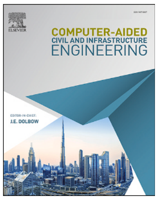
Computer-Aided Civil and Infrastructure Engineering 44 (2026) 100054

Contents lists available at ScienceDirect

Computer-Aided Civil and Infrastructure Engineering

journal homepage: www.sciencedirect.com/journal/computer-aided-civil-and-infrastructure-engineering



Research article

Physics-informed long short-term memory network with data folding for efficient site seismic response prediction

Yongxin Wu[a], Zhanpeng Yin[b], Juncheng Wang[a], Yue Hou[c,*], Haodong Shang[a], Houle Zhang[a]

[a] *State Key Laboratory of precision blasting, Hohai University, No. 1, Xikang Road, Nanjing, 210024, China*
[b] *Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, No. 1, Xikang Road, Nanjing, 210024, China*
[c] *Department of Civil Engineering, Faculty of Science and Engineering, Swansea University, Swansea, United Kingdom*

ARTICLE INFO

Keywords:
Physics-informed neural networks
Long short-term memory
Seismic response prediction
Data folding
Data augmentation
KiK-net

ABSTRACT

Accurate prediction of site seismic response is essential for earthquake engineering and seismic design. Numerical simulation methods, although physically rigorous, become computationally intensive when soils exhibit complex nonlinear behavior and are sensitive to constitutive model selection and parameter calibration. Data-driven deep learning models can approximate nonlinear mappings efficiently, yet they lack built-in physical constraints and risk producing predictions that violate fundamental mechanics when extrapolating beyond the training domain. This study presents a physics-informed deep long short-term memory (LSTM) framework for efficient and accurate site seismic response prediction. The framework enforces physical consistency by adding kinematic derivative relationships as soft constraints in the loss function and improves training efficiency by using Data folding modules. A targeted data augmentation strategy addresses measurement noise and signal variability in recorded data. Comprehensive validation on numerically simulated events and on KiK-net recordings shows the effectiveness of the methodology. On numerical data, the physics-informed model reaches a 97.98% confidence level within ±2% normalized error and reduces training time by more than 60%. On recorded data, the enhanced model with data augmentation reaches an 88.84% confidence level and a response spectrum correlation of 0.951, which supports reliable prediction of frequency content for engineering use. The framework provides an efficient and physically consistent solution for site response prediction with implications for seismic hazard assessment and structural design.

1. Introduction

Earthquakes can cause substantial casualties and economic disruption. As seismic energy travels from the source region toward the ground surface, the incoming motion is reshaped by local geological and geotechnical conditions, which may modify its spectral content, intensity, and effective duration (Tian et al., 2025; Zhang et al., 2023). These site-dependent modifications can increase the demand imposed on engineered structures (Zhang et al., 2024a; Lu et al., 2025) and may also induce severe geotechnical hazards, including soil liquefaction and earthquake-triggered landslides (Tsai & Li, 2024; Li et al., 2025). This amplification and the potential for secondary hazards increase the uncertainty and risk to structural safety. Given these complexities, accurate prediction of site seismic response has become crucial for modern earthquake engineering and seismic design.

In engineering practice, site response is commonly evaluated through numerical procedures, ranging from performance-based equivalent-linear frameworks to fully nonlinear time-domain analyses (Ansal et al., 2024; Bantis et al., 2025). The equivalent-linear method is widely adopted for its computational efficiency, although its linear elastic assumption limits predictive capability under strong ground motions (Bouckovalas et al., 2017; Ciancimino et al., 2018). Fully nonlinear time-history simulations provide higher physical fidelity but at substantially greater computational cost (Yan & Zhang, 2023). Moreover, prediction accuracy depends on constitutive model selection and parameter calibration (Chen et al., 2023; Qiu et al., 2023), as well as on subsurface heterogeneity (Huang et al., 2021). These computational and parametric challenges become bottlenecks in large-scale regional seismic hazard assessments (Ganji & Seylabi, 2023; Zhang et al., 2019).

Data-driven methods have therefore been explored as complementary tools for site-response and ground-motion related prediction tasks. Neural networks can approximate nonlinear mappings and exploit

---

[*] Correspondence author.
E-mail address: yue.hou@swansea.ac.uk (Y. Hou).

https://doi.org/10.1016/j.cacaie.2026.100054
Received 15 January 2026; Accepted 14 March 2026
Available online 15 April 2026
1093-9687/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

patterns in large datasets, making them suitable for problems where repeated simulations are prohibitive (Fonseca et al., 2003). Machine learning techniques have been applied to scalar ground-motion prediction and nonparametric spatial modeling of seismic attenuation, with broad learning networks offering computationally efficient alternatives to deep stacking architectures (Khosravikia & Clayton, 2021; Kuok & Yuen, 2020). Extending scalar prediction to full time-history modeling requires architectures that capture sequential dependence. Recurrent networks, particularly LSTMs, address this need by modeling temporal evolution in seismic response records from downhole arrays and infrastructure systems (Li et al., 2023a; Zhang et al., 2024b). Multi-source input integration and hybrid recurrent frameworks further improve generalization across diverse site conditions and ground-motion characteristics (Li et al., 2024; Fayaz et al., 2021). Deep learning has also been applied to rapid peak seismic response prediction of underground structures and to spatiotemporal deformation prediction of water-related concrete structures (Hu et al., 2024; Gu et al., 2025). Transfer learning strategies enable cross-structure knowledge sharing for seismic damage diagnosis and response prediction (Xu & Noh, 2021; Pak & Paal, 2025). In the frequency domain, real-time response spectrum estimation and deep learning-based modal combination approaches have demonstrated promising accuracy for engineering applications (Fayaz & Galasso, 2023; Kim et al., 2024). Reduced-order modeling techniques provide additional tools for computational efficiency (Zhong et al., 2025; Huhn et al., 2023), and deep learning has been demonstrated for rapid post-event regional damage assessment (Lu et al., 2021). The broad learning paradigm has been further extended to Bayesian probabilistic modeling with optimized architecture configuration, enabling uncertainty quantification alongside point predictions (Kuok & Yuen, 2021). Despite these advances, purely data-driven models remain difficult to interpret and may become unreliable when extrapolation is required (Hu et al., 2023; Li et al., 2023b). Without explicit physical structure, a model may fit observations while violating basic dynamic behavior, a risk pronounced in site-response applications where recordings of rare, large earthquakes are limited.

Physics-informed neural networks (PINNs) have been proposed to mitigate these issues and to reduce the dependence on large labeled datasets (Raissi et al., 2019). In PINNs, the governing equations of the underlying process are enforced during training through equation-based residual terms. This design promotes compliance with physical laws and introduces an interpretable constraint on the hypothesis space (Zhu et al., 2019). Prior knowledge-infused neural networks have demonstrated efficient structural performance assessment through few-shot incremental learning, reducing the reliance on extensive labeled data while improving generalization under complex conditions (Chen et al., 2024). The integration of structural dynamic equations and state evolution relationships into neural network architecture has led to substantial improvements in the model's capacity to learn seismic response features from limited samples (Zhang et al., 2020a, 2020b). Similarly, hybrid modeling frameworks that combine deep residual networks with conventional time integration algorithms have demonstrated enhanced accuracy and physical consistency in modeling nonlinear structural behavior while substantially reducing the need for training data (Wang et al., 2023). Furthermore, the recently developed CM-PINNs framework introduces nonlinear constitutive constraints, enabling efficient and accurate modeling of nonlinear seismic response under small-sample conditions and further expanding the application potential of PINNs in structural seismic prediction (Wu et al., 2025).

Despite these advances in structural dynamic modeling, the application of PINNs to seismic response time history prediction under complex site conditions remains in its infancy. Existing studies often focus on simplified physical models or idealized boundary conditions, which are insufficient to fully characterize the nonlinear, multi-scale, and strongly coupled features of actual site response. For instance, Chen et al. incorporated physical constraints into a BiLSTM architecture to improve response spectrum prediction accuracy (Chen et al., 2025). Although promising, this approach primarily targets frequency-domain features. It falls short of meeting the high-precision requirements for full-time history modeling of ground motions, indicating a need for further development.

This study proposes an improved physics-constrained long short-term memory (LSTM) framework to address these challenges, enabling efficient, accurate, and physically consistent prediction of site seismic response time histories. The framework enforces kinematic consistency among the predicted displacement, velocity, and acceleration responses by embedding their derivative relationships as soft constraints in the loss function, thereby guiding the model toward physically plausible solutions without requiring explicit knowledge of site-specific material parameters. An innovative data folding and unfolding strategy is introduced to overcome computational bottlenecks associated with long-duration seismic records, which significantly improves training efficiency without compromising predictive performance. The effectiveness of the proposed method is validated using both recorded and simulated datasets from the IBRH13 station of the Japanese KiK-net strong-motion network, enabling a comprehensive evaluation of model performance across various scenarios.

The paper is organized as follows. Section 2 presents the proposed physics-informed LSTM framework and Section 3 describes dataset preparation for numerical simulations and KiK-net recordings. Section 4 reports numerical validation and efficiency analyses. Section 5 presents recorded-data validation and introduces the noise-aware augmentation strategy. Section 6 concludes the paper.

2. A physics-informed LSTM framework of seismic response prediction

Traditional numerical simulation methods provide physically consistent results but are computationally expensive and require detailed soil characterization. In contrast, data-driven approaches lack physical interpretability and may violate fundamental physical laws. These limitations motivate physics-informed deep learning frameworks that couple efficiency with physical consistency.

2.1. Governing equation of site seismic response

The site equation of motion is discretized as:

$$\mathbf{M}\ddot{\boldsymbol{u}}(t)+\mathbf{C}\dot{\boldsymbol{u}}+\mathbf{f}_{int}(\boldsymbol{u}(t))=\mathbf{f}_{ext}(t) \tag{1}$$

where $\mathbf{u}(t)$is the nodal displacement vector, **M**and **C**are the mass and damping matrices, $\mathbf{f}_{\mathrm{int}}(\cdot)$denotes the internal restoring force, and $\mathbf{f}_{\mathrm{ext}}(t)$is the external seismic excitation. Because soil response is nonlinear and path dependent, $\mathbf{f}_{\mathrm{int}}$is governed by the constitutive behavior and largely controls the overall site response. In data-driven modeling, enforcing physics-based constraints can restrict the admissible solution space and promote physically consistent predictions.

Soil exhibits highly nonlinear and path-dependent behavior under dynamic loading, manifesting through backbone curve nonlinearity, hysteretic energy dissipation, and strain-rate dependency. The internal restoring force $\mathbf{f}_{int}$ is governed by the constitutive behavior of the soil, which plays a key role in determining the overall site response. Directly embedding Eq. (1) as a loss constraint would require accurate **M, C**, and $\mathbf{f}_{int}$, which are inaccessible in practice: for data generated by nonlinear constitutive solvers, restoring forces remain internal to the iterative process; for field recordings, subsurface parameters are inherently uncertain. Constructing a simplified surrogate equation (e.g., decomposing surface acceleration into input and relative components) would introduce additional predicted variables, extra loss terms, and a larger parameter space, increasing training cost without guaranteed accuracy gains.

In contrast, kinematic derivative relationships ($\dot{\boldsymbol{u}}$=d$\boldsymbol{u}$/dt, $\ddot{\boldsymbol{u}}$=d$\dot{\boldsymbol{u}}$/dt) are universal identities independent of constitutive models or site-specific parameters. Although adopting these constraints expands the output from $d = 1$ to $d = 3$, reliable ground-truth displacement and velocity can be obtained by numerically integrating the measured acceleration, providing known supervision for the expanded outputs. The resulting kinematic constraints act as a physically grounded regularizer that reduces the feasible solution space and enforces internal consistency in frequency content and energy distribution.

2.2. Framework overview

PINNs incorporate physical knowledge into training, typically by enforcing governing laws through the loss function. Building on this idea, a physics-informed deep learning framework is developed for site-specific seismic response prediction, embedding domain knowledge in both the network architecture and the objective function.

As shown in Fig. 1, the framework is organized into four modules: (1) a data module for numerical simulations and field recordings, with preprocessing for each dataset and augmentation for recordings only (noise injection, scaling, and time shift); (2) a deep LSTM predictor equipped with folding-unfolding operations for efficient long-sequence learning; (3) a physics-informed loss module that enforces kinematic consistency via a hybrid objective; and (4) an output module for time and frequency domain evaluation. LSTM networks (Hochreiter & Schmidhuber, 1997) were introduced to mitigate vanishing gradients in recurrent neural networks and to model long-range dependencies in sequential data (Sherstinsky, 2020). The proposed framework exploits the temporal modeling capability of LSTMs and enforces kinematic relations as soft constraints during training to obtain physically plausible solutions. The framework is elaborated through four key components in the following.

(1) Deep LSTM network Architecture

As shown in Fig. 1, the network comprises an input layer, three LSTM layers, three fully connected (FC) layers, and an output layer. ReLU is applied between successive layers. Each LSTM layer uses 128 hidden units. This depth and width align with configurations reported in seismic time-series studies (Li et al., 2023a; Zhang et al., 2024b), where two to four recurrent layers with 64–256 units are common; Section 5.2 confirms that reducing the hidden size to 64 still yields satisfactory accuracy on recorded data. The FC layers have 64, 32, and d units, where d is the number of predicted targets. Without physical constraints, the network outputs only surface acceleration ($d = 1$); with physical constraints, displacement, velocity, and acceleration are predicted jointly ($d = 3$).

The LSTM update equations are summarized below. At each time step i, the previous hidden state h_{i-1} and the current input x_i are concatenated and passed through four gating mechanisms. The forget gate determines which information to discard from the previous cell state:

$$f_i = \sigma\left(W_f \cdot [h_{i-1},\ x_i] + b_f\right) \tag{2}$$

The input gate controls which new information is admitted:

$$I_i = \sigma\left(W_I \cdot [h_{i-1},\ x_i] + b_I\right) \tag{3}$$

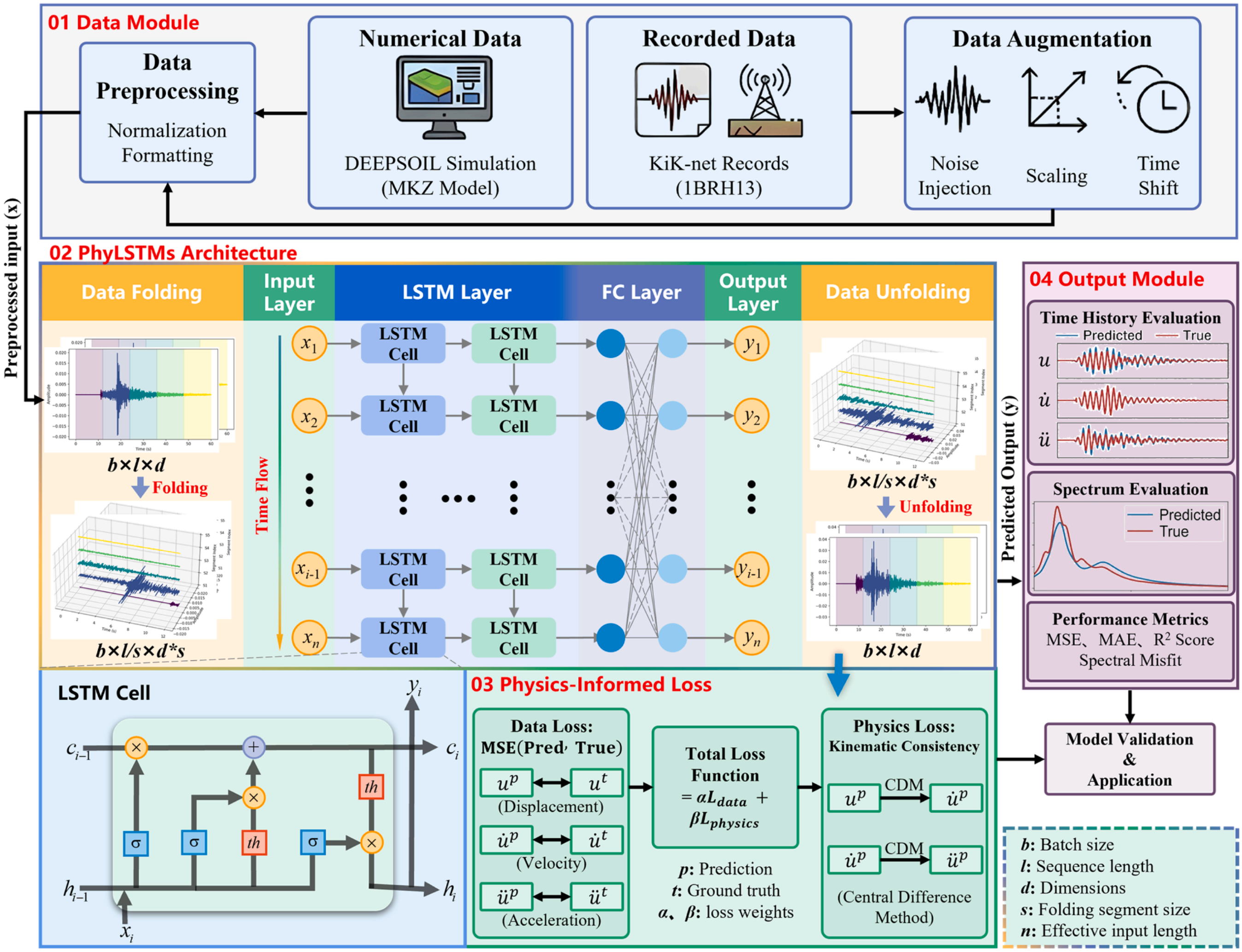


Fig. 1. Architecture of the physics-informed deep LSTM network framework for seismic site response prediction.

A candidate memory vector is computed via a hyperbolic tangent activation:

$$\widetilde{C}_i = \tanh(W_C \cdot [h_{i-1},\ x_i] + b_C) \tag{4}$$

The output gate regulates how much of the updated cell state is exposed as the hidden output:

$$o_i = \sigma(W_o \cdot [h_{i-1},\ x_i] + b_o) \tag{5}$$

The cell state is updated by combining the retained previous memory with the gated candidate:

$$c_i = f_i \times c_{i-1} + I_i \times \widetilde{C}_i \tag{6}$$

Finally, the hidden state, which also serves as the layer output, is obtained by passing the updated cell state through a tanh activation modulated by the output gate:

$$y_i = h_i = o_i \times \tanh(c_i) \tag{7}$$

where $x_i$and $y_i$are the input and output at time step $i(i = 1,\ldots,n)$, and h_{i-1} and C_{i-1} denote the previous hidden state and cell state. The gate outputs are f_i, I_i, and o_i. The nonlinearities $\sigma(\cdot)$ and tanh $(\cdot)$ are used as in Eqs. (2)-(7). For each gate $p \in \{f, I, C, o\}$, the corresponding weight matrix and bias are W_p and b_p.

(2) Input Data Folding

The data module contains two data branches: DEEPSOIL-simulated responses and KiK-net recordings. Each dataset is normalized and formatted independently, and models are trained on each dataset separately to account for their distinct characteristics.

For long seismic time series, computing and storing hidden states at each time step results in significant computational cost and memory consumption. Traditional approaches often employ down-sampling techniques to reduce computational burden; however, this strategy can introduce baseline drift in seismic waveforms, which is difficult to correct and may compromise the accuracy of subsequent analyses. To address this issue while avoiding the pitfalls of conventional down-sampling, a data folding strategy is adopted prior to feeding seismic records into the network. While LSTM models cannot parallelize operations along the temporal dimension, they can efficiently perform parallel computations along the feature dimension.

The original ground motion time history data are formatted as a three-dimensional tensor of size $b \times l \times d$, where b is the number of samples, l is the sequence length, and d is the number of feature channels. A folding operation reshapes this tensor into $b \times (l/s) \times (d \times s)$ by grouping every l/s consecutive samples ($s = 5$ in this study) along the temporal axis into a single feature vector. The reshape is bijective: each sample maps one-to-one to an element in the folded tensor and can be exactly recovered by the inverse operation. No data value is altered, discarded, or reordered, so the full sampling rate and waveform fidelity are strictly preserved. Unlike down-sampling, which discards samples and may introduce aliasing or baseline drift, folding retains the complete waveform. The LSTM then processes l/s time steps, each carrying $d \times s$ features; consecutive samples are handled jointly within one step while the hidden state propagates sequentially across all l/s steps.

(3) Output Data Unfolding

The hidden representations from the final LSTM layer pass through a fully connected layer that maps the spatiotemporal features to the prediction target space, producing a predicted tensor of size $b \times l \times d$ consistent with the folded input structure. An unfolding operation then reshapes this tensor back to $b \times l \times d$ before the loss function is evaluated. Both fold and unfold are parameter-free reshape operations that reside inside the differentiable computational graph. Gradients propagate from the full-length (l-step) loss through the unfold, back through all LSTM hidden states, and through the fold. The optimizer therefore receives gradient signals that reflect the complete temporal structure of the prediction error, and the fold-unfold pair introduces no information loss to gradient-based training.

(4) Physics-Informed Optimization

Following the rationale in Section 2.1, the framework adopts kinematic derivative relationships in place of the full equation of motion as physics-based soft constraints, and constructs a hybrid loss function that integrates data-driven and physics-based terms.

The data-driven loss term, L_{data}, measures discrepancies between predictions and ground truth of displacement, velocity, and acceleration:

$$L_u^D = \mathrm{MSE}(u^p,\ u^t) \tag{8}$$

$$L_{\dot{u}}^D = \mathrm{MSE}(\dot{u}^p,\ \dot{u}^t) \tag{9}$$

$$L_{\ddot{u}}^D = \mathrm{MSE}(\ddot{u}^p,\ \ddot{u}^t) \tag{10}$$

$$L_{data} = \alpha_1 L_u^D + \alpha_2 L_{\dot{u}}^D + \alpha_3 L_{\ddot{u}}^D \tag{11}$$

where $u^p, i^p$are the predicted displacement, velocity, and acceleration, and $u^t, i^t$are the corresponding ground truth values. The weights α_1, α_2, α_3 compensate for the magnitude differences among the three quantities. Acceleration is chosen as the reference (α_3=1) because it is the primary engineering quantity in seismic design and typically has the largest numerical magnitude, serving as a natural normalization anchor. During the first 25% of training epochs, only L_{data} is optimized and the relative magnitudes of its three components are recorded. The weights α_1 and α_2 are then set inversely proportional to the observed ratios, ensuring comparable contributions from all components. This adaptive calibration avoids manual tuning and adjusts automatically to each dataset.

The physics-based loss term, $L_{physics}$, penalizes deviations in the numerical time derivatives of predictions, ensuring consistency with kinematic relationships:

$$L_{\dot{u}^P = MSE(\mathrm{CDM}(u^p),\ \dot{u}^p)} \tag{12}$$

$$L_{\ddot{u}}^P = MSE\big(\mathrm{CDM}(\dot{u}^p),\ \ddot{u}^p\big) \tag{13}$$

$$L_{physics} = \beta_1 L_{\dot{u}}^P + \beta_2 L_{\ddot{u}}^P \tag{14}$$

where CDM denotes the central difference method, computing $\mathrm{CDM}(y_i) = (y_{i+1} - y_{i-1})/(2\Delta t)$ for uniformly sampled data with step Δt.

CDM provides second-order accuracy $O(\Delta t^2)$ and evaluates each derivative pointwise from neighboring values. As a purely algebraic operation on discrete samples, CDM is independent of the physical system that produced the signal and applies equally to linear and nonlinear responses. Sequential time-stepping schemes such as Newmark-β accumulate truncation error along the full sequence; for the 12 000-step recordings in this study, such cumulative drift would systematically bias the derivative estimates that enter the physics loss. First-order forward differencing achieves only $O(\Delta t)$ accuracy and yields asymmetric estimates. Input data are pre-filtered at 0.5–20 Hz and the neural-network forward pass provides implicit smoothing, so differentiation noise remains bounded; the adaptive physics-loss weights (β_1, β_2) further prevent residual noise from dominating optimization. The weighting parameters are specified as β_2=1 and β_1=α_2, ensuring that the physical constraint is incorporated consistently with its role in the data-fidelity term.

Within this framework, the LSTM network is extended by embedding prior physical knowledge, resulting in a PhyLSTM model that constrains predictions using kinematic differential equations. The final training objective is expressed as:

$$Loss\ =\ L_{data}\ +\ L_{physics} \tag{15}$$

Jointly optimizing the data-driven and physics-based loss terms enables the model to leverage deep learning's expressive power while maintaining physical validity, and the resulting model exhibits improved robustness and generalization.

3. Construction of the ground motion dataset

A high-quality dataset is essential for efficient training and reliable validation of deep learning models. This study establishes two types of ground motion datasets: one derived from recorded data and the other from numerical simulations, supporting subsequent modeling that integrates both data-driven and physics-informed approaches.

3.1. Data selection and processing

The recorded ground motions were obtained from the Japanese strong-motion observation network KiK-net (Aoi, 2000). Thompson et al. provided the methodology for station selection (Thompson et al., 2012). Sixteen representative stations were selected based on their approach. Each station exhibits approximately one-dimensional site response. The IBRH13 station was chosen as the primary research site for several reasons. The station has well-defined geological conditions, abundant recorded data, and stable instrument operation.

A total of 1000 two-component ground motion records were selected from the IBRH13 station between January 2000 and December 2020. The selection required a 100 Hz sampling frequency, with surface peak ground acceleration (PGA) between 1 gal and 100 gal The data spanned focal depths from 0 to 150 km and moment magnitudes (Mw) from 2 to 8. The epicentral distances ranged from 0 to 200 km, covering a broad spectrum of ground motion characteristics. Only the east-west component of both bedrock and surface data were retained for consistency in input-output dimensions.

Raw data often include low-frequency noise and baseline drift due to instrument tilt. A baseline correction was performed by subtracting the mean of the pre-event interval (0–20 s) and applying a 4th-order Butterworth band-pass filter (0.5–20 Hz) to remove noise. The 0.5–20 Hz passband covers the natural-period range of most engineered structures (0.05–2 s) and follows standard practice in engineering seismology. Components below 0.5 Hz, corresponding to periods exceeding 2 s, are excluded; long-period site effects at those frequencies fall outside the scope of the present study.

3.2. Construction of the recorded dataset

After standard preprocessing (detrending, baseline correction and band-pass filtering), 1000 ground-motion records were organized into structured input-output pairs, with bedrock acceleration time histories as inputs and the corresponding surface acceleration time histories as outputs. For the controlled numerical dataset used, each record was truncated to the first 60 s (6000 samples at 100 Hz), which retains the dominant energy content and is sufficient for demonstrating the model's performance under idealized conditions. For the recorded dataset, a longer window of 120 s (12,000 samples at 100 Hz) is adopted. This extended window better preserves late-arriving phases and long-period content and reduces the risk of truncation-induced baseline artifacts that can impair post-processing and interpretation of real recordings.

Fig. 2 presents the distributions of key seismic characteristics, including focal depth, magnitude, epicentral distance, and PGA, showing good consistency between training and test sets and ensuring data representativeness.

3.3. Construction of the numerically dataset

Although empirical data are representative, the input–output coupling is complex and obscures nonlinear amplification mechanisms. A DEEPSOIL-generated dataset was therefore constructed to help the model learn idealized physical processes.

The modeling results, shown in Fig. 3, divide the soil profile into seven layers based on publicly available geological information from KiK-net and detailed research by Romain Meite (Meite et al., 2020).

The MKZ nonlinear constitutive model (Yi, 2010; Mei et al., 2020) was adopted to simulate shear modulus degradation behavior:

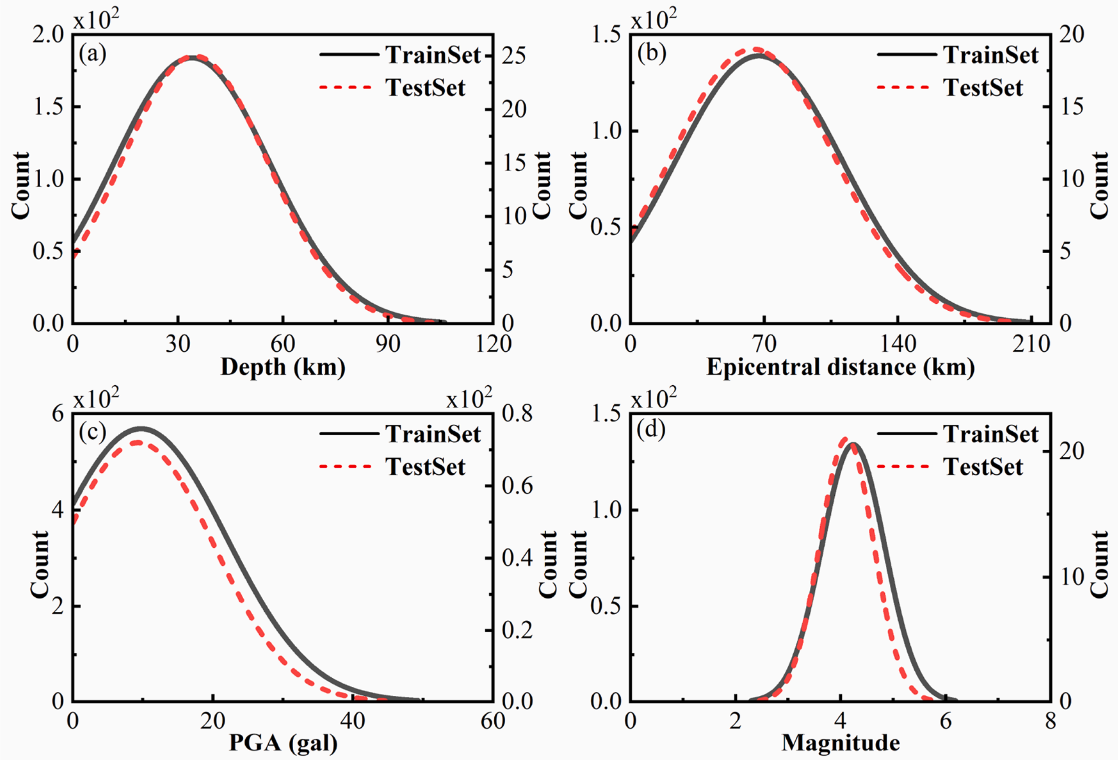


Fig. 2. Distribution of seismic parameters in training and test datasets: (a) focal depth, (b) epicentral distance, (c) PGA, and (d) magnitude.

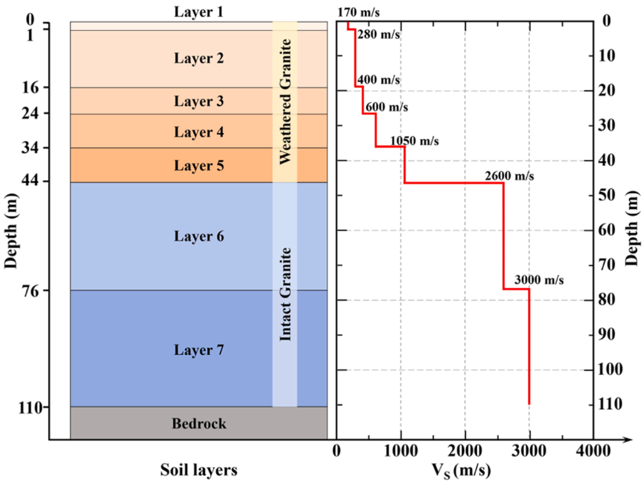


Fig. 3. Stratigraphy and shear-wave velocity (Vs) profile at the IBRH13 site.

$$\tau = \frac{G_{max}\cdot\gamma}{1+\beta\left(\frac{\gamma}{\gamma_r}\right)^s} \tag{16}$$

where $\tau$is the shear stress, $G_{\max}$denotes the initial shear modulus, $\gamma$is the shear strain, $\gamma_r$represents the reference shear strain, and $\beta$and $s$are empirical fitting parameters. The reference shear strain $\gamma_r$is computed as:

$$\gamma_r = s_{ref}\left(\frac{\sigma_v{}'}{\sigma_{ref}}\right)^b \tag{17}$$

where $s_{ref}$is defined at σ_{ref}, σ_{ref}denotes the reference effective stress, and $\sigma_v{}'$ is the vertical effective stress. In this study, the exponent $b$is set to 0.

The parameters adopted for each soil layer follow the study by Romain Meite et al. (Meite et al., 2020), as summarized in Table 1. The 1000 processed bedrock acceleration time histories were utilized as input for nonlinear dynamic response analyses, obtaining corresponding surface displacement, velocity, and acceleration responses. Fig. 4 illustrates a representative ground motion event, demonstrating the site amplification effect and spectral changes.

Both datasets were split into training, validation, and test subsets using an 8:1:1 ratio. The numerically simulated dataset provides reference benchmarks and ideal solution samples for subsequent physics-informed deep learning models, supporting the evaluation of model physical consistency and generalization capability.

4. Numerical validation

The numerically generated training, validation, and test datasets described in Section 3 were used to train and evaluate three models. The baseline LSTM was trained using data loss and the PhyLSTM employed physics-inform loss. PhyLSTMs extended PhyLSTM by incorporating input data folding and output data unfolding.

Table 1
Soil layer parameters for the DEEPSOIL model at IBRH13 site.

Layer	Thickness (m)	Density (kg/m^3)	Vs (m/s)	σ_{ref} (Mpa)	γ_r (%)	β	s
Layer1	1	1770	170	0.18	0.0382	1.5	0.86
Layer2	15	1910	280	0.18	0.0232	1.5	0.86
Layer3	8	2080	400	0.18	0.0163	1.5	0.86
Layer4	10	2080	600	0.18	0.0109	1.5	0.86
Layer5	10	2330	1050	0.18	0.0062	1.5	0.86
Layer6	32	2590	2600	0.18	0.0025	1.5	0.86
Layer7	24	2590	3000	0.18	0.0022	1.5	0.86

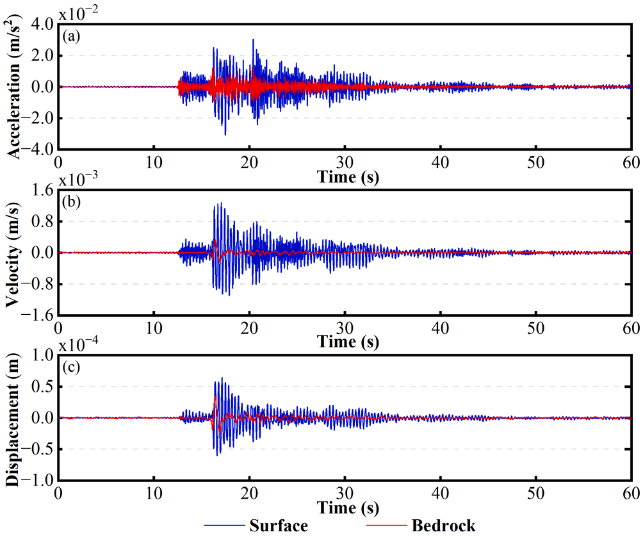


Fig. 4. Comparison of typical bedrock and surface responses: (a) acceleration, (b) velocity, and (c) displacement.

Training was conducted for 6000 epochs with a learning rate of 1 × $10^{-4}$and a batch size of 100 using the Adam optimizer (Kingma & Ba, 2014). All models were built in Python using PyTorch (Paszke et al., 2019) and trained on a PC with an Intel Core i7–11700 K (3.60 GHz) and an NVIDIA RTX 3070.

Four standard metrics were used for model evaluation: mean absolute error (MAE), root mean square error (RMSE), mean peak error (MeanPE), and Pearson correlation coefficient (R). MAE and RMSE quantify the average and root-mean-square prediction errors respectively. MeanPE measures the average relative error of peak ground acceleration across test samples. R measures the linear correlation between predicted and target time histories and ranges from 0 (no correlation) to 1 (perfect correlation).

Fig. 5(a)-(c) presents the training and validation losses of all three models on a logarithmic scale. The LSTM model shows loss stabilization after approximately 1000 epochs, with validation loss remaining higher than training loss, indicating a degree of overfitting. Both PhyLSTM and PhyLSTMs models exhibit faster convergence and achieve lower loss values. The PhyLSTMs model converges rapidly within the first 100 epochs and consistently maintains the lowest overall loss, demonstrating superior generalization performance. The PhyLSTMs model also achieves the highest training efficiency, with an average epoch time of only 3.6 s, substantially faster than the LSTM at 9.7 s and PhyLSTM at 11.2 s.

Numerical datasets were employed to train and test the three models for predicting surface acceleration responses under seismic loading. The normalized prediction errors on the test sets were statistically analyzed to assess model performance. Fig. 6 presents the probability density functions (PDFs) of the normalized prediction errors, with a detailed view of the 98% confidence interval.

Fig. 6 shows that the LSTM model exhibits a pronounced peak near zero with relatively narrow tails, indicating concentrated prediction errors around zero. A small number of significant errors appear in the tails with low probabilities, suggesting occasional but limited deviations. The PhyLSTM and PhyLSTMs models display similarly centered error distributions with even lower incidences of extreme deviations, reflecting enhanced robustness and accuracy. The proportions of normalized errors within the ±2% confidence interval are 96.32% for LSTM, 97.10% for PhyLSTM, and 97.98% for PhyLSTMs. These results represent consistent but modest improvements of 0.81 and 1.66 percentage points for PhyLSTM and PhyLSTMs respectively. The primary value of the physics-informed constraints on numerical data lies not in these aggregate metrics but in the substantial reduction of extreme

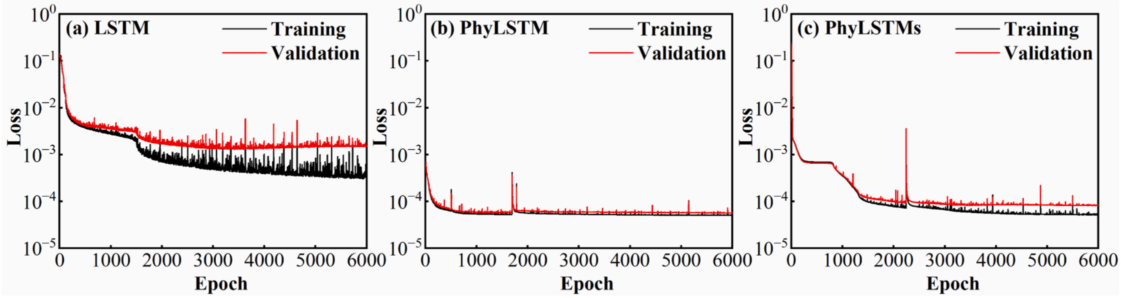


Fig. 5. Training and validation loss histories of the three models on synthetic dataset: (a) LSTM, (b) PhyLSTM, and (c) PhyLSTMs.

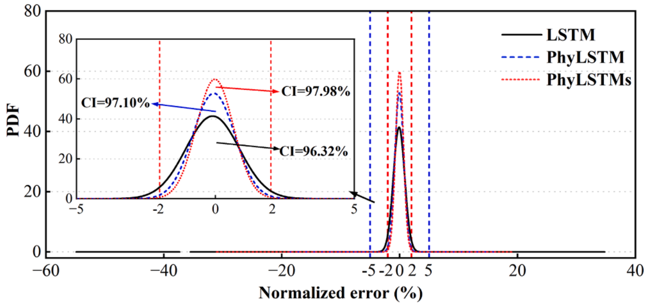


Fig. 6. Probability density functions of normalized prediction errors within 98% confidence interval for three models.

prediction errors: Table 3 shows that the worst-case relative peak error drops from 30.06% (LSTM) to 9.38% (PhyLSTM) and 10.23% (PhyLSTMs), an improvement critical for engineering reliability. A more pronounced overall benefit emerges on recorded data (Section 5), where the combined framework achieves a 15.4% RMSE reduction relative to the baseline.

Performance was assessed using the metrics defined in Section 4.1. Table 2 indicates similar error levels on the training and test sets for all models.

Compared with the LSTM model, the PhyLSTM model demonstrates lower prediction errors across the test set, with MAE reduced by 12.63% and RMSE by 13.07%. The correlation coefficient R is marginally higher, indicating improved predictive performance through the integration of physics-informed constraints.

The PhyLSTMs model shows additional improvements over PhyLSTM, with MAE decreasing by 13.25% and RMSE by 10.08%. The correlation coefficient R also increases slightly. The high R values (>99%) across all models reflect the deterministic nature of the numerical dataset: input and output are linked by a fixed constitutive model and soil profile with no measurement noise or three-dimensional scattering. Training and test sets are split by earthquake event in an 8:1:1 ratio with no shared input records, so data leakage is excluded. The contrast with the considerably lower R (66.16%) on recorded data in Section 5 further confirms that the numerical dataset does not artificially inflate performance but rather provides a controlled baseline for isolating the effect of each framework component.

Table 2
Predictive performance comparison of different models.

Model	Dataset	MAE	RMSE	R (%)
LSTM	Train	8.81×10^{-3}	6.55×10^{-2}	99.77
	Test	9.53×10^{-3}	7.19×10^{-2}	99.69
PhyLSTM	Train	7.19×10^{-3}	5.85×10^{-2}	99.82
	Test	8.31×10^{-3}	6.25×10^{-2}	99.79
PhyLSTMs	Train	7.42×10^{-3}	4.80×10^{-2}	99.88
	Test	7.16×10^{-3}	5.62×10^{-2}	99.82

Six representative samples from the test set were selected to illustrate model behavior at the two performance extremes; a full statistical summary covering all 100 test samples is provided in Appendix A. The best-performing sample for each model is denoted as L-B (LSTM), PL-B (PhyLSTM), and PLs-B (PhyLSTMs), while the worst-performing samples are denoted as L-W, PL-W, and PLs-W. Figs. 7 and 8 present comprehensive comparisons between predicted and target ground surface acceleration responses for these six cases. Each figure is arranged in a 3 × 3 grid: model-wise columns, with rows for acceleration records in the time domain, residuals, and spectral response. To facilitate a direct amplitude comparison between residuals and the original responses, the residual panels (d–f) are plotted using the same y-axis limits as the corresponding time-history panels (a–c); therefore, residual curves may appear close to zero when prediction errors are small.

Fig. 7 illustrates the best-performing samples of each model. The predicted acceleration time histories almost perfectly overlap with the target curves throughout the entire duration, with residuals remaining close to zero. The corresponding response spectra confirm high agreement between predicted and target results across the frequency domain. Fig. 8 displays the worst-performing samples. For PL-W and PLs-W cases, the predicted curves still closely follow the target with only minor deviations near peak values and relatively small residuals. For L-W case, the predicted curve exhibits significant discrepancies, particularly at the initial peak, with noticeably larger residuals. The response spectra in worst-case scenarios demonstrate that PhyLSTM and PhyLSTMs models provide more robust and accurate spectral predictions compared to the conventional LSTM model.

Table 3 presents a detailed comparison of PGA values for the six representative samples, including target and predicted peak values with corresponding absolute and relative errors.

Both PhyLSTM and PhyLSTMs models reduce peak prediction errors significantly compared to the baseline LSTM model, especially in worst-case scenarios. The relative errors for PhyLSTM and PhyLSTMs are much lower than LSTM in these challenging cases, demonstrating that physics-informed constraints substantially improve accuracy and robustness of peak response predictions.

Across the test set, the LSTM model produced an average relative peak error of 2.11%, while PhyLSTM and PhyLSTMs achieved 1.62% and 1.60% respectively. These improvements represent relative reductions of 23.38% and 24.14%, substantiating that incorporating physical knowledge enhances predictive performance in seismic response analysis.

The results validate that physics-informed constraints improve both accuracy and robustness of seismic ground motion predictions, particularly for peak responses and spectral characteristics. The PhyLSTMs model delivers superior performance and training efficiency, validating the data folding strategy as a practical approach. Once trained, inference for a single ground-motion scenario completes in milliseconds, a throughput comparable to equivalent-linear frequency-domain analysis while retaining the capacity to capture nonlinear response features that the linear-elastic assumption of the equivalent-linear method cannot represent.

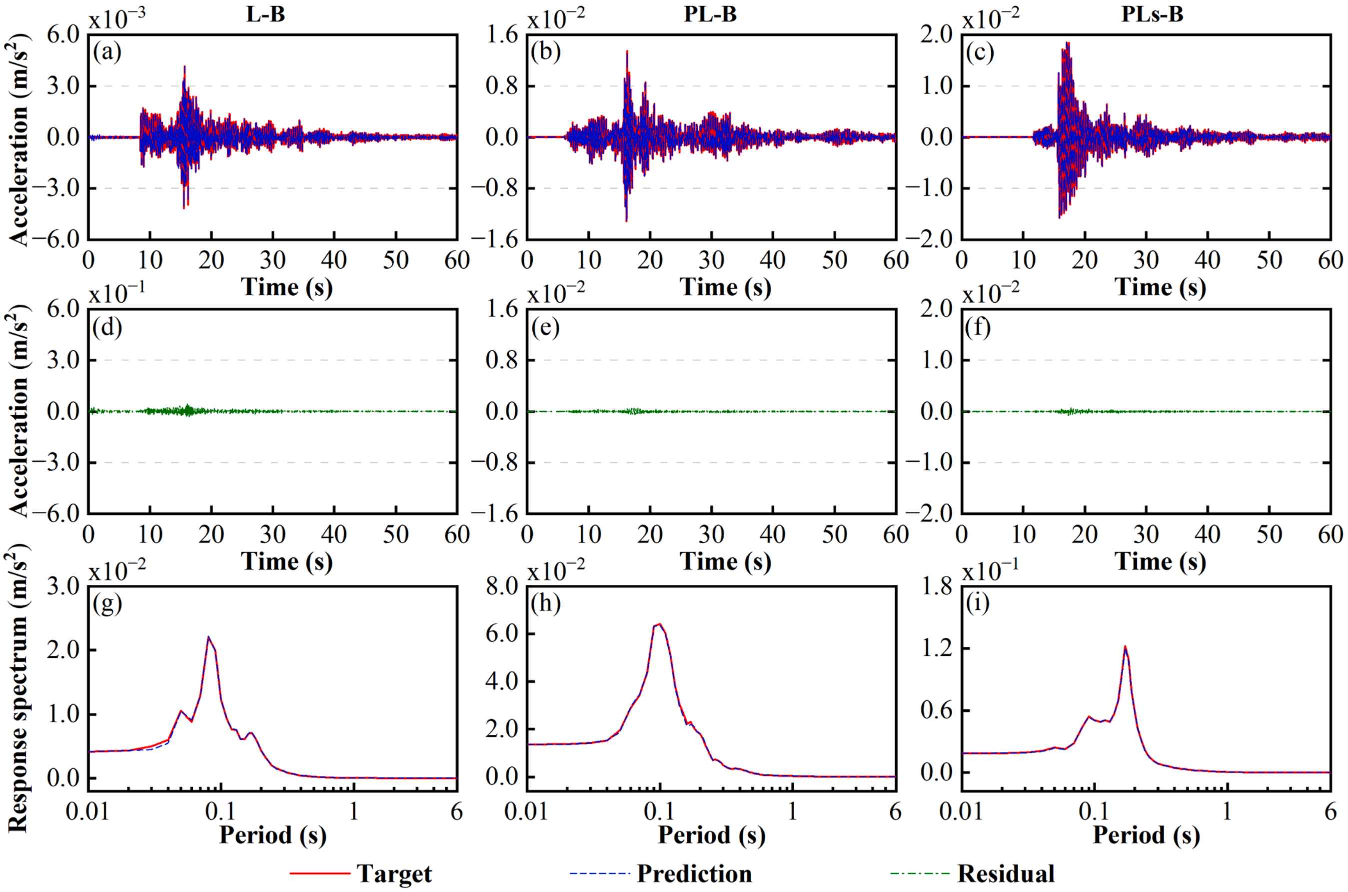


Fig. 7. Best numerical-data cases for three models: (a–c) acceleration time histories, (d–f) residuals, and (g–i) response spectra.

5. Application to recorded ground motions

Real earthquake records differ from numerical simulations and pose specific challenges. Measurement devices add high-frequency noise during low-amplitude phases, decreasing the signal-to-noise ratio in weak motions. Wind, traffic, and human activity contaminate pre-event and post-event segments. Standard baseline correction removes part of the drift, but residual low-frequency drift and instrument tilt still bias velocity and displacement. Heterogeneous three-dimensional geology at many sites exceeds the reach of one-dimensional models. These factors can push models to learn noise patterns instead of physical relationships, necessitating a targeted data augmentation strategy to improve robustness.

5.1. Noise-aware data augmentation strategy

Four perturbation techniques are used only during training to improve robustness. Validation and test sets remain unchanged to protect data integrity. Each original record is expanded into several augmented variants to increase diversity in the training set.

(1) Gaussian Noise Injection

Zero-mean Gaussian noise is added to input acceleration time histories to simulate instrumental noise. The noise standard deviation increases progressively across augmentation iterations as follows:

$$\sigma_n = 0.02 \times \frac{k}{K} \tag{18}$$

Here, k indexes the augmentation step and $K = 5$ denotes the total number of augmentations. The noise sequence is sampled as $n_g[i] \sim \mathcal{N}(0, \sigma_n^2)$. The procedure produces progressively stronger noise realizations.

(2) Amplitude Scaling

Random scaling factors account for uncertainties in earthquake magnitude and site amplification effects. The scaling factor α_s is defined as:

$$\alpha_s = 1 + U(-0.05,\ 0.05) \tag{19}$$

where $U(-0.05, 0.05)$ is a uniform random variable within ±5%.

(3) Temporal Shifting

Integer time shifts were used to represent variations in wave propagation paths and trigger timing. The augmented sequence after shifting was defined as:

$$x_s[i] = x[i - \Delta t] \tag{20}$$

where the shift $\Delta t \in \{-2, -1, 0, 1, 2\}$ corresponds to ±0.02 s.

(4) Low-Frequency Perturbations

Smooth low-frequency noise represents baseline uncertainties and long-period drift. This noise is generated by applying a moving average filter to zero-mean Gaussian noise with amplitude 0.01 as follows:

$$n_l[i] = \frac{1}{\omega} \sum_{j=i-\frac{\omega}{2}}^{i+\frac{\omega}{2}} \varepsilon[j],\ \ \varepsilon[j] \sim N\left(0, 0.01^2\right) \tag{21}$$

where $\omega = 5$ is the filter window size.

The augmentation operations are applied in the following order: temporal shifting, Gaussian noise injection, low-frequency perturbation, and amplitude scaling. The final augmented sequence is expressed as:

$$x_{aug} = \left(x_s + n_g + n_l\right) \times \alpha_s \tag{22}$$

Because the perturbation magnitudes are small relative to the signal (noise σ ≤ 0.02, scaling ±5%, low-frequency amplitude 0.01), the

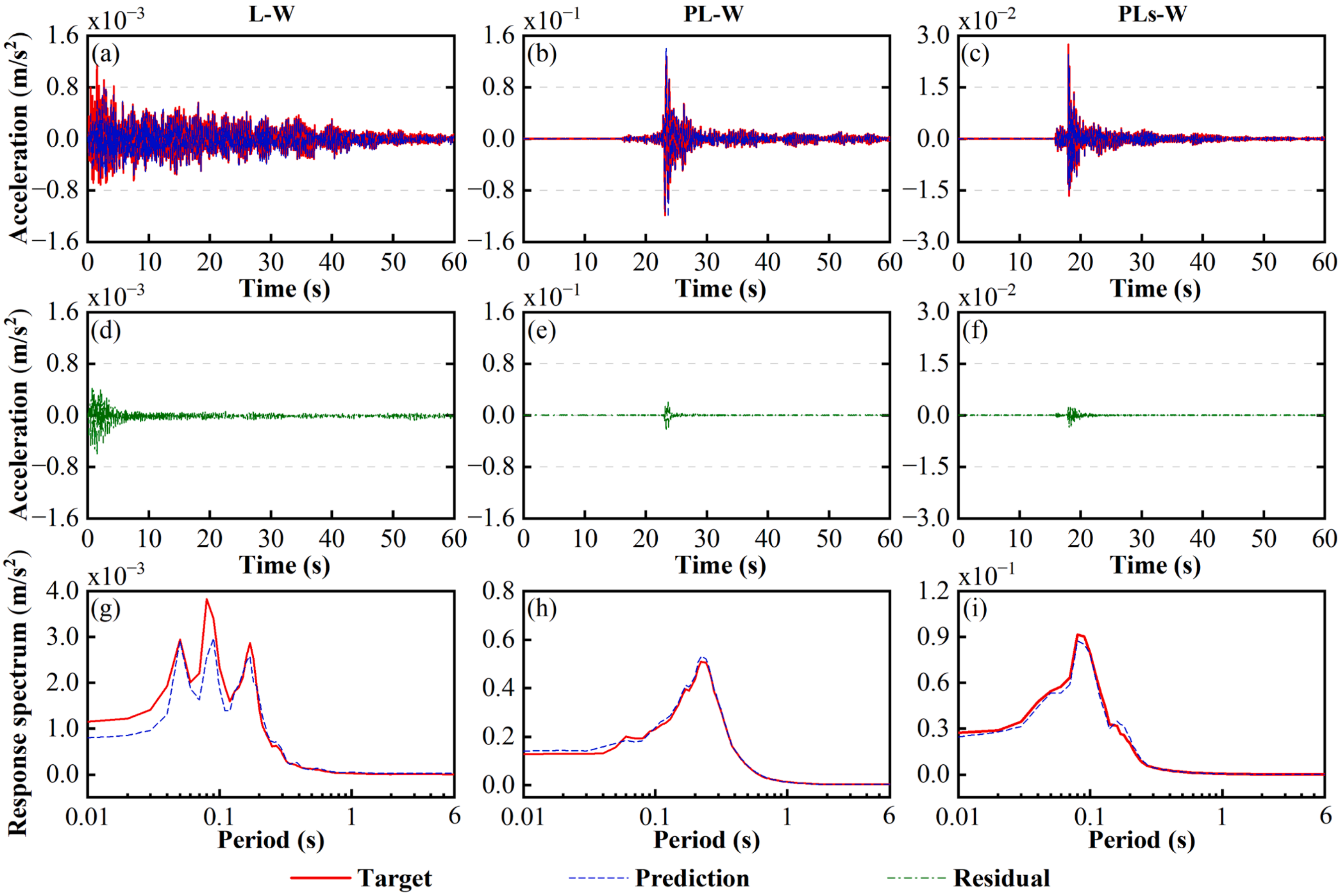


Fig. 8. Worst numerical-data cases for three models: (a–c) acceleration time histories, (d–f) residuals, and (g–i) response spectra.

Table 3
Peak prediction errors for representative samples from three models.

Sample ID	Target Peak (m/s^2)	Predicted Peak (m/s^2)	Absolute Error (m/s^2)	Relative Error (%)
L-B	4.16×10^{-3}	4.16×10^{-3}	1.02×10^{-6}	0.03
L-W	1.15×10^{-3}	8.02×10^{-4}	3.45×10^{-4}	30.06
PL-B	1.35×10^{-2}	1.35×10^{-2}	1.05×10^{-5}	0.03
PL-W	1.28×10^{-1}	1.40×10^{-1}	1.21×10^{-2}	9.38
PLs-B	1.86×10^{-2}	1.85×10^{-2}	4.00×10^{-6}	0.01
PLs-W	2.74×10^{-3}	2.46×10^{-3}	2.79×10^{-4}	10.23

commutation error between the additive and multiplicative operations is negligible, and the fixed application order does not introduce measurable bias.

After augmentation, all samples are combined with the original data to create an expanded training set. Velocity and displacement are computed through numerical integration for input to the PINN framework. This approach introduces realistic variability while maintaining physical consistency and improves model robustness for recorded ground motion prediction. Each perturbation type targets a general characteristic of earthquake recordings: instrumental noise, source-magnitude uncertainty, propagation-path variability, and baseline drift. The strategy design is network-independent, although the specific parameter values (e.g., σ_*n* range, scaling bounds) were calibrated for KiK-net and may require adjustment for networks with different signal-to-noise profiles.

5.2. *Recorded-Data validation*

Given the increased complexity of real earthquake records, modified training parameters were employed: 10,000 epochs, batch size of 150, and hidden layer size of 64. Table 4 presents a comprehensive performance comparison across four model configurations.

Table 4 shows that both data augmentation and physics-informed constraints individually improve model performance relative to the baseline LSTMs model. The baseline model achieves an 88.09% confidence level and 56.65% correlation. The LSTMs-E model reaches an 88.76% confidence level with MAE and RMSE reductions of 12.4% and 6.4%, demonstrating that data augmentation enhances robustness by exposing the network to realistic variations in recorded earthquake data. The PhyLSTMs model achieves a comparable 88.12% confidence level with more substantial error reductions of 13.5% for MAE and 10.3% for RMSE, indicating that embedding kinematic consistency constraints guides the model toward physically plausible predictions and reduces systematic errors in temporal evolution.

The combined PhyLSTMs-E configuration achieves the best overall performance, with an 88.84% confidence level and a 66.16% correlation coefficient, representing a 16.8% improvement over the baseline. Error

Table 4
Performance comparison of different model configurations on recorded ground motion dataset.

Model	CI95% (%)	MeanPE (%)	R (%)	MAE (× 10^{-3})	RMSE (× 10^{-1})
LSTMs	88.09	32.54	56.65	1.93	9.05
LSTMs-E	88.76	28.59	59.52	1.69	8.47
PhyLSTMs	88.12	30.51	58.55	1.67	8.12
PhyLSTMs-E	88.84	25.02	66.16	1.56	7.66

Note: The suffix "-E" denotes models enhanced with data augmentation strategies.

reductions reach 19.2% for MAE and 15.4% for RMSE, while the average peak error decreases to 25.02%, the lowest among all configurations. These results confirm that data augmentation and physics-informed constraints address complementary aspects: the former enhances robustness to data variability, while the latter ensures adherence to physical laws. Such combination leads to synergistic benefits in model performance. Compared to the 97.98% confidence level on numerical data, the lower performance on recorded data reflects the inherent complexity of real earthquake records. Fig. 9 provides a detailed statistical analysis of prediction errors and correlations across all test samples.

Fig. 9(a) shows normalized errors within the 95% confidence interval. All models are centered near zero with light tails, indicating minimal bias and few extreme errors; PhyLSTMs-E is the narrowest and most concentrated near zero, followed by LSTMs-E and PhyLSTMs, implying increasing stability. Fig. 9(b) presents correlation-coefficient distributions: LSTMs has the widest spread (mainly 0.4–0.7), LSTMs-E and PhyLSTMs shift density toward 0.6–0.8, and PhyLSTMs-E is most concentrated in 0.7–0.9 with clustering near R = 0.8, indicating the best waveform fidelity and temporal consistency. To illustrate these statistical trends with specific examples, Fig. 10 examines the best and worst prediction cases.

Fig. 10 compares the best and worst cases selected by peak error on the test set using time histories and response spectra. The best case (R-B) yields time-domain and spectral correlations of 0.894 and 0.993, respectively. The worst prediction case (R-W) shows a lower time-domain correlation of 0.823. Still, it maintains a high response spectrum correlation of 0.986, indicating that although some waveform details differ, the frequency content remains well captured. Across all 100 test samples, the average response spectrum correlation reaches 0.951, significantly surpassing the average time-domain correlation of 0.662. This demonstrates that the PhyLSTMs-E model reliably predicts frequency content and spectral characteristics essential for engineering applications, providing dependable seismic demand parameters for structural design and performance evaluation.

The recorded events at IBRH13 span PGA values of 1–100 gal, corresponding to weak-to-moderate shaking and low-to-intermediate strain levels. Deeper layers at this site have high shear-wave velocities (Vs > 1 000 m/s for Layers 5–7 in Table 1), so the observed nonlinearity remains modest. The numerical dataset, generated with the MKZ constitutive model in DEEPSOIL, incorporates shear-modulus degradation and hysteretic damping. The high prediction accuracy on that dataset (Table 2) confirms that the network can capture nonlinear input-output mappings when they are present in the training data. Strong-motion events with PGA exceeding 100 gal and the associated large-strain nonlinearity fall outside the validation range of this study and are discussed as a priority direction in Section 6.

The gap between the average time-domain correlation (R = 0.662) and the response-spectrum correlation (0.951) reflects the inherent difficulty of matching phase content in real recordings. Time-domain R is sensitive to small phase shifts: a timing offset of a single sample (0.01 s at 100 Hz) can substantially reduce R while leaving the spectral shape nearly unchanged. Three-dimensional wave-propagation effects, scattering, and residual instrument noise introduce phase variability that a one-dimensional transfer model cannot fully reproduce. From an engineering standpoint, seismic design codes rely on response spectra rather than point-by-point waveform matching, so the high spectral correlation is the more relevant metric for practical applications.

5.3. Sensitivity analysis of phylstms-e

Two ablation studies were conducted to examine the accuracy-efficiency trade-off: the effect of augmentation factor K at fixed folding granularity F = 5, and the impact of folding granularity F at K = 1 without augmentation. Tables 5a and 5b present the computational cost and accuracy metrics for different parameter configurations.

Table 5(a) shows that model accuracy improves progressively as the augmentation factor increases from K = 1 to K = 5. K = 5 achieves the lowest RMSE of 0.766 and the highest correlation coefficient of 66.16%, representing the optimal accuracy configuration. K = 3 attains nearly identical RMSE of 0.766 while reducing epoch time by 38% from 3.30 s to 2.05 s, offering an attractive balance between accuracy and computational efficiency. Beyond K = 5, further increases yield diminishing returns or slight performance degradation, indicating that excessive data augmentation may introduce noise. K = 5 is preferable when maximum waveform fidelity is required, while K = 3 provides an efficient alternative for resource-constrained applications.

Table 5b presents an ablation study on folding granularity, where F = 1 serves as the unfolded baseline. Data folding accelerates training substantially: F = 5 achieves an epoch time of 0.66 s versus 13.81 s at F = 1, a 21-fold speedup. Accuracy remains stable for moderate folding levels ($F \le 8$): CI95% ranges from 88.12% to 88.61% and RMSE from 0.797 to 0.812, with the unfolded baseline and the folded configurations showing comparable prediction quality. This stability confirms that the fold-unfold pipeline preserves the information content of the original sequence, consistent with the theoretical analysis in Section 2.2. Performance deteriorates at F = 20, where CI95% drops to 87.92% and RMSE rises to 0.840. At this level the LSTM processes only $l/20$ time steps (600 for the 12 000-sample recorded dataset), which limits the number of sequential gate updates available to capture long-range temporal dependencies. The correspondingly high per-step feature dimension ($d \times 20$) also increases the input-to-hidden mapping burden relative to the fixed hidden-layer size. An optimal folding factor balances temporal resolution against per-step feature load; F = 5 lies within this favorable range, reducing computational cost while maintaining prediction quality. The combined setting of F = 5 and K = 5 achieves a balanced accuracy-efficiency profile that supports practical seismic response prediction. Cross-architecture experiments with CNN-

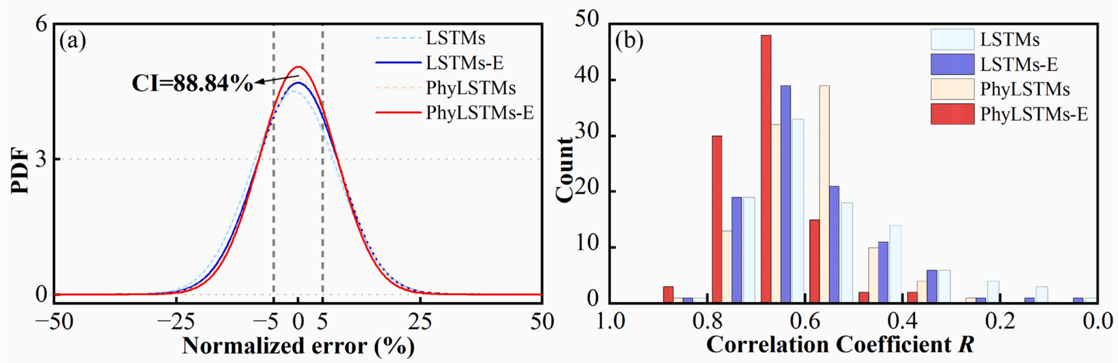


Fig. 9. Performance evaluation on recorded dataset: (a) normalized error distribution and (b) correlation analysis for four model configurations.

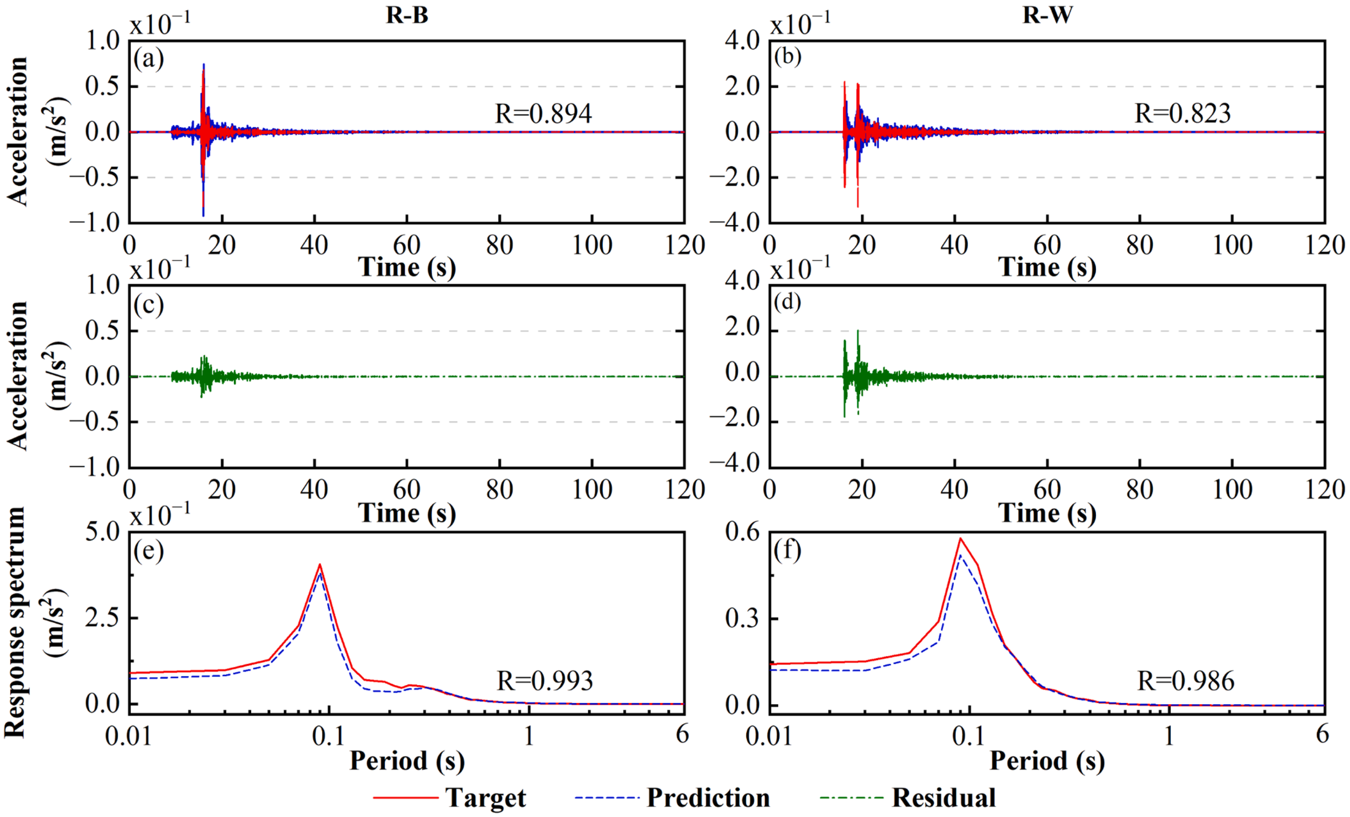


Fig. 10. Best and worst recorded-data cases for PhyLSTMs-E: (a,b) time histories, (c,d) residuals, and (e,f) response spectra.

Table 5a
Effect of augmentation factor K at F = 5 (PhyLSTMs-E).

K	Time per Epoch (s)	CI95% (%)	AvgPE (%)	R (%)	MAE (× 10^{-3})	RMSE (× 10^{-1})
1	0.66	88.12	30.51	58.55	1.67	8.12
2	1.35	88.71	24.99	62.72	1.59	7.87
3	2.05	88.84	23.50	65.04	1.57	7.66
5	3.30	88.84	25.02	66.16	1.56	7.66
7	4.45	88.90	26.40	64.95	1.58	7.70
10	6.76	88.81	24.62	65.70	1.61	7.74

Table 5b
Effect of folding granularity F at K = 1 (PhyLSTMs).

F	Time per Epoch (s)	CI95% (%)	AvgPE (%)	R (%)	MAE (× 10^{-3})	RMSE (× 10^{-1})
1	13.81	88.51	28.27	61.27	1.61	7.97
2	7.93	88.61	28.74	60.90	1.59	7.97
3	0.97	88.29	32.90	59.48	1.65	8.04
5	0.66	88.12	30.51	58.55	1.67	8.12
8	0.48	88.32	31.67	59.47	1.64	8.03
10	0.41	88.24	32.65	59.06	1.69	8.07
20	0.32	87.92	41.21	54.87	1.75	8.40

Transformer, CNN-LSTM, and Frequency-Domain LSTM backbones further confirm that the framework is architecture-agnostic (Appendix B).

6. Conclusions

This research presents a physics-informed deep LSTM framework for site seismic response prediction. The method incorporates kinematic derivative relationships as soft constraints and employs a data folding module to accelerate training. A noise-aware data augmentation strategy addresses measurement noise and signal variability in recorded data. The framework was validated on both numerically simulated and recorded earthquake data from the IBRH13 station of the KiK-net network under weak-to-moderate shaking conditions (PGA 1–100 gal), yielding the following findings.

(1) Embedding kinematic derivative constraints in the loss function improves physical consistency and reduces MAE and RMSE by more than 12% on numerical data, with the most pronounced gains in extreme-case peak predictions where worst-case relative error drops from 30% to below 11%.
(2) The data folding strategy accelerates training by more than twenty times on recorded data without sacrificing accuracy, as confirmed by ablation studies showing that the fold-unfold pipeline preserves full waveform information.
(3) On recorded ground motions, the combined framework with noise-aware augmentation achieves a response spectrum correlation of 0.951 across test samples, demonstrating reliable frequency content prediction for engineering applications.
(4) Data folding, noise-aware augmentation and kinematic constraints are architecture-independent and readily transferable to other sequence models.

The proposed framework provides an efficient and physically consistent solution for site-response prediction. Validation is currently limited to a single KiK-net station under weak-to-moderate shaking; future work will extend the approach to multiple stations with diverse stratigraphy and stronger ground motions, and will further enhance physical fidelity by embedding nonlinear soil constitutive relationships into the loss function.

Funding information

National Natural Science Foundation of China, Grant/ Number: 42377140

Declaration of generative AI and ai-assisted technologies in the writing process

During the preparation of this work, the authors used Claude4 to improve the language. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Code availability

The data and codes used in this paper will be publicly available on GitHub at https://github.com/yxwuhhu/PhyLSTMs after the paper is published.

CRediT authorship contribution statement

Yongxin Wu: Writing – original draft, Conceptualization, Methodology, Resources, Funding acquisition. **Zhanpeng Yin:** Writing – original draft, Formal analysis, Data curation, Software, Validation, Visualization. **Juncheng Wang:** Methodology, Investigation, Validation. **Yue Hou:** Conceptualization, Writing – review & editing, Visualization. **Haodong Shang:** Writing – review & editing, Investigation, Supervision. **Houle Zhang:** Formal analysis, Data curation, Visualization.

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.

Acknowledgments

The authors would like to acknowledge the support of the National Natural Science Foundation of China under Grant No. 42377140.

Appendix A. Full test-set relative peak error distributions

Fig. A1 summarizes the relative peak error distributions over the full numerical test set (100 samples). The median relative peak errors for LSTM, PhyLSTM, and PhyLSTMs are 1.28%, 0.95%, and 0.93%, respectively, indicating that both physics-informed variants achieve slightly lower typical peak-error levels than the purely data-driven LSTM. Notably, the LSTM results exhibit a much heavier upper tail with several large outliers; the maximum relative peak error reaches 30.06% (corresponding to the L–W case reported in Table 3). In contrast, the physics-informed models substantially suppress extreme peak errors, reducing the maximum to 9.38% for PhyLSTM and 10.23% for PhyLSTMs. The representative samples listed in Table 3 are selected to align with the distribution extremes (minimum/maximum), ensuring they reflect the full performance range rather than being selectively chosen.

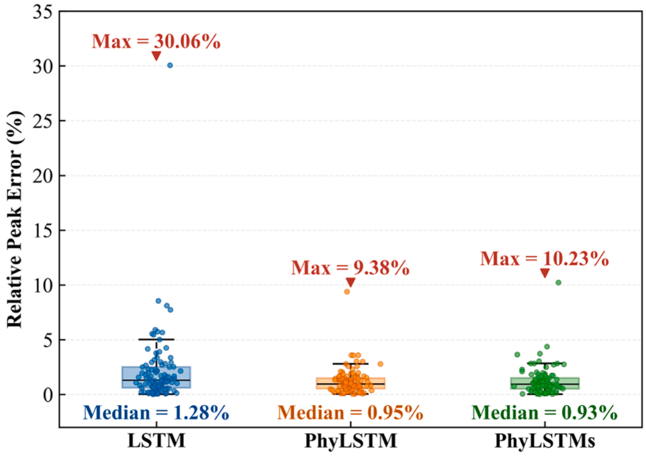


Fig. A1. Boxplot comparison of relative peak errors across all 100 numerical test samples for LSTM, PhyLSTM, and PhyLSTMs models.

Appendix B. Cross-architecture comparison

To evaluate the generalizability of the proposed framework, we benchmarked three alternative backbone architectures against the LSTM baseline on the recorded dataset: CNN-Transformer (1D CNN + Transformer encoder), CNN-LSTM (1D CNN + LSTM), and frequency-domain LSTM (FD-LSTM). All models were identically configured using the proposed Data Folding mechanism and kinematic constraints (F = 5, hidden size 64, Adam optimizer, lr=2 × 10^{-4}). The "-E" suffix denotes models incorporating the 5 × data augmentation strategy.

Table B1
Cross-architecture comparison on the recorded dataset.

Model	MAE ($\times 10^{-3}$)	RMSE ($\times 10^{-1}$)	R (%)	Training Time (s)
LSTM	1.63	8.05	59.96	6279
LSTM-E	1.56	7.66	66.16	33,000
CNN-LSTM	1.66	7.93	62.15	8366
CNN-LSTM-E	1.62	7.70	64.65	38,465
CNN-Trans	1.89	9.55	41.07	75,434
FD-LSTM	2.31	9.79	30.19	4473

As shown in Table B1, both the LSTM and CNN-LSTM achieve competitive accuracy among the non-augmented configurations. Furthermore, applying data augmentation consistently improves performance for both architectures, with CNN-LSTM-E closely approaching the baseline LSTM-E. This successful cross-architecture implementation demonstrates that the proposed Data Folding, noise-aware augmentation, and kinematic constraints are not strictly reliant on a specific backbone. Instead, they are inherently architecture-independent and readily transferable to other suitable sequence models.

However, this transferability requires an algorithmically compatible foundation. The CNN-Transformer significantly underperforms because its $O(n^2)$ attention mechanism struggles to generalize on the limited 800-sample dataset. Similarly, the FD-LSTM yields the lowest accuracy, as its global FFT operation assumes signal stationarity, which fundamentally conflicts with the non-stationary nature of seismic ground motions.

References

Ansal, A., Tönük, G., & Sadeghzadeh, S. (2024). Site response analysis in performance based approach[J]. *Soil Dynamics and Earthquake Engineering, 178*, Article 108480.

Aoi, S. (2000). New strong-motion observation network, KiK-net. *Eos Transactions AGU, 81*, F863.

Bantis, J., Miranda, E., & Heresi, P. (2025). Simplified site response analysis for regional seismic risk assessments. *Soil Dynamics and Earthquake Engineering, 188*, Article 109022.

Bouckovalas, G. D., Tsiapas, Y. Z., Zontanou, V. A., & Kalogeraki, C. G. (2017). Equivalent linear computation of response spectra for liquefiable sites: The spectral envelope method. *Journal of Geotechnical and Geoenvironmental Engineering, 143*(4), Article 04016115.

Chen, Z., Huang, D., & Wang, G. (2023). Large-scale ground motion simulation of the 2016 Kumamoto earthquake incorporating soil nonlinearity and topographic effects. *Earthquake Engineering & Structural Dynamics, 52*(4), 956–978.

Chen, S. Z., Feng, D. C., & Taciroglu, E. (2024). Prior knowledge-infused neural network for efficient performance assessment of structures through few-shot incremental learning. *Computer-Aided Civil and Infrastructure Engineering, 39*(13), 1928–1945.

Chen, S., Hu, X., Jiang, W., Wang, S., Chen, X., & Li, X. (2025). Data-physical fusion deep learning for site seismic response using KiK-net records. *Earthquake Engineering & Structural Dynamics, 54*(3), 993–1008.

Ciancimino, A., Foti, S., & Lanzo, G. (2018). Stochastic analysis of seismic ground response for site classification methods verification. *Soil Dynamics and Earthquake Engineering, 111*, 169–183.

Fayaz, J., & Galasso, C. (2023). A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. *Computer-Aided Civil and Infrastructure Engineering, 38*(1), 87–103.

Fayaz, J., Xiang, Y., & Zareian, F. (2021). Generalized ground motion prediction model using hybrid recurrent neural network. *Earthquake Engineering & Structural Dynamics, 50*(6), 1539–1561.

Fonseca, D. J., Navaresse, D. O., & Moynihan, G. P. (2003). Simulation metamodeling through artificial neural networks. *Engineering Applications of Artificial Intelligence, 16* (3), 177–183.

Ganji, H. T., & Seylabi, E. (2023). An LSTM RNN proposal for surrogate modeling the dynamic response of buried structures to earthquake plane waves in soil half-spaces. *Computers and Geotechnics, 164*, Article 105796.

Gu, H., Li, Y., Fang, Y., Wang, Y., Yu, Y., Wei, Y., … Chen, Y. (2025). Environmental-aware deformation prediction of water-related concrete structures using deep learning. . *Computer-Aided Civil and Infrastructure Engineering, 40(15), 2130–2151. Computer-Aided Civil and Infrastructure Engineering,, 40*(15), 2130–2151.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hu, Y., Tsang, H. H., Lam, N., & Lumantarna, E. (2023). Physics-informed neural networks for enhancing structural seismic response prediction with pseudo-labelling. *Archives of Civil and Mechanical Engineering, 24*(1), 7.

Hu, J., Wen, W., Zhang, C., Zhai, C., Pei, S., & Wang, Z. (2024). Rapid peak seismic response prediction of two-story and three-span subway stations using deep learning method. *Engineering Structures, 300*, Article 117214.

Huang, D., Sun, P., Jin, F., & Du, C. (2021). Topographic amplification of ground motions incorporating uncertainty in subsurface soils with extensive geological borehole data. *Soil Dynamics and Earthquake Engineering, 141*, Article 106441.

Huhn, Q. A., Tano, M. E., Ragusa, J. C., & Choi, Y. (2023). Parametric dynamic mode decomposition for reduced order modeling. *Journal of Computational Physics, 475*, Article 111852.

Khosravikia, F., & Clayton, P. (2021). Machine learning in ground motion prediction. *Computers & Geosciences, 148*, Article 104700.

Kim, T., Kwon, O. S., & Song, J. (2024). Deep learning-based response spectrum analysis method for building structures. *Earthquake Engineering & Structural Dynamics, 53*(4), 1638–1655.

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), arXiv:1412.6980.

Kuok, S. C., & Yuen, K. V. (2020). Broad learning for nonparametric spatial modeling with application to seismic attenuation. *Computer-Aided Civil and Infrastructure Engineering, 35*(3), 203–218.

Kuok, S. C., & Yuen, K. V. (2021). Broad bayesian learning (BBL) for nonparametric probabilistic modeling with optimized architecture configuration. *Computer-Aided Civil and Infrastructure Engineering, 36*(10), 1270–1287.

Li, L., Huang, X., Chen, S., Wu, T., Mei, L., Long, W., & Xiao, Y. (2023). Study on strategies for reducing training samples for accurate estimation of wind-induced structural response of LSTM networks. *Journal of Wind Engineering and Industrial Aerodynamics, 238*, Article 105421.

Li, L., Jin, F., Huang, D., & Wang, G. (2023). Soil seismic response modeling of KiK-net downhole array sites with CNN and LSTM networks. *Engineering Applications of Artificial Intelligence, 121*, Article 105990.

Li, L., Jin, F., Huang, D., & Wang, G. (2024). Multi-input integrative neural network for soil seismic response modeling at KiK-net downhole array sites. *Earthquake Engineering & Structural Dynamics, 53*(10), 3165–3183.

Li, Y., Wang, R., Ma, H., & Zhang, J. M. (2025). Rising groundwater table due to restoration projects amplifies earthquake induced liquefaction risk in Beijing. *Nature Communications, 16*(1), 1466.

Lu, X., Xu, Y., Tian, Y., Cetiner, B., & Taciroglu, E. (2021). A deep learning approach to rapid regional post-event seismic damage assessment using time-frequency distributions of ground motions. *Earthquake Engineering & Structural Dynamics, 50*(6), 1612–1627.

Lu, H., Zhang, N., Pan, J., Zhang, Y., & Dai, D. (2025). Dynamic interaction between an asymmetric V-shaped canyon and a nearby structure under oblique incident SH waves. *Soil Dynamics and Earthquake Engineering, 190*, Article 109178.

Mei, X., Olson, S. M., & Hashash, Y. M. (2020). Evaluation of a simplified soil constitutive model considering implied strength and pore-water pressure generation for one-dimensional (1D) seismic site response. *Canadian Geotechnical Journal, 57*(7), 974–991.

Meite, R., Wotherspoon, L., Kaklamanos, J., McGann, C. R., & Hayden, C. (2020). Sensitivity of 1-D ground motion predictions to analysis codes and material models using KiK-net vertical arrays. *Soil Dynamics and Earthquake Engineering, 133*, Article 106113.

Pak, H., & Paal, S. G. (2025). A real-time structural seismic response prediction framework based on transfer learning and unsupervised learning. *Engineering Structures, 323*, Article 119227.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems, 32*, 8024–8035.

Qiu, Z., Prabhakaran, A., Zhou, Y. G., & Elgamal, A. (2023). A practical three-dimensional plasticity model for cyclic degradation of soil in earthquake loading applications. *Earthquake Engineering & Structural Dynamics, 52*(12), 3835–3852.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, Article 132306.

Thompson, E. M., Baise, L. G., Tanaka, Y., & Kayen, R. E. (2012). A taxonomy of site response complexity. *Soil Dynamics and Earthquake Engineering, 41*, 32–43.

Tian, Y., Liu, S., Chen, S., & Xu, Z. (2025). A computational framework for site-city interaction analysis considering seismic wave propagation effects and its application to Beijing CBD. *Soil Dynamics and Earthquake Engineering, 196*, Article 109492.

Tsai, C. C., & Li, P. C. (2024). Quantifying near-fault motion effects on soil liquefaction through effective stress site response analysis. *Soil Dynamics and Earthquake Engineering, 183*, Article 108779.

Wang, T., Li, H., Noori, M., Ghiasi, R., Kuok, S. C., & Altabey, W. A. (2023). Seismic response prediction of structures based on Runge-Kutta recurrent neural network with prior knowledge. *Engineering Structures, 279*, Article 115576.

Wu, Y., Yin, Z., Gao, Y., Yang, S., & Hou, Y. (2025). Constitutive model-constrained physics-informed neural networks framework for nonlinear structural seismic response prediction. *Computer Methods in Applied Mechanics and Engineering, 443*, Article 118079.

Xu, S., & Noh, H. Y. (2021). PhyMDAN: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning. *Mechanical Systems and Signal Processing, 151*, Article 107374.

Yan, J., & Zhang, Y. (2023). In Y. Yang, S. N. Raman, B. Yuan, & Z. Xu (Eds.), *Advances in frontier research on engineering structures, 2* pp. 117–127). CRC Press.

Yi, F. (2010). Nonlinear cyclic characteristics of soils. In D. O. Fratta, A. J. Puppala, & B. Muhunthan (Eds.), GeoFlorida 2010: Advances in analysis, modeling & design (pp. 2601–2610). ASCE.

Zhang, R., Chen, Z., Chen, S., Zheng, J., Büyüköztürk, O., & Sun, H. (2019). Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures, 220*, 55–68.

Zhang, R., Liu, Y., & Sun, H. (2020). Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling. *Engineering Structures, 215*, Article 110704.

Zhang, R., Liu, Y., & Sun, H. (2020). Physics-informed multi-LSTM networks for metamodeling of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering, 369*, Article 113226.

Zhang, N., Lu, H., Gao, Y., Dai, D., Zhang, Y., & Pak, R. Y. (2023). Effect of a symmetric V-shaped canyon on the seismic response of an adjacent building under oblique incident SH waves. *Earthquake Engineering & Structural Dynamics, 52*(6), 1861–1883.

Zhang, N., Wang, L., Zhang, Y., Lu, H., Dai, D., Cai, G., & Gao, Y. (2024). Effect of a V-shaped canyon on the seismic response of a bridge under oblique incident SH waves. *Earthquake Engineering & Structural Dynamics, 53*(1), 496–514.

Zhang, X., Xie, X., Tang, S., Zhao, H., Shi, X., Wang, L., … Xiang, P. (2024). High-speed railway seismic response prediction using CNN-LSTM hybrid neural network. *Journal of Civil Structural Health Monitoring, 14*(5), 1125–1139.

Zhong, Q. M., Feng, D. C., & Chen, S. Z. (2025). Multi-fidelity enhanced few-shot time series prediction model for structural dynamics analysis. *Computer Methods in Applied Mechanics and Engineering, 434*, Article 117583.

Zhu, Y., Zabaras, N., Koutsourelakis, P. S., & Perdikaris, P. (2019). Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics, 394*, 56–81.