \beta = 0.238$
ω' (Hz/s)	Logistic	$\mu = -0.0516, \sigma = 0.0336$
$\sqrt{I_a}$	Exponential	$\mu = 0.00551$
D_{5-95}	Beta	$\alpha = 4.15, \beta = 5.15$

uniform distribution $E \sim \mathcal{U}(180, 220)$ (GPa).

The uncertain input parameter vector of this numerical example, includes also the stochastic ground motion model parameters $\omega_{mid}, \omega', I_a$, and D_{5-95} which follow the distributions indicated in figure 1. For this case study the IRF filter damping ratio ζ_f is considered constant and equal to 0.1 for all synthetic earthquake accelerograms. Moreover, the high correlation between the D_{5-95} and t_{mid} variables admits the usage of a linear relationship between this two variables (see figure 3). Thus, using a least square fitting t_{mid} is considered always equal to $0.3075 D_{5-95}$.

Summarizing, in total five independent input random variables with known pdfs are considered for this metamodeling problem (see table 2).

4.1 Simulation experiments

A total number of 200 simulations is conducted ($K = 200$) for a corresponding number of input random vector realizations $\boldsymbol{\xi}_k (k = 1, 2, \ldots, 200)$ sampled by the LHS method. The values of the sampled variables along with their histograms are shown in figure 4.

For each simulation experiment, the shear frame model is excited by a synthetic ground motion accelerogram applied in the x-axis direction (sampling frequency $f_s = 40$ Hz; 1000 number of samples for each accelerogram). The Peak Ground Acceleration (PGA) values of the synthetic accelerograms varies between 0.16 and 0.67 g with all the levels of this range being capable of exciting the nonlinear dynamics of the FE model (for instance see figure 5(d)). The acceleration of the FE model measured at node 12, along with the corresponding shear stress and top floor displacement recordings for the first three simulations are also shown in figure 5.
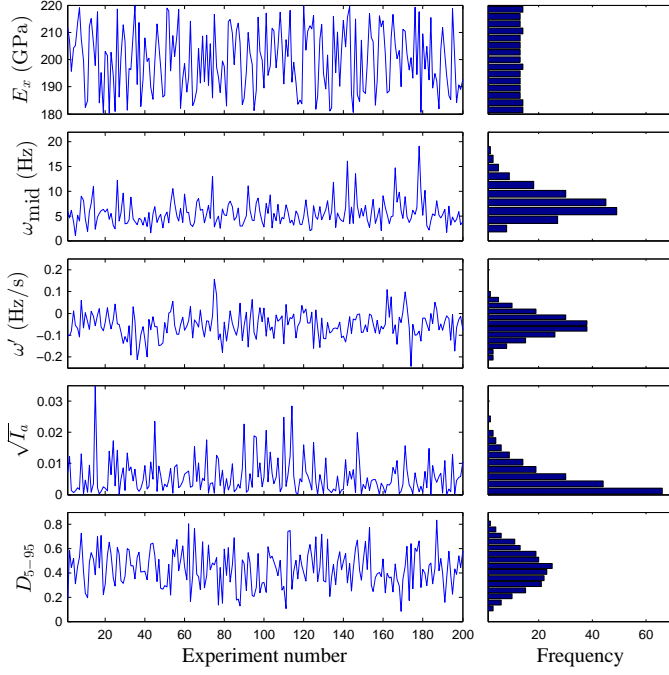
Figure 4: The input random vector realizations for the 200 simulations conducted.

4.2 PC-NARX metamodel identification

For the metamodel structure selection problem PC-NARX models with $n_a = n_b = 10$ are considered, while the appropriate nonlinear terms $g_i(\boldsymbol{z}[t])$ are searched among polynomial functions of the following form:

$$g_i(\boldsymbol{z}[t]) = z_{j_1}^{\ell_1}[t] \cdot z_{j_2}^{\ell_2}[t]$$

with $\ell_1, \ell_2 = 0, \ldots 3$, $\ell_1 + \ell_2 \leq 3$, and $\boldsymbol{z}[t] = [\, y[t-1], \ldots, y[t-10], x[t], x[t-1], \ldots, x[t-10]\,]^{\mathsf{T}}$. The application of the OFR algorithm for the selection of the most significant terms leads to the selection of the following 46 linear and nonlinear terms:

$y[t-1], \ldots, y[t-10], x[t], x[t-1], \ldots, x[t-10],$

$y[t-1] \cdot y^2[t-2], \ldots, y[t-1] \cdot y^2[t-10],$

$y^2[t-1] \cdot y[t-2], \ldots, y^2[t-1] \cdot y[t-10],$

$y^3[t-1], y^3[t-2], y^3[t-3].$

The GA optimization procedure outlined in section 2.2.2 is then utilized for the selection of the appropriate PC basis functions. The initial search space is defined by the truncated set of multivariate Legendre polynomials basis functions of maximum total degree equal to three – it is noted that the input random variables listed in table 2 are first transformed into standard uniform variables by using the inverse cumulative density functions of the corresponding pdfs. The minimization of the BIC criterion through the GA leads to the selection of nine basis functions.

The multi-indices vectors of these functions are given in table 3. It should be observed that according to these results the PC-NARX metamodel parameters θ_i and thus the dynamic properties of the corresponding numerical model seem to be highly sensitive to the input variable ω_{mid} which defines the dominant fre-
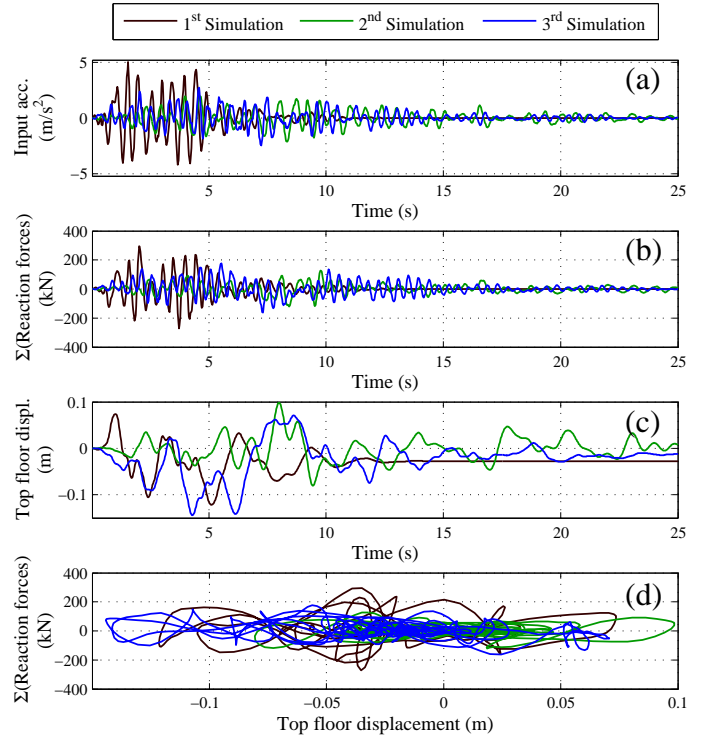

Figure 5: Input-output data from the first three simulation experiments: (a) Input excitation signals x_1^T, x_2^T and x_3^T, (b) node 12 acceleration response signals y_1^T, y_2^T and y_3^T, (c) top floor displacement measured at node 12, and (d) the shear stresses versus top floor displacements.

Table 3: Multi-indices vectors of the selected PC basis functions.

	$\xi_1(E)$	$\xi_2(\omega_{\mathrm{mid}})$	$\xi_3(\omega')$	$\xi_4(I_a)$	$\xi_5(D_{5-95})$
$\boldsymbol{d}(1)$	0	0	0	0	0
$\boldsymbol{d}(2)$	1	0	0	0	0
$\boldsymbol{d}(3)$	0	1	0	0	0
$\boldsymbol{d}(4)$	0	0	0	0	1
$\boldsymbol{d}(5)$	1	0	0	1	0
$\boldsymbol{d}(6)$	0	1	0	0	1
$\boldsymbol{d}(7)$	0	2	0	0	0
$\boldsymbol{d}(8)$	1	2	0	0	0
$\boldsymbol{d}(9)$	0	3	0	0	0

quency of the synthetic ground motion signal used for the excitation of the FE model. The significant sensitivity to the Young modulus variables comes also as no surprise since it is directly related to the dynamics of the modeled structure. On the other hand, the results of table 3 indicate no sensitivity of the metamodel parameters θ_i on the rate of change of the synthetic ground motion frequency ω'. Finally, interesting is the result that the PC basis functions indicate low sensitivity to the input variable $\sqrt{I_a}$ which controls the intensity of the ground motion. However, this fact may attributed to the relatively narrow PGA range of the synthetic earthquakes.

The estimated, based on the PE method, PC-NARX metamodel is also successfully validated with respect to the assumptions on the residual sequences (see equation 6) through the auto-correlation and cross-correlation functions. The uncrosscorrelatedness between the residual sequences $e_k^T(k = 1, \ldots, K)$ may be illustrated through the almost diagonal covariance matrix $\mathrm{cov}(\hat{\boldsymbol{e}}_{k_1}, \hat{\boldsymbol{e}}_{k_2})$ depicted in figure 6.
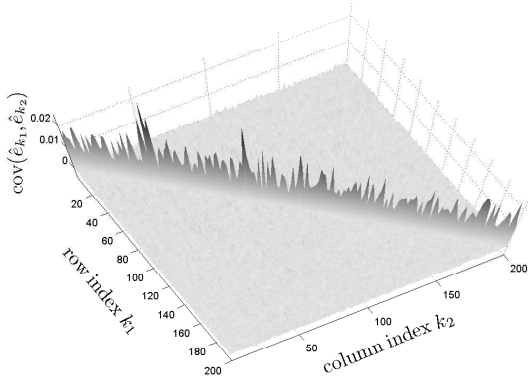
Figure 6: Covariance matrix between the estimated PC-NARX model residual sequences $\hat{e}_k(k = 1, \ldots, K)$.
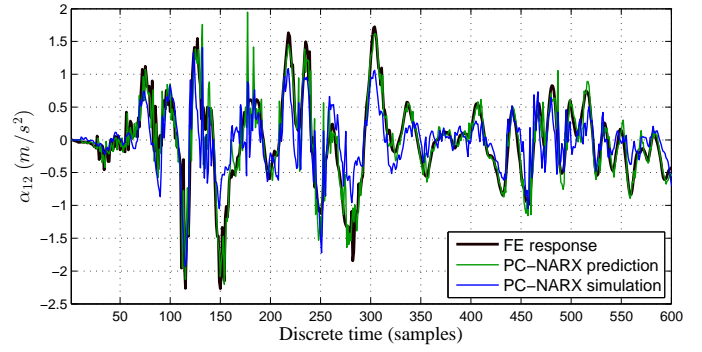

Figure 7: The dynamic response of the shear frame FE model and the corresponding PC-NARX based one-step-ahead-predictions and simulation signals for the El Centro earthquake time history loading.

The performance of the estimated metamodel is assessed through its application for both prediction and simulation of the dynamic response of the FE model excited by the N-S component of the El Centro earthquake accelerogram. The Young modulus of the shear frame for this test case is selected as $E = 200$ GPa, while the rest of the input random variables are estimated from the actual accelerogram as $\omega_{mid} = 3.8227$ (Hz), $\omega' = -0.0202$ (Hz/s), $\sqrt{I_a} = 0.0188$ and $D_{5-95} = 0.4544$ (sampling frequency equal to 40 Hz; 2688 samples).

The predictions of the PE-based and the simulations of the SE-based estimated PC-NARX metamodels are contrasted to the dynamic response of the numerical model in figure 7 (first 600 samples; for the SE method: Levenberg-Marquardt algorithm, *lsqnonlin* MATLAB function, TolFun $= 1 \times 10^{-6}$, TolX $= 1 \times 10^{-9}$). As it may be observed, the estimated metamodel is capable of reproducing the dynamic response of the numerical model with good accuracy. It is worth noting that these results are obtained for an excitation signal for which the assumptions of $\zeta_f = 0.1$, and $t_{mid} = 0.3075 D_{5-95}$ are no longer valid. Finally, it should be added that the PC-NARX based simulated response was calculated more than 100 times faster than that of the FE model.

5   CONCLUSIONS

This work introduces a methodology for the meta-modeling of large scale numerical models with uncertain parameters and under stochastic excitation. The metamodel utilized is based on Polynomial Chaos Nonlinear ARX models with parameters which are random variables depending on the input random variables, with this dependency being described by their expansion of a properly selected polynomial chaos basis.

The method is applied for the metamodling problem of a five-storey shear frame FE model subject to synthetic earthquake ground motion accelerograms. Input random variables are used to describe the uncertainties of both the material properties of the simulated structure and the input excitation signals. The estimation of the PC-NARX model is based on a set of 200 input-output simulation data sets with the results being very promising.

REFERENCES

Bathe, K.-J. (2009). Finite element method. In B. Wah (Ed.), *Wiley Encyclopedia of Computer Science and Engineering*, pp. 1253–1264. Wiley & Sons, Inc.

Blatman, G. & B. Sudret (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics 25*(2), 183–197.

Chen, S. & S. Billings (1989). Modelling and analysis of non-linear time series. *International Journal of Control 50*(6), 2151–2171.

Gholizadeh, S. & E. Salajegheh (2009). Optimal design of structures subjected to time history loading by swarm intelligence and an advanced metamodel. *Computer Methods in Applied Mechanics and Engineering 198*, 2936–2949.

Helton, J. & F. Davis (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety 81*(1), 23–69.

Kopsaftopoulos, F. & S. Fassois (2012). A stochastic functional model based method for vibration based damage detection, localization, and magnitude estimation. *Mechanical Systems and Signal Processing in press*, 1–19.

Ljung, L. (1999). *System Identification: Theory for the User* (2nd ed.). Prentice Hall PTR.

PEER (2012). Peer ground motion database: http://peer.berkeley.edu/peer_ground_motion_database.

Rezaeian, S. & A. D. Kiureghian (2010). Simulation of synthetic ground motions for specified earthquake and site characterics. *Earthquake Engineering and Structural Dynamics 39*(10), 1155–1180.

Soize, C. & R. Ghanem (2004). Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing 26*(2), 395–410.

Spiridonakos, M. & E. Chatzi (2012). Metamodeling of structural systems through polynomial chaos arx models. In *International Conference on Uncertainty in Structural Dynamics (USD2012), Leuven, Belgium, September 17-19, 2012*.

Wagner, S. & J. Ferris (2007). A polynomial chaos approach to arma modeling and terrain characterization. In *SPIE 6564, Modeling and Simulation for Military Operations II, 65640M*.

Wei, H. & S. Billings (2009). Improved parameter estimates for non-linear dynamical models using a bootstrap method. *International Journal of Control 82*(11), 1991–2001.