Physics-informed multi-LSTM networks for metamodeling of nonlinear structures

Ruiyang Zhang[a], Yang Liu[b], Hao Sun[a,c,*]

[a] *Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA*
[b] *Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115, USA*
[c] *Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, USA*

Abstract

This paper introduces an innovative physics-informed deep learning framework for metamodeling of nonlinear structural systems with scarce data. The basic concept is to incorporate available, yet incomplete, physics knowledge (e.g., laws of physics, scientific principles) into deep long short-term memory (LSTM) networks, which constrains and boosts the learning within a feasible solution space. The physics constraints are embedded in the loss function to enforce the model training which can accurately capture latent system nonlinearity even with very limited available training datasets. Specifically for dynamic structures, physical laws of equation of motion, state dependency and hysteretic constitutive relationship are considered to construct the physics loss. In particular, two physics-informed multi-LSTM network architectures are proposed for structural metamodeling. The satisfactory performance of the proposed framework is successfully demonstrated through two illustrative examples (e.g., nonlinear structures subjected to ground motion excitation). It turns out that the embedded physics can alleviate overfitting issues, reduce the need of big training datasets, and improve the robustness of the trained model for more reliable prediction with extrapolation ability. As a result, the physics-informed deep learning paradigm outperforms classical non-physics-guided data-driven neural networks.
© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Numerical simulations are widely utilized for structural analysis and design of complex engineering systems. Many successful computational implementations have been achieved in last several decades for analyzing structural integrity and capacity subjected to dynamic loading. For example, finite element method (FEM) is one of the most popular simulation-based methods for structural dynamic analysis with extensive applications in civil [1,2], mechanical [3,4], and aeronautical engineering [5,6]. Despite recent advances in computational power (e.g. high-performance computing clusters or facilities), dramatically growing complexity of numerical models still demands prohibitively heavy computation for complex, large engineering problems with nonlinear hysteretic behaviors under dynamic loads. In addition, the computational cost excessively increases especially when numerous simulations

---

are required to account for the optimization [7,8] and stochastic uncertainties of external loads (e.g., Monte Carlo simulations [9–11] or incremental dynamic analysis (IDA) [12–14] of nonlinear structural systems for fragility/reliability analysis).

To address the aforementioned challenge, researchers have explored the use of metamodels to replace the original time-consuming simulation in order to reduce the computational burden. A metamodel is essentially the reduced-fidelity surrogate model of a high-fidelity model, used to describe the input–output relationship of a system. Traditionally, regression and response surface methodology (RSM) are widely used for metamodeling [15–17] which are based on the polynomial lease-square fitting. These techniques allow fast computation; however, the accuracy is often insufficient for complex systems due to their simplicity and the well-known limitations of using second-order polynomials for approximating highly nonlinear behaviors [18]. Kriging [19], radial basis functions [20], polynomial chaos expansions [21], and support vector regression [22] have also been proposed as metamodeling techniques with applications to uncertainty quantification. A review of application of these methods for metamodeling of some engineering systems can be found in [18]. For the engineering design of dynamic structures and mechanical systems, structural optimization and model updating have been extensively studied and used to simulate structural behaviors [23–25]. However, it generally requires excessive computational efforts on calibrating the model especially when the model is of high fidelity with a large number of parameters. To reduce the computational efforts, model order reduction techniques (e.g., proper orthogonal decomposition [26] and equivalent reduction expansion [27]) have been developed to establish reduced-fidelity metamodels to approximate the high-fidelity models of complex engineering systems [28–30]. Nevertheless, the majority of these methods are generally limited to linear or low-order nonlinear systems under stationary conditions, which makes applying these approaches to model highly nonlinear structures intractable.

Recently, artificial neural networks (ANNs) have been proven to be a powerful metamodeling tool and approximator [31,32], which often outperforms conventional metamodeling techniques in terms of both prediction accuracy and capability of capturing underlying nonlinear input–output relationship for complex systems [33]. Researchers have successfully implemented shallow ANNs (e.g., with only a few layers) for metamodeling structural systems under static and dynamic loading during the past decade [34–36]. However, due to the simple architecture, shallow ANNs have distinct limitations in modeling time series of complex nonlinear dynamical systems. Thanks to the state-of-the-art advances in artificial intelligence (AI), recent studies have shown that deep learning (e.g., convolutional neural network (CNN) [37] and recurrent neural network (RNN) [38,39]) are a promising approach to establish metamodels for fast prediction of time history response of dynamical systems [36,40–42] and material constitutive modeling [43,44]. For example, Zhang et al. [41] successfully developed a deep long short-term memory (LSTM) network for modeling of nonlinear seismic response of structures with large plastic deformation. However, training a reliable deep learning model requires massive (sufficient) data that must contain rich input–output relationship, which typically cannot be satisfied in most engineering problems. Particularly, the "black-box" model highly depends on the representative quality of the labeled data that it is fed in, leading to low accuracy and generalizability outside available data (training/validation datasets). Even with rich data, the trained metamodel is uninterpretable and of no physical sense. Furthermore, grand challenges arise when available data is highly incomplete, scarce and/or noisy, e.g., due to (1) "synthetic": limited number of computationally intensive simulations of the high-fidelity model for training data generation, or (2) "sensing": limited number of recordings, limited number of sensors, low signal-to-noise ratio, and incompleteness of measured state variables. A potential solution to overcome this limitation is to incorporate scientific principles (e.g., partial differential equations, boundary conditions) into deep neural networks to reduce the violation of the embedded physical laws [42,45–50]. To address the aforementioned issues, we develop physics-informed multi-LSTM networks for metamodeling of nonlinear structures and show applications to buildings under earthquake excitation. The major contribution and novelty of this work are to develop such a framework that embeds available, yet incomplete, physics information (e.g., general form of equation of motion, state dependency and hysteretic constitutive relationship) to weakly supervise the deep LSTM networks, which will constrain and boost the learning within a feasible solution space. Such metamodels possess salient features that include (1) clear interpretability with physics meaning, (2) superior generalizability with robust inference, and (3) excellent capability of dealing with less rich data.

This paper is organized as follows. Section 2 introduces two physics-informed multi-LSTM network architectures for structural metamodeling, e.g., the physics-reinforced double-LSTM (e.g., PhyLSTM^2) and the physics-reinforced triple-LSTM (e.g., PhyLSTM^3). In Section 3, the performance of PhyLSTM^2 and PhyLSTM^3 is verified through

a steel moment-resisting frame with rate-independent hysteresis. Section 4 presents another numerical example to compare PhyLSTM[2] and PhyLSTM[3] for metamodeling of a nonlinear system with rate-dependent hysteresis. Section 5 summarizes the conclusions. The data and codes used in this paper will be publicly available on GitHub at https://github.com/zhry10/PhyLSTM after the paper is published.

2. Physics-informed multi-LSTM network for metamodeling

Metamodeling of structural systems aims to develop reduced-fidelity (or reduced-order) models that effectively capture underlying nonlinear input–output behaviors. A metamodel can be trained on datasets obtained from high-fidelity simulation or actual system sensing. For better illustration, we consider a building-type structure and hypothesize the earthquake dynamics is governed by the reduced-fidelity nonlinear equation of motion (EOM):

$$\mathbf{M\ddot{u}} + \underbrace{\mathbf{C\dot{u}} + \lambda\mathbf{Ku} + (1-\lambda)\mathbf{Kr}}_{\mathbf{h}} = -\mathbf{M\Gamma}a_g \tag{1}$$

where $\mathbf{M}$ is the mass matrices; $\mathbf{C}$ is the damping matrices; $\mathbf{K}$ is the stiffness matrices; $\mathbf{u}$, $\dot{\mathbf{u}}$, and $\ddot{\mathbf{u}}$ are the relative displacement, velocity, and acceleration vector to the ground; $\mathbf{r}$ is an auxiliary non-observable hysteretic parameter (or called hysteretic displacement); $\lambda \in (0, 1]$ is the ratio of post-yield stiffness to pre-yield (elastic) stiffness; a_g represents the ground acceleration; $\mathbf{\Gamma}$ is the force distribution vector; $\mathbf{h}$ represents the total nonlinear restoring force. The EOM essentially maps the ground motion a_g to structural response $\mathbf{u}$, $\dot{\mathbf{u}}$, $\ddot{\mathbf{u}}$ and $\mathbf{r}$. By normalizing Eq. (1) based on $\mathbf{M}$, the governing equation can be rewritten in a more general form as

$$\ddot{\mathbf{u}} + \mathbf{g} = -\mathbf{\Gamma}a_g \tag{2}$$

where $\mathbf{g}(t) = \mathbf{M}^{-1}\mathbf{h}(t)$ is the mass-normalized restoring force and $\mathbf{g}(t) = \mathcal{G}(\mathbf{Z}(t))$ with $\mathcal{G}$ being an *unknown latent function*. Here, $\mathbf{Z}$ denotes the state space (SS) variable that includes the displacement $\mathbf{u}$, the velocity $\dot{\mathbf{u}}$, and the hysteretic parameter $\mathbf{r}$, namely, $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}^T = \{\mathbf{u}, \dot{\mathbf{u}}, \mathbf{r}\}^T$. Developing mathematically closed-form of a nonlinear reduced-fidelity model based on physics (e.g., a parsimonious form of $\mathbf{g}$) is intractable especially when the nonlinearity is complex, implicit, and of high order.

In nonlinear time history analysis of building-type structures under seismic excitation, a fast prediction of the state space variable $\mathbf{Z}$ is of our primary interest. An effective metamodel could establish an efficient and accurate mapping from the seismic input to nonlinear structural response, e.g., $a_g \xrightarrow{\text{metamodel}} \mathbf{Z}$. Our recent study showed that LSTM is a powerful deep learning approach for sequence-to-sequence input–output relationship modeling and thus holds strong promise to serve as a metamodel [41]. However, to train an LSTM-based metamodel, it is essential to have complete state measurement of $\mathbf{Z}$ for a given seismic input a_g (e.g., response data of $\mathbf{u}$, $\dot{\mathbf{u}}$ and $\mathbf{r}$ should be all measured). This is particularly intractable and challenging because the auxiliary hysteretic parameter $\mathbf{r}$ is typically non-observable and latent which cannot be extracted from large-scale high-fidelity model simulations or from actual system sensing. Yet, predicting such a nonlinear parameter is very important since it reflects the macroscopic nonlinearity of the system (with attributes from local nonlinearity) and relates to the internal hysteretic restoring force. These evidences illustrate that a direct application of a deep learning approach (e.g., LSTM) to establish the metamodel is inapplicable for the above mentioned problem. To address this fundamental challenge, we develop an innovative physics-informed deep learning paradigm (e.g., multi-LSTM networks constrained by physics) for metamodeling of nonlinear structural systems, which systematically maps a_g to the full state $\mathbf{Z}$ given incomplete data (e.g., $\mathbf{r}$ is not measured). In the following subsections, we introduce the basic concept and algorithm architectures of the proposed new paradigm.

2.1. LSTM network

We first introduce the fundamental algorithm architecture of deep LSTM networks for sequence-to-sequence modeling [41], which consist of multiple hidden layers (including both LSTM layers and fully connected layers) in addition to the input and output layers as shown in Fig. 1(a). The deep LSTM network maps the input sequence to the output sequence pairwise in the temporal space ($\tau = 1, 2, \ldots, t$). To implement the deep LSTM network trained with multiple datasets, both the input and output sequences must be formatted as three-dimensional arrays,
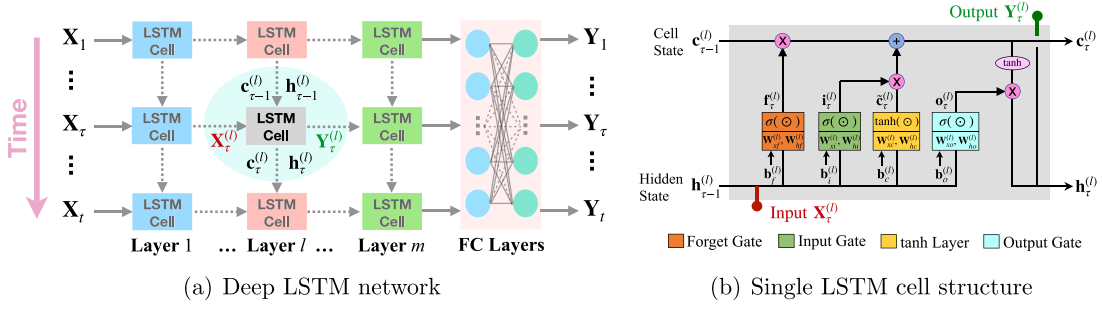
(a) Deep LSTM network   (b) Single LSTM cell structure

Fig. 1. Schematic of deep LSTM networks: (a) architecture of a deep LSTM network with m LSTM layers and multiple fully-connected layers for sequence-to-sequence modeling; (b) architecture of a typical LSTM cell of the lth layer at time t, which consists of cell input $\mathbf{X}_t^{(l)}$, cell output $\mathbf{Y}_t^{(l)}$, cell state $\mathbf{c}_t^{(l)}$, hidden state $\mathbf{h}_t^{(l)}$, and four gate variables $\{\mathbf{f}_t^{(l)}, \mathbf{i}_t^{(l)}, \tilde{\mathbf{c}}_t^{(l)}, \mathbf{o}_t^{(l)}\}$.

where the entries are the samples (e.g., independent datasets) in the first dimension, the time steps in the second dimension, and the input or output features/channels in the third dimension.

Each LSTM layer contains a suite of LSTM cells as shown in Fig. 1. Each LSTM cell, which is very similar to the neural node in classical neural networks, contains an independent set of weights and biases shared across the entire temporal space within the layer. The LSTM cell consists of four interacting units, including an internal cell, an input gate, a forget gate, and an output gate. The internal cell memorizes the cell state at the previous time step through a self-recurrent connection. The input gate controls the flow of input activation into the internal cell state. The output gate regulates the flow of output activation into the LSTM cell output. The forget gate scales the internal cell state, enabling the LSTM cell to forget or reset the cell's memory adaptively. Controlled by the input/forget/output gates in each LSTM cell, the cell state can selectively propagates valuable information along the temporal sequence to capture the long short-term time dependence in a dynamical system. Let us denote, at the time step t ($t = 1, \ldots, n$, where n is the total number of time steps) and within the lth LSTM network layer, the input state to the LSTM cell as $\mathbf{x}_t^{(l)}$, the forget gate as $\mathbf{f}_t^{(l)}$, the input gate as $\mathbf{i}_t^{(l)}$, the output gate as $\mathbf{o}_t^{(l)}$, the cell state memory as $\mathbf{c}_t^{(l)}$, and the hidden state output as $\mathbf{h}_t^{(l)}$. At the previous time step $t-1$, we denote the cell state memory as $\mathbf{c}_{t-1}^{(l)}$ and the hidden state output as $\mathbf{h}_{t-1}^{(l)}$. The relationship among these defined variables can be described by the equations as follows (also see Fig. 1(b) for schematic illustration):

$$\mathbf{f}_t^{(l)} = \sigma \left(\mathbf{W}_{xf}^{(l)} \mathbf{x}_t + \mathbf{W}_{hf}^{(l)} \mathbf{h}_{t-1} + \mathbf{b}_f^{(l)} \right) \tag{3}$$

$$\mathbf{i}_t^{(l)} = \sigma \left(\mathbf{W}_{xi}^{(l)} \mathbf{x}_t + \mathbf{W}_{hi}^{(l)} \mathbf{h}_{t-1} + \mathbf{b}_i^{(l)} \right) \tag{4}$$

$$\tilde{\mathbf{c}}_t^{(l)} = \tanh \left(\mathbf{W}_{xc}^{(l)} \mathbf{x}_t + \mathbf{W}_{hc}^{(l)} \mathbf{h}_{t-1} + \mathbf{b}_c^{(l)} \right) \tag{5}$$

$$\mathbf{o}_t^{(l)} = \sigma \left(\mathbf{W}_{xo}^{(l)} \mathbf{x}_t + \mathbf{W}_{ho}^{(l)} \mathbf{h}_{t-1} + \mathbf{b}_o^{(l)} \right) \tag{6}$$

$$\mathbf{c}_t^{(l)} = \mathbf{f}_t^{(l)} \odot \mathbf{c}_{t-1}^{(l)} + \mathbf{i}_t^{(l)} \odot \tilde{\mathbf{c}}_t^{(l)} \tag{7}$$

$$\mathbf{h}_t^{(l)} = \mathbf{o}_t^{(l)} \odot \tanh \left(\mathbf{c}_t^{(l)} \right) \tag{8}$$

where $\mathbf{W}_{\alpha\beta}^{(l)}$ (with $\alpha = \{x, h\}$ and $\beta = \{f, i, c, o\}$) denotes the weight matrices corresponding to different inputs (e.g., $\mathbf{x}_t^{(l)}$ or $\mathbf{h}_t^{(l)}$) within different gates (e.g., input gate, forget gate, tanh layer or output gate as shown in Fig. 1(b)), while $\mathbf{b}_\beta^{(l)}$ represents the corresponding bias vectors; the superscript l denotes the lth layer of the LSTM network. For example, $\mathbf{W}_{xf}^{(l)}$ and $\mathbf{W}_{hf}^{(l)}$ are the weight matrices corresponding to input vectors $\mathbf{x}_t$ or $\mathbf{h}_t$, respectively, within the forget gate. Here, $\tilde{\mathbf{c}}_t^{(l)}$ denotes a vector of intermediate candidate values created by a tanh layer shown in Fig. 1(b); σ is the logistic sigmoid function; tanh is the hyperbolic tangent function; $\odot$ denotes the Hadamard product (element-wise product). The complex connection mechanism within each LSTM cell makes the deep LSTM network powerful in sequence modeling, which the fully connected layers are beneficial for mapping the temporal feature maps to the corresponding output space.
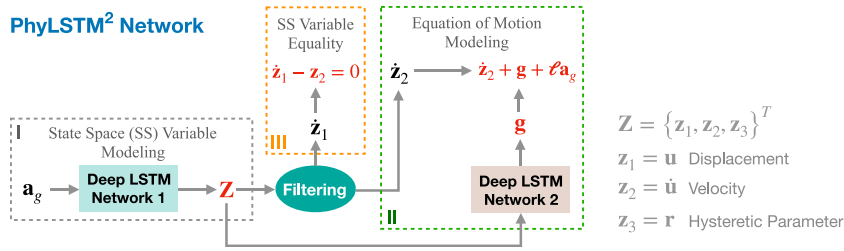
Fig. 2. The proposed PhyLSTM^2 network architecture. PhyLSTM^2 consists of two deep LSTM networks for modeling state space variables and nonlinear restoring force. The LSTM networks are interconnected through a tensor differentiator which calculates the derivative of state space variables.

2.2. PhyLSTM^2

The deep LSTM network introduced in the previous subsection is purely based on data and cannot be used to model latent variables (e.g., $\mathbf{r}$) which are not measured in data. To address this issue, we leverage available physics information (e.g., governing equations, states dependency) and encode it into the network architecture. The basic concept is to use one deep LSTM network (see Fig. 1(a) [41]) to model the sequence-to-sequence input–output relationship inter-connected, via a central finite difference filter-based numerical differentiator, with another one/two LSTM network(s) to model the physics. As a result, the multiple connected LSTM networks form a "one-network" architecture.

Firstly, we introduce the formulation and algorithm architecture of physics-informed double-LSTM network for structural metamodeling (PhyLSTM^2) as shown in Fig. 2, which consists of three components, including two deep LSTM networks and a tensor differentiator. To illustrate the concept, we first assemble the structural response to a group of state space variables, *v.i.z.*, $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}^T = \{\mathbf{u}, \dot{\mathbf{u}}, \mathbf{r}\}^T$, each of which has same number of n sample points ranging from t_1 to t_n, and use one deep LSTM network to establish nonlinear mapping from the ground motion a_g to the response $\mathbf{Z}$ (see Box I in Fig. 2), e.g., $\mathbf{Z} = \text{LSTM1}(a_g; \boldsymbol{\theta}_1)$ where $\boldsymbol{\theta}_1$ denotes the trainable weights and biases of LSTM1. With the available training data $\{\mathbf{u}_d, \dot{\mathbf{u}}_d\}^T$ (note that $\mathbf{r}$ is an immeasurable latent variable), we can formulate the "data loss function" of LSTM1, written as,

$$\mathcal{J}_d(\boldsymbol{\theta}_1) = \sum_{i=1}^{n_m} \left\| \mathbf{z}_1^{(i)}(\boldsymbol{\theta}_1) - \mathbf{u}_d^{(i)} \right\|_2^2 + \left\| \mathbf{z}_2^{(i)}(\boldsymbol{\theta}_1) - \dot{\mathbf{u}}_d^{(i)} \right\|_2^2 \tag{9}$$

where n_m is the number of measurement (data) samples. The differentiation will be realized through finite difference-based filtering, which produces derivatives of $\mathbf{Z}$, namely, $\dot{\mathbf{Z}} = \{\dot{\mathbf{z}}_1, \dot{\mathbf{z}}_2, \dot{\mathbf{z}}_3\}^T = \{\dot{\mathbf{u}}, \ddot{\mathbf{u}}, \dot{\mathbf{r}}\}^T$. By default, we have the SS variable equality condition $\dot{\mathbf{z}}_1 - \mathbf{z}_2 \to 0$ (see Box III in Fig. 2), leading to the "equality loss function":

$$\mathcal{J}_e(\boldsymbol{\theta}_1) = \sum_{i=1}^{n_c} \left\| \dot{\mathbf{z}}_1^{(i)}(\boldsymbol{\theta}_1) - \mathbf{z}_2^{(i)}(\boldsymbol{\theta}_1) \right\|_2^2 \tag{10}$$

where n_c is the number of collocation samples. A second LSTM network is then used to map the response $\mathbf{Z}$ to the mass-normalized restoring force $\mathbf{g}$ (see Box II in Fig. 2), e.g., $\mathbf{g} = \text{LSTM2}(\mathbf{Z}(\boldsymbol{\theta}_1); \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_2$ denotes the trainable weights and biases of LSTM2. Concerning the governing equation in Eq. (2), e.g., $\dot{\mathbf{z}}_2 + \mathbf{g} + \boldsymbol{\Gamma} a_g \to 0$, we obtain the "governing loss function" as

$$\mathcal{J}_g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{i=1}^{n_c} \left\| \dot{\mathbf{z}}_2^{(i)}(\boldsymbol{\theta}_1) + \mathbf{g}^{(i)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \boldsymbol{\Gamma} a_g \right\|_2^2 \tag{11}$$

A logical connection of the components in Boxes I, II and III thereby forms the proposed PhyLSTM^2 network, which can be trained by solving the following optimization problem through a standard training algorithm (e.g., gradient descent technique [51]):

$$\{\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2\} = \underset{\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}}{\arg\min} \, \mathcal{J}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \tag{12}$$
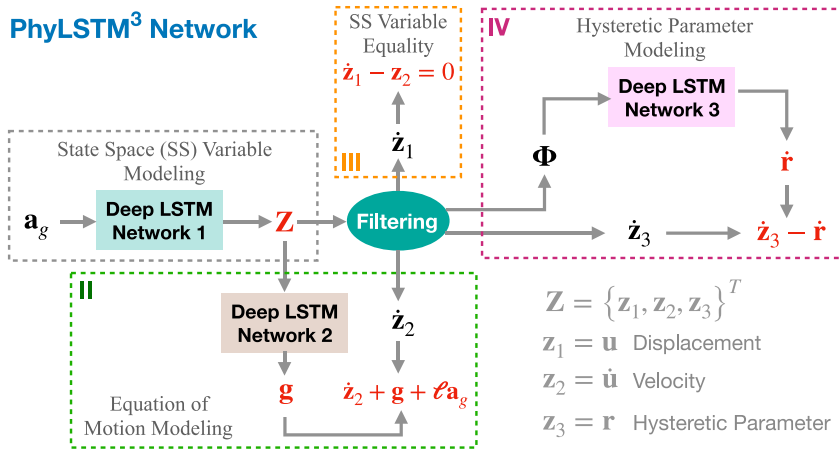
Fig. 3. The proposed PhyLSTM^3 network architecture. PhyLSTM^3 network consists of three deep LSTM networks for modeling state space variables, restoring force, and hysteretic parameter. Here, $\boldsymbol{\Phi}$ is a library of system variables, e.g., inspired from the Bouc–Wen model [52]. The LSTM networks are interconnected through a tensor differentiator which calculates the derivative of state space variables.

where $\mathcal{J}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the total loss function composed of both data loss and physics loss, given by

$$\mathcal{J}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \alpha \mathcal{J}_d(\boldsymbol{\theta}_1) + \beta \mathcal{J}_e(\boldsymbol{\theta}_1) + \gamma \mathcal{J}_g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \tag{13}$$

Here, α, β and γ are user-defined weight coefficients for convergence control (e.g., inversely proportional to the magnitude of each term; or for simplicity $\alpha = \beta = \gamma = 1$). The aim here is to optimize the network parameters $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ for both deep LSTM networks such that PhyLSTM^2 can interpret the measurement data while satisfying the physics constraints. Note that the equality condition and the governing equation should hold for any collocation samples that only consist of generic earthquake records with different magnitudes and frequency contents. This will essentially enhance the capability of LSTM1 for modeling the underlying nonlinear input–output relationship within a physically feasible solution space. Note that both LSTM networks in the proposed PhyLSTM^2 architecture used in this study have three LSTM layers and two fully-connected layers.

2.3. PhyLSTM^3

For dynamic systems with complex rate-dependent hysteretic behavior (e.g., dependent on $\dot{\mathbf{r}}$), the governing equation in Eq. (2) can be augmented by another nonlinear differential equation of the hysteretic parameter $\mathbf{r}$, expressed as,

$$\begin{cases} \ddot{\mathbf{u}} + \mathbf{g} = -\boldsymbol{\Gamma} a_g \\ \dot{\mathbf{r}} = f(\boldsymbol{\Phi}) \end{cases} \tag{14}$$

where f is a nonlinear function and $\boldsymbol{\Phi}$ is a library of system variables. For instance, the Bouc–Wen model [52] takes $\boldsymbol{\Phi} = \{\Delta \dot{\mathbf{u}}, |\Delta \dot{\mathbf{u}}|, \mathbf{r}, |\mathbf{r}|^{n-1}, |\mathbf{r}|^n\}^T$ to model the nonlinear hysteresis, where $\Delta \dot{\mathbf{u}}$ denotes the inter-story velocity vector. A simplified version of the library reads $\boldsymbol{\Phi} = \{\Delta \dot{\mathbf{u}}, \mathbf{r}\}^T$ if a priori knowledge is unknown. Therefore, we propose to augment the PhyLSTM^2 network by introducing another deep LSTM network to model the differential equation of $\mathbf{r}$ (see Box **IV** in Fig. 3), e.g., $\dot{\mathbf{r}} = \text{LSTM3}(\boldsymbol{\Phi}(\boldsymbol{\theta}_1); \boldsymbol{\theta}_3)$, where $\boldsymbol{\theta}_3$ denotes the trainable weights and biases of LSTM3. This essentially forms the PhyLSTM^3 network architecture as shown in Fig. 3, with four components, including three deep LSTM networks and a tensor differentiator. Similar to PhyLSTM^2, the other two LSTM networks are used to model the state space variables $\mathbf{Z}$ and the mass-normalized restoring force $\mathbf{g}$, respectively. The "hysteretic loss function" can then be obtained:

$$\mathcal{J}_h(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3) = \sum_{i=1}^{n_c} \left\| \dot{\mathbf{r}}^{(i)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3) - \dot{\mathbf{z}}_3^{(i)}(\boldsymbol{\theta}_1) \right\|_2^2 \tag{15}$$

(a) Plan view of the prototype building   (b) Illustration of the 3-story MRF-DBF structure
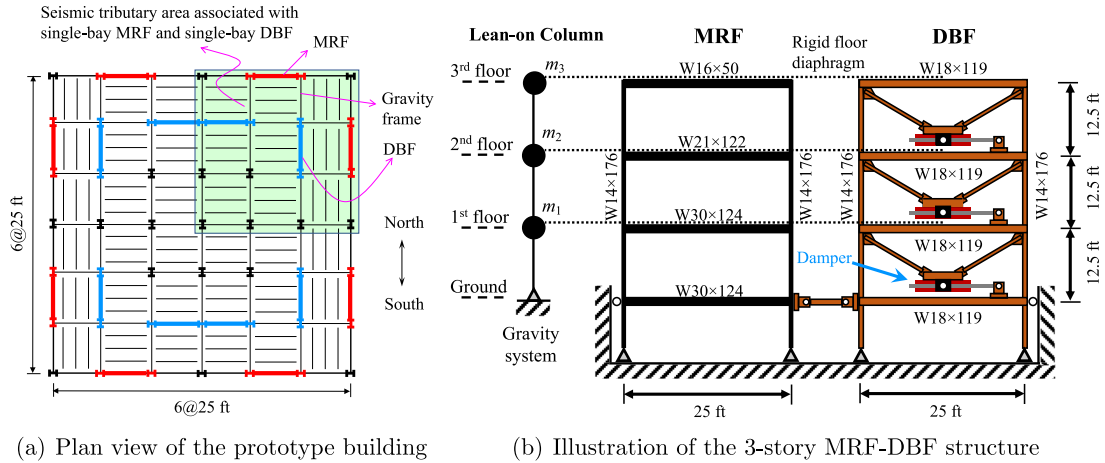
Fig. 4. The 3-story steel MRF building.

The tensor differentiator calculates the derivative of the state space outputs $\{\dot{\mathbf{z}}_1, \dot{\mathbf{z}}_2, \dot{\mathbf{z}}_3\}$ so that the physics constraints can be well constructed. Note that the PhyLSTM^3 network can be trained by optimizing the trainable parameters:

$$\{\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\theta}}_3\} = \underset{\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\}}{\arg\min} \left[\mathcal{J}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \eta \mathcal{J}_h(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3) \right] \tag{16}$$

where η is also a user-defined weight coefficient (e.g., $\eta = 1$ for simplicity). In PhyLSTM^3, the physics loss enforces the satisfactory of physics constraints including the SS variable equality ($\dot{\mathbf{z}}_1 - \mathbf{z}_2 \to 0$), equation of motion ($\ddot{\mathbf{u}} + \mathbf{g} + \boldsymbol{\Gamma} a_g \to 0$), and the hysteretic parameter equation ($\dot{\mathbf{z}}_3 - \dot{\mathbf{r}} \to 0$). Note that PhyLSTM^3, as a generalization of PhyLSTM^2, is, in theory, more powerful in metamodeling of highly nonlinear structures. This will be verified in the numerical example section.

3. Numerical validation: 3-story moment resisting frame

The proposed physics-informed multi-LSTM networks are firstly validated for metamodeling of a highly nonlinear structural system under seismic excitation. In this example, synthetic data (e.g., nonlinear time–history response) of a 3-story steel moment resisting frame (MRF) are generated by numerical simulation. We test the performance of the proposed PhyLSTM^2 and PhyLSTM^3 networks for seismic metamodeling of such a structure and compare them with the classical deep LSTM network. Both PhyLSTM^2 and PhyLSTM^3 map the ground motion a_g to the full state space response $\{\mathbf{u}, \dot{\mathbf{u}}, \mathbf{r}\}^T$ (see Figs. 2 and 3), while LSTM can only predict $\{\mathbf{u}, \dot{\mathbf{u}}\}^T$ (see Fig. 1(a)), given measured displacements and velocities. Note that, as mentioned previously, the hysteretic parameter $\mathbf{r}$ is a non-observable latent variable. The network training has been performed in the Python environment using TensorFlow [53] which is a popular and well documented open source symbolic math library for machine learning applications developed by Google Brain Team. It offers flexible data flow architecture enabling high-performance training of various types of neural networks on a variety of platforms (CPUs, GPUs, TPUs). Simulations in this paper are performed on a workstation with 28 Intel Core i9-7940X CPUs and 2 NVIDIA GTX 1080Ti GPU cards.

We test and validate the proposed methodology on a full scale 3-story office building. The prototype building adopted from Dong et al. [54] is assumed to be on a stiff site in Pomona, California. Fig. 4(a) shows the plan view of the building. The overall dimensions of the prototype structure are 45.7 m (150 ft) by 45.7 m (150 ft) in plan and 11.43 m (37.5 ft) in elevation. The structural system of the building includes a lateral resisting system, a damping system, and a gravity load system. The lateral resisting system consists of 8 identical single-bay moment resisting frames (MRFs). The damping system consists of 8 single-bay frames with nonlinear viscous dampers and associated bracing, termed as damped braced frames (DBFs). The gravity load system includes the uniformly distributed gravity frames in plan. The floor is assumed to be rigid, and thus the MRFs, DBFs, and the gravity system are assumed to deform together in each horizontal direction. Due to the symmetry of the prototype building, only one quarter of the floor plan within the seismic tributary area as shown in Fig. 4(a) is considered, forming the prototype structure
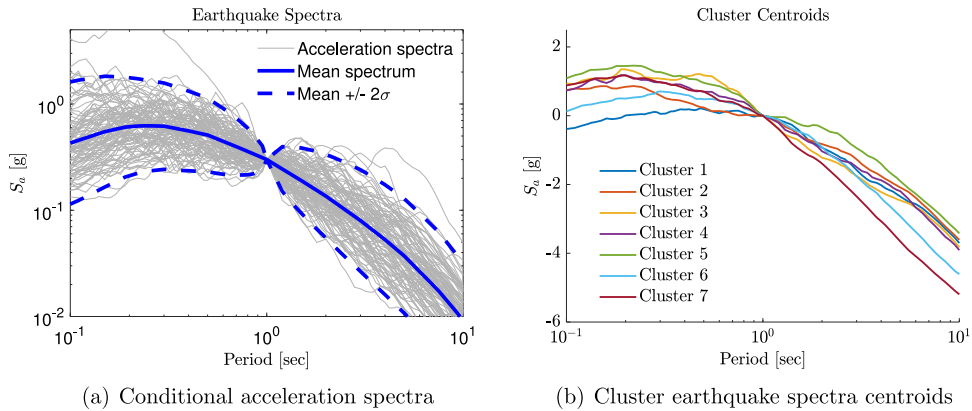
(a) Conditional acceleration spectra   (b) Cluster earthquake spectra centroids

Fig. 5. Suite of earthquake records used in this study.

investigated in this study. The 3-story prototype structure shown in Fig. 4(b) consists of a single-bay MRF, an associated single-bay DBF, and the gravity load system with associated seismic mass. The horizontal displacement at the ground level is restrained, and the columns are fixed at the base level. The design details of this structure can be found in the reference [54].

To generate the training/validation datasets, the prototype structure shown in Fig. 4(b) is modeled by the nonlinear computational platform, RT-Frame2D, developed in an *embedded* function under the MATLAB/Simulink environment [55,56]. To preserve stability for nonlinear dynamic analysis, an explicit unconditionally-stable integration scheme is adopted [57]. A concentrated plasticity model is employed for the nonlinear beam–column elements in RT-Frame2D, assuming that yielding occurs at the element ends. A bilinear moment–curvature hysteresis material model, with kinematic hardening and a post yielding ratio of 2.5%, is applied. Panel zone elements are used to model the shear deformation and the uniform bending deformation of the MRF panel zones. The element properties include the linear flexural rigidity (EI), axial rigidity (EA), shear rigidity (GA) and yield curvature κ. Mass is assigned as 4.78×10^5 kg and 5.17×10^5 kg distributed over beam elements at the first/second and third floor respectively for global mass matrix assembling. The gravity load system is represented by the lean-on column, which is modeled by elastic beam–column elements. The seismic mass is lumped and the gravity load is applied at each floor level on the lean-on column so that P-Δ effects are included in the nonlinear analysis. The lean-on column is connected to the MRF using a rigid diaphragm. The inherent damping ratios of the first two modes are assigned as 2% using Rayleigh damping. This does not account for energy dissipation from inelastic response of the MRF, which is included directly within the nonlinear elements. The natural frequencies are 1.02 Hz, 3.61 Hz, and 8.32 Hz for the first three modes. More details of the numerical modeling can be found in [55,56].

A synthetic database, consisting of nonlinear time–history responses of the structure (e.g., $\{\mathbf{u}_d, \dot{\mathbf{u}}_d\}^T$), is generated, under excitation of a suite of 97 earthquake records selected from the PEER strong motion database [58] in the area of Pomona, California (latitude, longitude = 34.0608° N, 117.7558° W) with a 10% probability of exceedance in 50 years. These ground motion records are selected using the earthquake selection and scaling tool developed by Baker and Lee [59] to match the target conditional spectrum which is conditional on a spectral value at a conditioning period of the fundamental natural frequency of the structure. The selected ground motion records are scaled such that the mean response spectrum matches the design spectrum of the prototype building. Fig. 5(a) shows the conditional acceleration spectra of all 97 selected earthquake records. To establish the database in this example, the incremental dynamic analysis (IDA) is conducted for each ground motion record with scaled intensities (e.g., amplitudes) to simulate different levels of structural damages and nonlinear responses composed of both elastic and plastic deformation, producing an ensemble of 806 total datasets for the prototype structure. Noteworthy, each dataset contains the input ground acceleration and output structural displacements, velocities, and mass-normalized restoring forces (not used in training and only used for testing the predictability of the trained metamodel). To generate the datasets for training the proposed physics-informed multi-LSTM networks, we select 7 representative ground motions based on unsupervised clustering [60] and take the corresponding 46 input–output pairs resulted from IDA as the training/validation datasets. The purpose of ground motion selection
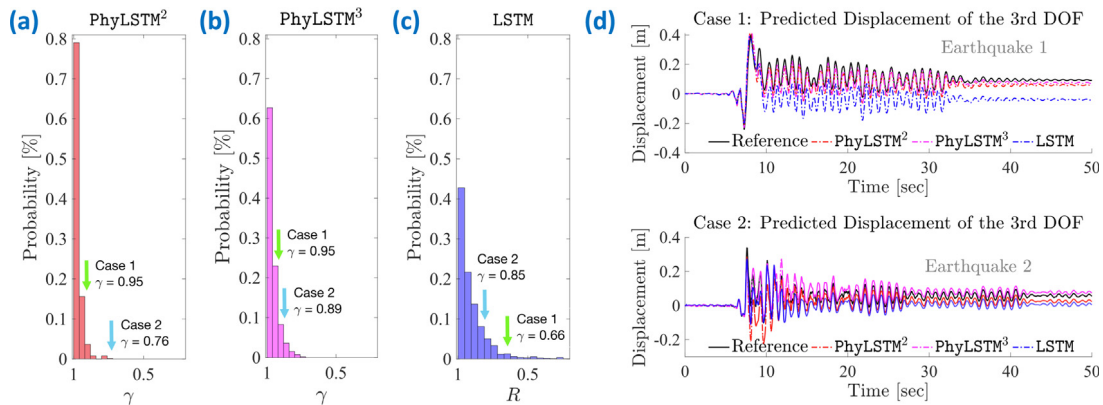
Fig. 6. Performance of PhyLSTM^2, PhyLSTM^3 and LSTM for prediction of nonlinear displacements of a 3-story MRF structure: (a)–(c) regression analyses where γ denotes the correlation coefficient, and (d) predicted displacements at the top floor under two unseen earthquake excitations randomly picked from the datasets for illustration purpose. Note that Case 1 denotes Earthquake 1 and Case 2 denotes Earthquake 2.

herein is to generate diversifying datasets for a limited, and ideally minimal, number. Since IDA is conducted for magnitude effects, the earthquake excitations are clustered based on the conditional spectral accelerations (S_a) shown in Fig. 5(a). Fig. 5(b) shows the identified 7 cluster centroids for the suite of 97 earthquakes using an unsupervised learning clustering algorithm [41,60]. Only one earthquake record that is closest to the cluster centroid is selected from each cluster for generating the training/validation datasets, while the rest are considered as the prediction dataset. A total of 760 IDA input–output pairs for the rest 90 unselected earthquakes are considered as the prediction/test datasets. Note that both training and validation datasets are considered as "known" where $\{a_g, \mathbf{u}_d, \dot{\mathbf{u}}_d\}^T$ are fully given for training/validating the PhyLSTM^2, PhyLSTM^3 and LSTM metamodels, while the prediction dataset is considered as "unknown ground truth" only for testing purpose.

All the training/validation datasets are reshaped to 3D arrays in order to be compatible with the data format for LSTM networks, e.g., the input and output sizes are [46, 10001, 1] and [46, 10001, 3]. A ratio of 0.8/0.2 is used for splitting training and validation datasets which are shuffled before each epoch to maximize feature learning from limited data. The datasets are fed into the LSTM network (see Fig. 1(a) or Box I in Figs. 2 and 3) to compute the data loss $\mathcal{J}_d$. A number of 200 earthquake samples in addition to the known earthquake records in the training/validation datasets are used as collocation samples for determining the physics losses (e.g., $\mathcal{J}_e$, $\mathcal{J}_g$, $\mathcal{J}_h$). Training the metamodels consists of two phases with different optimization algorithms. In pre-training, Adam (Adaptive Momentum Estimation) is selected as the optimizer with a learning rate of 0.001 and a decay rate of 0.0001 [51] for a total number of 1×10^4 epochs. The pre-trained model is further tuned using L-BFGS optimizer which is a quasi-Newton, gradient-based optimization algorithm [61]. The network parameters (weights and biases, e.g., $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$) are updated iteratively through back propagation such that the loss function defined in Eq. (13) or Eq. (16) is minimized. The trained network (e.g., with the minimum validation loss value) is then used as the metamodel to predict structural displacements, velocities, and restoring forces under unknown/unseen ground motions. Note that the speedup of the trained metamodel for response inference/estimation is $>10^3$ times faster compared to the conventional FEM simulation, making the model more suitable for fragility analysis.

Fig. 6 shows the performance of the three networks (e.g., PhyLSTM^2, PhyLSTM^3 and LSTM) for prediction of nonlinear displacements of the 3-story MRF structure. Fig. 6(a)–(c) summarize regression analysis of the predicted displacement time histories across all 760 testing datasets. It can be observed that the majority of the correlation coefficients (denoted as γ) for both PhyLSTM^2 and PhyLSTM^3 are greater than 0.9, indicating very accurate prediction. Clearly, the proposed physics-informed multi-LSTM approaches are much more robust and produce more accurate prediction compared to classical LSTM without embedded physics. The worst scenario for LSTM corresponds the correlation coefficient $\gamma = 0.25$ which is much lower compared to PhyLSTM^2 with $\gamma = 0.74$ and PhyLSTM^3 with $\gamma = 0.76$. Fig. 6(d) shows predicted displacement time histories at the top floor under two example earthquakes, with the corresponding correlation coefficients marked in the regression plot for PhyLSTM^2 ($\gamma = 0.95$ and 0.76), PhyLSTM^3 ($\gamma = 0.95$ and 0.89), and LSTM ($\gamma = 0.66$ and 0.85). The PhyLSTM^2 prediction, with
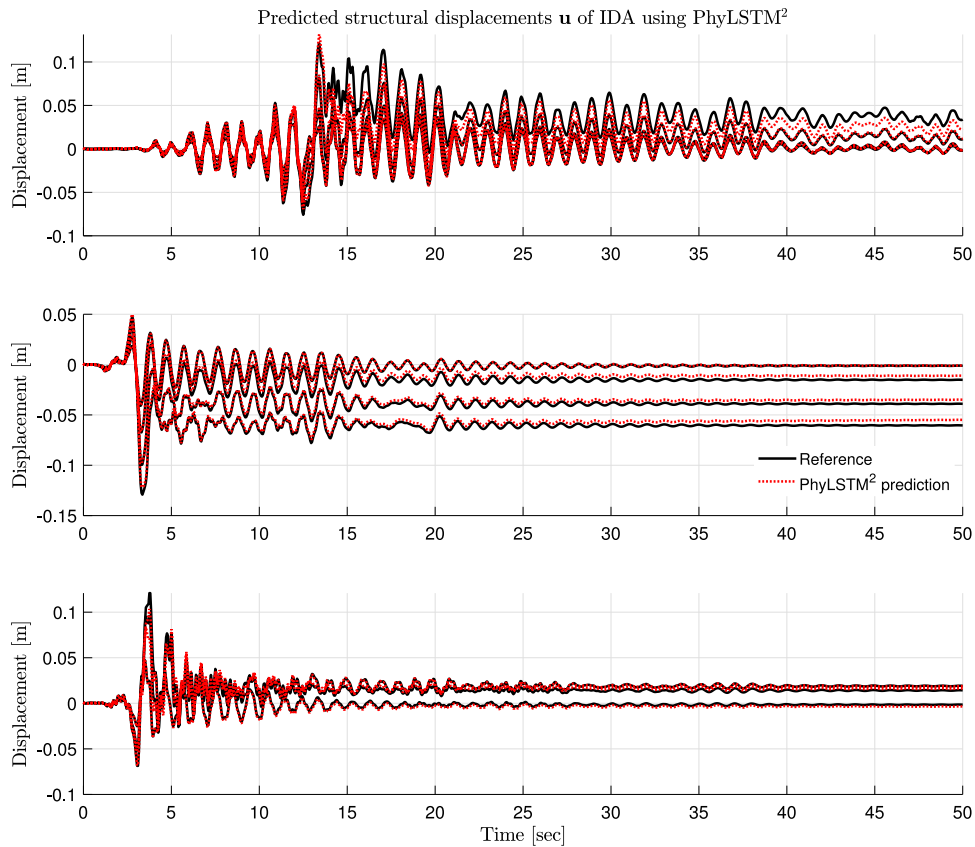
Fig. 7. PhyLSTM^2-predicted IDA displacements at the 3rd floor for three example unseen earthquakes with varying intensities (e.g., magnitudes).

$\gamma = 0.95$, matches the reference well in magnitudes, phases, as well as residual drifts that reflect plastic deformation as shown in Fig. 6(d). Note that the prediction displacement time histories for $\gamma > 0.95$ are not shown since the predicted displacements have an excellent match with the ground truth. Even for the case with less satisfactory prediction (e.g., $\gamma = 0.76$), the PhyLSTM^2 approach is still able to reasonably well predict the displacement time histories using very limited training data. Similar prediction performance is observed for the PhyLSTM^3 metamodel. The predicted structural displacements using LSTM are also presented in Fig. 6(d). Although the predicted peak magnitudes and phases of displacements relatively well match the reference, the residual drifts (e.g., plastic deformation) cannot be accurately predicted by LSTM. This indicates that it is intractable to learn the complex hysteretic behavior purely from data in training especially when available datasets are limited. In summary, both PhyLSTM^2 and PhyLSTM^3 outperform LSTM, while PhyLSTM^2 produces slightly better prediction compared with PhyLSTM^3. Note that the nonlinear hysteresis of this structure is rate-independent (e.g., independent on $\dot{r}$) such that PhyLSTM^2 is more capable of modeling the latent nonlinearity given its parsimonious architecture compared with PhyLSTM^3. The favorable performance of PhyLSTM^2, for example, is further illustrated in Fig. 7, which shows the predicted IDA displacements in comparison with the ground truth under excitation of the same earthquake but with varying intensities. It is seen that, although the input earthquakes are scaled linearly, the trained metamodel is capable of capturing and distinguishing the nonlinear structural responses, indicating that the trained metamodel has extrapolation ability and learns the underlying dynamics/physics instead of just regression of the training data.

Fig. 8 presents the result of predicted velocities by PhyLSTM^2, PhyLSTM^3 and LSTM, respectively. It turns out the velocities are much easier to learn and can be accurately predicted even using LSTM, because velocity time histories have less complex behaviors such as residuals. Nevertheless, PhyLSTM^2 and PhyLSTM^3 still provide better prediction accuracy compared with the data-driven LSTM. Another advantage of physics-informed multi-LSTM networks is that the latent state (e.g., the hysteretic parameter $\mathbf{r}$ resulting from LSTM1 or the nonlinear
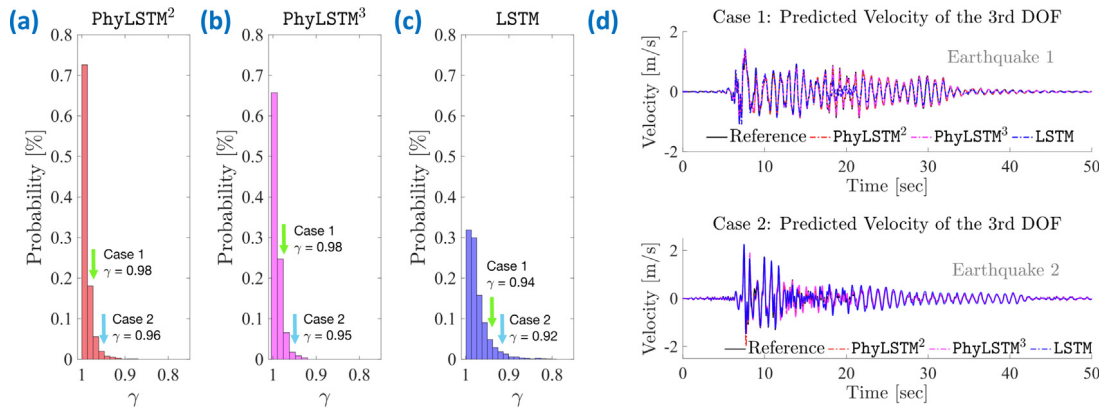
Fig. 8. Performance of PhyLSTM^2, PhyLSTM^3 and LSTM for prediction of velocities of a 3-story MRF structure: (a)–(c) regression analyses where γ denotes the correlation coefficient, and (d) predicted velocities at the top floor under two unseen earthquake excitations randomly picked from the datasets for illustration purpose. Note that Case 1 denotes Earthquake 1 and Case 2 represents Earthquake 2.
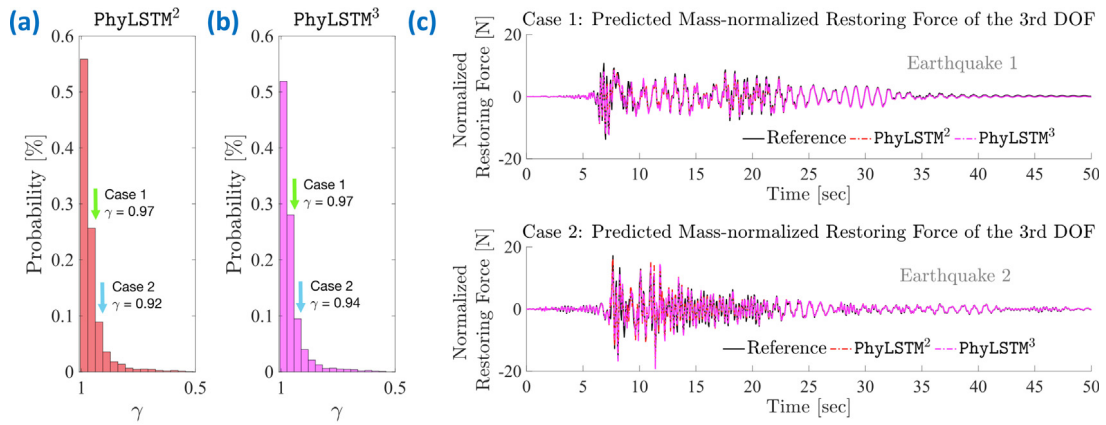

Fig. 9. Performance of PhyLSTM^2 and PhyLSTM^3 for prediction of the mass-normalized restoring forces **g**: (a)–(b) regression analyses where γ denotes the correlation coefficient, and (d) predicted mass-normalized restoring forces at the top floor under two unseen earthquake excitations randomly picked from the datasets for illustration purpose. Note that without the measurements of **g**, the physics-informed multi-LSTM approaches are able to predict the latent nonlinear restoring force while LSTM fails to predict it without measurement in training.

restoring force **g** from LSTM2, as shown in Figs. 2 and 3) can be predicted even though no measurement of the state is available for training. This can be realized by the physical knowledge encoded in the network. For example, Fig. 9 shows the predicted mass-normalized restoring force using PhyLSTM^2 and PhyLSTM^3 given no measurements of which in training. This is a mission impossible by classical data-driven LSTM networks. Note that the time history examples shown in Figs. 6, 8, and 9 are subjected to the same set of ground motion excitations for better comparison. This example clearly illustrates the accuracy and robustness of the proposed physics-informed multi-LSTM metamodels compared with the classical data-driven LSTM. From the aforementioned results, we can also conclude that, with physics constraints, the proposed physics-informed multi-LSTM metamodels are capable of learning and recognizing hidden patterns obeying given governing laws from very limited data.

4. Numerical validation: Bouc–Wen hysteresis model

We herein consider a nonlinear system with rate-dependent hysteresis (e.g., dependent on $\dot{\mathbf{r}}$) as described in Eq. (14) and compare the capability of PhyLSTM^2 and PhyLSTM^3 for complex hysteresis modeling. The Bouc–Wen model [52,62] is adopted for showcase, in which, for the ith degree-of-freedom (DOF), the rate-dependent
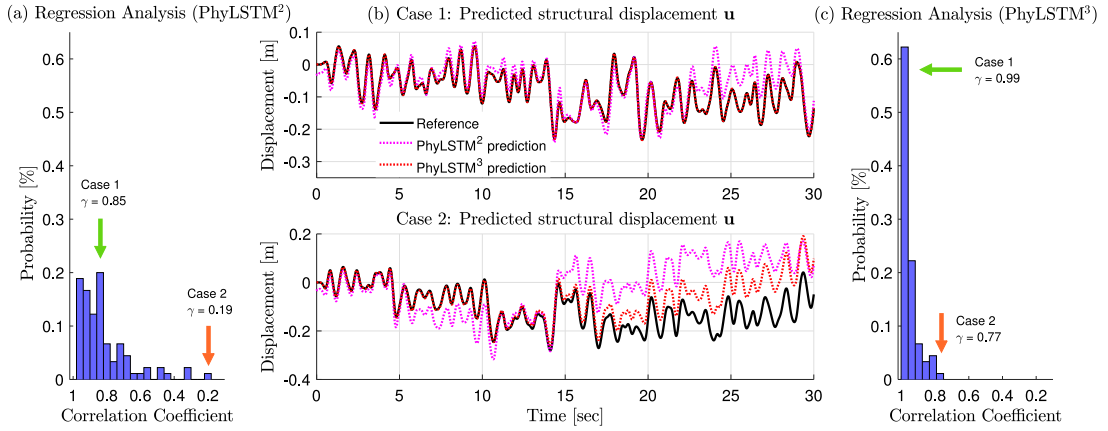
Fig. 10. Prediction performance of displacement u using PhyLSTM^2 and PhyLSTM^3: (a) regression analysis for PhyLSTM^2; (b) two examples of predicted displacement time histories; and (c) regression analysis for PhyLSTM^3.

hysteresis is expressed as [63]:

$$\dot{r}_i = \Delta\dot{u}_i - \alpha_i|\Delta\dot{u}_i||r_i|^{n_i-1}r_i - \beta_i\Delta\dot{u}_i|r_i|^{n_i} \tag{17}$$

where $\Delta\dot{u}_i$ is the relative velocity between $(i-1)$th and ith DOF, denoted as $\Delta\dot{u}_i = \dot{u}_i - \dot{u}_{i-1}$ for $i \geq 2$ and $\Delta\dot{u}_i = \dot{u}_1$ if $i = 1$; α_i, β_i and n_i are the nonlinear parameters of the Bouc–Wen model. In this example, a single DOF (SDOF) Bouc–Wen model is used with the following parameters: $m = 500$ kg, $c = 0.35$ kNs/m, $k = 25$ kN/m, $\alpha = 2$, $\beta = 2$ and $n = 3$. The natural frequency of the system is 1.13 Hz. The parameter λ in Eq. (1) is assumed as 0.5. A synthetic database, consisting of 100 samples (e.g., independent seismic sequences), was generated by numerical simulation for the SDOF nonlinear system excited by random band-limited white noise (BLWN) ground motions with different magnitudes. Each simulation was executed up to 30 s with a sampling frequency of 50 Hz resulting in 1501 data points for each record. All datasets are formatted to required 3D arrays for PhyLSTM^2 and PhyLSTM^3. Only 10 datasets with BLWN input and corresponding structural displacement and velocity responses are randomly selected and considered as "known" datasets for training/validation (with a split ratio of 0.8/0.2), while the rest are considered as "unknown" datasets to test the prediction performance of trained metamodels. 50 additional collocation samples (e.g., BLWN input records only) are used to guide the model training with physics constraints.

The network configuration for this example is given as follows: each LSTM network in PhyLSTM^2 and PhyLSTM^3 has two LSTM layers and one FC layer, which turns out to be sufficient to train an accurate model. The PhyLSTM^2 and PhyLSTM^3 models are first pre-trained using the Adam optimizer [51] with a learning rate of 0.001 for 5000 epochs and with a learning rate of 0.0001 for another 5000 epochs. Then the L-BFGS optimizer [61] is used to enhance the pre-trained model until the default convergence criteria is triggered. We take $\Phi = \{\Delta\dot{u}, r\}^T$ as the simplified library of basis functions for hysteresis modeling.

Fig. 10 summarizes the performance of both PhyLSTM^2 and PhyLSTM^3 for prediction of nonlinear displacement time histories of the SDOF Bouc–Wen model under unseen BLWN excitations. Comparing the regression analysis shown in Fig. 10(a) and (c) for PhyLSTM^2 and PhyLSTM^3 respectively, it can be clearly seen that PhyLSTM^3 ensures a larger probability of correlation coefficients close to one, demonstrating a better prediction performance. Besides, the accuracy for the worst scenario using PhyLSTM^3 ($\gamma = 0.77$) is much higher in contrast to PhyLSTM^2 ($\gamma = 0.19$), indicating that PhyLSTM^3 is a more robust and stable approach for nonlinear rate-dependent hysteresis modeling. Fig. 10(b) shows two examples of predicted displacement time histories using PhyLSTM^2 and PhyLSTM^3 with the corresponding correlation coefficients of $\gamma = 0.85$ and $\gamma = 0.99$ for Case 1 and $\gamma = 0.19$ and $\gamma = 0.77$ for Case 2. The mass-normalized restoring force g can be perfectly predicted (with $\gamma \approx 1$) using the proposed PhyLSTM^3 as shown in Fig. 11 even though no measurement is available in training. The hysteresis of this nonlinear system can also be well estimated by the trained PhyLSTM^3 metamodel as depicted in Fig. 12 which presents two examples of u-g curves (e.g., predicted displacement v.s. predicted restoring force). To further test the robustness of
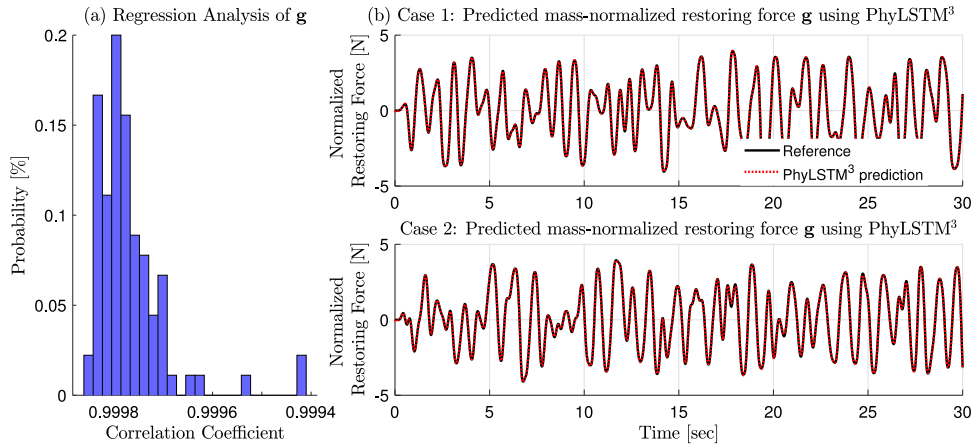
Fig. 11. PhyLSTM³-predicted mass-normalized restoring force: (a) regression analysis; and (b) predicted time histories.
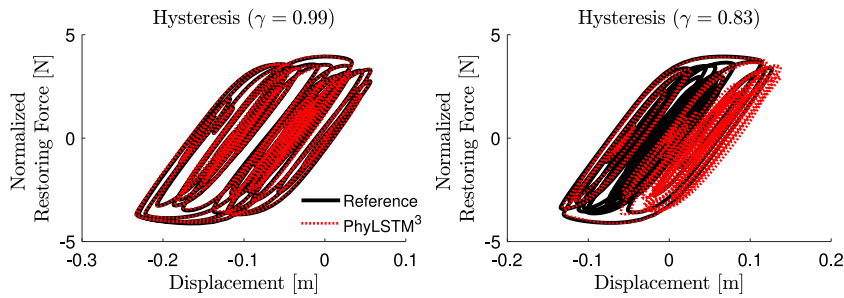

Fig. 12. Examples of predicted hysteresis curves of nonlinear restoring force versus displacement using the proposed PhyLSTM³.
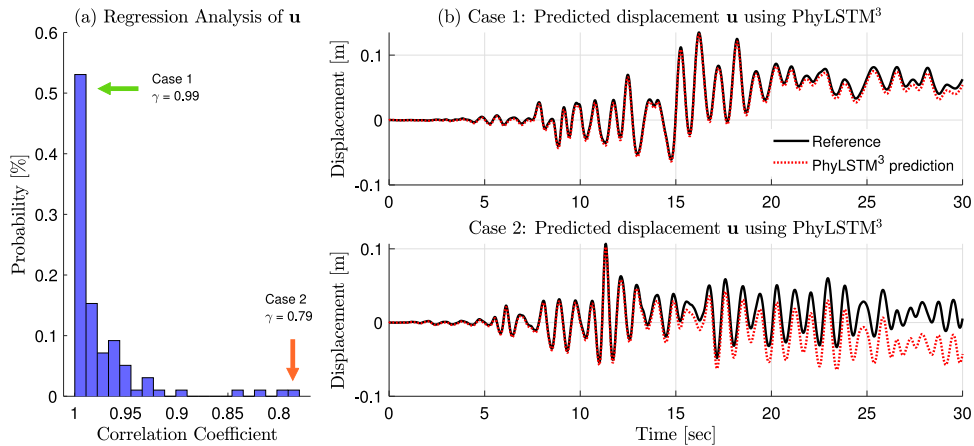

Fig. 13. Predicted displacements of the SDOF Bouc–Wen model under unseen earthquake records using the PhyLSTM³ metamodel trained by BLWN excitation data.

the proposed approach, the PhyLSTM³ metamodel trained by BLWN excitation data is employed to predict structural responses subjected to the suite of 97 ground motions used in the previous example. Fig. 13(a) summarizes the overall prediction performance over all 97 records using PhyLSTM³, as a result, with the majority (e.g., > 95%) of correlation coefficients greater than 0.9. Fig. 13(b) shows two example time histories of predicted structural

displacement with $\gamma = 0.99$ and $\gamma = 0.79$ (e.g., the worst scenario). In general, this clearly demonstrates the robustness of PhyLSTM[3] in metamodeling of nonlinear hysteretic system.

5. Conclusions

This paper presents a novel physics-informed deep learning paradigm for metamodeling of nonlinear structural systems with showcase of predicting nonlinear structural seismic responses. In particular, two architectures of physics-informed multi-LSTM networks (e.g., PhyLSTM[2] and PhyLSTM[3]) are presented for representation learning of sequence-to-sequence features from limited data enhanced by available physics. The laws of physics are taken as extra constraints, encoded in the network architecture, and embedded in the overall loss function to enforce the model training in a feasible solution space. In such way, the trained metamodel can accurately capture structural dynamics even with very scarce training/validation data. Another distinction of the proposed networks is that they can accurately model non-observable, latent nonlinear state variables (e.g., hysteretic parameter or nonlinear restoring force), where measurement is unavailable. The performance of PhyLSTM[2] and PhyLSTM[3] is demonstrated through two numerical examples (e.g., a 3-story MRF structure and a SDOF Bouc–Wen model). Numerical results illustrate that the physics-informed multi-LSTM models outperform the classical non-physics-guided data-driven LSTM network in terms of robustness and prediction accuracy. For nonlinear systems with rate-independent hysteresis, PhyLSTM[2] is more capable of modeling the latent nonlinearity given its parsimonious architecture compared with PhyLSTM[3]; however, for the system with rate-dependent hysteresis, PhyLSTM[3] is more powerful and produces much more accurate prediction thanks to its explicit modeling of the rate-dependent hysteresis using a differential equation. In general, the proposed PhyLSTM[2] and PhyLSTM[3] metamodels possess salient features that include (1) clear interpretability with physics meaning, (2) superior generalizability/extrapolability with robust inference, and (3) excellent capability of dealing with less rich data. It turns out that the embedded physics can provide constraints to the network outputs, alleviate overfitting issues, reduce the need of big training datasets, and thus improve the robustness of the trained model for more reliable prediction. Though the proposed metamodeling approaches are presented in the context of structural seismic response prediction, they can be easily extended to develop metamodels for other types of structural systems, where the physics-informed multi-LSTM network architectures should be adapted by changing the physics part as needed. In addition, we will test the scalability of the proposed methodology on developing metamodels in a higher-dimensional parameter space (e.g., the number of state variables is larger than 100) in the future.

Declaration of competing interest

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

Acknowledgment

References

[1] K. Phuvoravan, E.D. Sotelino, Nonlinear finite element for reinforced concrete slabs, J. Struct. Eng. 131 (4) (2005) 643–649.
[2] L. Kwasniewski, H. Li, J. Wekezer, J. Malachowski, Finite element analysis of vehicle–bridge interaction, Finite Elem. Anal. Des. 42 (11) (2006) 950–959.
[3] R. Smit, W. Brekelmans, H. Meijer, Prediction of the mechanical behavior of nonlinear heterogeneous systems by multi-level finite element modeling, Comput. Methods Appl. Mech. Engrg. 155 (1–2) (1998) 181–192.
[4] F. Migliavacca, L. Petrini, M. Colombo, F. Auricchio, R. Pietrabissa, Mechanical behavior of coronary stents investigated through the finite element method, J. Biomech. 35 (6) (2002) 803–811.
[5] C. Santiuste, X. Soldani, M.H. Miguélez, Machining FEM model of long fiber composites for aeronautical components, Compos. Struct. 92 (3) (2010) 691–698.
[6] Z. Kapidžić, L. Nilsson, H. Ansell, Finite element modeling of mechanically fastened composite-aluminum joints in aircraft structures, Compos. Struct. 109 (2014) 198–210.
[7] C.L. Bottasso, F. Campagnolo, A. Croce, S. Dilli, F. Gualdoni, M.B. Nielsen, Structural optimization of wind turbine rotor blades by multilevel sectional/multibody/3D-FEM analysis, Multibody Syst. Dyn. 32 (1) (2014) 87–116.

[8] R. Zhang, B.M. Phillips, Cyber-physical approach to the optimization of semiactive structural control under multiple earthquake ground motions, Comput.-Aided Civ. Infrastruct. Eng. 34 (5) (2019) 402–414.

[9] M. Papadrakakis, N.D. Lagaros, Reliability-based structural optimization using neural networks and Monte Carlo simulation, Comput. Methods Appl. Mech. Engrg. 191 (32) (2002) 3491–3507.

[10] J. Zhang, C. Wan, T. Sato, Advanced Markov chain Monte Carlo approach for finite element calibration under uncertainty, Comput.-Aided Civ. Infrastruct. Eng. 28 (7) (2013) 522–530.

[11] M. Pisaroni, F. Nobile, P. Leyland, A continuation multi level Monte Carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics, Comput. Methods Appl. Mech. Engrg. 326 (2017) 20–50.

[12] D. Vamvatsikos, C.A. Cornell, Incremental dynamic analysis, Earthq. Eng. Struct. Dyn. 31 (3) (2002) 491–514.

[13] D. Vamvatsikos, M. Fragiadakis, Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty, Earthq. Eng. Struct. Dyn. 39 (2) (2010) 141–163.

[14] L. Tirca, O. Serban, L. Lin, M. Wang, N. Lin, Improving the seismic resilience of existing braced-frame office buildings, J. Struct. Eng. 142 (8) (2015) C4015003.

[15] S. Durieux, H. Pierreval, Regression metamodeling for the design of automated manufacturing system composed of parallel machines sharing a material handling resource, Int. J. Prod. Econ. 89 (1) (2004) 21–30.

[16] G.E. Box, N.R. Draper, Empirical Model-Building and Response Surfaces, John Wiley & Sons, 1987.

[17] A.I. Khuri, S. Mukhopadhyay, Response surface methodology, Wiley Interdiscip. Rev. Comput. Stat. 2 (2) (2010) 128–149.

[18] T.W. Simpson, J. Poplinski, P.N. Koch, J.K. Allen, Metamodels for computer-based engineering design: survey and recommendations, Eng. Comput. 17 (2) (2001) 129–150.

[19] J.P. Kleijnen, Kriging metamodeling in simulation: A review, European J. Oper. Res. 192 (3) (2009) 707–716.

[20] M.F. Hussain, R.R. Barton, S.B. Joshi, Metamodeling: radial basis functions, versus polynomials, European J. Oper. Res. 138 (1) (2002) 142–154.

[21] M.D. Spiridonakos, E.N. Chatzi, Metamodeling of dynamic nonlinear structural systems through polynomial chaos NARX models, Comput. Struct. 157 (2015) 99–113.

[22] S.M. Clarke, J.H. Griebsch, T.W. Simpson, Analysis of support vector regression for approximation of complex engineering analyses, J. Mech. Des. 127 (6) (2005) 1077–1087.

[23] J.M. Brownjohn, P.-Q. Xia, Dynamic assessment of curved cable-stayed bridge by model updating, J. Struct. Eng. 126 (2) (2000) 252–260.

[24] B. Moaveni, J.P. Conte, F.M. Hemez, Uncertainty and sensitivity analysis of damage identification results obtained using finite element model updating, Comput.-Aided Civ. Infrastruct. Eng. 24 (5) (2009) 320–334.

[25] H. Sun, R. Betti, A hybrid optimization algorithm with Bayesian inference for probabilistic model updating, Comput.-Aided Civ. Infrastruct. Eng. 30 (8) (2015) 602–619.

[26] Q. Du, M. Gunzburger, Model reduction by proper orthogonal decomposition coupled with centroidal voronoi tessellation, in: Proc. Fluids Engineering Division Summer Meeting, FEDSM2002-31051, ASME, 2002.

[27] M. Papadopoulos, E. Garcia, Improvement in model reduction schemes using the system equivalent reduction expansion process, AIAA J. 34 (10) (1996) 2217–2219.

[28] Z. Bai, Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, Appl. Numer. Math. 43 (1–2) (2002) 9–44.

[29] M. Guo, J.S. Hesthaven, Data-driven reduced order modeling for time-dependent problems, Comput. Methods Appl. Mech. Engrg. 345 (2019) 75–99.

[30] Z. Zhang, M. Guo, J.S. Hesthaven, Model order reduction for large-scale structures with local nonlinearities, Comput. Methods Appl. Mech. Engrg. 353 (2019) 491–515.

[31] S. Chen, S. Billings, Neural networks for nonlinear dynamic system modelling and identification, Int. J. Control 56 (2) (1992) 319–346.

[32] C. Tianping, C. Hong, Approximations of continuous functions by neural networks with application to dynamic system, IEEE Trans. Neural Netw. 4 (6) (1993) 910–918.

[33] D.J. Fonseca, D.O. Navaresse, G.P. Moynihan, Simulation metamodeling through artificial neural networks, Eng. Appl. Artif. Intell. 16 (3) (2003) 177–183.

[34] W. Ying, W. Chong, L. Hui, Z. Renda, Artificial neural network prediction for seismic response of bridge structure, in: 2009 International Conference on Artificial Intelligence and Computational Intelligence, Vol. 2, IEEE, 2009, pp. 503–506.

[35] N.H. Christiansen, J.B. Høgsberg, O. Winther, Artificial neural networks for nonlinear dynamic response simulation in mechanical systems, in: NSCM-24, 2011.

[36] N.D. Lagaros, M. Papadrakakis, Neural network based prediction schemes of the non-linear seismic response of 3D buildings, Adv. Eng. Softw. 44 (1) (2012) 92–115.

[37] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, Handb. Brain Theory Neural Netw. 3361 (10) (1995) 1995.

[38] L. Medsker, L.C. Jain, Recurrent Neural Networks: Design and Applications, CRC press, 1999.

[39] D.P. Mandic, J. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, John Wiley & Sons, Inc., 2001.

[40] R.-T. Wu, M.R. Jahanshahi, Deep convolutional neural network for structural dynamic response estimation and system identification, J. Eng. Mech. 145 (1) (2018) 04018125.

[41] R. Zhang, Z. Chen, S. Chen, J. Zheng, O. Büyüköztürk, H. Sun, Deep long short-term memory networks for nonlinear structural seismic response prediction, Comput. Struct. 220 (2019) 55–68.

[42] R. Zhang, Y. Liu, H. Sun, Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling, Eng. Struct. 215 (2020) 110704.

[43] K. Wang, W. Sun, A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning, Comput. Methods Appl. Mech. Engrg. 334 (2018) 337–380.

[44] K. Wang, W. Sun, Meta-modeling game for deriving theory-consistent, microstructure-based traction–separation laws via deep reinforcement learning, Comput. Methods Appl. Mech. Engrg. 346 (2019) 216–241.

[45] M. Raissi, Deep hidden physics models: Deep learning of nonlinear partial differential equations, J. Mach. Learn. Res. 19 (1) (2018) 932–955.

[46] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.

[47] L. Sun, H. Gao, S. Pan, J.-X. Wang, Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data, Comput. Methods Appl. Mech. Engrg. 361 (2020) 112732.

[48] X. Yang, S. Zafar, J.-X. Wang, H. Xiao, Predictive large-eddy-simulation wall modeling via physics-informed neural networks, Phys. Rev. Fluids 4 (3) (2019) 034602.

[49] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, P. Perdikaris, Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, J. Comput. Phys. 394 (2019) 56–81.

[50] G. Kissas, Y. Yang, E. Hwuang, W.R. Witschey, J.A. Detre, P. Perdikaris, Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks, Comput. Methods Appl. Mech. Engrg. 358 (2020) 112623.

[51] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations (ICLR), 2014.

[52] Y.-K. Wen, Method for random vibration of hysteretic systems, J. Eng. Mech. Div. 102 (2) (1976) 249–263.

[53] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016, arXiv preprint arXiv:1603.04467.

[54] B. Dong, R. Sause, J.M. Ricles, Seismic response and damage of reduced-strength steel MRF structures with nonlinear viscous dampers, J. Struct. Eng. 144 (12) (2018) 04018221.

[55] N.E. Castaneda Aguilar, Development and validation of a real-time computational framework for hybrid simulation of dynamically-excited steel frame structures, 2012.

[56] N. Castaneda, X. Gao, S.J. Dyke, Computational tool for real-time hybrid simulation of seismically excited steel frame structures, J. Comput. Civ. Eng. 29 (3) (2013) 04014049.

[57] C. Chen, J.M. Ricles, Development of direct integration algorithms for structural dynamics using discrete control theory, J. Eng. Mech. 134 (8) (2008) 676–683.

[58] B. Chiou, R. Darragh, N. Gregor, W. Silva, NGA project strong-motion database, Earthq. Spectra 24 (1) (2008) 23–44.

[59] J.W. Baker, C. Lee, An improved algorithm for selecting ground motions to match a conditional spectrum, J. Earthq. Eng. 22 (4) (2018) 708–723.

[60] R. Zhang, J. Hajjar, H. Sun, Machine learning approach for sequence clustering with applications to ground motion selection, J. Eng. Mech. 146 (6) (2020) 04020040, (accepted).

[61] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Program. 45 (1–3) (1989) 503–528.

[62] R. Zhang, B.M. Phillips, S. Taniguchi, M. Ikenaga, K. Ikago, Shake table real-time hybrid simulation techniques for the performance evaluation of buildings with inter-story isolation, Struct. Control Health Monit. 24 (10) (2017) e1971.

[63] T. Sato, K. Qi, Adaptive H_∞ filter: its application to structural identification, J. Eng. Mech. 124 (11) (1998) 1233–1240.