



HAL
open science

A Kriging-NARX Model for Uncertainty Quantification of Nonlinear Stochastic Dynamical Systems in Time Domain

Biswarup Bhattacharyya, Eric Jacquelin, Denis Brizard

► **To cite this version:**

Biswarup Bhattacharyya, Eric Jacquelin, Denis Brizard. A Kriging-NARX Model for Uncertainty Quantification of Nonlinear Stochastic Dynamical Systems in Time Domain. Journal of Engineering Mechanics - ASCE, American Society of Civil Engineers, 2020, 146 (7), 66p. 10.1061/(ASCE)EM.1943-7889.0001792 . hal-02969425

HAL Id: hal-02969425

<https://hal.archives-ouvertes.fr/hal-02969425>

Submitted on 16 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Kriging-NARX model for uncertainty quantification of nonlinear stochastic dynamical systems in time domain

Biswarup Bhattacharyya¹,

Eric Jacquelin²,

and Denis Brizard³

1 ABSTRACT

2 A novel approach, referred to as sparse Kriging-NARX (KNARX), is proposed in the
3 present paper for the uncertainty quantification of nonlinear stochastic dynamical systems.
4 It combines the Nonlinear AutoRegressive with eXogenous (NARX) input model with the
5 high fidelity surrogate model Kriging. The sparsity in the proposed approach is introduced in
6 the NARX model by reducing the number of polynomial bases using the least angle regression
7 (LARS) algorithm. Sparse KNARX captures the non-linearity of a problem by the NARX
8 model, whereas the uncertain parameters are propagated using the Kriging surrogate model
9 and, further, LARS makes the model efficient. The accuracy and the efficiency of the sparse
10 KNARX is measured through uncertainty quantification applied to three nonlinear stochastic
11 dynamical systems. The time dependent mean and standard deviation are predicted for all
12 the numerical examples. Instantaneous stochastic response characteristics and maximum
13 absolute response are also predicted. All the results are compared with the full scale MCS
14 results and a mean error is calculated for all the numerical problems to measure the accuracy.
15 All the results show excellent agreement with the MCS results in a very limited computational

¹Corresponding author, Ph.D. student, Univ Lyon, Université Claude Bernard Lyon 1, IFSTTAR, LBMC UMR_T9406, F69622, Lyon, France. E-mail: biswarup.bhattacharyya@ifsttar.fr

²Professor, Univ Lyon, Université Claude Bernard Lyon 1, IFSTTAR, LBMC UMR_T9406, F69622, Lyon, France. E-mail: eric.jacquelin@univ-lyon1.fr

³Researcher, Univ Lyon, Université Claude Bernard Lyon 1, IFSTTAR, LBMC UMR_T9406, F69622, Lyon, France. E-mail: denis.brizard@ifsttar.fr

16 cost. Along with this, the efficiency of the sparse KNARX is also measured by the CPU time
17 and the required number of surrogate model evaluations. In all instances, sparse KNARX
18 outperforms other state-of-the-art methods which justifies the applicability of this model for
19 nonlinear stochastic dynamical systems.

20 **Keywords:** NARX model, Kriging, Nonlinear stochastic dynamical system, LARS, Uncer-
21 tainty quantification

22 INTRODUCTION

23 Uncertainty is an inherent property for any type of real world engineering problem which
24 may arise due to material heterogeneity, variability in dimensions, external forces. It is
25 evident from various studies (Grigoriu 1996; Lucor et al. 2004; Kundu and Adhikari 2014;
26 Chakraborty and Chowdhury 2015; Mai and Sudret 2017) that these sources of uncertainties
27 are always present in a dynamical system. Now, if the input quantities are uncertain, it is
28 obvious that the response characteristics would also be uncertain. The propagation of those
29 uncertain input parameters through a dynamical system and the prediction of uncertain
30 response characteristics are the main objectives of this paper.

31 The most used method for uncertainty quantification of any engineering problem is Monte
32 Carlo simulation (MCS) (Sheppard 1969; Muscolino et al. 2003). Mostly, MCS is used to pre-
33 dict the stochastic response characteristics for any type of problem and often result found by
34 MCS is used as the principal result for any problem. However, the accuracy of MCS greatly
35 depends on the number of simulations because MCS requires a large number of model evalu-
36 ations to predict an accurate result for a complex engineering problem. Thus, for a nonlinear
37 dynamical system, this method is highly time consuming and very much computationally
38 expensive. To overcome the issues of MCS, several surrogate models have been developed
39 by the researchers (Lucor and Karniadakis 2004; Lucor et al. 2004; Gerritsma et al. 2010;
40 Kundu and Adhikari 2014; Luchtenburg et al. 2014; Mai and Sudret 2017) in the last few
41 decades for stochastic dynamical systems. These surrogate models are the combination of
42 MCS and some kind of polynomials which maintain a trade-off between efficiency and accu-

43 racy such that the stochastic response behavior can be predicted with higher accuracy and
44 lower computational cost. In this regards, one class of surrogate model widely known as
45 polynomial chaos expansion (PCE) (Xiu and Karniadakis 2002) has been used extensively in
46 the last few decades for the uncertainty propagation in stochastic dynamical systems (Lucor
47 et al. 2004; Wan and Karniadakis 2005; Kundu and Adhikari 2014; Jacquelin et al. 2015;
48 Jacquelin et al. 2017; Ozen and Bal 2017).

49 PCE was proposed only to propagate Gaussian input random variables in (Wiener 1938)
50 which was further improved in (Xiu and Karniadakis 2002) to account the other types of
51 random variables. This method is also known as generalized PCE (gPCE). Most of the
52 methods developed till date have utilized gPCE as the basic tool for propagating uncertainties
53 in the stochastic dynamical systems. Therefore, Wan and Karniadakis (Wan and Karniadakis
54 2005) developed a multi-element gPCE to capture the non-linearity of the problem. This
55 method decomposes the domain into multiple segments (when the variance of the stochastic
56 response becomes too worse) such that the non-linearity and the probability density function
57 (PDF) of the response can be captured efficiently. A similar method has been developed in
58 (Gerritsma et al. 2010) and (Luchtenburg et al. 2014) using some different criteria. However,
59 for a system of ordinary differential equations (ODE) having three random variables, it is
60 shown in (Gerritsma et al. 2010) that PCE needs to be constructed at every time-step. On
61 the other hand, a different type of development has been made in (Maitre et al. 2010) and
62 (Mai and Sudret 2017) by rescaling the time domain which is often called time warping
63 PCE. However, it can be seen clearly in (Mai and Sudret 2017) that the time warping
64 PCE cannot predict the variance of the stochastic response properly in the later time for
65 nonlinear dynamical system with a few random variables. In contrast, a surrogate model has
66 been recently proposed (Spiridonakos and Chatzi 2015; Mai et al. 2016) by combining PCE
67 and Nonlinear Auto-Regressive with eXogenous input (NARX) (Billings et al. 1989; Chen
68 et al. 1990; Wei and Billings 2009; Billings 2013) model to account for the dynamic behavior
69 and the non-linearity of the problem. In this method, the non-linearity of the problem is

70 taken care by the NARX model while the uncertain parameters of the dynamical system are
71 propagated by PCE. However, the proposed model still uses high degree polynomials even
72 for nonlinear single degree of freedom (SDOF) dynamical systems (Mai et al. 2016) which
73 increases the computational cost drastically.

74 Another class of surrogate model widely known as Gaussian process regression or, Kriging
75 surrogate model (Krige 1951; Santner et al. 2003) has proved the efficiency in predicting
76 the stochastic response characteristics for high-dimensional problems (Mukhopadhyay et al.
77 2017; Bhattacharyya 2018). However, if one wants to predict the time-dependent stochastic
78 response behavior of an uncertain dynamical system, a new Kriging model is required to fit
79 for each time-instances and almost seems the limitation like MCS. It is also mentioned in
80 (Chakraborty and Chowdhury 2015; Chatterjee and Chowdhury 2017) that for a non-linear
81 dynamical systems, Kriging requires a large number of model evaluations. However, the
82 efficiency in predicting the stochastic response behavior for the engineering system having
83 spatial random parameters has been well established through several researches (Simpson
84 et al. 2001; Tong et al. 2015; Sugai et al. 2015; Huang et al. 2016; Lu et al. 2018), but the use
85 of Kriging surrogate model is still questionable for nonlinear stochastic dynamical systems.
86 Therefore, several improvements have been made by the researchers (Kersaudy et al. 2015;
87 Chakraborty and Chowdhury 2017; Bayarri et al. 2007; Chatterjee and Chowdhury 2018)
88 to alleviate the issue of the curse of dimensionality and account for the non-linearity of
89 the response behavior, but still these methods are suffering to describe the time dependent
90 response behavior of random non-linear dynamical systems efficiently. To address this issue,
91 a surrogate model is constructed in the present paper in a similar way to (Mai et al. 2016)
92 by combining the NARX model with the Kriging surrogate model. Therefore, the issue
93 of capturing the high order non-linearity of a stochastic dynamical system can be resolved
94 by utilizing the NARX model and the uncertainties are propagated by the Kriging. Indeed,
95 NARX model has been investigated by very few researchers in context of stochastic dynamical
96 systems (Spiridonakos and Chatzi 2015; Mai et al. 2016; Worden et al. 2018) as compared

97 to system identification tool for deterministic nonlinear dynamical systems (Billings et al.
98 1989; Wei and Billings 2009; Xie et al. 2009; Zhang and Li 2015).

99 The rest of the paper is organized as follows. A brief description of Kriging and NARX
100 model are given in section 2 and section 3 respectively. Section 4 introduces the proposed
101 model along with an algorithm to implement it for dynamical systems. Further, the appli-
102 cability of the proposed approach for dynamical systems is illustrated through three typical
103 nonlinear stochastic dynamical systems in section 5 and section 6 describes the important
104 conclusion drawn from the present study.

105 KRIGING

106 Kriging is a well established surrogate model technique which combines a regression
107 function with a Gaussian process. Kriging was first introduced in geostatistics (Matheron
108 1963) and afterwards it was used for the analysis of computer experiments (Sacks et al. 1989).
109 One of the important characteristics of Kriging is that it predicts the quantity of interest in
110 a finite region by an unbiased estimator. A brief description of Kriging is presented in this
111 section.

112 Construction of model

113 Consider $\Xi = \{\xi_1, \xi_2, \dots, \xi_d\}$ as the d -dimensional input random variables in the probabil-
114 ity space \mathcal{P} , where $\Xi \in \mathcal{D} \subset \mathbb{R}^d$. For N number of sample points, $Y = \{y(\Xi_1), y(\Xi_2), \dots, y(\Xi_N)\}^T$
115 are the corresponding function evaluations. In Kriging, the performance function for a real-
116 ization is given by:

$$\mathcal{M}(\Xi) = \beta^T \phi(\Xi) + Z(\Xi) \quad (1)$$

117 In Equation 1, the first part of the expression $\beta^T \phi(\Xi)$ represents the regression part of the
118 model which can also be written as:

$$\beta^T \phi(\Xi) = \sum_{i=1}^P \beta_i \phi_i(\Xi) \quad (2)$$

119 $\phi_i(\Xi)$ are the polynomial basis functions which are constructed with the normalized input
120 variables and β_i are the corresponding coefficients of the basis functions. P represents the
121 total number of terms in the basis function which depends on the degree of the polynomial.
122 According to the form of the polynomial, Kriging model has several variants in the literature
123 (Mukhopadhyay et al. 2017). A universal Kriging model is utilized in the present work
124 (Equation 2). The second part of Equation 1 defines the Gaussian process with mean zero
125 and the process covariance is

$$\text{cov}[Z(\Xi_i), Z(\Xi_j)] = \sigma_Z^2 \mathcal{R}(\Xi_i, \Xi_j); \quad i, j = 1, 2, \dots, N \quad (3)$$

126 where, Ξ_i and Ξ_j are two different sample points and σ_Z^2 is the process variance. $\mathcal{R}(\Xi_i, \Xi_j)$ is
127 the auto-correlation function between two sample points. A variety of auto-correlation func-
128 tions have been adopted by the researchers (Kaymaz 2005; Bhattacharyya 2018; Sacks et al.
129 1989) such as linear, exponential or Gaussian. The mostly used Gaussian auto-correlation
130 function is utilized in the present paper which is given by:

$$\mathcal{R}(\Xi_i, \Xi_j) = \prod_{k=1}^d \exp[-\theta_k(\xi_{i,k} - \xi_{j,k})^2] \quad (4)$$

131 where θ_k is the hyper-parameter of the auto-correlation function. The Kriging model param-
132 eters β and σ_Z , and the hyper-parameter θ_k can be determined by the maximum likelihood
133 estimation (MLE) (Sacks et al. 1989). The procedure of computing the parameters is ex-
134 plained in (Sacks et al. 1989; Santner et al. 2003; Kaymaz 2005).

135 Prediction

136 Consider any untried sample point for prediction $\Xi_0 \in \mathbb{R}^d$. The function value at the
137 new sample Ξ_0 is predicted by the best linear unbiased predictor (BLUP). The BLUP and
138 the predicted variance are given by:

$$\hat{\mathcal{M}}(\Xi_0) = w^T(\Xi_0)Y \quad (5)$$

$$= \hat{\beta}^T \phi(\Xi_0) + r^T(\Xi_0) \mathcal{R}^{-1} \left(\hat{\beta}^T \phi(\Xi_0) \right) \quad (6)$$

$$\sigma_{\hat{M}}^2(\Xi_0) = \hat{\sigma}_Z^2 \left(1 - \begin{bmatrix} \phi^T(\Xi_0) & r^T(\Xi_0) \end{bmatrix} \begin{bmatrix} 0 & F^T \\ F & \mathcal{R} \end{bmatrix}^{-1} \begin{bmatrix} \phi(\Xi_0) \\ r(\Xi_0) \end{bmatrix} \right) \quad (7)$$

139 where, $r(\Xi_0) = \{\mathcal{R}(\Xi_0, \Xi_1), \mathcal{R}(\Xi_0, \Xi_2), \dots, \mathcal{R}(\Xi_0, \Xi_N)\}$ is the correlation matrix between
 140 the new untried sample point and the initial samples, and F is the matrix of basis function
 141 at the initial sample points of dimension $N \times P$. \mathcal{R} is the correlation matrix for the initial
 142 samples. Further, the unknown parameters $\hat{\beta}$ and σ_Z^2 are calculated as:

$$\hat{\beta} = (F^T \mathcal{R}^{-1} F)^{-1} F^T \mathcal{R}^{-1} Y \quad (8)$$

$$\hat{\sigma}_Z^2 = \frac{1}{N} (Y - F \hat{\beta})^T \mathcal{R}^{-1} (Y - F \hat{\beta}) \quad (9)$$

143 Once the unknown parameters using the Kriging model are estimated, the prediction of
 144 the response at the untried point can be made easily by utilizing Equation 6.

145 **NARX MODEL**

146 **Overview of NARX model**

147 A brief overview of NARX model is presented in this section. The time dependent
 148 response characteristics of a dynamical system at current time instance can be predicted by
 149 the responses of some previous time instances and the excitation of some previous and current
 150 time instances through a suitable NARX model (Chen and Billings 1989; Billings 2013). If
 151 we consider a dynamical system having the time dependent excitation, then according to the

152 NARX model, the dynamical system can be expressed as:

$$y(t) = \mathcal{F}[z(t)] + \varepsilon(t) \quad (10)$$

153 where, $z(t) = \{x(t), x(t - \Delta t), x(t - 2\Delta t), \dots, x(t - n_{x_m}\Delta t), y(t - \Delta t), y(t - 2\Delta t), \dots,$
 154 $y(t - n_{y_m}\Delta t)\}^T$ is the vector having all the lagged system excitation and response compo-
 155 nents which forms the time-dependent auto-regressive response model $\mathcal{F}[\bullet]$. $\varepsilon(t)$ is the
 156 residual of the NARX model which is supposed to be a normally independent distributed
 157 (NID) process with zero mean and Δt is the time-step chosen for the NARX model. n_{x_m} and
 158 n_{y_m} are the maximum time lags for the excitation and the response quantity respectively. It
 159 is obvious that the underlying form of the function must be nonlinear to capture the strong
 160 non-linearity of a dynamical system. Consequently, different types of functions have been
 161 used by the researchers such as polynomial (Cantelmo and Piroddi 2010), wavelet (Billings
 162 and Wei 2005), sigmoid function (Sjöberg et al. 1995), RBF (Li et al. 2005) and neural net-
 163 work (Tsunghan Lin et al. 1996). Out of these, the effectiveness of polynomial function has
 164 already been proved in the literature (Leontaritis and Billings 1985; Cantelmo and Piroddi
 165 2010; Cheng et al. 2011). Therefore, a linear-in-parameter form of polynomial has been used
 166 in the present paper which is represented by

$$\mathcal{F}[z(t)] = \sum_{i=1}^M \varphi_i \psi_i[z(t)] \quad (11)$$

167 In the above equation, M is the total number of terms in the polynomial basis function,
 168 $\psi_i[z(t)]$ are the polynomial basis functions and φ_i are the corresponding coefficients of the
 169 NARX model. The polynomial basis function for the NARX model is dependent on the
 170 time varying excitation and response. Consequently, the polynomial basis matrix and the

171 coefficient matrix for a particular sample point k are represented by:

$$\Psi_k(t, \Xi_k) = \{\psi_1[z_k(t, \Xi_k)], \psi_2[z_k(t, \Xi_k)], \dots, \psi_M[z_k(t, \Xi_k)]\}^T; \quad \Psi_k \in \mathbb{R}^{M \times 1} \quad (12)$$

$$\varphi^k = \{\varphi_1, \varphi_2, \dots, \varphi_M\}; \quad \varphi^k \in \mathbb{R}^{1 \times M} \quad (13)$$

172 Now, if we discretize the total time T in N_t number of time instances i.e. $t \in \{t_1, t_2, \dots, t_{N_t}\}$,
 173 then for N_t number of time-steps, the polynomial basis function matrix can be given by
 174 $\Psi_k \in \mathbb{R}^{M \times N_t}$.

175 The procedure of choosing the type of polynomial for $\mathcal{F}[z(t)]$ and estimating the corre-
 176 sponding coefficients are described in the next section.

177 **Model formulation and parameter estimation of the NARX model**

178 The polynomial basis function for the NARX model can be formulated by two variables,
 179 i.e. the excitation and the response of the dynamical system, either as an independent
 180 variable basis function (Spiridonakos and Chatzi 2015) or as a composition of both variables
 181 (Mai et al. 2016) with certain maximum degree of the polynomial basis. For a nonlinear
 182 system, it is always important to use the interaction terms (Billings 2013). Hence, a similar
 183 kind of polynomial basis function has been adopted in the present paper. Another important
 184 issue with the polynomial basis function is the selection of the maximum time lags n_{x_m} and
 185 n_{y_m} for the excitation and the response respectively. The maximum time lags are selected
 186 according to (Spiridonakos and Chatzi 2015; Mai et al. 2016) in the present paper, which
 187 are two-times the number of DOF of the system.

188 On the other hand, the computation of the NARX coefficients is one of the challenging
 189 tasks in the construction of the NARX model. The coefficients of the NARX model can be
 190 computed easily by ordinary least square method (for the k -th sample point). However, it
 191 has been found often in the literature (Blatman and Sudret 2011; Spiridonakos and Chatzi

2015) that all the terms in the polynomial basis do not get involved in predicting the response characteristics of the system, and reducing the degree of the polynomial may reduce the accuracy of prediction. Thus, it is important to capture the important terms in the polynomial basis function which are solely responsible for the response behavior of a system. Recently, the important NARX basis terms were identified using genetic algorithm in (Spiridonakos and Chatzi 2015). However, due to the form of NARX model, the important terms can also be identified with the popularly used regularized least square method, least angle regression (LARS) (Efron et al. 2004). Consequently, LARS is used in the present paper for selecting the important terms in the polynomial basis.

201 Sparse NARX model

202 The sparse NARX model is selected in a similar way to (Mai et al. 2016) in the present
 203 paper. At first, it is required to select the samples exhibiting high non-linearity. The
 204 responses having highly non-linear behavior can be selected by the measurement of non-
 205 linearity of the response (Spiridonakos and Chatzi 2015) or, specifying some threshold value
 206 for the response series (Mai et al. 2016). In the present paper, a combination of both
 207 the mentioned methods (Spiridonakos and Chatzi 2015; Mai et al. 2016) has been utilized.
 208 Firstly, the response versus restoring force is plotted arbitrarily without performing any
 209 simulations (keeping other parameters constant at their mean values). The intensity of
 210 the response is increased upto a certain limit till the nonlinear behavior is noticed in the
 211 force-displacement relationship. From the restoring force curve, the threshold value for a
 212 response series can be selected easily by observing the starting point of the non-linearity of
 213 the response series. Imposing the threshold value would reduce the number of samples to
 214 $N_1 < N$. Now, for each of the N_1 samples (e.g. for the k -th sample), a full NARX model is
 215 formulated as follows:

$$y(t, \Xi_k) = \varphi^k \Psi_k(\Xi_k) \quad (14)$$

216 where, Ψ_k is the matrix having all the M terms of the NARX model polynomial basis for
 217 all the time-steps with dimension $M \times N_t$ which is formulated by the current and previous
 218 excitation, and previous responses. $y(t, \Xi_k)$ is the response series for the k -th sample point.
 219 At this step, the coefficient vector φ^k is the unknown.

220 **Remark 1:** One should discretize the total time T with a suitable time-step Δt (small
 221 enough to capture the dynamics correctly). Besides, the time-step plays a vital role in case of
 222 NARX model. In the present paper, we have used the same time-step for the time integration
 223 and the construction of the NARX model.

224 The LARS algorithm is then employed to select the important terms for each of the N_1
 225 full NARX models.

226 **Remark 2:** It is evident that N_1 different NARX models are found at this step. The
 227 most important terms for all the NARX models are found using LARS algorithm. However,
 228 it may happen that more than one NARX model contains similar terms in the polynomial
 229 basis matrix. Therefore, only the N_2 different NARX models are selected at this step.

230 The unique NARX models are selected from the N_1 number of NARX models which
 231 may further reduce the number of NARX models to $N_2 \leq N_1$. Hence, for each of the N_2
 232 sparse NARX models, the coefficients of all the N initial samples are computed by ordinary
 233 least square. The response series of the initial N samples are then re-constructed using
 234 the coefficients and the selected polynomial bases, and the error for each sample point is
 235 predicted as follows:

$$\epsilon_k = \frac{\sum_{i=1}^{N_t} [y(t_i, \Xi_k) - \hat{y}(t_i, \Xi_k)]^2}{\sum_{i=1}^{N_t} [y(t_i, \Xi_k) - \bar{y}(\Xi_k)]^2} \quad (15)$$

236 where, $\bar{y}(\Xi_k)$ is the time average mean of the k -th response series which is given by:

$$\bar{y}(\Xi_k) = \frac{1}{N_t} \sum_{i=1}^{N_t} y(t_i, \Xi_k) \quad (16)$$

237 The mean predicted error for all the sample points by a particular sparse NARX model
 238 out of N_2 sparse NARX models is given by:

$$\bar{\epsilon} = \frac{1}{N} \sum_{k=1}^N \epsilon_k \quad (17)$$

239 The finally selected sparse NARX model is the one having the predicted mean error for
 240 N sample points lower than some threshold value. In the present paper, the threshold value
 241 of the mean error is imposed as 1×10^{-3} for all the examples.

242 **Remark 3:** It is of utter importance to mention that if the predicted mean errors for
 243 two different NARX model are found same (or, lower than the threshold value) then the
 244 sparse NARX model having less number of terms in the polynomial basis is selected as the
 245 final sparse NARX model.

246 KRIGING-NARX MODEL

247 Consider a dynamical system having some uncertain input parameters $\Xi = \{\xi_1, \xi_2, \dots, \xi_d\}$,
 248 then the time dependent response of the system can be represented by a NARX model as:

$$y(t, \Xi_k) = \sum_{i=1}^M \varphi_i(\Xi_k) \psi_i[z_k(t)]; \quad k = 1, 2, \dots, N \quad (18)$$

249 It is observed from Equation 18 that the coefficients of the NARX model are dependent
 250 on the sample points which does not make the model stochastic in nature. To get the
 251 independent coefficients of the NARX model, the NARX coefficients are represented by the
 252 Kriging surrogate model as given in Equation 1:

$$\varphi_i(\Xi) = \beta_i^T \phi(\Xi) + Z_i(\Xi); \quad i = 1, 2, \dots, M \quad (19)$$

253 Further, the model is constructed by the Kriging surrogate model as discussed in section 2.

254 Therefore, the full KNARX model is expressed as:

$$y(t, \Xi) = \sum_{i=1}^M (\beta_i^T \phi(\Xi) + Z_i(\Xi)) \psi_i[z(t)] \quad (20)$$

255 The coefficient of the regression part, β^T and the Gaussian process part $Z(\Xi)$ are depen-
 256 dent on the number of terms in the NARX model, whereas the polynomial basis function
 257 of the Kriging model for all the NARX coefficients remain the same as the basis is purely
 258 dependent on the uncertain input parameters.

259 **Remark 4:** It should be noted that the Kriging model should be calibrated M times (M
 260 is the total number of terms for a full NARX model) because the coefficients of the NARX
 261 model for N samples corresponding to a polynomial basis act as the single response quantity
 262 for the Kriging model.

263 The response quantity of a dynamical system at some untried sample points Ξ_0 can be
 264 predicted by BLUP as mentioned in Equation 6 in accordance with the full NARX model in
 265 an auto-regressive manner.

$$\hat{y}(t, \Xi_0) = \sum_{i=1}^M \left[\hat{\beta}_i^T \phi(\Xi_0) + r^T(\Xi_0) R^{-1} \left(\hat{\beta}_i^T \phi(\Xi_0) \right) \right] \psi_i[z(t)] \quad (21)$$

266 Equation 21 helps in predicting the time dependent response characteristics of a dynam-
 267 ical system at large number of sample points having d dimensional random input variables.
 268 It is evident from Remark 4 that the computational cost increases with the increase of num-
 269 ber of terms in the NARX polynomial basis. For that reason, the sparse NARX model as
 270 discussed in section 3 is used in accordance with the Kriging surrogate model in the present
 271 paper. Consequently, the number of terms for the NARX model is decreased to $M_1 < M$
 272 which ultimately reduces the number of Kriging model calibration (M_1) and computational
 273 cost. The algorithm for constructing the sparse KNARX model is provided in Table 1.

274 NUMERICAL APPLICATIONS

275 The sparse KNARX model as described in the previous section has been utilized for

276 uncertainty quantification of three typical nonlinear stochastic dynamical systems. For each
 277 of the examples, the accuracy of the method is measured using the predicted mean error
 278 as given in Equation 17 (Mai et al. 2016) (taking MCS as the reference) and the value of
 279 the coefficient of correlation R^2 . Uncertainty quantification for all the problems is made
 280 by predicting the time dependent mean and standard deviation of the responses. Further,
 281 the PDFs of the responses are also predicted at some time instances. The computational
 282 efficiency is measured through the computational cost (CPU time), the number of surrogate
 283 model calibration and the number of initial sample points N . The first problem is also solved
 284 by the Kriging model to measure the efficiency and the accuracy of the sparse KNARX model
 285 over the Kriging. For the first two examples, the sparse KNARX predicted results are also
 286 compared with the recently proposed sparse PCE-NARX surrogate model (Mai et al. 2016).

287 It is well known that a suitable sample point (often known as experimental design point)
 288 generation strategy is required to use for the generation of initial samples in step 2 of Table 1.
 289 Consequently, one of the most widely used sampling strategies, Latin hypercube sampling
 290 (LHS) (Mai et al. 2016; Chatterjee and Chowdhury 2017) has been utilized for the initial N
 291 number of sample points generation.

292 **Duffing oscillator**

293 A non-linear duffing oscillator is considered as the first example for the illustration of the
 294 proposed model. The governing differential equations of the duffing oscillator are given by:

$$\ddot{y}(t) + 2\zeta\omega\dot{y}(t) + \omega^2 [y(t) + \varepsilon y^3(t)] = x(t) \quad (22)$$

$$x(t) = A \sin(\omega_x t) \quad (23)$$

295 In Equation 22, ω represents the natural frequency considering the undamped linear struc-
 296 ture ($\varepsilon = 0, \zeta = 0$), ζ is the damping ratio and ε controls the non-linearity of the system.
 297 $x(t)$ denotes the excitation part of the dynamical system which is considered sinusoidal in

298 this case (see Equation 23). The initial conditions are $y(0) = 0$ and $\dot{y}(0) = 0$. The numer-
 299 ical integration has been performed for $T = 30$ s with a time-step of $\Delta t = 0.01$ s through
 300 MATLAB solver *ode45*. All the parameters of the duffing oscillator are considered uncertain
 301 (i.e. $\Xi = \{\omega, \zeta, \varepsilon, A, \omega_x\}$) (Mai 2016). The distribution types along with the parameters of
 302 the distribution are listed in Table 2.

303 Time dependent stochastic displacement of the duffing oscillator is predicted by MCS,
 304 Kriging, sparse PCE-NARX (Mai et al. 2016) and sparse KNARX model. MCS has been
 305 performed with 3×10^4 sample points and $N = 50$ samples have been generated by LHS
 306 for the prediction of the stochastic response behavior using Kriging. The Kriging model
 307 has been calibrated for each time-step (i.e. 3001 times). The step by step procedure of
 308 constructing the sparse KNARX model (according to Table 1) is described below:

- 309 1. For the duffing oscillator, $d = 5$ (see Table 2).
- 310 2. The sparse KNARX model is constructed using $N = 25$ LHS points.
- 311 3. For the construction of the sparse KNARX and the sparse PCE-NARX model, it is
 312 important to specify the threshold value for capturing the nonlinear response series
 313 according to step 3 of Table 1. Full NARX models are to be constructed on those
 314 samples which satisfy the criterion of threshold value. For that reason, the displacement
 315 $y(t)$ versus restoring force $f_s = \omega^2 [y(t) + \varepsilon y^3(t)]$ has been plotted in Figure 1 (keeping
 316 other parameters constant at their mean values). From Figure 1, it is clear that the
 317 displacement behaves almost linearly for $y(t) \in [-0.045 \text{ m}, 0.045 \text{ m}]$ and beyond this
 318 region, the displacement is highly nonlinear. Consequently, to capture the highly
 319 nonlinear response series, the threshold value for the nonlinear displacement is set as
 320 $\max(|y(t)|) > 0.045 \text{ m}$. Thus, only those samples will be picked for the construction
 321 of NARX model which falls beyond the region $y(t) \in [-0.045 \text{ m}, 0.045 \text{ m}]$.
- 322 4. The displacement series are then obtained at the initial $N = 25$ samples. Out of all
 323 the response series, the displacement versus restoring force for two different samples

324 (13-th and 25-th sample point) are plotted in Figure 2. It is clearly seen that the 13-th
 325 sample point exhibits higher order non-linearity than the 25-th sample point. It should
 326 be noted that the region of displacement value is almost restricted in $[-0.02 \text{ m}, 0.02 \text{ m}]$
 327 for the 25-th sample point which is far less than the specified threshold value. On
 328 the other hand, the displacement intensities are far beyond the threshold value and
 329 behaves nonlinearly for the 13-th sample point.

330 5. After imposing the above-specified threshold criterion, only $N_1 = 2$ samples are selected
 331 as the highly nonlinear response series. Thus, only 2 full NARX models are required
 332 for the duffing oscillator using the sparse KNARX model ($N = 25$). The basis function
 333 for the full NARX model is chosen as:

$$\psi_i [z(t)] = x^{l_i} (t - n_{x_i} \Delta t) y^{m_i} (t - n_{y_i} \Delta t) \quad (24)$$

334 where, x and y are the excitation and the response of the duffing oscillator respectively.

335 6. The maximum time lags for the excitation and the response are taken as $n_{x_m} = 2$ and
 336 $n_{y_m} = 2$ respectively due to 1-DOF system. $n_x \in \{0, 1, 2\}$ and $n_y \in \{1, 2\}$ represent
 337 all the time lags for the excitation and the response respectively. $l \in \{0, 1\}$ and
 338 $m \in \{0, 1, 2, 3\}$ are the degrees for the excitation and the response respectively, with a
 339 maximum degree of the polynomial 3 (i.e. $l_i + m_i \leq 3$), due to the cubic non-linearity
 340 of the system. A total 22 number of terms have been obtained in the polynomial
 341 basis matrix utilizing all the above-mentioned criterion which depicts $M = 22$ and
 342 $i = 1, 2, \dots, M$ in Equation 24.

343 7. Therefore, 2 full NARX models are constructed using the basis function as given in
 344 Equation 24, and they have 22 terms in the NARX polynomial basis matrix.

345 8. The sparse NARX models are then constructed using the LARS algorithm by selecting
 346 the most important terms in the bases for both the full NARX models.

- 347 9. The sparse NARX models are then retained which have unique set of polynomial bases
348 and for the duffing oscillator, both the sparse NARX models are found unique. The
349 polynomials selected by the LARS algorithm for both the NARX models are listed in
350 Table 3.
- 351 10. The coefficients for all the 25 samples are obtained with the ordinary least square for
352 both the sparse NARX model. For this step, explicitly $N \times N_2$ number of ordinary
353 least squares is performed.
- 354 11. Then, the response series are reconstructed again in a recursive manner using the
355 coefficients and the polynomial bases.
- 356 12. $\bar{\epsilon}$ is then computed by both the sparse NARX models for the N samples.
- 357 13. $\bar{\epsilon}$ for both the sparse NARX models is found below the threshold value. Hence, the
358 Model 2 of Table 3 is chosen as the best sparse NARX model (according to Remark 3)
359 and the predicted mean error for the $N = 25$ samples is found as $\bar{\epsilon} = 1.68 \times 10^{-7}$.
- 360 14. The coefficient vector $\varphi_i(\Xi) \in \mathbb{R}^{N \times 1}; i = 1, \dots, 8$ corresponding to each of the polyno-
361 mials are uncertain. These 8 coefficient vectors are considered as the uncertain response
362 quantities for the Kriging model. Hence, the unknown parameters of the Kriging model
363 $\hat{\beta}$, $\hat{\sigma}_Z^2$ and θ_k are computed by the MLE for all the 8 NARX coefficients separately.
364 Consequently, 8 Kriging models are calibrated for converting the model stochastic.
- 365 15. The same 3×10^4 MCS samples are used here for the prediction by the surrogate model.
- 366 16. The 8 NARX coefficient vectors for the 3×10^4 samples are then predicted using the
367 8 calibrated Kriging models by BLUP.
- 368 17. Further, the predicted coefficients and the selected polynomials are used to predict the
369 response series auto-regressively for the 3×10^4 MCS samples.

370 The same $N = 25$ LHS samples are also utilized for uncertainty quantification using
 371 the sparse PCE-NARX (Mai et al. 2016) model. Here, the sparse adaptive PCE model
 372 is constructed using the UQLab module (Marelli and Sudret 2018). All the parameters
 373 of the sparse PCE model are taken similar to (Mai et al. 2016) for the present com-
 374 putation. Therefore, the maximum interaction is chosen as 2 and the q-norm is taken
 375 as 1 with the degree of the polynomials varying between 1 and 20 (Marelli and Sudret
 376 2018; Mai et al. 2016). However, the predicted response is not converged using the sparse
 377 PCE-NARX model with the same number of samples. For that reason, the number of
 378 samples for the sparse PCE-NARX model is increased to $N = 35$ and the time depen-
 379 dent responses are predicted with the sparse PCE-NARX model. Here, the final sparse
 380 NARX model is found having $M_1 = 9$ terms in the NARX polynomial basis which are
 381 $\{y(t - \Delta t), y(t - 2\Delta t), y^2(t - \Delta t), y^3(t - \Delta t), y^3(t - 2\Delta t), x(t), x(t - 2\Delta t), x(t)y(t - \Delta t),$
 382 $x(t - 2\Delta t)y^2(t - 2\Delta t)\}$.

383 To reduce further the computational cost, another study has been performed using less
 384 number of initial samples with $N = 21$ (for sparse KNARX). In this case, 3 samples are
 385 retained initially based on the criterion of threshold value of the response series and the 3
 386 full NARX models are formulated using the basis function as mentioned in Equation 24.
 387 Further, the sparse NARX models are found by applying the LARS algorithm and all the
 388 three models are found as the unique sparse NARX models. Therefore, the coefficients for
 389 the $N = 21$ samples are found by the ordinary least square method and then, all the 21
 390 response series are reconstructed using the coefficients of the sparse NARX models. The
 391 finally selected sparse NARX model is the Model 1 of Table 3 and the predicted mean error
 392 of the selected sparse NARX model is found as $\bar{\epsilon} = 1.80 \times 10^{-7}$ for the 21 samples. Further,
 393 the coefficients corresponding to the sparse NARX polynomial bases are modelled using the
 394 Kriging surrogate model. On the contrary, the sparse PCE-NARX model is not able to
 395 predict the response with less samples.

396 The time varying stochastic response characteristics are predicted by the time dependent

397 mean and standard deviation which are plotted in Figure 3. The figure clearly depicts the
 398 efficiency (initial number of sample points N) and the accuracy of the sparse KNARX model
 399 over the Kriging model. The sparse KNARX has predicted better results with fewer number
 400 of model evaluations ($N = 21$) and the sparse PCE-NARX model is incapable to predict a
 401 result even using $N = 25$ samples. To illustrate the accuracy of the sparse KNARX, the
 402 instantaneous response characteristics are also plotted. The scatter diagrams and the PDFs
 403 of the predicted response are plotted in Figure 4 at three different time instance (10s, 20s
 404 and 30s). The accuracy metrics of the instantaneous response characteristics are listed in
 405 Table 4, they clearly suggest that the sparse KNARX is highly efficient as compared to the
 406 Kriging model. The accuracy of the sparse KNARX model is also comparatively higher using
 407 less model evaluations than the sparse PCE-NARX model.

408 The stochastic absolute maximum displacement $\max(|y(t)|)$, plotted in Figure 5, ulti-
 409 mately measures the safety corridor for the dynamical system. It is seen clearly that the
 410 sparse KNARX outperforms the Kriging and the sparse PCE-NARX with very less num-
 411 ber of sample points in predicting the PDF of $\max(|y(t)|)$, and the accuracy of the sparse
 412 KNARX is given by the R^2 value in Table 5 which is very close to 1.0 by $N = 25$.

413 Further, the accuracy of the overall model is computed by the mean error (Equation 17)
 414 for the predicted responses and the error for the predicted $\max(|y(t)|)$. The efficiency of
 415 the sparse KNARX has already been shown by the initial number of sample points. An
 416 accurate result is predicted by the sparse KNARX even with much less number of model
 417 evaluations ($N = 21$) as compared to the other methods. The efficiency of the sparse
 418 KNARX is also measured by the number of surrogate model calibrations (n_K) for a method
 419 and by the CPU time. All the accuracy and the efficiency measurement metrics are reported
 420 in Table 5. Table 5 is suggesting that the sparse KNARX model outperforms the Kriging
 421 and the sparse PCE-NARX in accuracy and efficiency. It is noticeable that the CPU time
 422 is lower with higher value of N for the sparse KNARX model due to less number of Kriging
 423 model calibration.

424 **Bouc-Wen oscillator**

425 A non-linear Bouc-Wen oscillator (Bouc 1967; Wen 1976) is investigated in this example.
 426 The governing differential equation for the Bouc-Wen oscillator is given by:

$$\begin{aligned} \ddot{y}(t) + 2\zeta\omega\dot{y}(t) + \omega^2[\rho y(t) + (1 - \rho)w(t)] &= -A \sin(\omega_x t) \\ \dot{w}(t) &= \gamma\dot{y}(t) - \alpha|\dot{y}(t)||w(t)|^{n-1}w(t) - \beta\dot{y}(t)|w(t)|^n \end{aligned} \quad (25)$$

427 where, ζ and ω are the damping ratio and the natural frequency of the oscillator respectively.
 428 $\rho = 0$, $\gamma = 1$, $\beta = 0$ and $n = 1$ are considered for this problem. The excitation of the oscillator
 429 is considered as $x(t) = A \sin(\omega_x t)$ and $w(t)$ is the hysteric displacement as given in (Wen
 430 1976). The initial conditions at rest are $y(0) = 0$, $\dot{y}(0) = 0$ and $w(0) = 0$. Similar to
 431 (Mai and Sudret 2017), 5 uncertain parameters are considered for the oscillator, which are
 432 $\Xi = \{\zeta, \omega, \alpha, A, \omega_x\}$. The type of distributions for all the uncertain parameters are given in
 433 Table 6.

434 The stochastic response is calculated for the oscillator in the time domain $t \in [0 \text{ s}, 30 \text{ s}]$
 435 with a time-step of $\Delta t = 0.005 \text{ s}$. As it has already been illustrated through the previous
 436 example, Kriging is unable to predict the stochastic response behavior for the non-linear
 437 stochastic duffing oscillator even with higher samples than the sparse KNARX. Therefore,
 438 Kriging has not been utilized from now onward. Similar to the previous example, the sparse
 439 KNARX model has been constructed with two different sizes of sample points $N = 40$ and
 440 $N = 10$. A different type of basis function has been considered for the Bouc-Wen oscillator
 441 according to (Mai et al. 2016), which is given by:

$$\psi_i[z(t)] = \{x^{l_i}(t - n_{x_i}\Delta t)|\dot{y}(t - \Delta t)|^{m_i}, \dot{y}^{l_i}(t - n_{y_i}\Delta t)|\dot{y}(t - \Delta t)|^{m_i}\} \quad (26)$$

442 In Equation 26, the basis function relies on the excitation and the velocity of the oscillator.
 443 Thus, the velocity is computed by the sparse KNARX model and further, the displacement of
 444 the system is obtained through numerical integration by utilizing the MATLAB solver *ode45*.
 445 Due to the hysteric displacement component, the bases are computed by two expressions as

446 given in Equation 26 which has already been considered in (Mai et al. 2016). $l \in \{0, 1\}$
447 and $m \in \{0, 1\}$ are considered for the oscillator as single degree is noticed for the excitation
448 and the velocity in Equation 25. $n_{x_m} = n_{y_m} = 4$ is considered as the problem can be
449 appraised as 2-DOF system due to an extra hysteric displacement part i.e. $n_x \in \{0, \dots, 4\}$
450 and $n_y \in \{1, \dots, 4\}$. A total 21 number of terms are found in the full NARX polynomial
451 basis matrix by utilizing Equation 26 and imposing all the mentioned conditions.

452 A threshold similar to the previous example was employed to detect the highly non-
453 linear samples. The threshold value for the response of the Bouc-Wen oscillator is cho-
454 sen as $\max(|\dot{y}(t)|) > 0.3 \text{ ms}^{-1}$. The threshold value reduces the number of samples from
455 $N = 40$ to $N_1 = 3$ which means only 3 full NARX models are required. Thus, 3 full
456 NARX models are formulated utilizing the basis function of Equation 26 and further the
457 LARS algorithm has been employed to make the full NARX models sparse. All the three
458 sparse NARX models are found unique in this step. Therefore, the coefficients for all the 40
459 initial samples are computed by the ordinary least square method for the 3 sparse NARX
460 models and the mean errors (Equation 17) are predicted by computing the time series in a
461 recursive manner. The finally selected sparse NARX model produces a mean error of $\bar{\epsilon} =$
462 1.10×10^{-5} for the 40 samples and contains 8 terms in the polynomial basis matrix which are
463 $\{ |\dot{y}(t - \Delta t)|, x(t), x(t - 4\Delta t), x(t - 4\Delta t) |\dot{y}(t - \Delta t)|, \dot{y}(t - \Delta t), \dot{y}(t - 4\Delta t), \dot{y}(t - \Delta t) |\dot{y}(t - \Delta t)|,$
464 $\dot{y}(t - 4\Delta t) |\dot{y}(t - \Delta t)| \}$.

465 A similar procedure is adopted by substantially reducing the initial number of sample
466 points to $N = 10$. 4 samples are found mostly non-linear based on the previous threshold
467 value. Out of the 4 full NARX models, 2 are found as unique sparse NARX models after
468 applying the LARS for detecting the most important terms. Therefore, one sparse NARX
469 model is selected based on the predicted mean error from both sparse NARX models. The
470 final sparse NARX model is found to have the same terms in the polynomial basis as the
471 previous one with $N = 40$.

472 The stochastic responses are also predicted here by the sparse PCE-NARX model (Mai

473 et al. 2016) (the sparse PCE model is constructed here with the same conditions as the
474 previous example). However, the sparse PCE-NARX is unable to predict the stochastic
475 response behavior with $N = 40$ number of model evaluations for the Bouc-Wen oscillator.
476 For that reason, the number of model evaluations is increased to $N = 50$ and the stochastic
477 responses are predicted with the same $M_1 = 8$ terms in the sparse NARX model as in the
478 sparse KNARX model with $N = 40$.

479 This study aims at quantifying the uncertainty associated with the response quantity.
480 The uncertain response characteristics are predicted for the displacement and the velocity
481 of the Bouc-Wen oscillator. The time dependent mean and the standard deviation of $y(t)$
482 and $\dot{y}(t)$ are shown in Figure 6 and Figure 7 respectively. The figures show that both the
483 statistical moments predicted by the sparse KNARX are following the MCS results with
484 utmost accuracy. However, the accuracies of the statistical moments are deteriorating from
485 the initial time-steps using the sparse PCE-NARX model even with more samples. The
486 scatter diagrams and the PDFs at three different time instances (10 s, 20 s and 30 s) are also
487 plotted in Figure 8 and Figure 9 for the displacement and the velocity respectively. It is
488 seen from all the figures that the sparse KNARX performs well in all instances with very few
489 samples. The error metrics for all the instantaneous displacements and velocities are given
490 in Table 7. An excellent accuracy in terms of error (ϵ) and R^2 value is noticed for all the
491 time instances by the sparse KNARX model.

492 For the prediction of the safety corridor under uncertainty, the scatter plots and the PDFs
493 of the maximum absolute displacement and velocity are plotted in Figure 10 and Figure 11
494 respectively. The worst prediction is noticed for the PDFs by the sparse PCE-NARX model,
495 whereas an excellent accuracy is observed in both the cases for the predicted maximum re-
496 sponses by the sparse KNARX model. These plots, which represent the uncertain maximum
497 response behavior, can be utilized to measure the safety margin of the system.

498 To observe the accuracy of the surrogate models, the mean of the displacement and
499 the velocity are plotted in state space by the MCS, the sparse PCE-NARX and the sparse

500 KNARX in Figure 12. It is evident from the figure that the accuracy in predicting the state
501 space behavior of the response is greatly achieved even using much less number of samples
502 ($N = 10$) by the sparse KNARX model whereas the sparse PCE-NARX predicted result is
503 the worst even with more samples. Similar to the previous example, the accuracy and the effi-
504 ciency measurement metrics are given in Table 8. The table is showing clearly that the sparse
505 KNARX has predicted results with higher accuracy by minimum computational cost. The
506 efficiency of the proposed sparse KNARX method is also observed over the recently proposed
507 method time warping PCE (Mai and Sudret 2017) which required $N = 100$ ($\gg N = 10$)
508 samples to predict the stochastic response behavior for the same example.

509 **A 2-DOF dynamical system**

510 Finally, a 2-DOF dynamical system (Mai et al. 2016) has been considered for the appli-
511 cability of the proposed sparse KNARX model. The dynamical system is shown in Figure 13
512 and the governing differential equation of the problem is:

$$\begin{aligned}
m_s \ddot{y}_1(t) &= -k_s [y_1(t) - y_2(t)]^3 - c [\dot{y}_1(t) - \dot{y}_2(t)] \\
m_u \ddot{y}_2(t) &= k_s [y_1(t) - y_2(t)]^3 + c [\dot{y}_1(t) - \dot{y}_2(t)] + k_u [x(t) - y_2(t)]
\end{aligned}
\tag{27}$$

513 where m_s is the sprung mass which is connected with the mass m_u by a non-linear spring
514 with stiffness k_s and a damper with damping coefficient c . k_u is a linear spring attached to
515 the ground having a sinusoidal displacement function $x(t) = A \sin(\omega_x t)$.

516 Similar to (Mai et al. 2016), all the parameters of the system are considered uncertain
517 i.e. $\Xi = \{k_s, k_u, m_s, m_u, c, A, \omega_x\}$. The mean and the standard deviation of all the uncertain
518 parameters are given in Table 9.

519 The main aim of this example is to predict the uncertain response $y_1(t)$ of mass m_s which
520 is attached with the non-linear spring. This system has been solved by the sparse KNARX
521 model with two different number of sample points ($N = 50$ and $N = 20$). As it has already
522 been seen from the previous two examples that the sparse PCE-NARX model is unable to
523 predict the stochastic response behavior of the dynamical systems properly, the sparse PCE-

524 NARX model is not utilized for this problem. It should be noted that the problem has been
 525 solved in the time domain $t \in [0 \text{ s}, 30 \text{ s}]$ with a time-step of $\Delta t = 0.01 \text{ s}$ by *ode45*. The initial
 526 conditions for the system at rest ($t = 0 \text{ s}$) are given by:

$$\begin{aligned}
 y_1(0) &= 0 \\
 \dot{y}_1(0) &= 0 \\
 y_2(0) &= 0 \\
 \dot{y}_2(0) &= 0
 \end{aligned} \tag{28}$$

527 For the construction of the NARX model, the NARX basis function is chosen similar to
 528 the first example. The NARX basis function for this problem is given by:

$$\psi_i[z(t)] = x^{l_i}(t - n_{x_i}\Delta t) y_1^{m_i}(t - n_{y_i}\Delta t) \tag{29}$$

529 In Equation 29, $l \in \{0, 1\}$ and $m \in \{0, 1, 2, 3\}$ with $l_i + m_i \leq 3$ are chosen due to cubic non-
 530 linearity of the system. $n_x \in \{0, 1, \dots, 4\}$, $n_y \in \{1, \dots, 4\}$ are chosen because the system
 531 is having 2-DOF. The full NARX model is constructed with this NARX basis function
 532 which has $M = 58$ terms. Initially, after investigating the response versus restoring force,
 533 a threshold criterion of responses has been decided as $\max(|y_1(t)|) > 1.2 \text{ m}$ which reduces
 534 the number of samples to $N_1 = 10$. Consequently, only 10 number of full NARX models
 535 are constructed by using the basis function as mentioned in Equation 29. The sparsity is
 536 introduced in this step by using the LARS algorithm on the 10 full NARX models to get
 537 the unique sparse NARX models which further reduces the number of unique sparse NARX
 538 models to $N_2 = 5$. Thus, 5 unique sparse models are used to get the coefficients of the sparse
 539 NARX models by ordinary least square for $N = 50$ samples. Finally, the mean error for each
 540 of the sparse NARX models is computed using Equation 17 and the finally selected sparse
 541 NARX model has mean error of $\bar{\epsilon} = 2.95 \times 10^{-4}$ with 5 terms in the NARX polynomial out
 542 of 58 which are $\{y_1(t - \Delta t), y_1(t - 4\Delta t), y_1^3(t - \Delta t), x(t - 4\Delta t), x(t - 4\Delta t) y_1^2(t - 4\Delta t)\}$.
 543 Therefore, only 5 number of Kriging models are required to calibrate for converting the sparse

544 NARX model stochastic which is highly efficient.

545 The same example has also been solved by reducing the initial number of sample points
546 with $N = 20$. For this case, initially, 3 samples are selected as the measure of non-linearity
547 and 3 full NARX models are constructed using the basis function as mentioned in Equa-
548 tion 29. Further, all the 3 NARX models are found unique sparse NARX model utilizing
549 the LARS. The predicted mean error for the 20 samples using the finally selected sparse
550 NARX model was $\bar{\epsilon} = 1.85 \times 10^{-6}$ with 5 terms in the polynomial. Here, the 5 terms
551 are $\{y_1(t - \Delta t), y_1(t - 4\Delta t), x(t), x(t - 4\Delta t), x(t - 4\Delta t)y_1^2(t - 4\Delta t)\}$. The same num-
552 ber of Kriging models are calibrated as the previous case to make the sparse NARX model
553 stochastic.

554 The time dependent mean and standard deviation of the displacement (y_1) and velocity
555 (\dot{y}_1) are predicted for this problem. The velocities are predicted by numerical differentiation
556 using the Newton's central difference scheme. The time dependent statistical characteristics
557 of displacement and velocity are plotted in Figure 14 and Figure 15 respectively. It is evident
558 from both the figures that the sparse KNARX performs very well even with very few sample
559 points.

560 Similar to the previous examples, the instantaneous predicted displacements by the sparse
561 KNARX are compared with the MCS results at three different time instances. The instan-
562 taneous scatter plots and the PDFs are plotted in Figure 16. The error of the predicted
563 responses at instantaneous time instances along with the R^2 values are given in Table 10.
564 The error for a particular time instance ($\epsilon_{y_1(t)}$) is computed using Equation 15. The results
565 show a high accuracy of the sparse KNARX in predicting the instantaneous response char-
566 acteristics. The accuracy and the efficiency are also noticed for the sparse KNARX at 30 s
567 ($N = 50, \epsilon_{y_1(t)} = 2.3 \times 10^{-3}$) as compared to the recently proposed sparse PCE-NARX
568 model ($N = 100, \epsilon_{y_1(t)} = 4.21 \times 10^{-3}$) (Mai et al. 2016). The accuracy of the stochastic
569 response behavior has also been checked by plotting the absolute maximum displacement
570 and the velocity in Figure 17 and Figure 18 respectively. A very accurate result is noticed

571 using sample points $N = 50$.

572 The mean state space behavior of the predicted responses are also plotted in Figure 19.
573 It is noticeable that the mean state space behavior is predicted well even with low number of
574 sample points ($N = 20$) by the sparse KNARX. The global mean error of the predicted time
575 series is measured by using Equation 17. Along with this, the accuracy in predicting the
576 maximum absolute displacement using different sample points are calculated using Equa-
577 tion 15 and R^2 value. All these results along with the efficiency measurement metrics of the
578 proposed sparse KNARX are listed in Table 11. Here, the efficiency of the sparse KNARX
579 model has not been emphasized by the number of surrogate model calibration (n_K), instead
580 it is greatly affected by the initial number of model evaluations (N). All the results show
581 sparse KNARX perform very well in all instances.

582 CONCLUSION

583 A novel method has been proposed in the present paper for uncertainty propagation
584 and quantification of nonlinear stochastic dynamical systems in the time domain. The well
585 known NARX model (a nonlinear system identification technique) has been coupled with
586 Kriging (a high fidelity surrogate model) to propose the method referred as KNARX. Further,
587 the sparsity in the NARX model has been introduced by utilizing the LARS algorithm to
588 formulate the sparse KNARX model.

589 The proposed sparse KNARX model has been used for uncertainty quantification on three
590 typical nonlinear stochastic dynamical systems: a 1-DOF duffing oscillator, a Bouc-Wen
591 oscillator and a 2-DOF dynamical system. Time dependent mean and standard deviation
592 are predicted for all the examples. Along with this, PDFs are plotted for instantaneous
593 response characteristics at some time instances and for maximum absolute response. All
594 the results have been compared with full scale MCS. The first example is also solved with
595 Kriging, whereas the first two examples are computed by the sparse PCE-NARX model (Mai
596 et al. 2016). The accuracy has been measured by a global mean error of the predicted model
597 whereas the efficiency has been measured by number of initial sample points, number of

598 surrogate model evaluations and CPU time. The results for the first example have shown
599 that sparse KNARX outperforms Kriging in all cases even with a very limited number of
600 initial samples. Consequently, Kriging has not been utilized for the last two examples. On
601 the other hand, the sparse PCE-NARX model predicted acceptable results by using a higher
602 number of sample points than the sparse KNARX for the first two examples. Along with
603 this, the efficiency of the sparse KNARX model is higher as compared to the sparse PCE-
604 NARX model considering the CPU time. All the results predicted by the sparse KNARX are
605 in excellent agreement with the full scale MCS results in all instances. The introduction of
606 sparsity has reduced substantially the number of Kriging model evaluations which ultimately
607 enhanced the efficiency of the sparse KNARX model. All the examples have also been solved
608 by reducing number of model evaluations to check the accuracy of the fitted model with less
609 sample points. All the results have been predicted very close to the MCS results even with
610 very few model evaluations for all the examples. This study suggests that reducing the
611 number of model evaluation does not affect the accuracy too much due to the fact that the
612 major non-linearity of the system has been captured by the sparse NARX model.

613 In conclusion, the proposed sparse KNARX outperforms the state-of-the-art methods
614 (Mai and Sudret 2017; Mai et al. 2016) in predicting the time dependent stochastic response
615 characteristics for nonlinear dynamical systems. The accuracy and the efficiency of the
616 proposed method has also been illustrated through several examples. The proposed sparse
617 KNARX model is limited to use for the nonlinear dynamical systems having parametric
618 uncertainties only. Therefore, this model is not useful for the dynamical systems with white
619 noise excitation for the prediction of time dependent statistical responses.

620 **DATA AVAILABILITY**

621 Some or all data, models, or code generated or used during the study are available from
622 the corresponding author by request.

- 623 • NARX model

624 • NARX polynomial basis functions

625 REFERENCES

- 626 Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J.,
627 Parthasarathy, R. J., Paulo, R., Sacks, J., and Walsh, D. (2007). “Computer model vali-
628 dation with functional output.” *The Annals of Statistics*, 35(5), 1874–1906.
- 629 Bhattacharyya, B. (2018). “A Critical Appraisal of Design of Experiments for Uncertainty
630 Quantification.” *Archives of Computational Methods in Engineering*, 25(3), 727–751.
- 631 Billings, S. and Wei, H.-L. (2005). “A New Class of Wavelet Networks for Nonlinear System
632 Identification.” *IEEE Transactions on Neural Networks*, 16(4), 862–874.
- 633 Billings, S. A. (2013). *Nonlinear System Identification: NARMAX Methods in the Time,*
634 *Frequency, and Spatio-Temporal Domains*. Wiley, first edition.
- 635 Billings, S. A., Chen, S., and Korenberg, M. J. (1989). “Identification of MIMO non-linear
636 systems using a forward-regression orthogonal estimator.” *International Journal of Con-*
637 *trol*, 49(6), 2157–2189.
- 638 Blatman, G. and Sudret, B. (2011). “Adaptive sparse polynomial chaos expansion based on
639 least angle regression.” *Journal of Computational Physics*, 230(6), 2345–2367.
- 640 Bouc, R. (1967). “Forced vibration of mechanical systems with hysteresis.” *Proceedings of*
641 *the Fourth Conference on Nonlinear Oscillation*, Prague, Czechoslovakia, 315.
- 642 Cantelmo, C. and Piroddi, L. (2010). “Adaptive model selection for polynomial NARX
643 models.” *IET Control Theory & Applications*, 4(12), 2693–2706.
- 644 Chakraborty, S. and Chowdhury, R. (2015). “Polynomial Correlated Function Expansion for
645 Nonlinear Stochastic Dynamic Analysis.” *Journal of Engineering Mechanics*, 141(3), 1–11.
- 646 Chakraborty, S. and Chowdhury, R. (2017). “An efficient algorithm for building locally
647 refined hpadaptive H-PCFE: Application to uncertainty quantification.” *Journal of Com-*
648 *putational Physics*, 351, 59–79.
- 649 Chatterjee, T. and Chowdhury, R. (2017). “An efficient sparse Bayesian learning framework
650 for stochastic response analysis.” *Structural Safety*, 68, 1–14.

651 Chatterjee, T. and Chowdhury, R. (2018). “Refined sparse Bayesian learning configuration
652 for stochastic response analysis.” *Probabilistic Engineering Mechanics*, 52, 15–27.

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

655 Chen, S., Billings, S. A., and Grant, P. M. (1990). “Non-linear system identification using
656 neural networks.” *International Journal of Control*, 51(6), 1191–1214.

657 Cheng, Y., Wang, L., Yu, M., and Hu, J. (2011). “An efficient identification scheme for a
658 nonlinear polynomial NARX model.” *Artificial Life and Robotics*, 16(1), 70–73.

659 Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). “Least angle regression.” *The
660 Annals of Statistics*, 32(2), 407–499.

661 Gerritsma, M., Steen, J.-B. V. D., Vos, P., and Karniadakis, G. (2010). “Time-dependent
662 generalized polynomial chaos.” *Journal of Computational Physics*, 229(22), 8333–8363.

663 Grigoriu, M. (1996). “Response of dynamic systems to poisson white noise.” *Journal of
664 Sound and Vibration*, 195(3), 375–389.

665 Huang, X., Chen, J., and Zhu, H. (2016). “Assessing small failure probabilities by AKSS:
666 An active learning method combining Kriging and Subset Simulation.” *Structural Safety*,
667 59, 86–95.

668 Jacquelin, E., Adhikari, S., Sinou, J.-J., and Friswell, M. (2015). “Polynomial chaos ex-
669 pansion in structural dynamics: Accelerating the convergence of the first two statistical
670 moment sequences.” *Journal of Sound and Vibration*, 356, 144–154.

671 Jacquelin, E., Dessombz, O., Sinou, J. J., Adhikari, S., and Friswell, M. I. (2017). “Polyno-
672 mial chaos-based extended Padé expansion in structural dynamics.” *International Journal
673 for Numerical Methods in Engineering*, 111(12), 1170–1191.

674 Kaymaz, I. (2005). “Application of kriging method to structural reliability problems.” *Struc-
675 tural Safety*, 27(2), 133–151.

676 Kersaudy, P., Sudret, B., Varsier, N., Picon, O., and Wiart, J. (2015). “A new surrogate
677 modeling technique combining Kriging and polynomial chaos expansions - Application to

678 uncertainty analysis in computational dosimetry.” *Journal of Computational Physics*, 286,
679 103–117.

680 Krige, D. (1951). “A Statistical Approach to Some Basic Mine Valuation Problems on
681 the Witwatersrand.” *Journal of the Chemical, Metallurgical and Mining Society of South
682 Africa*, 52(6), 119 – 139.

683 Kundu, A. and Adhikari, S. (2014). “Transient Response of Structural Dynamic Systems
684 with Parametric Uncertainty.” *Journal of Engineering Mechanics*, 140(2), 315–331.

685 Leontaritis, I. J. and Billings, S. A. (1985). “Input-output parametric models for non-linear
686 systems Part II: stochastic non-linear systems.” *International Journal of Control*, 41(2),
687 329–344.

688 Li, K., Peng, J.-X., and Irwin, G. (2005). “A fast nonlinear model identification method.”
689 *IEEE Transactions on Automatic Control*, 50(8), 1211–1216.

690 Lu, C., Feng, Y.-W., Liem, R. P., and Fei, C.-W. (2018). “Improved Kriging with ex-
691 tremum response surface method for structural dynamic reliability and sensitivity analy-
692 ses.” *Aerospace Science and Technology*, 76, 164–175.

693 Luchtenburg, D. M., Brunton, S. L., and Rowley, C. W. (2014). “Long-time uncertainty
694 propagation using generalized polynomial chaos and flow map composition.” *Journal of
695 Computational Physics*, 274, 783–802.

696 Lucor, D. and Karniadakis, G. E. (2004). “Adaptive Generalized Polynomial Chaos for
697 Nonlinear Random Oscillators.” *SIAM Journal on Scientific Computing*, 26(2), 720–735.

698 Lucor, D., Su, C. H., and Karniadakis, G. E. (2004). “Generalized polynomial chaos and
699 random oscillators.” *International Journal for Numerical Methods in Engineering*, 60(3),
700 571–596.

701 Mai, C. V. (2016). “Polynomial chaos expansions for uncertain dynamical systems Appli-
702 cations in earthquake engineering.” Ph.D. thesis, Chair of Risk, Safety & Uncertainty
703 Quantification, ETH Zurich, Switzerland.

704 Mai, C. V., Spiridonakos, M. D., Chatzi, E. N., and Sudret, B. (2016). “Surrogate modeling

705 for stochastic dynamical systems by combining nonlinear autoregressive with exogenous
706 input models and polynomial chaos expansions.” *International Journal for Uncertainty*
707 *Quantification*, 6(4), 313–339.

708 Mai, C. V. and Sudret, B. (2017). “Surrogate models for oscillatory systems using sparse
709 polynomial chaos expansions and stochastic time warping.” *SIAM/ASA Journal on Un-*
710 *certainty Quantification*, 5(1), 540–571.

711 Maitre, O. P. L., Mathelin, L., Knio, O., and Hussaini, M. (2010). “Asynchronous Time
712 Integration for Polynomial Chaos Expansion of Uncertain Periodic Dynamics.” *Discrete*
713 *Continuum Dynamic Systems - Series A*, 28(1), 199–226.

714 Marelli, S. and Sudret, B. (2018). “UQLab user manual – Polynomial chaos expansions.”
715 *Report no.*, Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich. Report
716 UQLab-V1.1-104.

717 Matheron, G. (1963). “Principles of geostatistics.” *Economic Geology*, 58(8), 1246–1266.

718 Mukhopadhyay, T., Chakraborty, S., Dey, S., Adhikari, S., and Chowdhury, R. (2017). “A
719 Critical Assessment of Kriging Model Variants for High-Fidelity Uncertainty Quantifica-
720 tion in Dynamics of composite Shells.” *Archives of Computational Methods in Engineering*,
721 24(3), 495–518.

722 Muscolino, G., Ricciardi, G., and Cacciola, P. (2003). “Monte Carlo simulation in the stochas-
723 tic analysis of non-linear systems under external stationary Poisson white noise input.”
724 *International Journal of Non-Linear Mechanics*, 38(8), 1269–1283.

725 Ozen, H. C. and Bal, G. (2017). “A dynamical polynomial chaos approach for long-time
726 evolution of SPDEs.” *Journal of Computational Physics*, 343, 300–323.

727 Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). “Design and Analysis of
728 Computer Experiments.” *Statistical Science*, 4(4), 409–423.

729 Santner, T., Williams, B., and Notz, W. (2003). *The design and analysis of computer exper-*
730 *iments*. Springer, 1 edition.

731 Sheppard, C. (1969). “Computer simulation of stochastic processes through model-sampling

732 (Monte Carlo) techniques.” *FEBS Letters*, 2, S14–S21.

733 Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F. (2001). “Kriging Models
734 for Global Approximation in Simulation-Based Multidisciplinary Design Optimization.”
735 *AIAA Journal*, 39(12), 2233–2241.

736 Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson,
737 H., and Juditsky, A. (1995). “Nonlinear black-box modeling in system identification: a
738 unified overview.” *Automatica*, 31(12), 1691–1724.

739 Spiridonakos, M. and Chatzi, E. (2015). “Metamodeling of dynamic nonlinear structural
740 systems through polynomial chaos NARX models.” *Computers & Structures*, 157, 99–113.

741 Sugai, M., Mori, Y., and Ogawa, K. (2015). “Application of Kriging method into practical
742 estimations of earthquake ground motion hazards.” *Journal of Structural and Construction
743 Engineering*, 80(707), 39–46.

744 Tong, C., Sun, Z., Zhao, Q., Wang, Q., and Wang, S. (2015). “A hybrid algorithm for reli-
745 ability analysis combining Kriging and subset simulation importance sampling.” *Journal
746 of Mechanical Science and Technology*, 29(8), 3183–3193.

747 Tsungnan Lin, Horne, B., Tino, P., and Giles, C. (1996). “Learning long-term dependencies
748 in NARX recurrent neural networks.” *IEEE Transactions on Neural Networks*, 7(6), 1329–
749 1338.

750 Wan, X. and Karniadakis, G. E. (2005). “An adaptive multi-element generalized polynomial
751 chaos method for stochastic differential equations.” *Journal of Computational Physics*,
752 209(2), 617–642.

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

755 Wen, Y.-K. (1976). “Method for Random Vibration of Hysteretic Systems.” *Journal of the
756 Engineering Mechanics Division*, 102(2), 249–263.

757 Wiener, N. (1938). “The homogeneous chaos.” *American Journal of Mathematics*, 60(4),
758 897–936.

- 759 Worden, K., Becker, W. E., Rogers, T. J., and Cross, E. J. (2018). “On the confidence bounds
760 of Gaussian process NARX models and their higher-order frequency response functions.”
761 *Mechanical Systems and Signal Processing*, 104, 188–223.
- 762 Xie, H., Tang, H., and Liao, Y.-H. (2009). “Time series prediction based on NARX neural
763 networks: An advanced approach.” *Eighth International Conference on Machine Learning
764 and Cybernetics*, Baoding, IEEE, 1275–1279 (jul).
- 765 Xiu, D. and Karniadakis, G. E. (2002). “The Wiener-Askey polynomial chaos for stochastic
766 differential equation.” *SIAM Journal on Scientific Computing Scientific Computing*, 24(2),
767 619–644.
- 768 Zhang, L. and Li, K. (2015). “Forward and backward least angle regression for nonlinear
769 system identification.” *Automatica*, 53, 94–102.

770 **List of Tables**

771	1	Algorithm for the sparse KNARX model	35
772	2	Uncertain parameters for the duffing oscillator	36
773	3	The polynomials selected by the LARS algorithm for the duffing oscillator	37
774	4	Accuracy of the surrogate models in predicting instantaneous response characteristics for the duffing oscillator	38
775			
776	5	Comparison of accuracy and efficiency of the surrogate models for the duffing oscillator	39
777			
778	6	Uncertain parameters for the Bouc-Wen oscillator	40
779	7	Accuracy of the surrogate models in predicting the instantaneous response characteristics for the Bouc-Wen oscillator	41
780			
781	8	Accuracy and efficiency of the surrogate models in computing $\dot{y}(t)$ for the Bouc-Wen oscillator	42
782			
783	9	Uncertain parameters for the 2-DOF dynamical system	43
784	10	Accuracy of sparse KNARX in predicting instantaneous response characteristics for the 2-DOF dynamical system	44
785			
786	11	Accuracy and efficiency of sparse KNARX in computing $y_1(t)$ for the 2-DOF dynamical system	45
787			

TABLE 1: Algorithm for the sparse KNARX model

-
1. Declare the d number of uncertain variables with the type of distribution.
 2. Generate N number of sample points for the uncertain variables.
 3. Draw the restoring force versus response curve to decide the threshold value for capturing nonlinear response series.
 4. Get the response of the dynamical system upto time T for each of the N samples.
 5. Select the samples and the response series having high order non-linearity using some threshold (according to step 3) on the response series. This step selects N_1 samples ($N_1 < N$).
 6. Decide the maximum time lags n_{x_m} and n_{y_m} for the excitation and response quantity, respectively. Decide the type of NARX polynomial basis function along with the maximum degrees also.
 7. For each of the N_1 samples, build the NARX model (see Equation 10 and 11) which has M number of terms.
 8. Select the most important terms for each of the N_1 NARX models using LARS (Efron et al. 2004).
 9. Select the N_2 different sparse NARX models (see Remark 2) having similar terms in the NARX polynomial basis ($N_2 \leq N_1$).
 10. Perform ordinary least square on N samples for the N_2 sparse NARX models to get the NARX coefficients of all the response series.
 11. Reconstruct the N response series using the coefficients computed in step 10 by all the N_2 sparse NARX models.
 12. Predict the mean error $\bar{\epsilon}$ for all the N_2 number of sparse NARX models using Equation 17.
 13. Select the most appropriate sparse NARX model having $\bar{\epsilon}$ lower than some threshold value (1×10^{-3} in the present paper) and less number of terms (M_1) in the polynomial.
 14. Calibrate $M_1 < M$ number of Kriging models using the NARX coefficients as the response parameter and the sample points generated in step 2 as the uncertain input quantities.
 15. Generate a large number of new untried samples for the prediction.
 16. Predict the NARX coefficients for the untried samples using Kriging models generated in step 14.
 17. Predict the response series, in a auto-regressive way (see Equation 21), at the untried samples generated in step 15 using the coefficients of step 16 and the M_1 number of NARX polynomial bases selected in step 13.
-

TABLE 2: Uncertain parameters for the duffing oscillator

Variable	Distribution type	Mean	Standard deviation	Unit
ω	Uniform	2π	$\frac{\pi}{\sqrt{3}}$	rad s^{-1}
ζ	Uniform	0.03	$\frac{0.015}{\sqrt{3}}$	—
ε	Uniform	100	$\frac{\sqrt{3}}{10}$	—
A	Normal	0.6	0.06	N
ω_x	Normal	1	0.1	rad s^{-1}

TABLE 3: The polynomials selected by the LARS algorithm for the duffing oscillator

Model 1 ($M_1 = 9$)	Model 2 ($M_1 = 8$)
$x(t)$	$x(t)$
$x(t - 2\Delta t)$	$x(t - 2\Delta t)$
$y(t - \Delta t)$	$y(t - \Delta t)$
$y(t - 2\Delta t)$	$y(t - 2\Delta t)$
$y^2(t - \Delta t)$	$y^2(t - \Delta t)$
$y^3(t - \Delta t)$	$y^3(t - \Delta t)$
$y^3(t - 2\Delta t)$	$y^3(t - 2\Delta t)$
$x(t - 2\Delta t)y^2(t - 2\Delta t)$	$x(t - 2\Delta t)y^2(t - 2\Delta t)$
$x(t - 2\Delta t)y(t - 2\Delta t)$	

TABLE 4: Accuracy of the surrogate models in predicting instantaneous response characteristics for the duffing oscillator

Method	N	Time instance	$\epsilon_{y(t)}$	R^2
Kriging	50	$t = 10$ s	337.8×10^{-3}	0.6622
Sparse PCE-NARX	35		9.6×10^{-3}	0.9904
Sparse KNARX	21		8.8×10^{-3}	0.9912
Sparse KNARX	25		4.6×10^{-3}	0.9954
Kriging	50	$t = 20$ s	926.9×10^{-3}	0.0731
Sparse PCE-NARX	35		4.4×10^{-3}	0.9956
Sparse KNARX	21		5.4×10^{-3}	0.9946
Sparse KNARX	25		3.1×10^{-3}	0.9969
Kriging	50	$t = 30$ s	1701.1×10^{-3}	—
Sparse PCE-NARX	35		2.3×10^{-3}	0.9977
Sparse KNARX	21		3.0×10^{-3}	0.9970
Sparse KNARX	25		2.0×10^{-3}	0.9980

TABLE 5: Comparison of accuracy and efficiency of the surrogate models for the duffing oscillator

Method	Accuracy			Efficiency	
	$\bar{\epsilon}$	$\epsilon_{\max(y(t))}$	$R^2_{\max(y(t))}$	n_K	CPU time
Kriging ($N = 50$)	1.4971	2100.7×10^{-3}	—	3001	530.95 s
Sparse PCE-NARX ($N = 35$)	1.1×10^{-3}	2.9×10^{-3}	0.9971	9	30.55 s
Sparse KNARX ($N = 21$)	1.2×10^{-3}	1.4×10^{-3}	0.9986	9	26.13 s
Sparse KNARX ($N = 25$)	5.89×10^{-4}	1.0×10^{-3}	0.9990	8	24.63 s
MCS	—	—	—	—	705.43 s

TABLE 6: Uncertain parameters for the Bouc-Wen oscillator

Variable	Distribution type	Mean	Standard deviation	Unit
ζ	Uniform	0.02	0.002	—
ω	Uniform	2π	0.2π	rad s^{-1}
α	Uniform	50	5	—
A	Uniform	1	0.1	N
ω_x	Uniform	π	0.1π	rad s^{-1}

TABLE 7: Accuracy of the surrogate models in predicting the instantaneous response characteristics for the Bouc-Wen oscillator

Method	N	Time instance	Displacement		Velocity	
			ϵ	R^2	ϵ	R^2
Sparse PCE-NARX	50	$t = 10$ s	660.0×10^{-3}	0.3370	44.0×10^{-3}	0.9557
Sparse KNARX	10		6.8×10^{-3}	0.9932	0.9×10^{-3}	0.9991
Sparse KNARX	40		3.8×10^{-3}	0.9962	0.5×10^{-3}	0.9995
Sparse PCE-NARX	50	$t = 20$ s	880.0×10^{-3}	0.1224	48.0×10^{-3}	0.9519
Sparse KNARX	10		8.6×10^{-3}	0.9914	1.0×10^{-3}	0.9990
Sparse KNARX	40		5.6×10^{-3}	0.9944	0.5×10^{-3}	0.9995
Sparse PCE-NARX	50	$t = 30$ s	1140.0×10^{-3}	—	50.0×10^{-3}	0.9503
Sparse KNARX	10		10.8×10^{-3}	0.9892	1.1×10^{-3}	0.9989
Sparse KNARX	40		7.8×10^{-3}	0.9922	0.5×10^{-3}	0.9995

TABLE 8: Accuracy and efficiency of the surrogate models in computing $\dot{y}(t)$ for the Bouc-Wen oscillator

Method	$\bar{\epsilon}$	Accuracy		Efficiency	
		$\epsilon_{\max(y(t))}$	$R_{\max(y(t))}^2$	n_K	CPU time
Sparse PCE-NARX ($N = 50$)	9.3×10^{-2}	2.6×10^{-1}	0.7414	8	27.40 s
Sparse KNARX ($N = 10$)	1.9×10^{-3}	4.6×10^{-3}	0.9954	8	19.12 s
Sparse KNARX ($N = 40$)	8.0×10^{-4}	1.9×10^{-3}	0.9981	8	21.72 s
MCS	—	—	—	—	1223.17 s

TABLE 9: Uncertain parameters for the 2-DOF dynamical system

Variable	Distribution type	Mean	Standard deviation	Unit
k_s	Normal	2000	200	N/m ³
k_u	Normal	2000	200	N m ⁻¹
m_s	Normal	20	2	kg
m_u	Normal	40	4	kg
c	Normal	600	60	N s m ⁻¹
A	Uniform	0.1	$\frac{0.01}{\sqrt{3}}$	m
ω_x	Uniform	2π	$\frac{0.2\pi}{\sqrt{3}}$	rad s ⁻¹

TABLE 10: Accuracy of sparse KNARX in predicting instantaneous response characteristics for the 2-DOF dynamical system

N	Time instance	$\epsilon_{y_1(t)}$	R^2
20	$t = 10$ s	11.6×10^{-3}	0.9884
50		1.1×10^{-3}	0.9989
20	$t = 20$ s	16.2×10^{-3}	0.9838
50		1.9×10^{-3}	0.9981
20	$t = 30$ s	16.2×10^{-3}	0.9838
50		2.3×10^{-3}	0.9977

TABLE 11: Accuracy and efficiency of sparse KNARX in computing $y_1(t)$ for the 2-DOF dynamical system

Method	Accuracy			Efficiency	
	$\bar{\epsilon}$	$\epsilon_{\max(y_1(t))}$	$R_{\max(y_1(t))}^2$	n_K	CPU time
Sparse KNARX ($N = 20$)	8.9×10^{-3}	15.4×10^{-3}	0.9846	5	7.76 s
Sparse KNARX ($N = 50$)	1.2×10^{-3}	1.7×10^{-3}	0.9983	5	22.62 s
MCS	—	—	—	—	466.34 s

788 **List of Figures**

789 1 Displacement versus restoring force plot for the duffing oscillator 48

790 2 Displacement versus restoring force plots of two different realizations for the
791 duffing oscillator; (a) 13-th sample, (b) 25-th sample 49

792 3 Statistical response characteristics of the duffing oscillator; (a) Mean, (b)
793 Standard deviation 50

794 4 Comparison of instantaneous response characteristics for the duffing oscillator
795 at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s,
796 (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s,
797 (f) PDF at $t = 30$ s 51

798 5 Comparison of predicted $\max(|y(t)|)$ for the duffing oscillator; (a) Scatter
799 plot, (b) PDF 52

800 6 Statistical response characteristics for the displacement of the Bouc-Wen os-
801 cillator; (a) Mean, (b) Standard deviation 53

802 7 Statistical response characteristics for the velocity of the Bouc-Wen oscillator;
803 (a) Mean, (b) Standard deviation 54

804 8 Prediction of instantaneous displacement characteristics for the Bouc-Wen
805 oscillator at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at
806 $t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at
807 $t = 30$ s, (f) PDF at $t = 30$ s 55

808 9 Prediction of instantaneous velocity characteristics for the Bouc-Wen oscilla-
809 tor at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s,
810 (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s,
811 (f) PDF at $t = 30$ s 56

812 10 Comparison of predicted $\max(|y(t)|)$ for the Bouc-Wen oscillator; (a) Scatter
813 plot, (b) PDF 57

814	11	Comparison of predicted $\max(\dot{y}(t))$ for Bouc-Wen oscillator; (a) Scatter plot, (b) PDF	58
815			
816	12	Mean trajectory of $y(t)$ and $\dot{y}(t)$ for the Bouc-Wen oscillator; (a) MCS, (b) Sparse PCE-NARX ($N = 50$), (c) Sparse KNARX ($N = 10$), (d) Sparse	
817		KNARX ($N = 40$)	59
818			
819	13	A 2-DOF dynamical system	60
820	14	Statistical response characteristics for displacement ($y_1(t)$) of the 2-DOF dynamical system; (a) Mean, (b) Standard deviation	61
821			
822	15	Statistical response characteristics for velocity ($\dot{y}_1(t)$) of the 2-DOF dynamical system; (a) Mean, (b) Standard deviation	62
823			
824	16	Prediction of instantaneous displacement characteristics for 2-DOF dynamical system at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at	
825		$t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at	
826		$t = 30$ s, (f) PDF at $t = 30$ s	63
827			
828	17	Comparison of predicted $\max(y_1(t))$ for the 2-DOF dynamical system; (a) Scatter plot, (b) PDF	64
829			
830	18	Comparison of predicted $\max(\dot{y}_1(t))$ for the 2-DOF dynamical system; (a) Scatter plot, (b) PDF	65
831			
832	19	Mean trajectory of $y_1(t)$ and $\dot{y}_1(t)$ for the 2-DOF dynamical system; (a) MCS, (b) Sparse KNARX ($N = 20$), (c) Sparse KNARX ($N = 50$)	66
833			

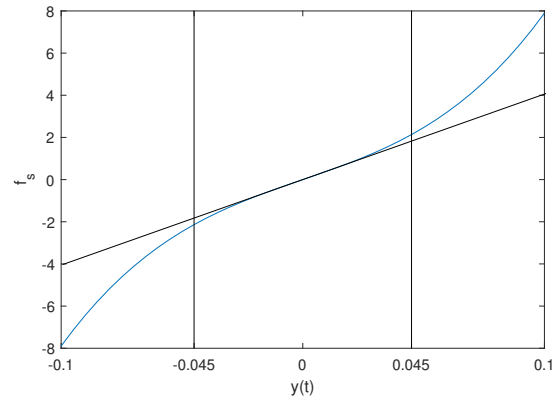
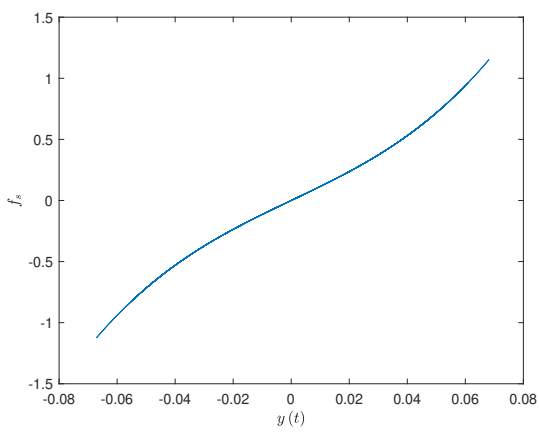
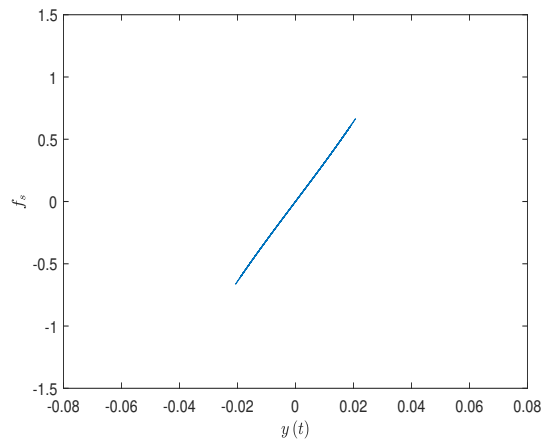


FIG. 1: Displacement versus restoring force plot for the duffing oscillator

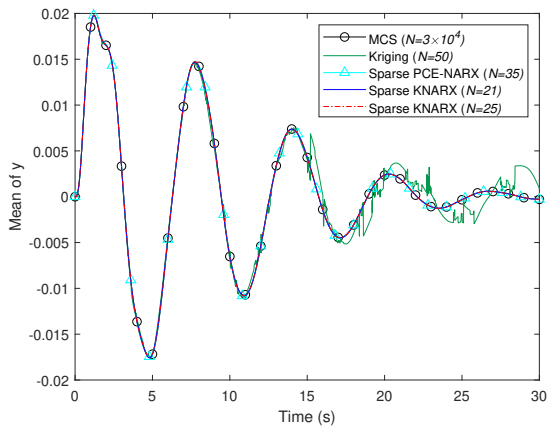


(a)

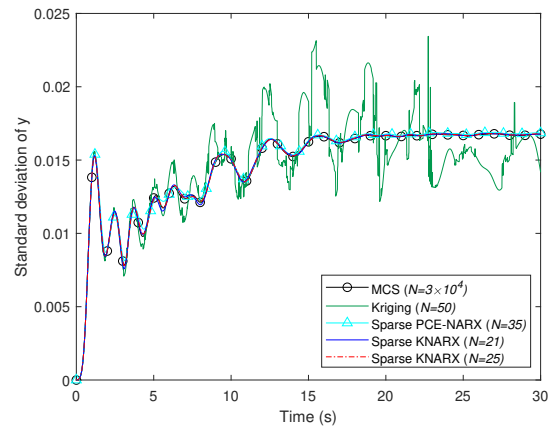


(b)

FIG. 2: Displacement versus restoring force plots of two different realizations for the duffing oscillator; (a) 13-th sample, (b) 25-th sample



(a)



(b)

FIG. 3: Statistical response characteristics of the duffing oscillator; (a) Mean, (b) Standard deviation

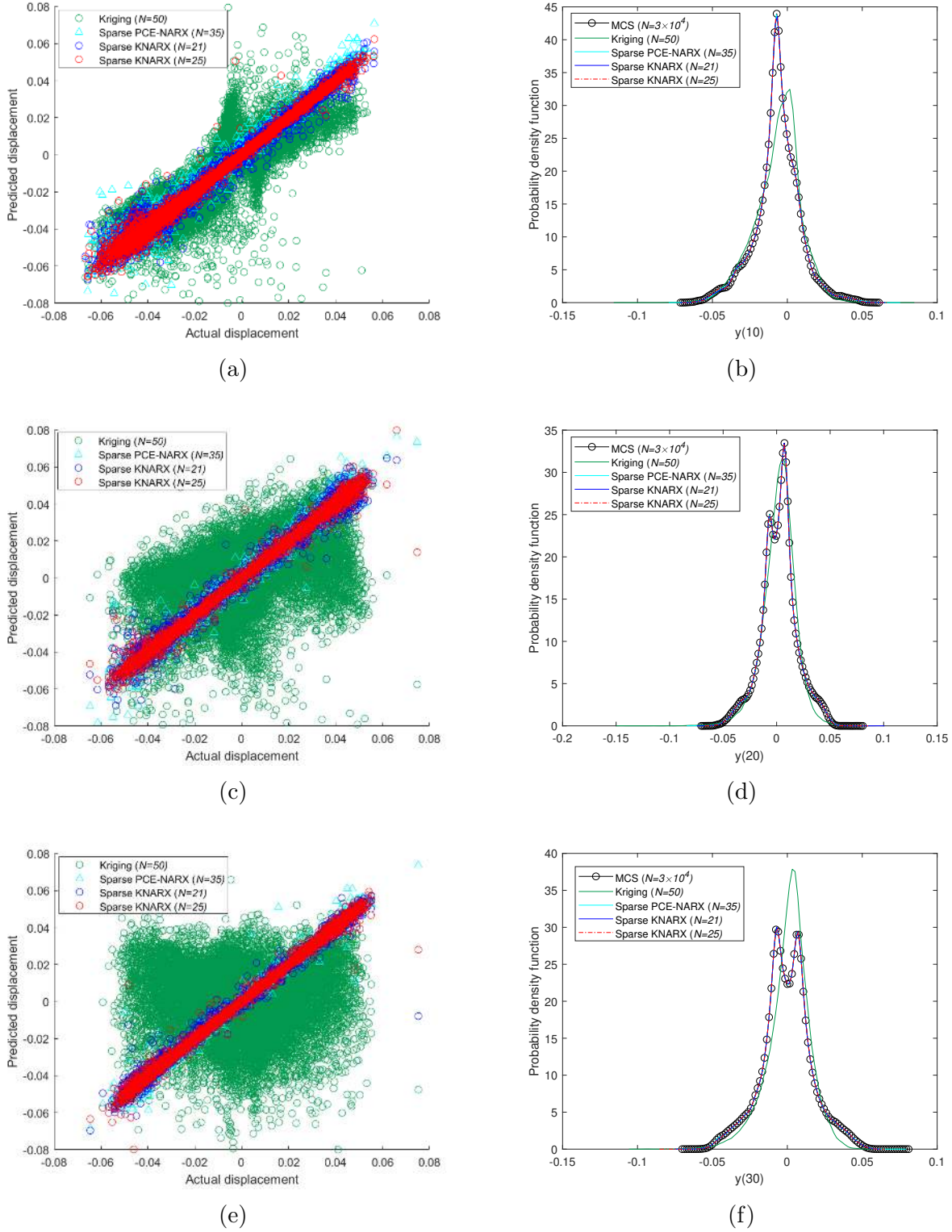


FIG. 4: Comparison of instantaneous response characteristics for the duffing oscillator at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s, (f) PDF at $t = 30$ s

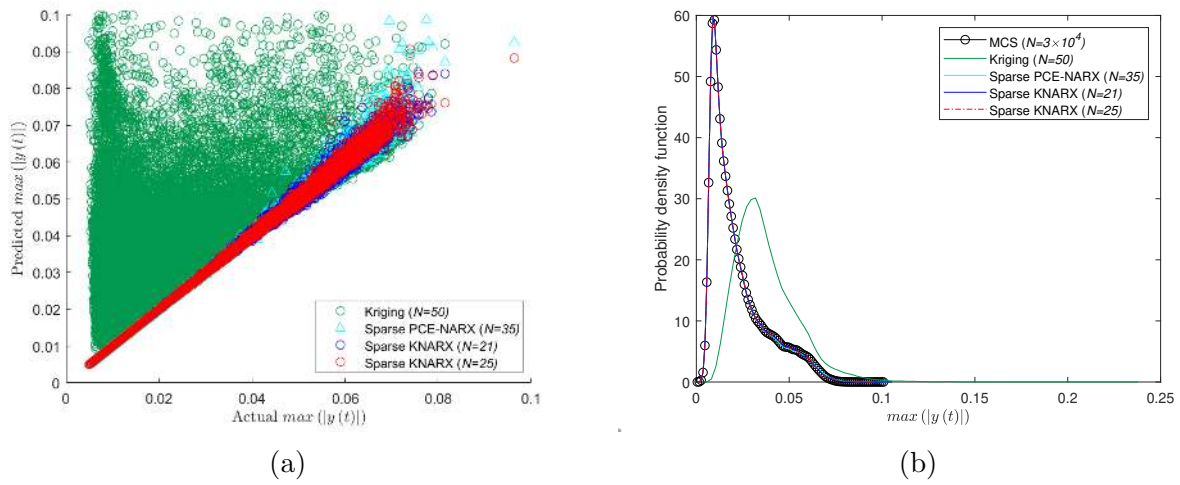
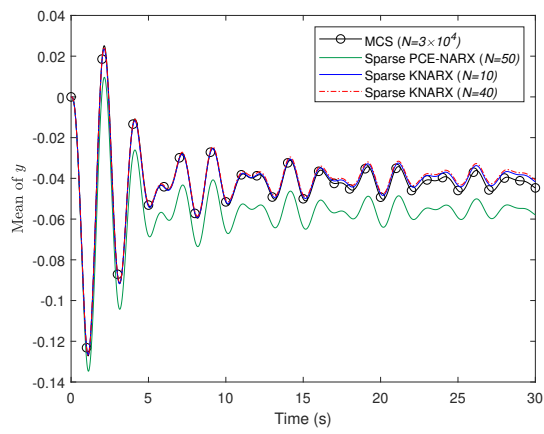
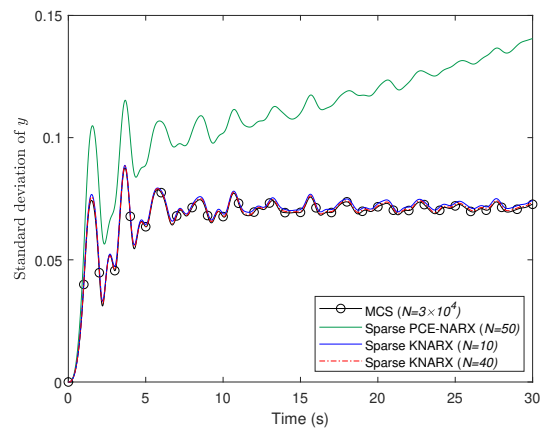


FIG. 5: Comparison of predicted $\max(|y(t)|)$ for the duffing oscillator; (a) Scatter plot, (b) PDF

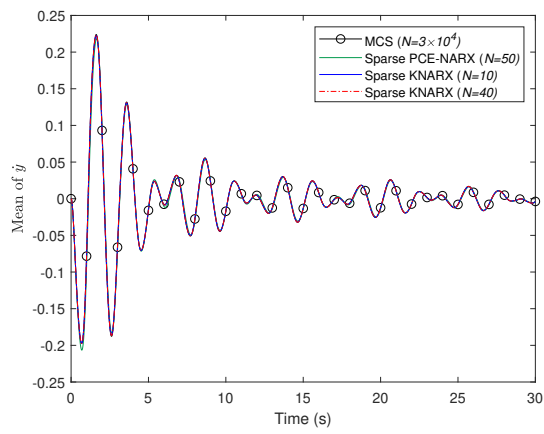


(a)

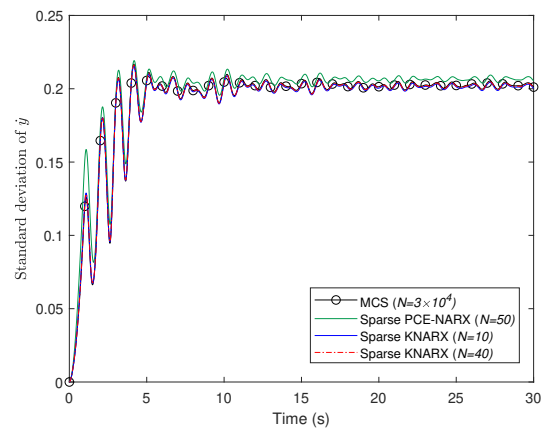


(b)

FIG. 6: Statistical response characteristics for the displacement of the Bouc-Wen oscillator; (a) Mean, (b) Standard deviation

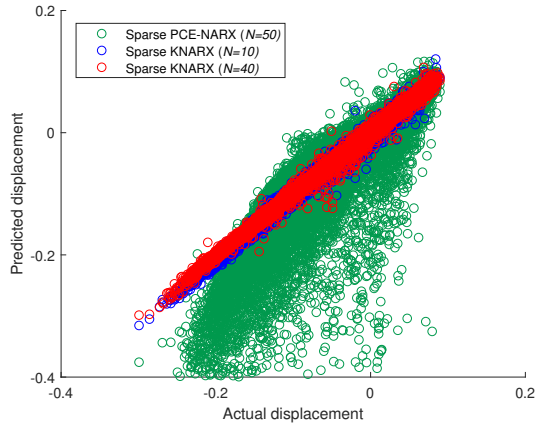


(a)

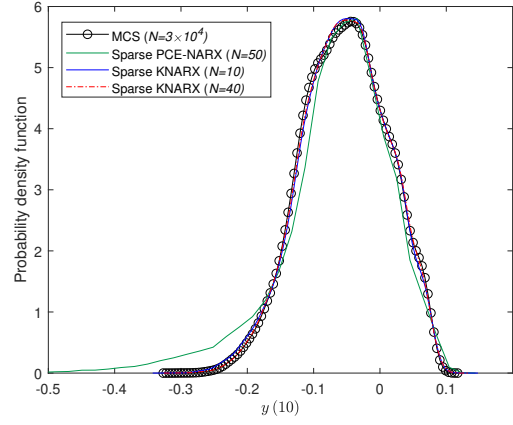


(b)

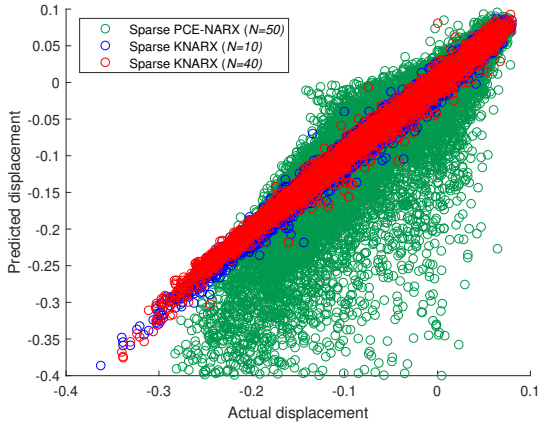
FIG. 7: Statistical response characteristics for the velocity of the Bouc-Wen oscillator; (a) Mean, (b) Standard deviation



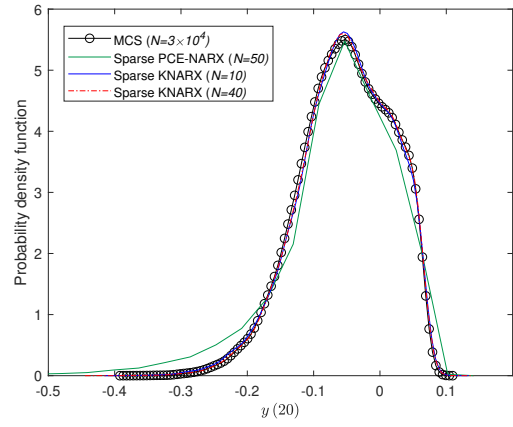
(a)



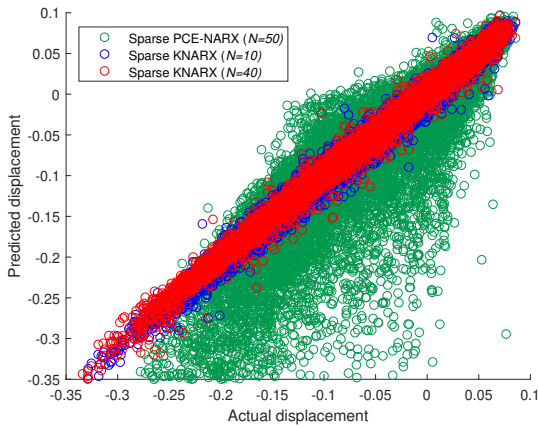
(b)



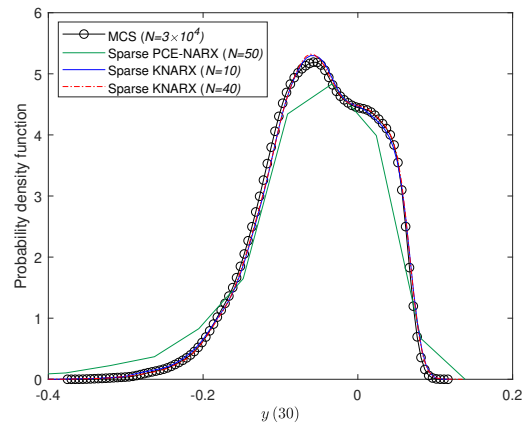
(c)



(d)



(e)



(f)

FIG. 8: Prediction of instantaneous displacement characteristics for the Bouc-Wen oscillator at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s, (f) PDF at $t = 30$ s

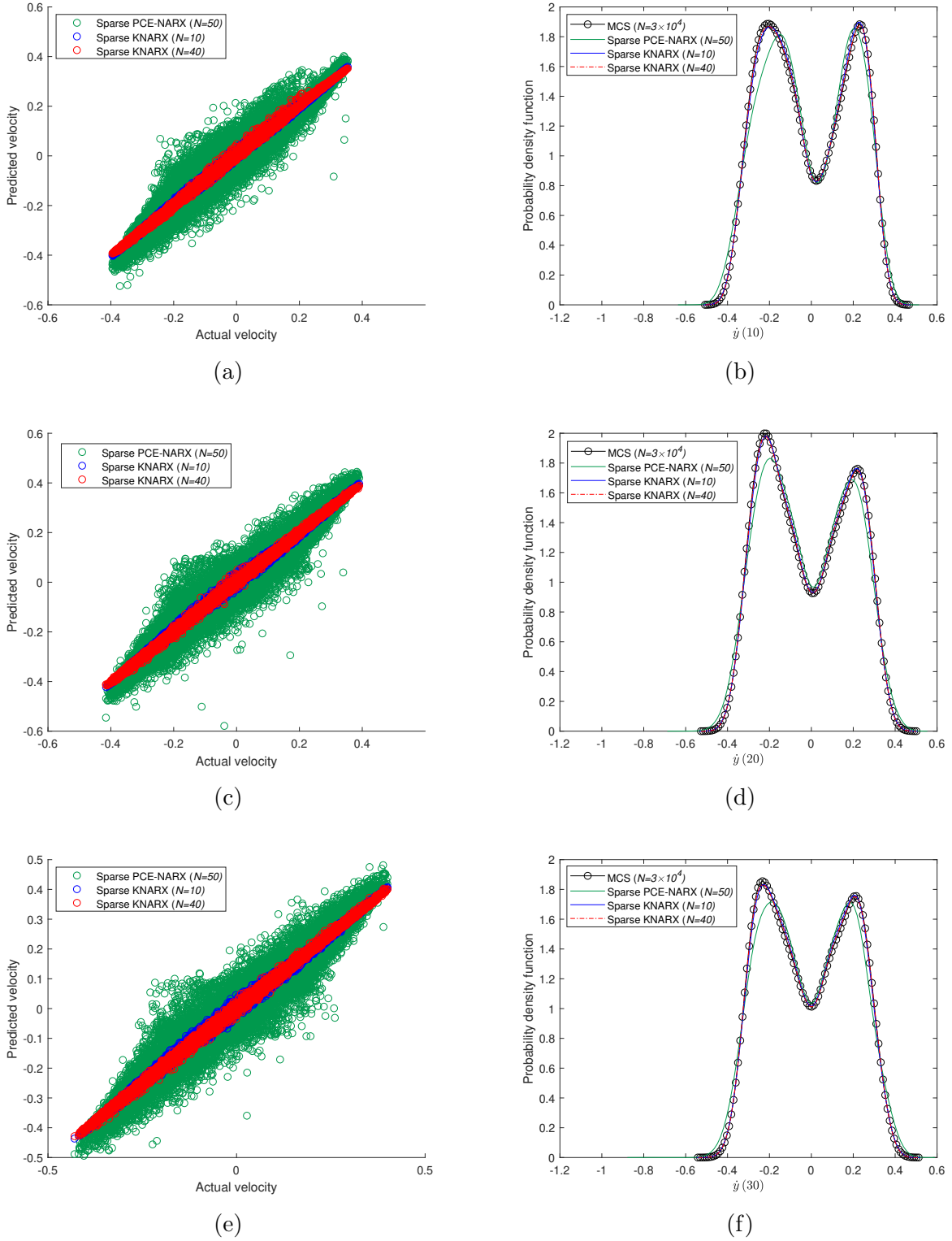
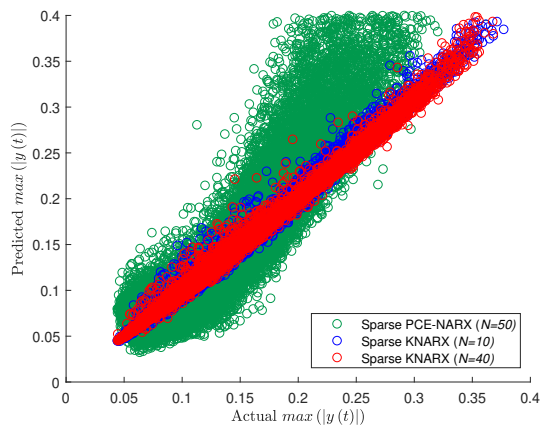
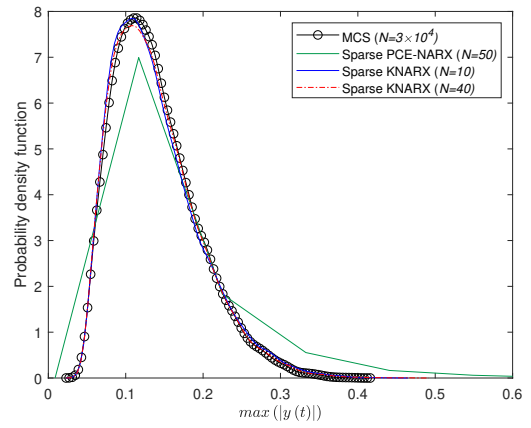


FIG. 9: Prediction of instantaneous velocity characteristics for the Bouc-Wen oscillator at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s, (f) PDF at $t = 30$ s

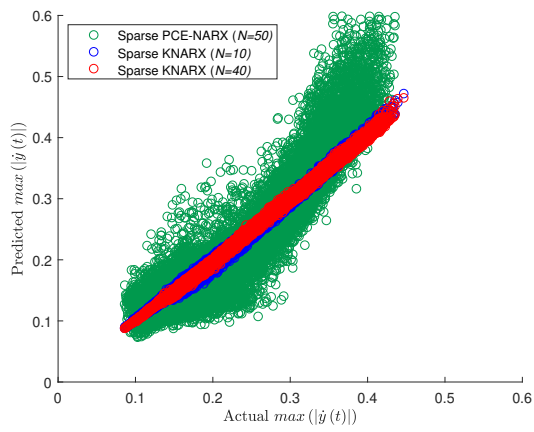


(a)

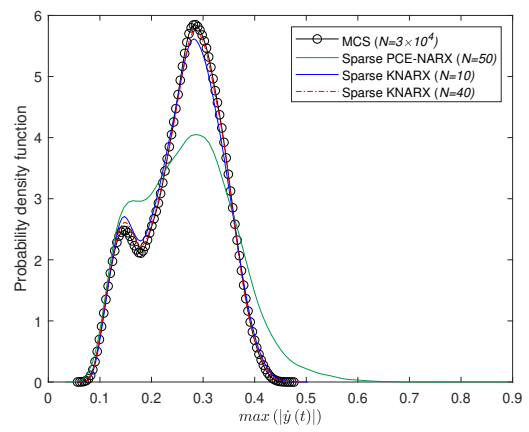


(b)

FIG. 10: Comparison of predicted $\max(|y(t)|)$ for the Bouc-Wen oscillator; (a) Scatter plot, (b) PDF

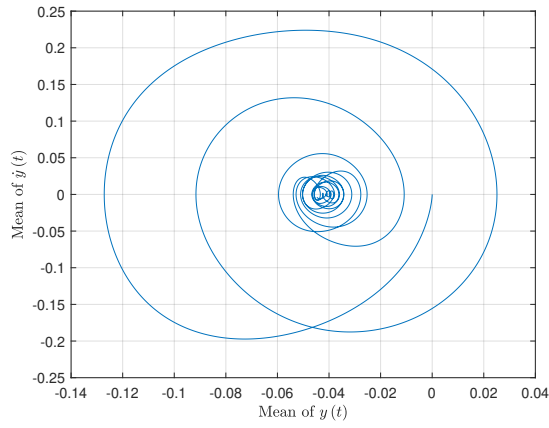


(a)

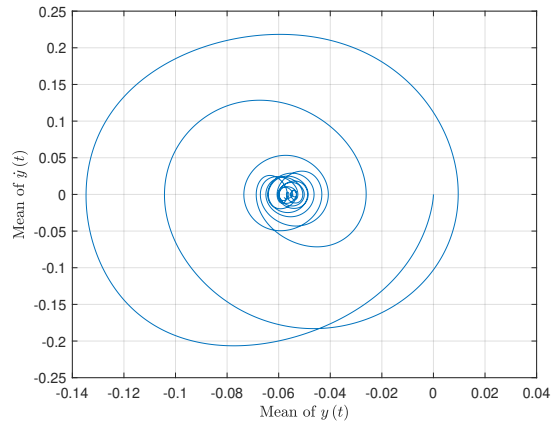


(b)

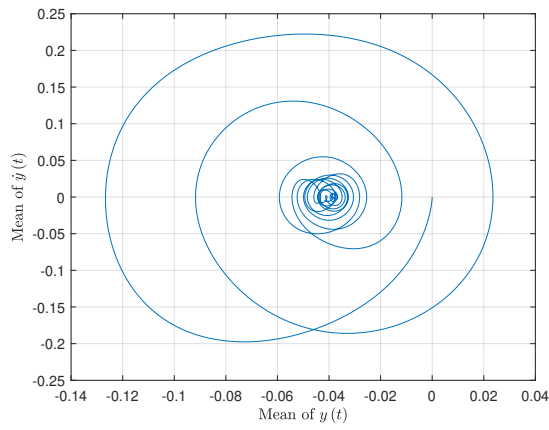
FIG. 11: Comparison of predicted $\max(|\dot{y}(t)|)$ for Bouc-Wen oscillator; (a) Scatter plot, (b) PDF



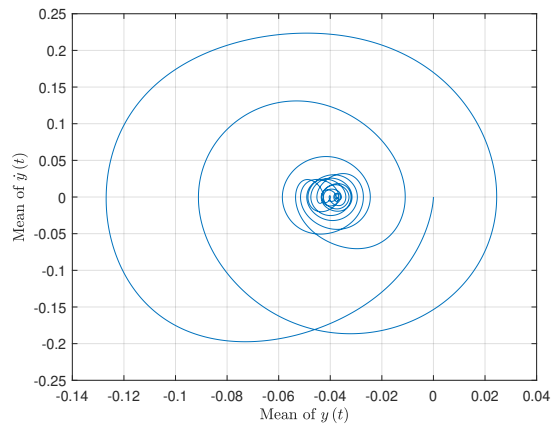
(a)



(b)



(c)



(d)

FIG. 12: Mean trajectory of $y(t)$ and $\dot{y}(t)$ for the Bouc-Wen oscillator; (a) MCS, (b) Sparse PCE-NARX ($N = 50$), (c) Sparse KNARX ($N = 10$), (d) Sparse KNARX ($N = 40$)

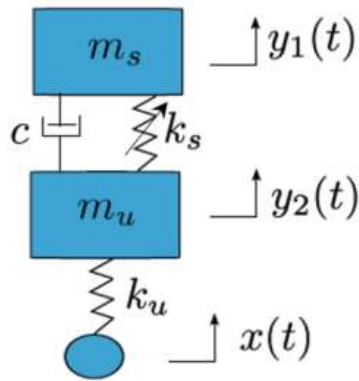
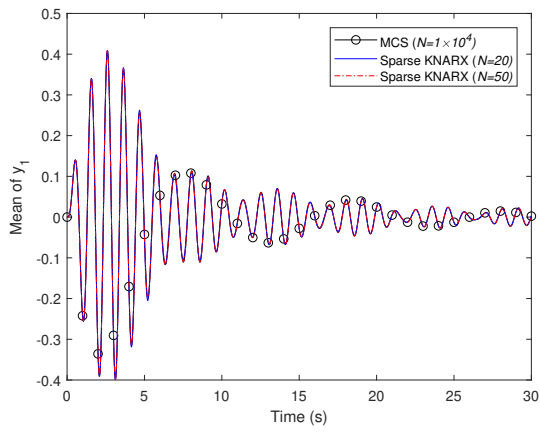
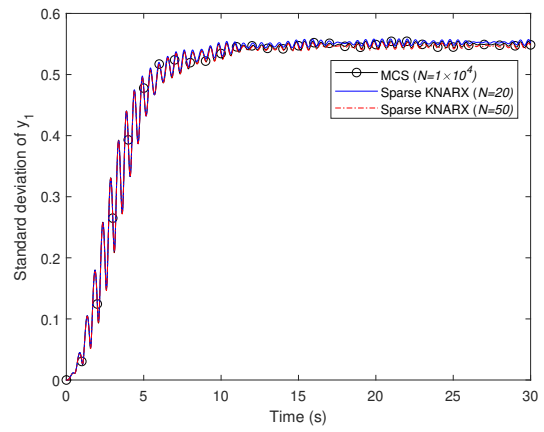


FIG. 13: A 2-DOF dynamical system

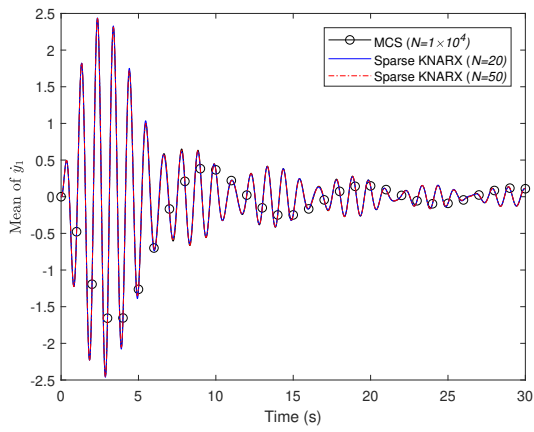


(a)

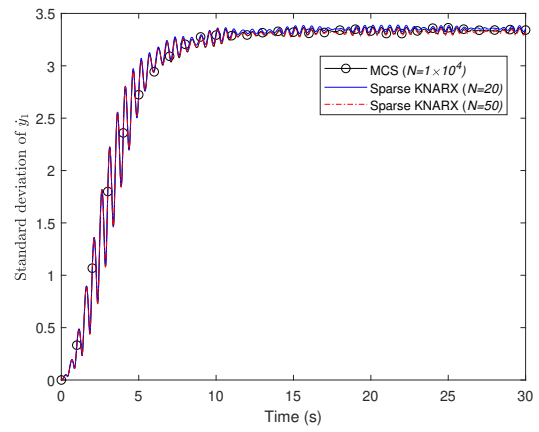


(b)

FIG. 14: Statistical response characteristics for displacement ($y_1(t)$) of the 2-DOF dynamical system; (a) Mean, (b) Standard deviation

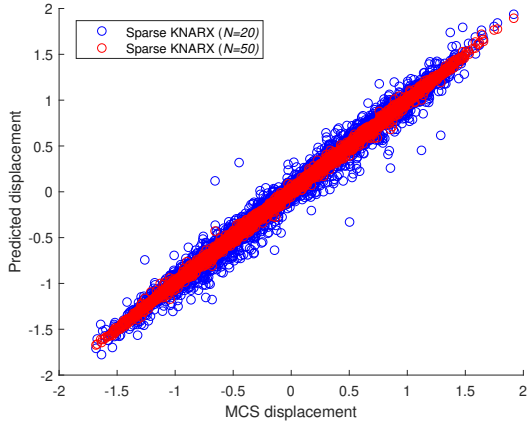


(a)

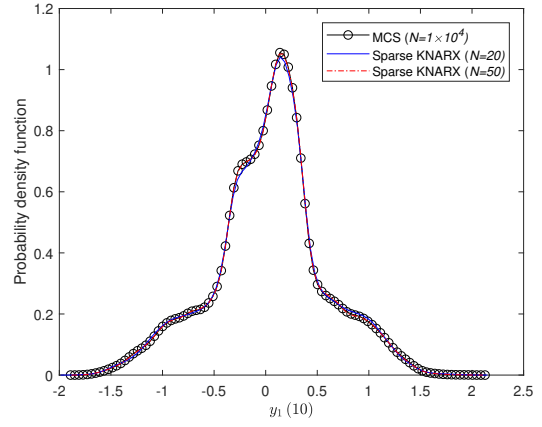


(b)

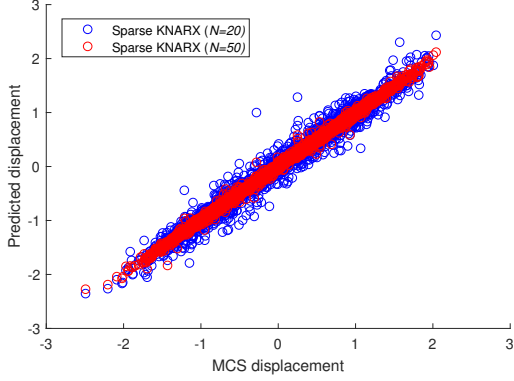
FIG. 15: Statistical response characteristics for velocity ($\dot{y}_1(t)$) of the 2-DOF dynamical system; (a) Mean, (b) Standard deviation



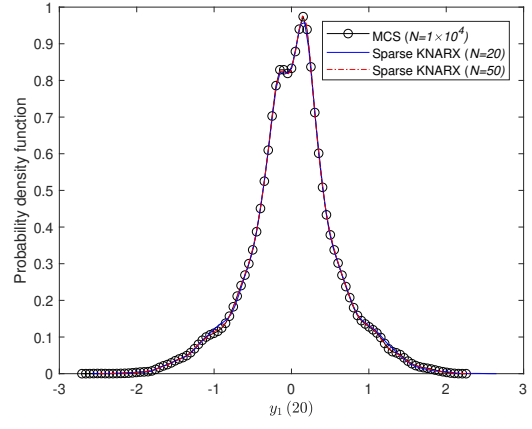
(a)



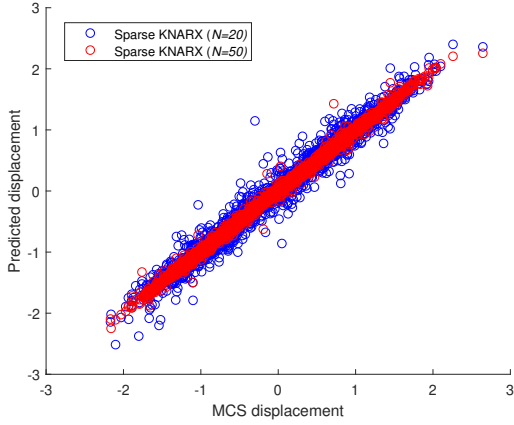
(b)



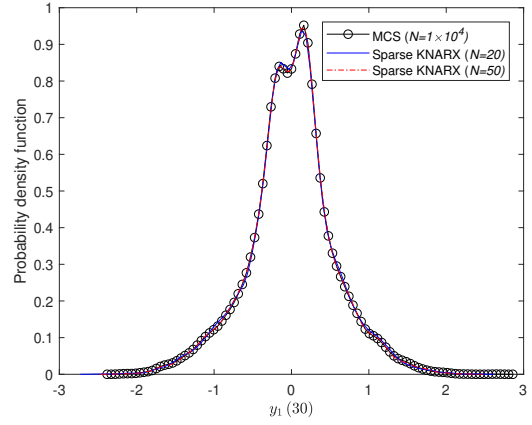
(c)



(d)



(e)



(f)

FIG. 16: Prediction of instantaneous displacement characteristics for 2-DOF dynamical system at different time instances; (a) Scatter plot at $t = 10$ s, (b) PDF at $t = 10$ s, (c) Scatter plot at $t = 20$ s, (d) PDF at $t = 20$ s, (e) Scatter plot at $t = 30$ s, (f) PDF at $t = 30$ s

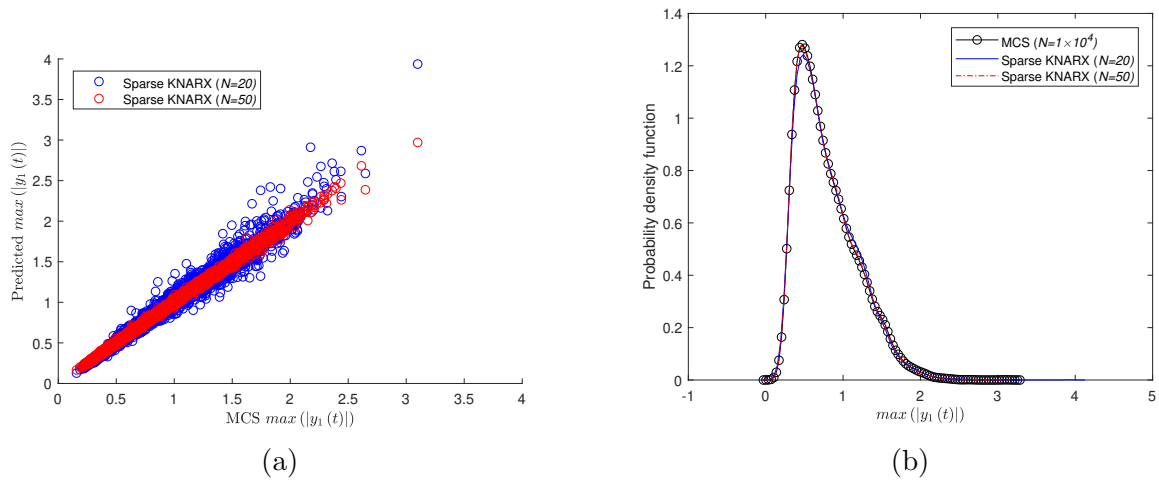


FIG. 17: Comparison of predicted $\max(|y_1(t)|)$ for the 2-DOF dynamical system; (a) Scatter plot, (b) PDF

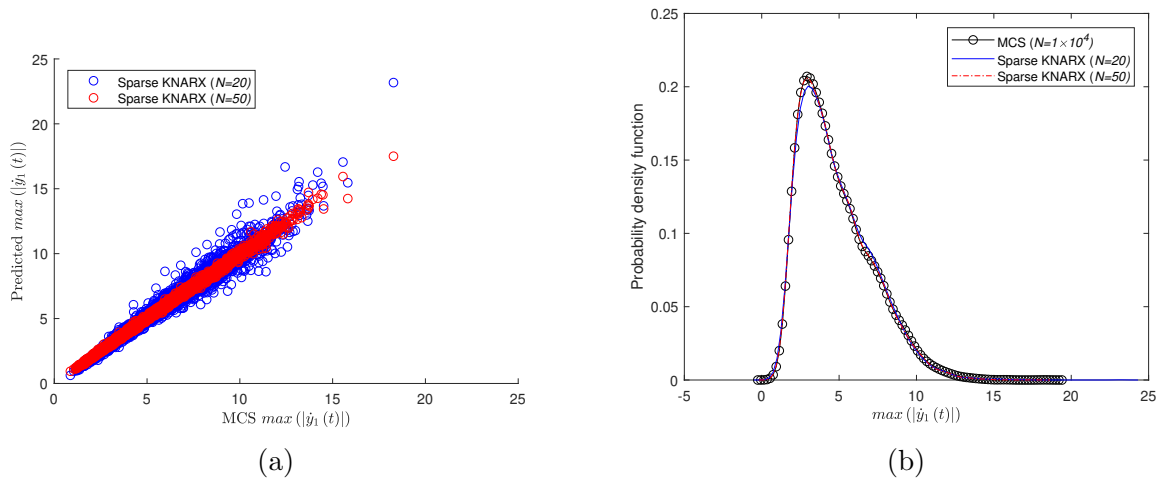
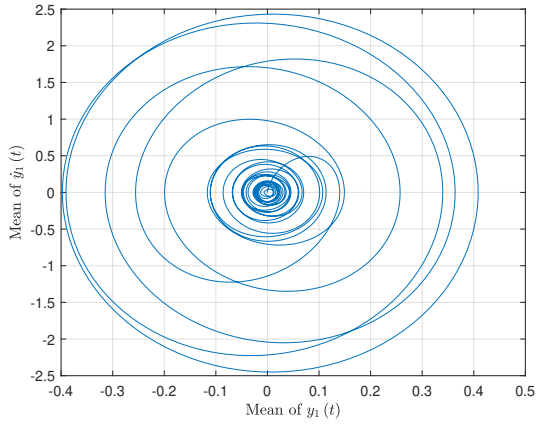
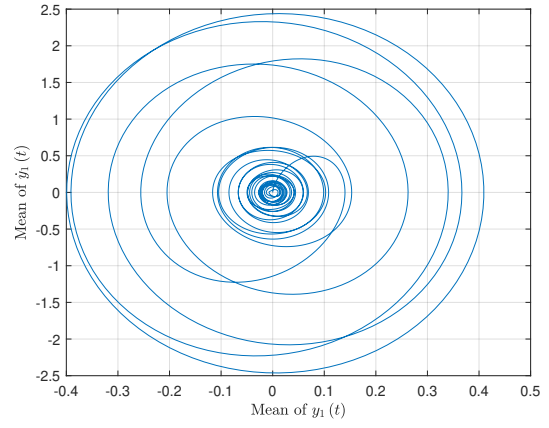


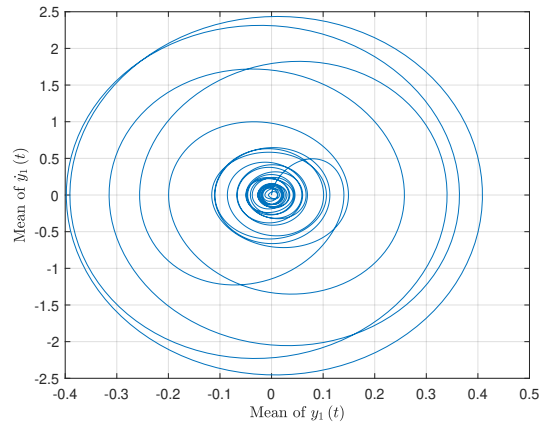
FIG. 18: Comparison of predicted $\max(|\dot{y}_1(t)|)$ for the 2-DOF dynamical system; (a) Scatter plot, (b) PDF



(a)



(b)



(c)

FIG. 19: Mean trajectory of $y_1(t)$ and $\dot{y}_1(t)$ for the 2-DOF dynamical system; (a) MCS, (b) Sparse KNARX ($N = 20$), (c) Sparse KNARX ($N = 50$)