Supplementary Information

A Physics-Informed Foundation Model for Rapid High-Fidelity Structural Response Prediction

Shiqiao Meng[1], Ying Zhou[1]*, Bingxu Liao[1], Mushi Chang[1], Tianshu Zhang[1], Abouzar Jafari[1], Abderrahim Djerrad[1]

[1]State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China
*yingzhou@tongji.edu.cn

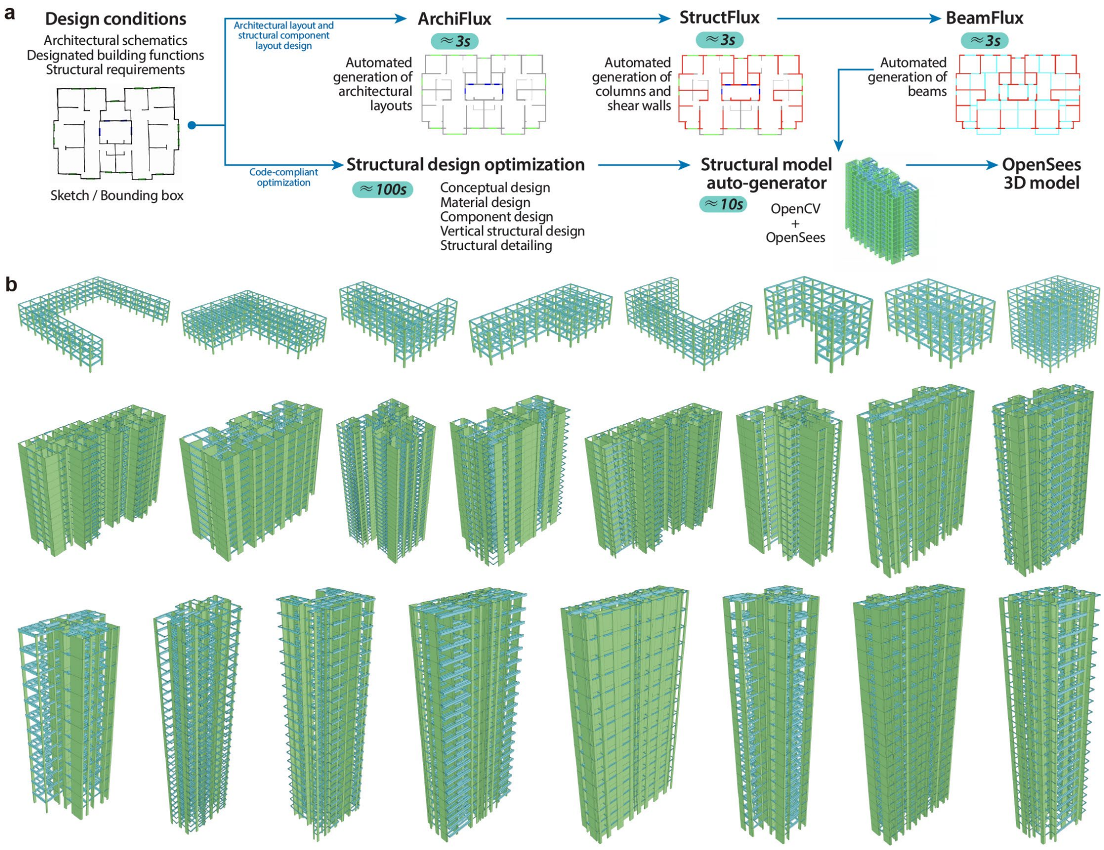
1. AI-Driven Structural Design and Automated Finite Element Model Generation

We introduce an automated pipeline for structural design and finite element model generation, integrating generative modeling, computer vision, and code-compliant optimization. The process begins with user-defined constraints-floor area, number of stories, functional zones, and seismic intensity—which are encoded into spatial prompts. These are processed by the Flux backbone, a rectified flow generative model augmented with ControlNet for conditioning, and fine-tuned into three specialized variants: ArchiFlux, StructFlux, and BeamFlux.

ArchiFlux synthesizes architectural layouts with semantic segmentation; StructFlux infers column and wall configurations; BeamFlux completes beam networks based on connectivity requirements. The models operate hierarchically, ensuring alignment between architectural intent and structural logic. Over 270,000 layouts were generated using a dataset of annotated plans and trained with RGB-encoded labels representing structural classes.

Post-processing converts color-encoded images into vectorized primitives via skeletonization, topological reconstruction, and chromatic classification. Structural elements are extracted as polylines, classified, and assembled into connected graphs.

These layouts are transformed into reinforced concrete finite element models through a rules-based design engine incorporating Chinese seismic codes[1,2]. Member dimensions, reinforcement ratios, and detailing are assigned analytically and filtered by strength, ductility, and drift criteria. Models failing code checks undergo a multi-stage optimization, adjusting geometry, materials, and configuration. Seismic compliance is verified through nonlinear time-history analysis. Persistent violations lead to rejection, while compliant designs are archived with full geometric, material, and seismic metadata. This pipeline enables high-throughput generation of code-compliant, structurally realistic models for large-scale simulation and learning. The entire generative workflow is schematically illustrated in Supplementary Figure 1.



Supplementary Figure 1 | AI-driven workflow for structural design, automated FE modeling, and synthetic building generation. Schematic of the end-to-end pipeline for AI-based structural design and automated finite element model generation. The process starts with the definition of design conditions, functional requirements, and structural constraints. Based on these inputs, three specialized deep learning models—ArchiFlux, StructFlux, and BeamFlux—were sequentially employed to generate the architectural layout (including doors, windows, and partitions), the spatial distribution of shear walls and columns, and the beam configuration, respectively. These models were fine-tuned from the Flux architecture using ControlNet, trained on a semantically annotated dataset derived from over 200 real-world architectural drawings, where structural elements were labeled by type using color-coded masks. ArchiFlux takes as input a conceptual sketch or spatial boundary and outputs the floorplan with openings and wall placements. StructFlux builds upon this to determine the arrangement of primary vertical load-resisting elements. BeamFlux generates beam layouts, with outputs post-processed using OpenCV to extract precise geometric placements. Detailed structural design follows, including material specification, reinforcement detailing, and optimization based on inter-story drift constraints. Iterative evaluations ensure compliance with performance criteria. The final structural system was automatically converted into a three-dimensional FE model using the Structural Model Auto-Generator, a tool developed by the authors, which integrates the generated layout with OpenSees for numerical analysis. The synthetic building dataset comprising tens of thousands of FE models used in this study was generated through this fully automated pipeline. **b**, Representative visualizations of diverse building structures within the synthetic building dataset. These samples illustrate the architectural and structural variability captured by the generative pipeline, encompassing a wide range of typologies.

1.1 Generation of Architectural Floor Plans and Structural Layouts Using Flux and ControlNet

This section presents the architectural framework and training strategy used to implement a generative structural design pipeline comprising the latent-diffusion backbone (Flux), its spatial conditioning extension (ControlNet), and three task-specific variants—ArchiFlux, StructFlux, and BeamFlux. Together, these models enable hierarchical structural layout synthesis, from schematic floor plan generation to detailed member-level configuration, using a compact expert-labelled dataset of architectural drawings. The generated layouts support the development of 270,000 structurally valid building models, which form the basis of all subsequent nonlinear finite-element simulations.

1.1.1 Rectified flow formulation of Flux

Let $\mathbf{x}_t \in \mathbb{R}^{C\times H\times W}$ denote the latent representation of a floor plan image at continuous time $t \in [0,1]$. In Flux model, synthesis is posed as a rectified flow problem rather than a denoising-diffusion process. A single noise sample $\mathbf{x}_t \sim \mathcal{N}(0,I)$ is transported to the data manifold at t=0 by integrating the ordinary differential equation

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{v}_\theta(x_t, t, c), \qquad \mathbf{x}_{t=1} \sim \mathcal{N}(0,I) \tag{1}$$

where $\mathbf{v}_\theta$ is a conditional velocity field parameterized by θ and driven by conditioning information c (text embeddings, spatial priors, or layout constraints)[3].

Flux is trained with the flow-matching loss, which aligns the predicted velocity with a reference velocity $\bar{\mathbf{v}}(x_t, t)$ derived from probability-flow theory:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{t\sim\mathcal{U}[0,1],\mathbf{x}_0\sim p_{\mathrm{data}}} \|\mathbf{v}_\theta(\mathbf{x}_t, t, c) - \bar{\mathbf{v}}(\mathbf{x}_t, t)\|_2^2 \tag{2}$$

where $\mathbf{x}_t$ is obtained by analytically sampling along the reference flow connecting $\mathbf{x}_1$ and $\mathbf{x}_0$. No denoiser or β_t variance schedule is required.

At test time, images are generated by discretizing Equation (1) with a fixed-step forward-Euler solver:

$$\mathbf{x}_{t-\Delta t} = \mathbf{x}_t - \Delta t \mathbf{v}_\theta(\mathbf{x}_t, t, c), \qquad \Delta t = \frac{1}{N} \tag{3}$$

where N is typically 10–50. Classifier-free guidance is applied by evaluating both conditional and unconditional velocities and linearly combining them to modulate prompt fidelity. This rectified-flow formulation eliminates the diffusion forward process and the noise-prediction objective altogether, enabling Flux 1 to achieve high-resolution, text-conditioned image synthesis with an order-of-magnitude fewer function evaluations than conventional denoising-diffusion models.

1.1.2 The model architecture of Flux

The generative core of Flux is a high-capacity Transformer architecture, FluxTransformer2D, which models conditional vector fields over compressed latent spatial grids. The network operates on a resolution of 64 × 64 latent patches, each corresponding to a 16×16 region in pixel space, and adopts a U-Net-style topology with deep hierarchical attention, parameter-efficient normalization, and cross-modal conditioning mechanisms. The model architecture consists of three principal stages—input embedding, a dual-stream encoder, and a single-stream fusion block—all implemented using transformer layers augmented with convolutional projections and residual attention.

Latent patch embedding: The input latent tensor $\mathbf{x}_t \in \mathbb{R}^{C\times H\times W}$, typically initialized from standard Gaussian noise at t=1, is first passed through a 1×1 convolutional projection layer that maps it into a 64-channel token space. This serves as the initial representation for subsequent attention-based processing.

Dual-stream encoder: The first stage comprises 19 Transformer blocks, each composed of residual self-attention and feedforward layers, collectively referred to as double stream blocks. Spatial positional encoding is implemented using rotary positional embedding[4] (RoPE) over a canonical patch grid (16,56,56), which is critical for maintaining alignment constraints and local symmetry—particularly important in structured design tasks such as architectural layout synthesis.

Single-stream fusion: To facilitate interaction between visual latents and textual conditions, the second stage consists of 38 Transformer blocks that integrate the image stream with textual embeddings via cross-attention. Prior to fusion, textual embeddings from CLIP[5] (ViT-L/14) and T5[6] encoders are projected into a shared 768-dimensional space and concatenated along the sequence dimension. Cross-modal joint attention is then computed over the fused sequence, with a model-wide joint embedding dimension of 4096.

Each Transformer block comprises a 24-head self-attention mechanism with 128-dimensional heads, followed by a feedforward network using SwiGLU[7] activation and a 4× expansion ratio (≈8192 units). RMSNorm[8] is applied after both attention and multi-layer perceptron (MLP) layers to ensure stable optimization and efficient gradient propagation at scale.

1.1.3 Spatial conditioning with ControlNet

To guide the generative process with structured spatial constraints, a ControlNet[9] module is integrated at each resolution stage. This mechanism adds a parallel, trainable path to the Flux backbone, enabling fine-grained control without modifying the original network weights. Specifically, a frozen encoder $\mathcal{F}_{locked}$ is combined with a trainable duplicate $\mathcal{F}_{train}$, modulated through a zero-initialized 1×1 convolution layer $\mathbf{W}_0$:

$$\mathbf{y} = \mathcal{F}_{\text{locked}}(\mathbf{x}_t) + \mathbf{W}_0 * \mathcal{F}_{\text{train}}(\mathbf{s}) \tag{4}$$

where $\mathbf{s}$ represents the spatial condition input (e.g., zoning boundaries or element masks). At inference, classifier-free guidance is used to adjust conditioning strength:

$$\boldsymbol{\varepsilon}_\theta^{\text{cfg}} = \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, \emptyset, t) + g[\boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, c, t) - \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, \emptyset, t)] \tag{5}$$

with g controlling the trade-off between fidelity and constraint adherence.

1.1.4 Hierarchical domain adaptation: ArchiFlux, StructFlux, and BeamFlux

A core methodological contribution of this work is the hierarchical adaptation of the Flux rectified-flow model for automated structural layout generation, realised through three task-specific variants: ArchiFlux, StructFlux, and BeamFlux. These models were fine-tuned from a shared Flux.1 backbone using a domain-specific dataset comprising 232 semantically annotated architectural drawings at 1024 × 1024 resolution. Each drawing encodes six structural and architectural categories via a fixed colour mapping: white (background), grey (infill walls), red (columns or shear walls), blue (beams), green (windows), and dark blue (doors).

Fine-tuning was performed using the ControlNet, which attaches a zero-initialised conditioning branch to the frozen backbone of Flux. Each model was trained for 100 epochs using the AdamW optimiser (learning rate 5×10^{-5}, batch size 4, mixed-precision). To preserve hierarchical consistency, earlier-stage models were frozen during the training of downstream variants.

ArchiFlux generates multi-class architectural floor plans from binary input masks representing building footprints.

StructFlux takes the output of ArchiFlux as input and predicts the spatial arrangement of vertical load-bearing components, including columns and shear walls.

BeamFlux receives the StructFlux layout and outputs the complete beam network defining the gravity-load system.

Each model was trained under consistent settings and evaluated using Euler-based integration over 100 steps. In total, the framework generated 270,000 structured floor plan layouts, including 150,000 frame structures, 60,000 frame–shear wall structures, and 60,000 shear wall structures, providing a large-scale, code-compliant dataset for downstream structural simulation and learning.

1.2 Post-processing of Image Data for Generated Structural Layout

To convert color-encoded raster images into high-fidelity finite element models, we developed a sophisticated pipeline that integrates state-of-the-art computer vision techniques with advanced structural modeling frameworks. This process ensures the accurate extraction, classification, and serialization of structural elements for subsequent analysis in OpenSees.

Image Preprocessing and Skeletonization: The initial step involves isolating structural elements based on their chromatic encoding: beams are represented in blue (R=0, G=0, B=255), and columns or shear wall boundaries in red (R=255, G=0, B=0). The images are converted to grayscale and binarized to distinguish foreground elements. Morphological operations, including dilation and directional opening, are applied to enhance the continuity of structural lines. Subsequently, the Guo–Hall thinning algorithm is employed to reduce the binary images to their skeletal representations, preserving the topology and connectivity of structural components.

Geometric Extraction and Topological Reconstruction: From the skeletonized images, maximal axis-aligned line segments are extracted using a flood-fill approach. To address fragmentation due to raster artifacts, adjacent segments within a 2-pixel proximity are merged. Intersections, including T-junctions and crosses, are detected to ensure nodal continuity. Endpoints within a 2-pixel threshold are unified to resolve minor discrepancies, and segments are straightened to enforce orthogonality, resulting in a clean and connected structural graph.

Structural Classification: Each line segment is sampled along its length, and RGB values are interpolated to determine the dominant structural class via Euclidean color distance. This classification distinguishes beams from columns or walls, enabling the segregation of structural elements into distinct categories for modeling purposes.

1.3 Automated Structural Design and Drift-Controlled Optimization for Synthetic Finite Element Models

To generate large-scale, code-compliant structural models from architectural layouts, we developed an automated design framework that transforms building skeletons produced by the Flux model into finite element (FE) representations. This pipeline performs structural member sizing, reinforcement design, and seismic compliance verification based on regulatory criteria, with iterative optimization steps if necessary. While the design success rate for any individual building may be relatively low, the extreme scalability of Flux-generated candidates ensures that a sufficiently large and diverse database of valid structural models can still be constructed.

1.3.1 Structural Design Workflow

Given a discrete building layout generated by the Flux model—comprising walls, beam spans, columns, and floor heights—a Python-based structural design pipeline initiates the assignment of concrete section sizes and reinforcement for all vertical and horizontal components. The design adheres to the relevant provisions of Chinese structural codes, including GB 50010-2010[1], and JGJ 3-2010[2], encompassing both member-level and system-level requirements.

The design process follows a hierarchical logic:

1) Initial Assignment: Cross-sectional dimensions and reinforcement ratios for beams, columns, and shear walls are computed using closed-form solutions embedded in Chinese design codes. Beams are evaluated for flexural and shear capacity; shear walls are divided into boundary and core zones, with axial and vertical reinforcement ratios set according to seismic intensity and structural importance class.
2) Serviceability and Strength Filtering: Each design is evaluated against key structural criteria, including: (i) flexural and shear strength compliance; (ii) strong-column–weak-beam requirements; (iii) span-to-depth and axial compression limits; and (iv) reinforcement ratios and detailing for seismic ductility. Members failing any check are marked for redesign or excluded during optimization.

3) Drift Control and Seismic Compliance Criteria: To ensure regulatory compliance under seismic actions, each structural model must satisfy inter-storey drift limits as prescribed in GB 50011-2010[1], with allowable thresholds explicitly defined by structural system type and seismic intensity. Specifically, the maximum inter-storey drift ratio (θ) must not exceed 1/550 under frequent earthquakes and 1/50 under rare earthquakes for frame structures; 1/800 and 1/100, respectively, for frame–shear wall systems; and 1/1000 and 1/120 for pure shear wall systems. These limits are rigorously evaluated through nonlinear time-history analysis or equivalent static procedures following model generation. If a given configuration violates the prescribed drift criteria, the pipeline engages a hierarchical optimization process that iteratively adjusts global building scale, material grade, wall thickness, and storey configuration. Models are reanalyzed after each adjustment cycle. Designs that consistently fail to meet the drift requirements after all permitted modifications are systematically excluded from the dataset. Owing to the high-throughput generative capability of the Flux model, this rejection-based strategy remains computationally efficient and enables the construction of a large, structurally diverse corpus of finite element models that conform to seismic design codes.

1.3.2 Optimization Strategy

For designs that violate strength or drift limits, a structured multi-stage optimization is employed. The pipeline sequentially adjusts (i) the global geometric scale to modify stiffness and seismic demand, (ii) material grades to higher concrete and reinforcement strengths, (iii) wall and column dimensions to enhance lateral and axial capacities, and (iv) the number of storeys within allowable height constraints. Each modified configuration is re-evaluated via finite element analysis. To control computational cost, the process is capped at a maximum of four iterations. Designs that remain noncompliant after all permitted adjustments are discarded.

1.3.3 Code Compliance and Structural Detailing

To enhance the fidelity and regulatory alignment of the generated finite element models, the design pipeline incorporates key structural detailing and verification procedures drawn from current seismic and concrete design codes. These procedures aim to ensure that each model not only meets basic strength and drift requirements but also captures essential features of realistic structural performance under seismic loading. At the member level, ductile detailing provisions are applied, including transverse reinforcement in beam and column plastic hinge zones, confinement zones at wall boundaries, and joint reinforcement satisfying strong-column–weak-beam requirements. Shear design checks are implemented to ensure that both flexural and shear capacities remain within allowable limits under design loading conditions. Structural responses are evaluated across a range of ultimate and serviceability load combinations, incorporating dead, live, and seismic actions, as well as accidental torsion effects to account for plan irregularities. Force and displacement envelopes derived from these combinations are used to guide member verification and ensure compliance with governing load effects. All design outputs, including geometry, material properties, reinforcement layouts, and seismic performance indicators, are systematically stored in a structured database. This facilitates traceability, supports downstream analysis, and enables integration with data-driven modeling efforts. Collectively, these enhancements improve the realism, robustness, and regulatory conformity of the synthetic structural models, while preserving the scalability required for large-scale generation.

1.3.4 Automatic Design Method of Steel Structures

To extend the synthetic finite element (FE) database beyond reinforced concrete systems, we developed an automatic design and modeling pipeline for steel moment-resisting frames. This pipeline transforms discretized building layouts into fully compliant three-dimensional steel frame models by performing section selection, stability checks, and nonlinear time-history analysis under seismic actions. The design framework incorporates member-level verification against Chinese steel design provisions (GB 50017-2017[10]) and seismic regulations (GB 50011-2010[1]), ensuring both strength and serviceability compliance. The methodology mirrors the high-throughput, rejection-based design strategy used for concrete systems, but adapts the workflow to capture the unique slenderness, buckling, and ductility characteristics of steel structures.

Structural Design Workflow: Beginning with sampled geometric attributes—including storey number, bay configuration, span dimensions, and inter-storey heights—a Python-based automated design engine systematically assigns structural profiles (H/I-sections, T-sections, L-sections, and C-sections) to beams and columns, selecting from standardized steel catalogues. The design workflow proceeds in a hierarchical manner, ensuring that section allocation follows code-based verification and structural consistency across member types.

1) Initial Sizing: Candidate sections are screened using tributary loads from floor slabs, live load combinations, and member self-weight. For beams, ultimate flexural demand (M_u) is checked against section capacity ($M_y = W_x f_y/\varphi_c$), while shear resistance and deflection limits under serviceability combinations are enforced. Lateral–torsional buckling is evaluated via the $\varphi(\lambda/\varepsilon_k)$ stability reduction factor, obtained by linear interpolation of tabulated values and extended with Appendix-D (GB 50017-2017[10]) closed-form solutions for high slenderness ratios. For columns, a top-down iterative procedure computes axial demands floor-by-floor, including tributary slab load and cumulative steel self-weight. Each candidate is verified against axial resistance $N_{\text{cap}} = \varphi f_y A$, where φ accounts for column slenderness effects.
2) Serviceability and Strength Filtering: Members that fail checks on bending, shear, buckling, or deflection are discarded, and heavier alternatives are trialed. The process enforces strong-column–weak-beam principles, ensuring system-level ductility under seismic excitation.
3) System-Level Assembly: Following the finalization of member sections, the structural configuration is instantiated as a three-dimensional OpenSees finite element model. Columns and beams are represented by fiber-based H/I-, T-, L-, and C-section elements, with elastic torsional stiffness incorporated through section aggregators. Rigid diaphragm constraints are imposed at each storey to enforce in-plane rigidity. Gravity loads are applied to establish the initial stress state, nodal masses are assigned based on half-element tributary contributions, and the fundamental modal properties are extracted to calibrate damping parameters for subsequent dynamic analyses.

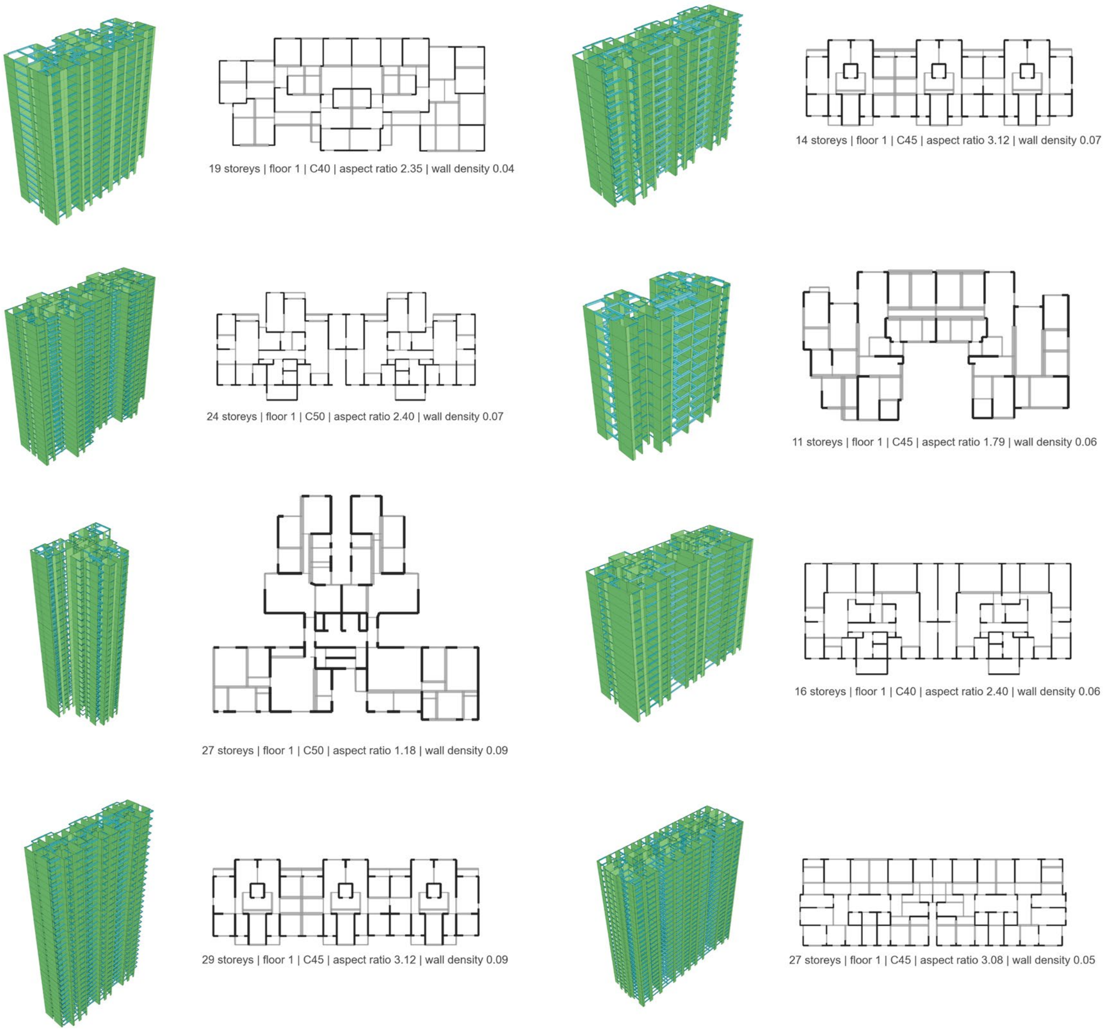
Drift Control and Seismic Compliance Criteria: Each designed steel structure is subjected to nonlinear time-history analysis under a suite of ground motions. The maximum inter-storey drift ratio is compared against the limits prescribed in GB 50011-20101 for steel frame structures, with thresholds of θ ≤ 1/550 under frequent earthquakes and θ ≤ 1/50 under rare earthquakes. Models exceeding these limits are flagged for redesign or excluded. The rejection-based strategy, supported by large-scale sampling of geometries, ensures that a structurally diverse corpus of drift-compliant steel frames is retained.

1.4 Descriptor-level validation of the synthetic structural dataset

Although Sections 1.1–1.3 establish the generative, topological, and code-based design pipeline used to construct the synthetic finite-element dataset, a further validation step is required to demonstrate that the accepted synthetic structural dataset is not merely algorithmically admissible, but also structurally realistic within the intended application domain. In this study, that domain is restricted to floor-defined multistorey building systems with explicit storey organization and lateral-force-resisting layouts, including frame, frame–shear wall, and shear wall structures. We therefore assessed the accepted synthetic structural dataset from three complementary perspectives: (i) qualitative structural realism, by visualizing representative plan-level and three-dimensional finite-element topologies; (ii) regulatory conformity, by quantifying code-normalized drift control across all accepted buildings; and (iii) descriptor-level overlap with the real-world building dataset used in this study. Together, these analyses were designed to test whether the accepted synthetic models occupy a physically and statistically plausible design space, rather than simply satisfying the internal logic of the generation pipeline.

1.4.1 Representative structural topologies and qualitative realism of the accepted synthetic structural dataset

First, the qualitative realism of the accepted synthetic structural dataset is illustrated in Supplementary Figure 2, which presents eight representative examples selected from the accepted synthetic dataset together with their corresponding plan layouts and three-dimensional finite-element renderings. These examples were chosen to span substantial variation in storey number, plan elongation, wall concentration, and global lateral-load-resisting organization, thereby avoiding an overly narrow or repetitive visual characterization of the dataset. Importantly, the resulting layouts are not arbitrary geometric sketches. They arise from a hierarchical generative process in which ArchiFlux synthesizes floor plans, StructFlux infers the spatial arrangement of walls and columns, and BeamFlux completes the beam network, after which the raster outputs are skeletonized, topologically reconstructed, classified into structural primitives, and assembled into connected structural graphs for downstream finite-element generation. The examples shown in Supplementary Figure 2 therefore visualize the end product of a pipeline that is already constrained by real annotated architectural drawings, structural connectivity rules, and code-based member design. The resulting models exhibit non-trivial wall distributions, asymmetric and elongated plans, variable stiffness organization, and realistic three-dimensional assemblages of beams, columns, slabs, and walls, indicating that the accepted synthetic structural dataset captures engineering-plausible topological diversity rather than degenerate or visually implausible abstractions.



Supplementary Figure 2 | Representative synthetic buildings illustrating topological and structural realism within the accepted synthetic structural dataset. Eight representative examples are shown, each including the plan layout and the corresponding three-dimensional finite-element model. The examples were selected to span substantial variation in storey number, plan elongation, wall distribution, and global lateral-load-resisting organization. Together, they illustrate that the accepted synthetic structural dataset contains structurally plausible floor-defined multistorey buildings with non-trivial plan-level topology, realistic three-dimensional assemblages of beams, columns, slabs, and walls, and a broad range of engineering-relevant geometric configurations.

1.4.2 Mechanical rationality of the accepted synthetic structural dataset under code-consistent seismic drift verification

Second, the code-level validity of the accepted synthetic structural dataset was assessed through explicit nonlinear seismic verification of the finalized three-dimensional finite-element models, and the results are summarized in Supplementary Figure 3. For each accepted building, the governing deformation demand was evaluated in terms of the normalized governing drift ratio, η_{drift}, obtained from OpenSees-based nonlinear time-history analysis (NLTHA) after completion of the automated design and optimization procedure. In accordance with structural design practice, gravity loads were first applied incrementally to establish the initial stress state of the structure, including the influence of axial compression on columns and walls, after which horizontal seismic excitation was imposed independently in the global X and Y directions. This sequential loading strategy is essential for reinforced-concrete and mixed lateral-force-resisting systems because inter-storey drift demand is strongly affected by the interaction between gravity-induced pre-stress, stiffness degradation, and nonlinear lateral deformation under dynamic loading.

The seismic input records were drawn from the dynamic loading database used in this study and screened according to the target seismic intensity level associated with the designed structure. In practical terms, the selected records were required to be representative, in a response-spectrum sense, of the code-consistent demand level for the corresponding structural category and seismic design condition. For each building, three ground motions were sampled from the admissible record set and applied separately in the X and Y directions, yielding multiple NLTHA verification cases for the same structural model. All input records were uniformly resampled to a time interval of Δt=0.02 s to ensure numerical consistency across the large-scale simulation campaign. This strategy allowed the drift evaluation to reflect not only one particular accelerogram, but a small suite of spectrally representative seismic inputs spanning different temporal and frequency characteristics.

The nonlinear dynamic response was computed using the three-dimensional OpenSees framework described in Section 2. Each model was assigned six degrees of freedom per node, with rigid diaphragm constraints imposed at every storey so that the floor slab acted as an in-plane rigid body while preserving translational and torsional response at the storey level. Energy dissipation was represented by Rayleigh damping calibrated to a 5% critical damping ratio in the first and third translational modes, thereby ensuring stable damping representation across both dominant and higher-mode response components. The equations of motion were integrated using the constant-average Newmark scheme with γ=0.5 and β=0.25, while equilibrium at each time step was enforced through full Newton–Raphson iteration. To ensure robustness in the nonlinear regime, the solution procedure employed adaptive recovery measures when convergence difficulties were encountered, including time-step subdivision and, where necessary, switching to a modified Newton strategy based on the initial stiffness. This analysis framework enabled highly robust large-scale NLTHA across the synthetic dataset.

For each verification analysis, the inter-storey drift ratio at storey i was computed as

$$\theta_i = \max_t \frac{|u_i(t) - u_{i-1}(t)|}{h_i} \tag{6}$$

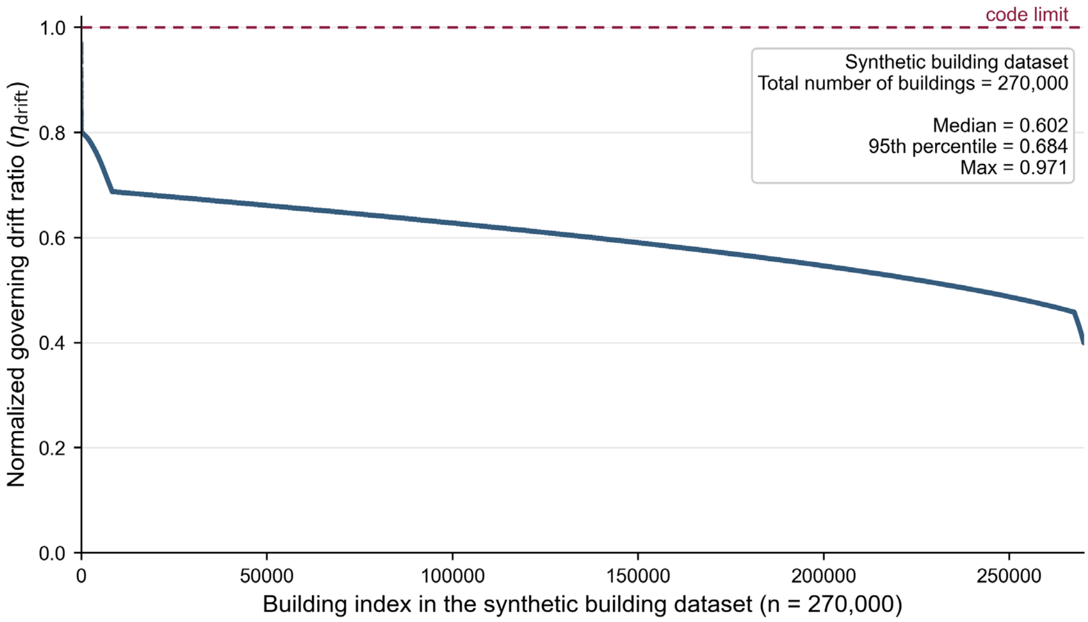
where $u_i(t)$ and $u_{i-1}(t)$ denote the lateral displacement histories of two consecutive storeys and h_i is the corresponding inter-storey height. The maximum value was taken over the full response history of the record, and the building-level governing drift demand was then defined as the maximum θ_i over all storeys and over both horizontal loading directions. This procedure follows standard structural-dynamics practice, in which the maximum inter-storey deformation demand is treated as the key serviceability and seismic-performance indicator because it directly reflects the concentration of lateral deformation, damage potential in vertical components, and compatibility with code-prescribed deformation limits.

To compare buildings with different structural systems within a common metric, the governing drift demand was normalized by the corresponding code-prescribed drift limit, yielding

$$\eta_{\mathrm{drift}} = \frac{\theta_{\max}}{\theta_{\mathrm{limit}}} \tag{7}$$

Following GB 50011[1], the limiting inter-storey drift ratios were taken as 1/550 under frequent earthquakes and 1/50 under rare earthquakes for frame structures, 1/800 and 1/100 for frame–shear wall systems, and 1/1000 and 1/120 for shear wall structures, respectively. Accordingly, $\eta_{\mathrm{drift}}<1$ indicates that the finalized structural model satisfies the deformation requirement for its corresponding structural typology and seismic loading level, whereas $\eta_{\mathrm{drift}} \geqslant 1$ indicates non-compliance and triggers redesign or rejection within the automated design pipeline. Because this normalized quantity is evaluated after full nonlinear reanalysis of the finalized finite-element model, it constitutes an explicit engineering verification metric rather than a geometric proxy or pre-screening heuristic.

As shown in Supplementary Figure 3, all retained buildings satisfy $\eta_{\mathrm{drift}}<1$, and the majority of the accepted synthetic structures remain appreciably below the governing drift threshold. This distribution indicates two important properties of the accepted synthetic structural dataset. First, the automated generation framework does not merely produce visually plausible layouts; it yields finite-element models whose deformation demand remains consistent with code-level seismic requirements under explicit NLTHA verification. Second, the retained models are not clustered immediately at the admissibility boundary, but exhibit a measurable reserve relative to the prescribed limit, indicating that the accepted dataset contains not only nominally compliant designs but also structurally stable configurations with realistic deformation margins. In this sense, Supplementary Figure 3 provides a direct bridge between the generative design pipeline and conventional structural-engineering acceptance criteria, thereby addressing the realism of the synthetic dataset from the perspective of dynamic seismic performance rather than geometry alone.



Supplementary Figure 3 | Code-normalized governing drift ratios of the accepted synthetic structural dataset. For each accepted synthetic building, the governing inter-storey drift ratio obtained from nonlinear time-history analysis in OpenSees was normalized by the corresponding code-prescribed drift limit, yielding the normalized governing drift ratio, η_{drift}. The verification analyses were performed on the finalized three-dimensional finite-element models under horizontal seismic excitation in the global X and Y directions, and the governing value was taken as the maximum code-normalized inter-storey drift ratio over all storeys and both horizontal directions. Values below unity indicate compliance with the prescribed drift limit for the corresponding structural system. All accepted buildings satisfy $\eta_{\mathrm{drift}}<1$, indicating that the accepted synthetic structural dataset is not only admissible within the generation pipeline, but also satisfies code-consistent drift constraints under nonlinear seismic verification. Source data are provided as a Source Data file.

1.4.3 Descriptor-level validation of the realism of the accepted synthetic structural dataset against the real-world building dataset

Third, we evaluated whether the accepted synthetic structural dataset overlaps in a statistically meaningful way with the real-world building dataset, which in this study comprises 694 as-built buildings and introduces substantially greater heterogeneity than the synthetic pretraining set. This comparison serves a specific validation purpose. The preceding sections establish that the accepted synthetic models are generated from annotated architectural plans, converted into structurally connected graphs, and subsequently filtered through code-based design and nonlinear seismic verification. However, code compliance alone does not guarantee that the retained dataset occupies a realistic region of the design space encountered in practice. A synthetic dataset may satisfy internal admissibility criteria while still being overly concentrated in a narrow configuration regime or, conversely, artificially scattered in regions that are only weakly represented in actual buildings. The descriptor-level comparison in Supplementary Figure 4 was therefore introduced to assess whether the retained synthetic dataset is not only feasible and code-compliant, but also statistically anchored to realistic building attributes within the intended application domain. In this study, that domain is restricted to floor-defined multistorey buildings, and the comparison was correspondingly limited to the 10–30-storey range.

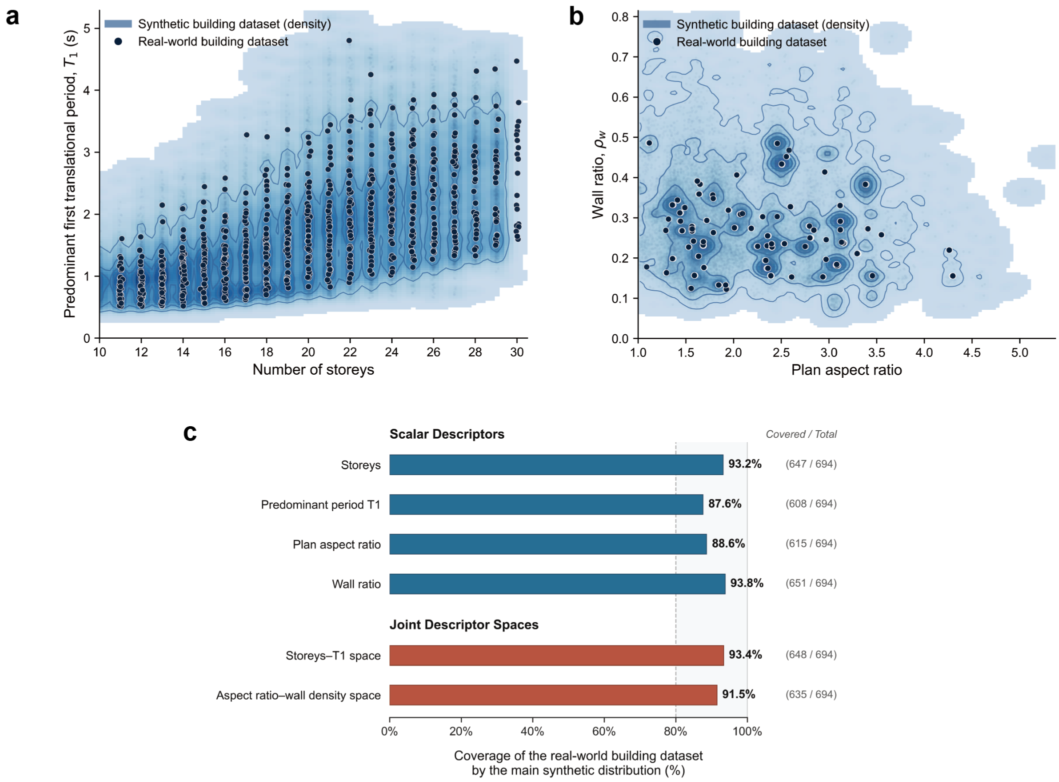
The descriptors used in Supplementary Figure 4 were selected to probe complementary aspects of structural realism. Supplementary Figure 4a examines overlap in the geometry--dynamics space defined by number of storeys and predominant first translational period, T_1, where T_1 is taken as the larger of the first translational periods in the global X and Y directions. These two quantities together provide a compact characterization of building scale and global lateral flexibility. Storey number reflects the vertical extent and broad typological class of the structure, whereas T_1 integrates the combined effects of mass, stiffness, height and lateral-force-resisting layout. Agreement in this space therefore indicates that the synthetic dataset reproduces not only the nominal height range of realistic buildings, but also the associated dynamic response regime expected from physically plausible structural systems.

Supplementary Figure 4b evaluates overlap in the plan--topology space defined by plan aspect ratio and wall ratio. Here, the wall ratio is defined as the area of the union of all valid wall polygons divided by the gross floor-plan area. The polygon union was computed before area calculation to avoid double-counting overlapping wall regions. This descriptor therefore measures the proportion of the floor plan occupied by wall elements, rather than the centreline-length density of walls. Plan aspect ratio distinguishes compact from elongated floor plates, whereas wall ratio quantifies the intensity of wall-based lateral-load-resisting organization within the plan. Consistent agreement within this descriptor space indicates that the synthetic dataset reproduces not only admissible plan geometries, but also realistic wall-layout intensity and lateral-force-resisting organization. Collectively, Supplementary Figure 4a and Supplementary Figure 4b assess whether the retained synthetic dataset aligns with the real-world building dataset across both global dynamic characteristics and engineering-relevant plan topology, rather than in either dimension independently.

In both descriptor spaces, the real-world building dataset occupies a largely overlapping region within the main synthetic distribution, indicating that the accepted synthetic structural dataset reproduces not only admissible geometry, but also plausible dynamic characteristics and plan-level structural organization. At the same time, the synthetic distribution extends modestly beyond the densest concentration of real-world points. This behaviour is both expected and desirable. It reflects the dataset-construction strategy adopted in the main manuscript, in which the synthetic design ranges were intentionally extended beyond typical empirical correlations to improve pretraining diversity and promote downstream generalization. The objective of the synthetic dataset is therefore not to replicate the exact empirical frequency of every descriptor in the real-world sample, which would risk overfitting the pretraining source to a limited observed set of buildings, but rather to provide a physically plausible, code-consistent, and moderately expanded representation of the admissible structural design space. In this context, strong overlap with the real-world dataset demonstrates that the synthetic models are centered on realistic structural regimes, whereas modest extension beyond the densest real-world cluster indicates that the dataset retains sufficient breadth to expose the predictive model to rare yet still plausible design configurations.

This interpretation is explicitly illustrated in Supplementary Figure 4c, which quantifies the extent to which the main synthetic distribution covers the real-world building dataset. For scalar descriptors, coverage is defined as the fraction of real-world buildings lying within the 5^{th}–95^{th} percentile interval of the synthetic distribution. For the two-dimensional joint descriptor spaces, coverage is defined as the fraction of real-world buildings falling within the 95% highest-density region (HDR) of the synthetic distribution. These metrics were deliberately chosen to evaluate whether real buildings reside within the core synthetic design space, rather than within an overly permissive envelope dominated by sparse synthetic extremes. This distinction is critical from both statistical and engineering perspectives. Using the full min–max range or the outermost occupied support of the synthetic dataset could lead to apparent agreement driven by a few isolated synthetic points, providing limited evidence that real-world buildings are embedded within the primary body of the accepted synthetic dataset. In contrast, the 5^{th}−95^{th} percentile interval and the 95% HDR assess whether real-world buildings lie within the dominant synthetic regime, which is the relevant criterion for evaluating the realism of the dataset as a pretraining resource. Accordingly, the high coverage values observed in Supplementary Figure 4c indicate that the accepted synthetic structural dataset effectively captures the principal geometric, dynamic, and plan-level characteristics of the real-world building dataset within the validated domain of floor-defined multistorey buildings. The strong overlap and high coverage observed here therefore provide direct evidence that the accepted synthetic dataset is both physically grounded and statistically relevant for realistic structural response prediction.

Taken together, Supplementary Figure 2-Supplementary Figure 4 delineate three complementary and mutually reinforcing facets of realism. Supplementary Figure 2 demonstrates that the accepted synthetic buildings yield finite-element realizations that are both visually coherent and topologically plausible. Supplementary Figure 3 establishes that these buildings adhere to code-consistent deformation limits under nonlinear seismic verification. Supplementary Figure 4 further indicates that, within the defined application domain, the descriptor-space distribution of the accepted synthetic structural dataset closely mirrors that of the real-world building population. Collectively, these results provide robust evidence that the accepted synthetic structural dataset constitutes a physically grounded and statistically representative pretraining resource for structural response prediction in floor-defined multistorey buildings.



Supplementary Figure 4 | Descriptor-space overlap between the accepted synthetic structural dataset and the real-world building dataset. a, Overlap in the geometry-dynamics space defined by number of storeys and predominant first translational period, T_1, with the synthetic training dataset shown as a smoothed density field and the real-world building dataset as individual points. T_1 denotes the larger of the first translational periods in the global X and Y directions. **b**, Overlap in the plan – topology space defined by plan aspect ratio and wall ratio. Plan aspect ratio is defined as the ratio of the larger to the smaller plan dimension. Wall ratio is defined as the area of the union of all valid wall polygons divided by the gross floor-plan area; polygon union was performed before area calculation to avoid double-counting overlapping wall regions. **c**, Coverage of the real-world building dataset by the main synthetic distribution. For scalar descriptors, coverage is defined using the 5^{th} – 95^{th} percentile interval of the synthetic distribution; for the two joint descriptor spaces, coverage is defined using the 95% highest-density region (HDR). Percentages indicate coverage, and values in parentheses denote covered versus total real-world buildings. Together, these analyses show substantial descriptor-level overlap between the accepted synthetic structural dataset and the real-world building dataset within the validated domain of floor-defined multistorey buildings. Source data are provided as a Source Data file.

2. Finite-Element Modelling and Nonlinear Time-History Analysis Based on OpenSees

2.1 Modelling Framework and Global Assumptions

All simulations were executed with OpenSees (v3.6.0) in a three-dimensional setting. Millimetre–Newton–second units were maintained throughout. Each model is instantiated with a three-dimensional builder, providing six degrees of freedom (DoF) per node. The global axes are defined as X and Y in-plane and Z vertical. Storey diaphragms are assumed rigid in-plane: all nodes belonging to a storey are slaved to a master node located at the geometric centroid of the slab. This constraint suppresses membrane distortion while still allowing vertical, torsional and out-of-plane motions.

2.2 Constitutive Formulations

To make the constitutive scope of the finite-element database explicit, the reinforced-concrete OpenSees models used in this study rely on path-dependent nonlinear inelastic constitutive formulations that admit irreversible response under cyclic loading, including unloading–reloading hysteresis, post-yield deformation, stiffness degradation or softening, and crack-closure effects where applicable. The constitutive laws adopted for beam–column members and wall panels are summarized in Supplementary Table 1 and illustrated in Supplementary Figure 5 and Supplementary Figure 6.

Concrete: Confined and unconfined concrete fibers in reinforced-concrete beams and columns were modelled using the OpenSees Concrete02 uniaxial material. This material was selected because it represents the main cyclic features required for the nonlinear beam–column response considered in this study, including nonlinear compressive response, post-peak softening, unloading–reloading stiffness degradation, limited tensile resistance after cracking and tension softening. In the finite-element models, each concrete material was specified through the Concrete02 parameter set

$$\boldsymbol{\eta}_{\mathrm{c}} = [f_{\mathrm{c}}', \varepsilon_{\mathrm{c0}}, f_{\mathrm{cu}}', \varepsilon_{\mathrm{cu}}, \lambda, f_{\mathrm{t}}, E_{\mathrm{ts}}] \tag{8}$$

Here, f_{c}' and $\varepsilon_{\mathrm{c0}}$ denote the peak compressive strength and the corresponding strain; f_{cu}' and $\varepsilon_{\mathrm{cu}}$ define the residual compressive strength and ultimate compressive strain; λ controls the unloading stiffness ratio; and f_{t} and E_{ts} define the tensile strength and tension-softening stiffness, respectively. Confined and unconfined concrete regions were assigned separate parameter sets according to their section location and reinforcement detailing. These parameters were passed directly to the OpenSees

Concrete02 material definition, and no custom concrete constitutive update was implemented in this study. Consequently, the concrete fibers used in the present beam−column formulation admit path-dependent and irreversible cyclic behaviour rather than purely reversible nonlinear elasticity.

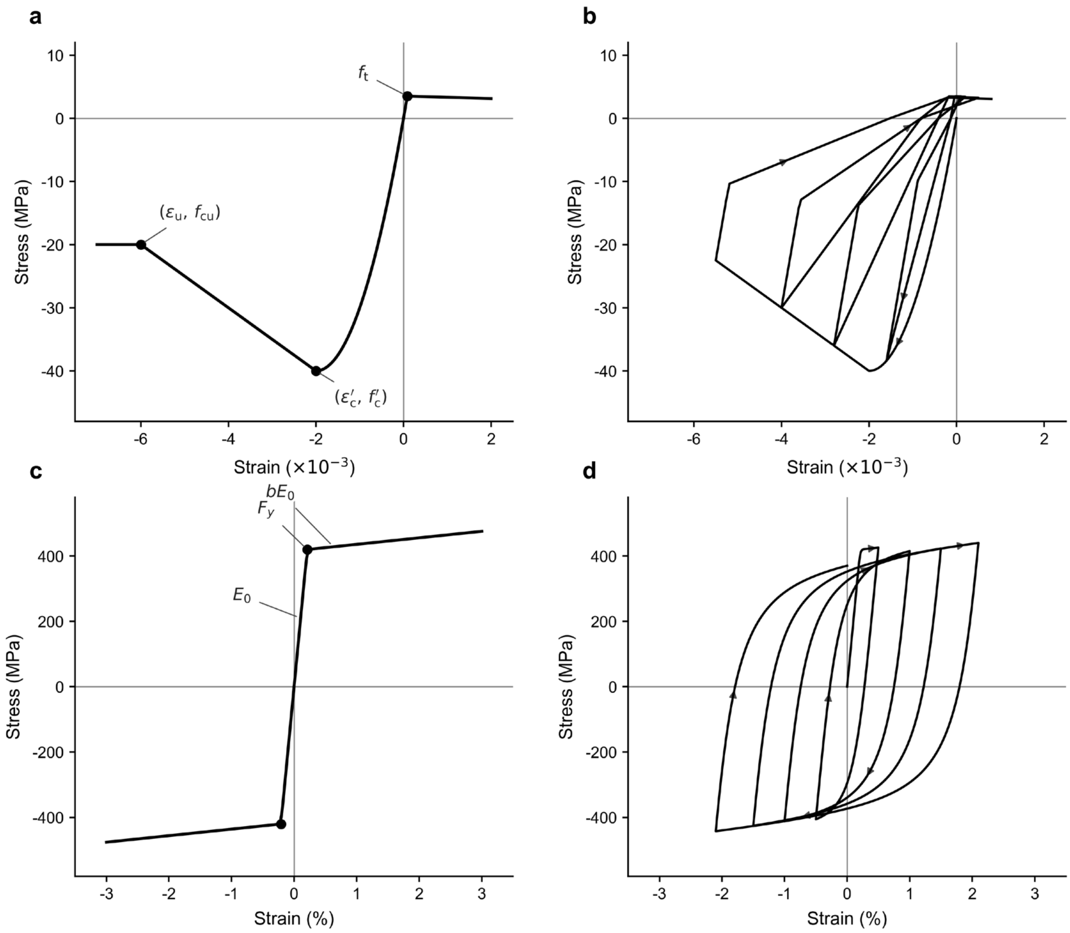
Reinforcing steel: Longitudinal bars, stirrups, and boundary steels in beam–column fiber sections follow the Giuffré–Menegotto–Pinto rule (Steel02). The stress update reads

$$\sigma_{\mathrm{s}} = bE_0\Delta\varepsilon + (1-b)\frac{E_0\Delta\varepsilon}{1+\left|\Delta\varepsilon/(\varepsilon_{\mathrm{y}}R_0)\right|^{R}} \tag{9}$$

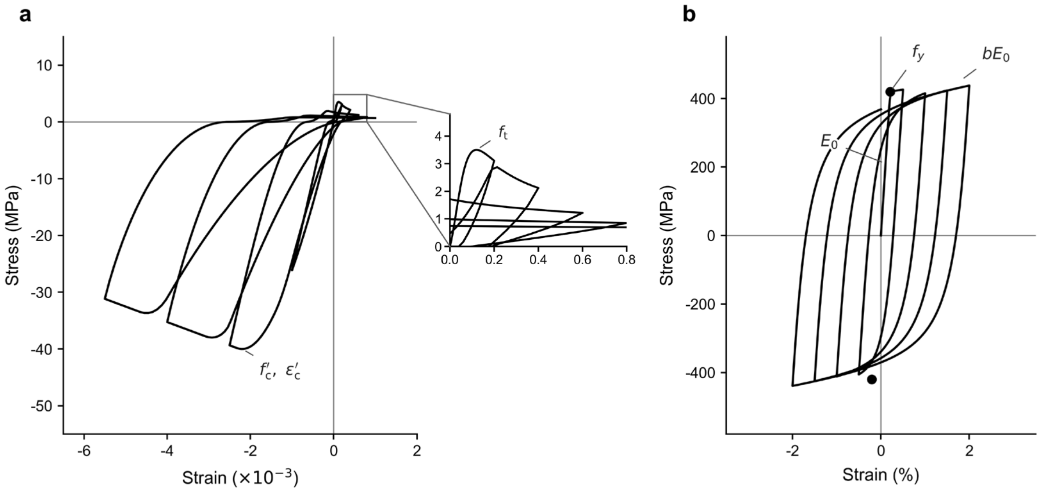
$$R = R_0[1 - c_1(|\sigma/E_0|)^{c_2}] \tag{10}$$

with E_0 the elastic modulus, b the strain-hardening ratio, ε_{y} the yield strain, R_0 a transition coefficient and R governs the curvature of the elastic-to-post-yield transition and the Bauschinger effect after strain reversal. The Steel02 formulation therefore represents yielding, smooth hysteretic reversal, reversal-dependent stiffness evolution, and post-yield hardening, enabling irreversible cyclic steel response within the fiber-based beam–column sections.

Concrete–steel coupling in wall panels: For reinforced-concrete shear walls, the wall-panel constitutive chain is represented by ConcreteCM + SteelMPF + FSAM. The concrete constituent of the wall panel is modeled using ConcreteCM, a hysteretic concrete constitutive law capable of representing cyclic compression and tension, progressive stiffness degradation, smooth unloading–reloading response, and gradual crack-closure effects. The orthogonal reinforcement in the wall panel is modeled using SteelMPF, i.e., the Menegotto–Pinto–Filippou steel formulation, which represents smooth transition between elastic and post-yield branches, reversal-dependent hysteresis, and optional isotropic hardening in tension and compression. In Supplementary Figure 6b, the two black markers schematically indicate the reversal-dependent branch-state pair used in the Menegotto–Pinto–Filippou construction, namely the strain-reversal state and the updated reference state that define the subsequent transition branch after load reversal. They are included to visualize the history-dependent update of the hysteretic response, rather than to denote fixed material points or independent constitutive parameters. These concrete and steel constituents are combined through the FSAM plane-stress reinforced-concrete panel model. In FSAM, the strain fields acting on the concrete and reinforcing steel components are assumed equal under a perfect-bond assumption, while the reinforcing steel develops uniaxial stresses along the bar directions and the concrete stress directions evolve with the state of cracking. The formulation further incorporates crack-dependent shear-transfer mechanisms through friction-based aggregate interlock and dowel action, controlled by the parameter α_{dow}. Consequently, the wall panels in the present study are modeled using a constitutive framework that admits path-dependent and irreversible cyclic response rather than purely reversible nonlinear behavior.



Supplementary Figure 5 | Constitutive laws adopted for concrete and reinforcing steel in fiber-based beam–column sections. a, Monotonic stress–strain envelope of the Concrete02 material used for confined and unconfined concrete fibers in reinforced-concrete beams and columns. The constitutive law is parameterized by the compressive strength f_{c}', strain at peak compressive stress $\varepsilon_{\mathrm{c}}'$, crushing strength f_{cu}, crushing strain ε_{u}, unloading stiffness ratio λ, tensile strength f_{t}, and tension-softening stiffness E_{ts}. **b**, Representative cyclic stress-strain response of Concrete02, illustrating unloading-reloading hysteresis, compressive softening, and limited tensile response under crack opening and closure. **c**, Monotonic stress-strain envelope of the Steel02 material used for reinforcing steel fibers, with initial elastic modulus E_0, yield strength F_{y}, and post-yield tangent bE_0. **d**, Representative cyclic stress–strain response of Steel02, illustrating smooth transition from the elastic to post-yield branch, strain hardening, and the Bauschinger effect under strain reversals. Together, these constitutive laws provide the material-level basis for the irreversible sectional response developed in the fiber-based nonlinearBeamColumn elements used in the finite-element database.



Supplementary Figure 6 | Constitutive laws adopted for reinforced-concrete wall panels. a, Representative cyclic stress-strain response of the ConcreteCM material used as the concrete constituent in reinforced-concrete wall panels, illustrating cyclic compression and tension, progressive stiffness degradation, smooth unloading – reloading behavior, and crack-closure effects. **b**, Representative cyclic stress – strain response of the SteelMPF material used for orthogonal wall reinforcement, illustrating smooth elastic-to-post-yield transition, strain hardening, and reversal-dependent hysteresis associated with the Menegotto-Pinto-Filippou formulation. The two black markers schematically denote the reversal-dependent branch-state pair used to construct a post-reversal transition branch, corresponding to the strain-reversal state and the updated reference state after load reversal. These markers are shown to clarify the history-dependent evolution of the hysteretic branch and should not be interpreted as fixed yield points or additional constitutive parameters. The F_y annotation indicates the positive-direction yield-stress scale in the schematic.

2.3 Structural Component Modeling

The modeling approaches adopted for all structural components in this study are summarized in Supplementary Table 1. Concrete regions were discretized using quadrilateral fiber patches, while longitudinal reinforcement was assigned via straight bar layers aligned along the member axis. To ensure numerically stable torsional behavior without interfering with flexural plasticity, all fiber sections were augmented with an elastic torsional spring, providing Saint-Venant stiffness in the out-of-plane direction.

Supplementary Table 1 | Finite-element modeling strategies for different structural components

Component	Element type	Section / constitutive formulation	Integration	Irreversible nonlinear behaviors represented
Columns	nonlinearBeam Column[11]	Fiber section with Aggregator and torsional stiffness. Concrete fibers are modeled using Concrete02;	5-point Gauss–Lobatto integration	Path-dependent sectional response under cyclic loading, including concrete compression nonlinearity, unloading – reloading

		reinforcing steel fibers are modeled using Steel02.		hysteresis, tensile softening, steel yielding, Bauschinger effect, and post-yield hardening.
Beams	nonlinearBeamColumn[11]	Fiber section with reduced reinforcement ratio to represent cracked inertia. Concrete fibers are modeled using Concrete02; reinforcing steel fibers are modeled using Steel02.	5-point Gauss–Lobatto integration	Path-dependent flexural response under cyclic loading, including cracking-induced stiffness reduction, concrete softening and hysteresis, steel yielding, smooth reversal, and post-yield hardening.
Slabs	Modeled implicitly	Represented through tributary mass assignment and uniform beam loads; in-plane floor constraint enforced through rigid diaphragm idealization.	/	Slabs are not modeled as explicit nonlinear load-resisting components; their primary role is to provide mass distribution and diaphragm constraint in the global response analysis.
Shear walls	E_SFI_MVLEM_3D[12] (24 DOF) with 20 vertical macro-fibres	Wall-panel concrete constituent modeled using ConcreteCM; wall-panel reinforcement constituent modeled using SteelMPF; reinforced-concrete panel behavior represented through FSAM; wall-level response assembled through E_SFI_MVLEM_3D, capturing axial – flexural – shear interaction.	Internal axial–flexural–shear coupling via macro-fiber integration	Path-dependent wall response under multidirectional cyclic loading, including cyclic concrete compression/tension, stiffness degradation, crack closure, nonlinear steel hysteresis, crack-dependent panel stress redistribution, and irreversible axial – flexural – shear coupling at wall level.

2.4 Steel Structure Modeling Method in OpenSees

Steel moment-resisting frames were modeled in OpenSees (v3.6.0) as three-dimensional systems with six degrees of freedom per node, using a millimetre–Newton–second unit convention. The global axes were defined with X and Y in-plane and Z vertical. Storey diaphragms were assumed rigid in-plane, implemented by constraining all floor nodes to a master node located at the geometric centroid of the slab. This assumption suppresses in-plane distortion while retaining vertical, torsional, and out-of-plane degrees of freedom. Foundation nodes at the base elevation were fully restrained to represent fixed-base conditions.

Section Families and Constitutive Laws: Three families of rolled sections were employed: (i) I-sections (W-shapes) for primary columns, beams, and girders; (ii) T-sections for edge members and collectors; and (iii) C-sections (channels) for secondary framing such as stringers and purlins. Each section was parameterized by overall depth, flange width, flange thickness, and web thickness, with properties drawn from standard steel catalogues and converted to consistent SI units. Member fibers adopted a bilinear kinematic hardening model (Hardening in OpenSees) with modulus E_s=200 GPa, yield stress f_y=355 MPa, zero isotropic hardening, and a kinematic modulus calibrated to match design post-yield stiffness.

Fiber Section Construction and Torsion Aggregation: All members were discretized using fiber sections. H/I-sections were divided into three quadrilateral patches (flange–web–flange), T-sections into two patches (flange + stem), and C-sections into three patches offset to one side of the web. Saint-Venant torsion was incorporated through an elastic torsional spring aggregated into the fiber section. The torsional stiffness *GJ* was introduced through an Aggregator section, ensuring realistic torsional behavior without interfering with flexural plasticity. Uniform fiber meshes (e.g., 16 × 4 for flanges, 16 × 2 for webs) were adopted across section families to maintain numerical stability.

Element Formulation and Member Assignment: All steel members were modeled using nonlinearBeamColumn elements with five-point Gauss–Lobatto integration. Columns were assigned storey-dependent I-sections (or C/T alternatives where appropriate), while beams and girders were specified per direction. Edge collectors and secondary members were modeled with T- and C-sections to capture realistic load transfer in peripheral regions. Local member axes were aligned to the global frame through geometric transformations, ensuring correct force projection under gravity and lateral loading. Joint regions were assumed rigid in the baseline dataset.

2.5 Boundary Conditions and Load

To simulate fixed-base conditions, all nodes at the foundation level were fully restrained in translation and rotation, thereby eliminating any rigid body motion. Prior to dynamic analysis, a static loading step was conducted to apply vertical gravity loads. These included both permanent actions (dead loads), such as the self-weight of structural and non-structural components, and imposed live loads representative of occupancy and usage demands. The vertical loads were applied incrementally and allowed to reach static equilibrium, thereby establishing a stable initial stress state for subsequent dynamic simulation. Following gravity loading, nonlinear time-history analysis was performed under horizontal ground acceleration records. The external excitation input was introduced directly after the completion of the static step, allowing the dynamic response to evolve from a load-consistent initial configuration. This sequential application of loading ensured that the interaction between axial pre-stress and lateral deformation demands—particularly critical in nonlinear regimes—was faithfully captured.

2.6 Non-Linear Time-History Analysis

The structural response of each structural model was simulated through nonlinear time-history analysis, capturing the full coupling between inertia, damping, and nonlinear restoring forces under ground motion excitation. All analyses accounted for geometric and material nonlinearity, as well as the influence of pre-applied gravitational loads on the dynamic evolution of the structure. The dynamic equilibrium equation governing the system is expressed as:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{R}\big(\mathbf{u}(t), \dot{\mathbf{u}}(t)\big) = -\mathbf{M}\mathbf{a}_g(t) \tag{11}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, and $\mathbf{R}$ denotes the nonlinear restoring force vector obtained from the current displacement and velocity state. The input $\mathbf{a}_g(t)$ represents the acceleration time history, applied as uniform horizontal excitation in either the global X or Y direction. To model energy dissipation, Rayleigh damping[13] was adopted in the form:

$$\mathbf{C} = \alpha_M \mathbf{M} + \beta_K^{\mathrm{curr}} \mathbf{K}_{\mathrm{tan}} \tag{12}$$

where α_M and β_K^{curr} are the mass- and stiffness-proportional damping coefficients, respectively, and $\mathbf{K}_{\mathrm{tan}}$ denotes the current tangent stiffness matrix. The coefficients were calibrated to achieve a 5% critical damping ratio for the first and third translational modes, ensuring consistent damping performance across both dominant and higher-mode responses. Specifically, the coefficients are given by:

$$\alpha_M = \frac{2\xi\omega_1\omega_3}{\omega_1 + \omega_3}, \beta_K^{\mathrm{curr}} = \frac{2\xi}{\omega_1 + \omega_3} \tag{13}$$

with ξ=0.05 and ω_1, ω_3 denoting the first and third circular frequencies extracted from eigenvalue analysis.

The equation of motion was advanced in time using the constant-average Newmark method, with parameters γ=0.5 and β=0.25, ensuring numerical stability even in the presence of significant nonlinearity. For a nominal time step of Δt=0.02s, equilibrium at each time increment was achieved via a full Newton–Raphson iteration, whereby the residual

$$\mathbf{\Phi} = \mathbf{M}\ddot{\mathbf{u}}_{n+1} + \mathbf{C}\dot{\mathbf{u}}_{n+1} + \mathbf{R}(\mathbf{u}_{n+1}) + \mathbf{M}\mathbf{a}_g^{n+1} \tag{14}$$

was driven to zero through successive updates of the displacement field. Convergence was declared when the energy norm of the residual satisfied:

$$\|\mathbf{\Phi}\|_2 \leq 10^{-8}\ \mathrm{N}\cdot\mathrm{mm} \tag{15}$$

within a maximum of six iterations. If convergence failed, an adaptive recovery procedure was invoked: the time step was successively halved to 0.5, 0.2, 0.1, or 0.05 of the original value. If required, the iteration strategy was switched to a modified Newton method using the initial stiffness. This hierarchical strategy ensured high solution robustness, enabling 99.9% of over two million simulations to complete without manual intervention.

2.7 Masonry and Base-Isolated Structures

In addition to the automatically designed steel frames described in Section 2.4, two categories of real-world structural systems were incorporated to broaden the scope of the zero-shot evaluation: masonry buildings and base-isolated buildings. These systems were not generated synthetically but instead modeled directly from as-built structural data in established commercial platforms.

Masonry structures: A total of 33 representative masonry buildings were modeled in Abaqus, yielding 73 nonlinear time-history analyses. Constructed directly from as-built architectural and structural drawings, these building models were incorporated into the zero-shot benchmark to evaluate the capacity of SeisGPT to generalize to structural systems characterized by material behaviors, stiffness distributions, and damage evolution mechanisms that diverge substantially from the primary synthetic pretraining corpus. As detailed in the main text, these masonry systems constitute one of three out-of-distribution structural categories used to rigorously assess cross-system generalization.

Rather than adopting an unspecified, black-box finite-element approach, the masonry benchmark employs an equivalent macro-scale modeling philosophy. This approach aligns conceptually with established nonlinear discrete-homogenized masonry strategies (e.g., Sharma et al.[14]), wherein the structural response is captured at the wall scale via an assemblage of macro-regions interconnected by nonlinear deformable links. The objective herein was not to replicate an exhaustive meso-scale homogenization workflow — involving explicit brick-mortar discretization and interface-level constitutive calibration—for each real building. Instead, the core mechanical fidelity and constitutive behavior were retained at the structural scale. Accordingly, each building was partitioned into macroscopic wall regions defined by storey levels, wall piers, spandrels, openings, and principal intersections. Floor diaphragms were explicitly modeled to enforce realistic mass distributions and floor-level kinematic compatibility. This idealization captures the global nonlinear seismic response while preserving the explicit representation of architectural openings, wall segmentation, stiffness discontinuities, and macroscopic force-transfer mechanisms.

At the constitutive level, nonlinear behavior is localized within equivalent macro-scale link components connecting adjacent wall regions. These links are governed by a damage-plasticity constitutive law within Abaqus. Following a nominal-stress representation analogous to the Concrete Damaged Plasticity framework, the constitutive response is formulated as:

$$\sigma = (1 - d)E_0(\varepsilon - \varepsilon_{pl}) \tag{16}$$

where σ is the nominal stress, E_0 is the initial elastic stiffness, ε is the total strain, ε_{pl} is the plastic strain, and $d \in [0,1]$ is a scalar damage variable dictating progressive stiffness degradation. Crucially, this formulation is not purely nonlinear elastic; it explicitly accommodates irreversible inelastic strain alongside damage-induced stiffness loss, thereby enabling path-dependent, brittle nonlinear behavior under both monotonic and cyclic loading regimes.

For each nonlinear macro-link j, the equivalent axial strain and stress measures are defined as:

$$\varepsilon_j = \frac{\Delta u_j}{L_j}, \sigma_j = \frac{F_j}{A_j} \tag{17}$$

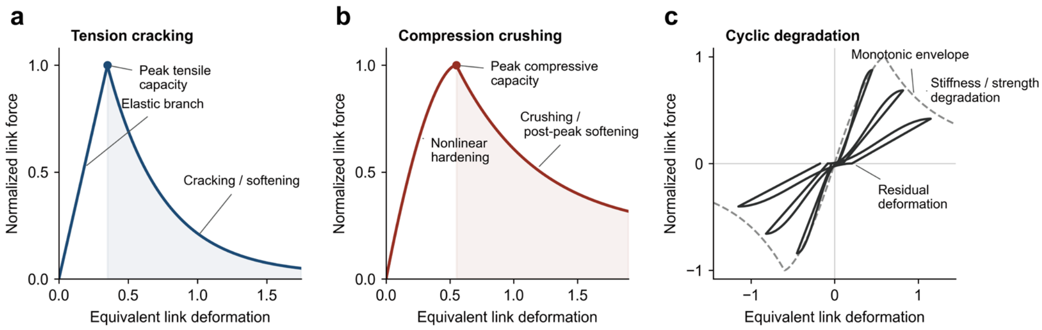
where Δu_j denotes the relative deformation projected onto the link axis, L_j is the effective link length, F_j is the axial force carried by the link, and A_j is the corresponding equivalent cross-sectional area. This equivalent continuum representation effectively consolidates wall-level cracking, crushing, unloading–reloading degradation, and stiffness loss into a computationally tractable set of nonlinear connectors without sacrificing the integrity of the building-scale force-transfer mechanisms. The corresponding link laws were rigorously calibrated to reflect equivalent macroscopic behavior derived from masonry material properties, structural drawings, established brittle failure typologies, and global stiffness consistency constraints. This modeling framework offers a highly favorable balance between constitutive realism and computational tractability for large-scale nonlinear dynamic studies. While fully detailed meso-scale models of masonry units and interfaces provide high mechanical fidelity, they demand a prohibitive number of constitutive parameters and computational resources. The adopted macro-scale strategy circumvents these limitations while preserving the critical phenomenological features of masonry failure — including crushing, frictional slip, fracture, and softening — without explicitly reproducing each meso-scale interface calculation step.

Consequently, the present benchmark fundamentally belongs to the family of nonlinear discrete-homogenized/damage-plasticity masonry models, standing in stark contrast to simplified purely elastic or purely bilinear elastoplastic continuum idealizations. Supplementary Figure 7 schematically illustrates the constitutive scope of these equivalent macro-links, detailing the tensile cracking and softening branch, the compressive crushing and post-peak degradation branch, and the cyclic stiffness/strength degradation coupled with irreversible deformation. These conceptual macro-scale response laws are designed to represent the dominant brittle nonlinear mechanisms governing global seismic response at the structural scale, rather than acting as direct meso-scale interface constitutive curves for individual mortar joints or brick-level failure surfaces. This clarification strictly defines the modeling family within which the zero-shot masonry results of SeisGPT should be interpreted.

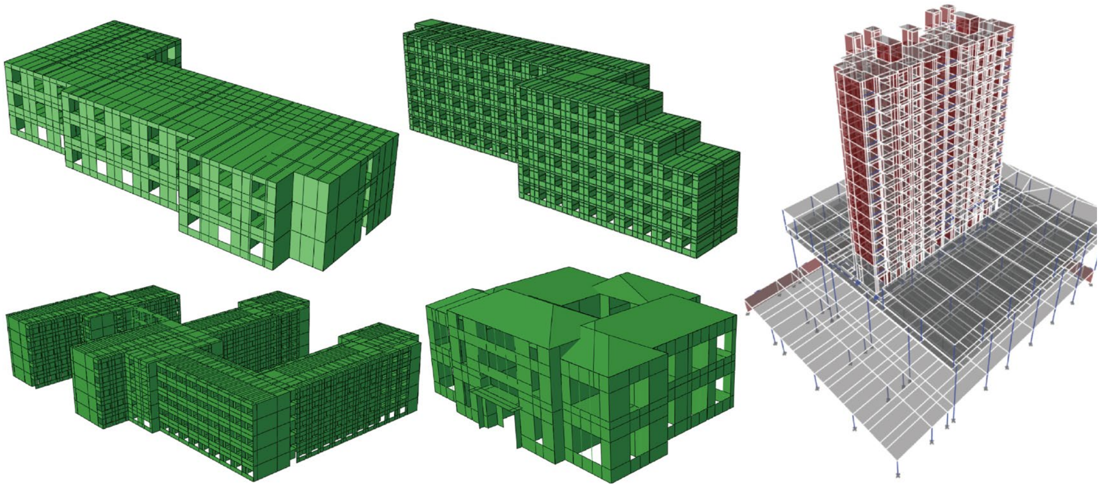
Prior to nonlinear dynamic analysis, gravity loads were applied to establish the initial stress state, consistent with the global nonlinear time-history procedure detailed in Section 2.6. Subsequently, horizontal ground acceleration records were imposed. The dynamic response was computed utilizing a robust nonlinear solution framework consistent with the broader structural database, incorporating Rayleigh damping, constant-average Newmark integration, Newton–Raphson equilibrium iterations, and adaptive time-step recovery for challenging nonlinear increments. Thus, while the masonry benchmark differs from the synthetic reinforced-concrete and steel systems in geometric typology and fundamental damage mechanisms, it remains fully integrated within the study's unified global dynamic-analysis protocol.

Base-isolated structures: Four real buildings with seismic isolation systems were modeled in ETABS, generating 33 nonlinear analyses. These structures include laminated rubber bearings and lead-core isolation devices located at the base isolation plane, together with reinforced concrete superstructures above. Supplementary Figure 8 illustrates one such ETABS model, showing both the isolation layer and the superstructure framing. The inclusion of these models allows evaluation of the model's predictive capacity under systems dominated by energy-dissipating isolation devices.

Together with the synthetic steel structure dataset, these Abaqus masonry and ETABS base-isolated buildings form a triad of structurally distinct systems that differ fundamentally in material properties, stiffness distributions, and energy-dissipation mechanisms. By incorporating real-world building models, the evaluation corpus provides a rigorous and heterogeneous benchmark for assessing SeisGPT's cross-system generalization capacity.



Supplementary Figure 7 | Conceptual constitutive response adopted for the nonlinear macro-links in the Abaqus-derived masonry benchmark. a, Representative tensile branch of the equivalent macro-link law, illustrating crack initiation, peak tensile resistance, and post-peak softening. **b**, Representative compressive branch of the equivalent macro-link law, illustrating nonlinear hardening, crushing, and post-peak degradation. c, Representative cyclic response of the equivalent macro-link law, illustrating stiffness degradation, strength deterioration, and irreversible deformation under load reversals. These curves are provided to clarify the constitutive scope of the masonry benchmark and should be interpreted as schematic macro-scale response laws for the equivalent nonlinear links, rather than as direct meso-scale brick–mortar interface constitutive curves.



Supplementary Figure 8 | Real-world masonry and base-isolated building models used for zero-shot evaluation. Abaqus-derived masonry building (left) and ETABS-derived base-isolated building (right).

3. Architectural Details of the SeisGPT Model

The SeisGPT architecture introduces a dynamics-consistent framework for modeling structural responses by integrating principles from structural mechanics with contemporary machine learning methodologies. At the core of this framework is the Spectral Duhamel–Green (SDG) Mixer block, which replaces conventional self-attention mechanisms with a structure-conditioned spectral mixing operator grounded in modal dynamics. Rather than relying on generic token-wise attention, the architecture performs information propagation within a structure-specific spectral basis derived from the underlying coupling operator. By drawing on Green-function–inspired response propagation, the model enables physically interpretable long-range interactions while maintaining robustness in dynamic structural response prediction.

3.1 Physical motivation of the Spectral Duhamel–Green Mixer

This section provides the structural-dynamics motivation for the Spectral Duhamel–Green (SDG) Mixer. The complete computational formulation of the SDG-Mixer used in the experiments is given in the Methods section of the main text; therefore, the Supplementary Information does not repeat the full architecture-level derivation. The purpose here is to clarify why a modal spectral propagation operator is an appropriate inductive bias for building response prediction.

For a linearized multi-degree-of-freedom structural system, the dynamic equilibrium equation can be written as

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{f}(t) \tag{18}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ denote the mass, damping, and stiffness operators, respectively; $\mathbf{u}(t)$ is the displacement vector; and $\mathbf{f}(t)$ is the generalized external excitation. In such systems, the response can be interpreted through Duhamel's integral, in which the structural response is obtained by convolving the external excitation with the impulse-response operator. Modal analysis further decomposes this impulse-response operator into components governed by natural frequencies, modal shapes, and damping ratios. This observation motivates replacing generic token mixing with a spectral propagation mechanism whose structure reflects the modal organization of dynamic response.

In SeisGPT, this physical idea is not used as a closed-form linear solver. Instead, it is embedded as an architectural prior inside a learnable neural operator. The model constructs a token-space structural coupling operator from the stiffness and mass descriptors, performs eigendecomposition in this token space, propagates latent features in the corresponding spectral basis, and maps the propagated features back to the token domain. The resulting SDG-Mixer is therefore a structure-conditioned token mixer: it retains the long-range propagation and resonance-aware organization of modal dynamics, while allowing learnable corrections to account for nonlinear finite-element response, stiffness degradation, hysteretic dissipation, discretization effects, and modelling mismatch.

3.2 Relation between modal propagation and the implemented SDG-Mixer

The implemented SDG-Mixer should be understood as a dynamics-inspired spectral mixing layer rather than a direct modal superposition solver. In conventional modal analysis, each mode evolves according to a damped oscillator whose phase advance and amplitude decay are controlled by the modal frequency and damping ratio. The SDG-Mixer follows this principle at the level of latent token features. After projection into the eigenbasis of the token-space coupling operator, paired latent channels are propagated by a damped rotation, where the phase advance, decay rate, and amplitude modulation are generated by lightweight structure-conditioned parameter networks. A bounded rational spectral correction is further applied as a frequency-dependent compensator, enabling the propagation operator to represent nonlinear departures from idealized modal response without directly modifying the eigenvalues of the coupling operator.

This design differs from standard self-attention in two important aspects. First, global information exchange is organized in a structural spectral basis derived from stiffness–mass information, rather than through unconstrained pairwise attention scores among tokens. This gives the mixing operation an explicit connection to the intrinsic dynamic organization of the structure, including modal coupling, long-range propagation, and resonance-sensitive feature interactions. Second, the propagation parameters are conditioned on building-specific structural descriptors, enabling the same pretrained model to adjust its latent dynamics across different floor numbers, stiffness distributions, and mass distributions. In this sense, the SDG-Mixer provides a physics-guided spectral mixing mechanism that bridges classical modal propagation and learnable nonlinear response approximation, while remaining sufficiently flexible to capture finite-element nonlinearities and structure-dependent modelling discrepancies.

3.3 Detailed model architecture and component specifications

The forward pass of SeisGPT involves several key stages: (i) temporal-to-feature projections of excitation and simplified response data, (ii) fusion of excitation and prior features, (iii) physics-informed structural encoding (PhySE), (iv) tokenization via floor-to-token lifting, (v) physics encoder to construct the token-space coupling operator, (vi) modal spectral propagation using SDG-Mixer blocks, and (vii) final token-to-floor decoding. At each stage, the features are transformed to capture both temporal dynamics and structural properties, enabling the model to learn both data-driven patterns and physics-guided interactions.

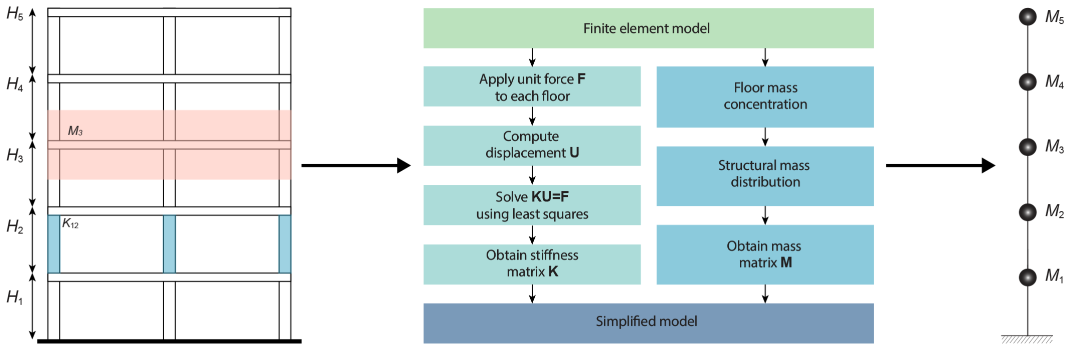
In this section, we provide a detailed breakdown of the individual components that constitute the SeisGPT architecture, emphasizing the mathematical foundations and the specific role of each module. These components are carefully designed to ensure the integration of both structural physics and machine learning principles to model complex dynamic behavior in building systems.

3.3.1 Simplified dynamic-response (SDR) prior: derivation, directional scope, and relation to the parent 3D FE model

The simplified dynamic-response (SDR) module used in SeisGPT should be interpreted as a reduced, direction-specific mechanics prior extracted from the parent three-dimensional finite-element (FE) model, rather than as an independent replacement for the full three-dimensional structural analysis. This distinction is important for the interpretation of torsional effects. As described in Section 2.1, all parent structural models in this study are formulated in OpenSees as three-dimensional systems with six degrees of freedom per node, and rigid-diaphragm constraints are imposed at each storey by slaving floor nodes to a master node located at the geometric centroid of the slab. Under this formulation, the floor slab is constrained against in-plane distortion while the parent FE system still retains storey-level translational, torsional, and out-of-plane kinematics. Consequently, for structurally eccentric or plan-irregular buildings, even nominally unidirectional lateral loading in the global X or Y direction may induce torsion-influenced storey deformation in the underlying three-dimensional FE response.

In the present implementation, however, the SDR prior is not constructed as a fully coupled reduced model with explicit storey-level coordinates $[u_x, u_y, \theta_z]$. Instead, it is extracted separately for the X and Y loading directions, yielding direction-specific effective mass and stiffness operators for translational floor response. For a given loading direction, unit lateral forces are applied sequentially to each storey of the parent three-dimensional FE model, and the resulting storey-level translational displacement fields are assembled. A dense, symmetric, full-bandwidth effective stiffness operator is then identified from the relation $\mathbf{F} = \mathbf{KU}$ by constrained least-squares fitting, while the corresponding lumped mass operator is obtained from the parent model through tributary floor-mass aggregation. Repeating this procedure independently in the global X and Y directions yields two direction-specific condensed surrogates, rather than a single fully coupled reduced-order system. This treatment is also consistent with the evaluation setting used in the main manuscript, where the responses in the X and Y directions are computed independently and only combined subsequently for visualization when needed.

The corresponding identification workflow is summarized schematically in Supplementary Figure 9. This figure clarifies that the simplified dynamic-response prior is not introduced as an independently idealized shear-building model; rather, it is extracted directly from the parent finite-element system through force–displacement probing and constrained operator identification, thereby preserving a mechanics-consistent reduced representation for subsequent neural inference.



Supplementary Figure 9 | Derivation of a simplified model from an FE model. The figure illustrates the process of deriving a simplified model from an FE model of a building structure. A unit force F_i was applied sequentially to each floor, and the resulting displacements U_{ij} were computed. The stiffness matrix $\mathbf{K}$, constrained to be symmetric, was obtained by solving $\mathbf{KU} = \mathbf{F}$ using the least squares method. The mass matrix $\mathbf{M}$ was directly extracted from the FE model, with floor masses concentrated at discrete nodes and the mass of structural components distributed to the adjacent upper and lower floors. Together, $\mathbf{K}$ and $\mathbf{M}$ define the simplified model used for structural analysis.

Accordingly, the present SDR formulation extracts orthogonal, direction-specific translational operators while the reduced representation is expressed solely in terms of translational floor coordinates. Rotational storey degrees of freedom, torsional inertia, and explicit cross-directional coupling operators such as $\mathbf{K}_{xy}$, $\mathbf{K}_{x\theta}$ and $\mathbf{K}_{y\theta}$ are not retained as independent reduced variables. The resulting SDR prior therefore represents directional translational dynamics rather than a fully coupled lateral–torsional reduced-order system. Quantities associated with explicit torsional modal behaviour, in-plane diaphragm distortion, and the distributed response field across the floor plan are consequently not represented as explicit reduced coordinates.

Despite this reduced coordinate structure, the resulting translational operators are not mechanically equivalent to those of an idealized uncoupled planar shear-building model. The operators are identified directly from the response of the parent three-dimensional finite-element system and therefore reflect the mechanical influence of degrees of freedom that are not explicitly retained in the reduced representation. This relationship can be clarified by considering a partitioned linearized system in which the retained translational coordinate in one loading direction is coupled with a set of additional coordinates that include transverse translation and storey rotation. If the full stiffness operator is written as,

$$\begin{bmatrix} \mathbf{F}_x \\ \mathbf{F}_r \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{xx} & \mathbf{K}_{xr} \\ \mathbf{K}_{rx} & \mathbf{K}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_r \end{bmatrix} \tag{19}$$

where $\mathbf{u}_r$ collects the coordinates not explicitly retained in the SDR representation, static condensation yields the effective directional operator

$$\mathbf{K}_x^{\mathrm{eff}} = \mathbf{K}_{xx} - \mathbf{K}_{xr}\mathbf{K}_{rr}^{-1}\mathbf{K}_{rx} \tag{20}$$

The influence of the eliminated coordinates enters the retained translational operator implicitly through the condensation process. Therefore, when the parent three-dimensional FE model exhibits torsion-influenced deformation due to plan eccentricity, stiffness irregularity, or higher-order coupling effects, part of that influence may be embedded in the identified directional translational operators. In this sense, the present SDR prior can be said to inherit torsion-related effects only implicitly, through their contribution to the parent three-dimensional translational displacement fields from which the reduced operators are identified.

It should be clarified that the SDR prior is introduced solely as a coarse, mechanics-informed input that provides a low-order, direction-specific approximation of floor-level translational responses under the prescribed excitation. It does not constitute the final prediction target of the model. SeisGPT is not trained to reproduce the SDR response itself. Instead, the downstream encoder–decoder architecture receives the SDR prior together with excitation histories and structural descriptors, while the model parameters are optimized against nonlinear time-history responses generated by the full three-dimensional OpenSees system. Consequently, although the SDR prior represents a reduced description that does not explicitly retain full torsional dynamics, the learned mapping remains supervised by high-fidelity responses of the parent three-dimensional finite-element model. Within this framework, the SDR module serves as a physics-guided initialization that facilitates representation learning, whereas the ultimate predictive objective remains the nonlinear three-dimensional structural response of the original OpenSees system.

3.3.2 Temporal-to-feature projection and per-floor latent embedding

The model takes as input the external excitation history $\mathbf{X}_{\text{eq}}$ and the SDR-prior response $\mathbf{X}_{\text{SDR}}$, both arranged as tensors of size $B \times F_{\max} \times T$, where B is the batch size, $F_{\max}$ is the padded floor dimension, and $T = T_m + T_p$ denotes the temporal input length used by the network. The inputs are padded to ensure a uniform size across buildings with different numbers of floors. A floor mask $\mathbf{m}_f \in \mathbb{R}^{B \times F_{\max} \times 1}$ is applied to ensure that padded floors contribute zero to all subsequent computations.

The first stage maps the excitation and SDR-prior sequences independently into floor-wise latent features through linear temporal projections applied along the time dimension. This operation converts each floor-level input sequence into a D-dimensional feature vector while preserving the floor index for subsequent structural encoding. Building-specific stiffness and mass information is introduced through the structural graph encoder and the physics encoder described in the following sections, rather than through a separate structural embedding module at this stage.

After floor-to-token lifting, a learnable token positional embedding is added to the fixed-token representation to distinguish latent token identities before SDG-Mixer propagation.

3.3.3 Two-stream gated fusion for excitation–prior integration

The core of SeisGPT's fusion mechanism involves combining the excitation data with the simplified prior structural response data. A gated fusion mechanism allows the model to dynamically learn the most relevant interactions between these two input streams. Let $\mathbf{H}_{\text{eq}}$ and $\mathbf{H}_{\text{SDR}}$ denote the floor-wise latent features obtained from the excitation and SDR-prior streams, respectively. The two streams are concatenated along the feature dimension and passed through a lightweight gating layer:

$$g = \text{softmax}(\mathbf{W}_g[\mathbf{H}_{\text{eq}} || \mathbf{H}_{\text{SDR}}] + b_g) \tag{21}$$

where $g = [g_{\text{eq}}, g_{\text{SDR}}]$ contains the normalized stream-wise gating weights. The fused representation is computed as

$$\mathbf{H}_{\text{fused}} = g_{\text{eq}} \odot \mathbf{H}_{\text{eq}} + g_{\text{SDR}} \odot \mathbf{H}_{\text{SDR}} + \mathbf{W}_t[\mathbf{H}_{\text{eq}} || \mathbf{H}_{\text{SDR}}] + b_t \tag{22}$$

Here, $\odot$ denotes element-wise multiplication, and the final linear transformation of the concatenated streams provides an additional learnable residual mixing term. This formulation allows the model to retain stream-specific information while learning cross-stream interactions between the imposed excitation and the SDR-prior response.

3.3.4 Implementation details of the structural graph encoder

The mathematical formulation of the physics-informed graph encoder is provided in the Methods section of the main text. To avoid duplication, this section focuses on implementation-level details used in all experiments.

For each building sample b, graph encoding was applied only to valid floor nodes, while padded floors were excluded using a binary floor mask. Let F_b denote the actual number of valid floors in building b, and let $F_{\max}$ denote the maximum number of floors used for padding across all samples within a batch. For each floor index $i \in \{1,2,\ldots,F_{\max}\}$, the binary floor mask $m_{f,b}^{i}$ is defined as

$$m_{f,b}^{i} = \begin{cases} 1, & i \le F_b, \\ 0, & i > F_b, \end{cases} \tag{23}$$

where $m_{f,b}^{i}$=1 indicates that the i-th floor corresponds to a physically valid floor node in building b, whereas $m_{f,b}^{i}$=0 indicates a padded (non-existent) floor introduced only for tensor alignment. All padded floor features were explicitly set to zero before temporal projection, structural encoding, and final output decoding.

In the implementation used for all reported SeisGPT experiments, the structural graph encoder was a four-layer GATv2-based graph module with 16 attention heads per layer. The hidden dimension was D=4096, and each attention head therefore used a 256-dimensional subspace before head concatenation. The scalar edge attribute was taken directly from the nonzero entries of the floor-level structural coupling matrix supplied to the graph encoder. Specifically, edges were constructed from the nonzero pattern of the input coupling matrix, and the corresponding matrix value was used as a one-dimensional edge feature. The graph encoder used residual updates with layer normalization, GELU activation and dropout rate 0.1 after each GATv2 layer. The output was then reassembled into a padded tensor of size $F_{\max}$ by appending zero vectors for invalid floors.

Unless otherwise stated, the SeisGPT configuration used in the main experiments adopted N_{tok}=256 latent tokens, latent dimension D=4096, two SDG-Mixer blocks, 32 SDG spectral heads, four graph-encoder layers, 16 graph-attention heads and dropout rate 0.1. The structural-injection gate was initialized at zero, so that the graph-derived structural stream was introduced progressively during training rather than dominating the token representation at initialization.

3.3.5 Floor-to-token lifting and latent tokenization strategy

After temporal projection and graph-based structural encoding, each building is represented by floor-wise latent features defined on its physical storeys. Because the number of floors varies across buildings, SeisGPT maps these floor-wise features into a fixed-size latent token space before applying the SDG-Mixer. This design decouples the variable physical floor coordinate from the fixed computational token coordinate, enabling one pretrained model to process buildings with different heights and structural layouts.

For building b, the masked floor-wise features obtained from the excitation stream, the SDR-prior stream and the graph-encoded structural stream are denoted as $\mathbf{H}_{\mathrm{eq},b}$, $\mathbf{H}_{\mathrm{SDR},b}$ and $\mathbf{H}_{\mathrm{str},b}$, respectively. Each feature tensor has size $F_{\max} \times D$ after padding and masking, where $F_{\max}$ is the padded floor dimension and D is the latent feature dimension. In the main experiments, D=4096. These floor-wise features are lifted to N_{tok} latent tokens using a shared learnable floor-to-token operator $\mathbf{W}_L$ with size $N_{\mathrm{tok}} \times F_{\max}$. For a generic floor-wise feature tensor $\mathbf{H}_b$, the lifting operation is written as

$$\mathbf{Z}_b = \mathbf{W}_L \mathbf{H}_b \tag{24}$$

where $\mathbf{Z}_b$ has size $N_{\mathrm{tok}} \times D$. In all main experiments, N_{tok}=256. The same lifting operator is used for the excitation-derived features, the SDR-prior features and the graph-encoded structural features, so that all input streams are represented on a common latent token grid.

The excitation and SDR-prior token streams are first combined through a lightweight gated fusion module. The graph-derived structural token stream is then introduced through a residual injection gate, giving

$$\mathbf{Z}_{0,b} = \mathrm{Fuse}\left(\mathbf{Z}_{\mathrm{eq},b}, \mathbf{Z}_{\mathrm{SDR},b}\right) + \tanh\left(\gamma\right) \mathbf{Z}_{\mathrm{str},b} \tag{25}$$

where γ is a learnable scalar initialized at zero. This initialization prevents the structural stream from dominating the token representation at the beginning of training, while allowing its contribution to increase adaptively during optimization. A learnable token positional embedding is subsequently added before the SDG-Mixer blocks. The resulting fixed-token representation $\mathbf{Z}_{0,b}$ provides a structure-conditioned latent state for spectral propagation.

It should be emphasized that this floor-to-token lifting is not a conventional modal transformation. Instead, it is a learnable mapping from the physical floor coordinate to a fixed latent token coordinate.

The mechanics-guided spectral basis used by the SDG-Mixer is introduced separately through the token-space stiffness–mass coupling operator described below.

3.3.6 Token-space stiffness–mass coupling and descriptor conditioning

The physics encoder constructs a building-specific token-space coupling operator from the unpadded stiffness and mass information of each structure. This operator provides the spectral coordinate system used by the SDG-Mixer and allows the fixed token representation to remain explicitly conditioned on the physical stiffness–mass characteristics of the original building.

For building b, let $\mathbf{K}_b$ denote the floor-level stiffness matrix and $\mathbf{m}_b$ denote the corresponding floor-mass vector after padded floors have been removed. The stiffness matrix is first symmetrized, and the floor masses are constrained to be positive for numerical stability. A mass-normalized stiffness operator is then computed as

$$\mathbf{D}_b = \mathrm{diag}(\mathbf{m}_b)^{-1/2}\mathbf{K}_b\mathrm{diag}(\mathbf{m}_b)^{-1/2} \tag{26}$$

This normalization expresses stiffness interactions on a mass-consistent scale and improves comparability across buildings with different floor-mass distributions. To control the numerical scale before projection to the token space, $\mathbf{D}_b$ is further normalized by its trace-averaged magnitude,

$$\widehat{\mathbf{D}}_b = \frac{\mathbf{D}_b}{\frac{|\mathrm{tr}(\mathbf{D}_b)|}{F_b} + \varepsilon_D} \tag{27}$$

where ε_D is a small positive constant. Because the learnable floor-to-token operator $\mathbf{W}_L$ is defined on the padded floor coordinate with maximum floor number $F_{\max}$, whereas each building has its own number of physical floors F_b, padded floors must be excluded from the construction of the structural coupling operator. For building b, we introduce a valid-floor embedding matrix $\mathbf{P}_b$, with size $F_{\max} \times F_b$, which maps the F_b physical floors into the padded floor coordinate used by the neural network. The corresponding building-specific lifting operator is

$$\mathbf{L}_b = \mathbf{W}_L\mathbf{P}_b \tag{28}$$

The token-space stiffness–mass coupling operator is then constructed from the normalized valid-floor stiffness–mass operator $\widetilde{\boldsymbol{D}}_b$ as

$$\mathbf{B}_{\mathrm{tok},b} = \mathbf{L}_b\widetilde{\boldsymbol{D}}_b\mathbf{L}_b{}^{\top} \tag{29}$$

This construction projects the physical stiffness–mass operator onto the fixed latent token coordinate while using only valid physical floors. Therefore, the SDG-Mixer can operate on a fixed-size spectral representation without allowing padded entries to affect the structural operator. Before spectral propagation, $\mathbf{B}_{\text{tok},b}$ is rescaled by its mean absolute magnitude and clipped to a bounded numerical range to avoid unstable spectral scales. Its eigendecomposition is then computed as

$$\mathbf{B}_{\text{tok},b} = \mathbf{U}_b \mathbf{\Lambda}_b \mathbf{U}_b^{\top} \tag{30}$$

The eigenvectors $\mathbf{U}_b$ define the spectral coordinate system of the token-space coupling operator, whereas the eigenvalues in $\mathbf{\Lambda}_b$ provide frequency-like structural coordinates for SDG-Mixer propagation. These quantities are not treated as finite-element mode shapes or physical natural frequencies. Instead, they form a learnable token-space spectral representation induced by the stiffness–mass information of the original structure.

In addition to the token-space coupling matrix, the physics encoder extracts a compact global descriptor from $\mathbf{m}_b$ and $\mathbf{D}_b$. This descriptor summarizes the mass distribution, stiffness scale, stiffness distribution, spectral range, and spectral conditioning of the structure. The mass-distribution terms are defined as

$$\mu_{m,b} = \frac{1}{F_b}\sum_{i=1}^{F_b} m_{b,i}, \ \sigma_{m,b} = \left[\frac{1}{F_b}\sum_{i=1}^{F_b}\left(m_{b,i} - \mu_{m,b}\right)^2\right]^{1/2} \tag{31}$$

The stiffness-scale and stiffness-distribution terms are computed from the mass-normalized stiffness operator as

$$\tau_{D,b} = \frac{|\mathrm{tr}(\mathbf{D}_b)|}{F_b}, \ \delta_{D,b} = \frac{1}{F_b}\sum_{i=1}^{F_b}\left|D_{b,ii}\right|, \ \phi_{D,b} = \frac{\|\mathbf{D}_b\|_F}{F_b} \tag{32}$$

Here, $\tau_{D,b}$ represents the trace-averaged stiffness scale, $\delta_{D,b}$ characterizes the average diagonal stiffness magnitude, and $\phi_{D,b}$ measures the overall stiffness-interaction strength. Let $\lambda_{b,1}$, $\lambda_{b,2}$,…, λ_{b,F_b} denote the eigenvalues of $\mathbf{D}_b$. The extremal spectral magnitudes and the spectral condition estimate are defined as

$$\lambda_{\max,b} = \max_{1\le i\le F_b}\left|\lambda_{b,i}\right|, \ \lambda_{\min,b} = \max\left(\min_{1\le i\le F_b}\left|\lambda_{b,i}\right|, \varepsilon_\lambda\right), \ \kappa_{D,b} = \frac{\lambda_{\max,b}+\varepsilon_\lambda}{\lambda_{\min,b}+\varepsilon_\lambda} \tag{33}$$

where ε_λ is a small positive constant for numerical stability. These terms encode the dominant spectral scale, the smallest resolved spectral component, and the conditioning of the stiffness spectrum. The final structural descriptor is written as

$$\mathbf{q}_b = [\log\left(1+\mu_{m,b}\right), \log\left(1+\sigma_{m,b}\right), \log\left(1+\tau_{D,b}\right), \log\left(1+\delta_{D,b}\right), \tag{34}$$

$$\log(1+\phi_{D,b}), \log(1+\lambda_{\max,b}), \log(1+\lambda_{\min,b}), \log(1+\kappa_{D,b})]$$

The logarithmic transformation reduces scale disparities among buildings while preserving the relative ordering of the underlying structural quantities. The descriptor is mapped to a 64-dimensional conditioning embedding through a two-layer multilayer perceptron,

$$\mathbf{d}_b = \mathrm{MLP}_{\mathrm{desc}}(\mathbf{q}_b) \tag{35}$$

The embedding $\mathbf{d}_b$, together with the pooled structural token representation, conditions the head-wise propagation parameters in the SDG-Mixer blocks. Therefore, the spectral propagation is governed not only by the token-space stiffness–mass coupling operator, but also by compact global descriptors of mass distribution, stiffness distribution, spectral range, and conditioning characteristics.

3.3.7 SDG-Mixer layer: Green-like spectral propagation with rational modal correction

Each SDG-Mixer layer applies Green-function–inspired propagation in the modal basis. Given the token features $\mathbf{Z} \in \mathbb{R}^{B\times N\times D}$, the features are projected into the modal basis using **U**, where the modal components are propagated through a damped-rotation scheme. The damping rate α and the phase shift θ are learned via lightweight parameter generators, conditioned on the modal frequencies ω_i and the structure descriptor *d*. The paired latent channels within each spectral component are updated through the damped-rotation operation described in the Methods section of the main text. This allows the model to adaptively modulate the damping and frequency characteristics of the propagation process according to the specific structural properties of each building, going beyond idealized proportional damping assumptions.

After the spectral components have been propagated and corrected in the building-specific spectral basis, the updated spectral features are mapped back to the latent token space through $\mathbf{U}_b$. This back-projection returns the dynamics-propagated representation to the feature space used by subsequent layers. The SDG-Mixer block is implemented with a pre-normalized residual structure to ensure stable training. A bounded residual gate, $\beta = \tanh(\rho)$, is initialized near zero so that the model can gradually increase the contribution of the spectral propagation pathway during optimization. The residual update is formulated as

$$\mathbf{Z}_{\mathrm{out},b} = \mathbf{Z}_b + \beta \cdot \mathrm{SDGMixer}(\mathrm{RMSNorm}(\mathbf{Z}_b); \mathbf{B}_{\mathrm{tok},b}; \mathbf{d}_b) \tag{36}$$

Here, $\mathbf{B}_{\mathrm{tok},b}$ is the building-specific token-space stiffness−mass coupling operator, and $\mathbf{d}_b$ is the structural descriptor embedding used to condition the head-wise propagation parameters. Thus, the quantity passed to the SDG-Mixer is not a generic attention bias, but a mechanics-guided spectral coupling operator derived from the valid-floor stiffness–mass representation of the building. This design allows the same pretrained decoder to adapt its latent propagation behaviour across different floor numbers, mass distributions and stiffness organizations while retaining a fixed token-space computation.

Following the SDG-Mixer layer, the features are passed through a feed-forward network (FFN) that refines the channel-wise nonlinear representations. The FFN utilizes the SwiGLU activation function, which has been shown to outperform traditional ReLU activations in certain settings.

3.3.8 Token-to-floor decoding and output masking

After propagation through the SDG-Mixer layers, the final token representations encode globally mixed, structure-conditioned features. These tokens are first projected onto the prediction horizon *P*, yielding

$$\mathbf{Y} = \mathbf{W}_D \mathbf{Y}_{\mathrm{tok}} \in \mathbb{R}^{B \times F_{\max} \times P} \tag{37}$$

where $\mathbf{Y}_{\mathrm{tok}}$ denotes the token-wise horizon projection and $\mathbf{W}_D$ is a learned token-to-floor mapping matrix. This linear operator redistributes the globally propagated spectral information back to physically interpretable floor-level responses.

Finally, a floor mask is applied to the reconstructed output to ensure that padded floors contribute zero to the prediction, thereby preserving structural consistency for variable-height buildings.

3.4 Numerical stability framework and mixed-precision design principles

The SeisGPT architecture adopts a mixed-precision computational strategy to balance efficiency with numerical robustness. The majority of tensor operations are executed in bfloat16 precision to reduce memory consumption and accelerate throughput, whereas numerically sensitive computations are selectively promoted to float32. In particular, matrix multiplications involving spectral operators, eigen-decomposition of the token coupling matrix, and modal token transformations are carried out in float32 to mitigate round-off accumulation, underflow, and overflow effects. This selective precision allocation enables the model to scale to large parameter counts while preserving stable forward propagation and gradient evaluation.

To further ensure numerical reliability, several stability-oriented design principles are embedded within the architecture. The spectral decomposition of the coupling operator B is performed in float32 to maintain orthogonality of eigenvectors and accuracy of eigenvalues, both of which are essential for well-conditioned modal propagation. Residual gating mechanisms, including the structural injection gate Υ and the SDG-Mixer gate, are initialized near zero, thereby constraining their initial contribution and promoting gradual adaptation during early training stages. This controlled residual activation reduces the risk of abrupt parameter updates that could destabilize optimization.

3.5 Building-specific few-shot adaptation with low-rank adaptation

Low-rank adaptation (LoRA) was used only for the building-specific few-shot adaptation experiments reported in Figure 4 of the main text. It was not used to obtain SeisGPT-Enhanced. SeisGPT-Enhanced was obtained by fine-tuning SeisGPT-Base on the real-building finite-element response dataset, whereas LoRA was subsequently applied to specialize this globally fine-tuned model to individual buildings using only a small number of response records.

For a pretrained weight matrix $\mathbf{W} \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$, LoRA represents the adapted weight as

$$\mathbf{W}_{\text{adapted}} = \mathbf{W} + \Delta\mathbf{W},\ \Delta\mathbf{W} = \frac{\alpha_{\text{LoRA}}}{r}\mathbf{B}_{\text{LoRA}}\mathbf{A}_{\text{LoRA}} \tag{38}$$

where $\mathbf{A}_{\text{LoRA}} \in \mathbb{R}^{r \times d_{\text{in}}}$, $\mathbf{B}_{\text{LoRA}} \in \mathbb{R}^{d_{\text{out}} \times r}$, $r \ll \min(d_{\text{in}}, d_{\text{out}})$ is the adaptation rank, and α_{LoRA} is the scaling factor. During building-specific adaptation, the original pretrained weights were frozen and only the low-rank matrices were optimized. This design substantially reduced the number of trainable parameters and allowed rapid per-building specialization.

Starting from SeisGPT-Enhanced, we evaluated four LoRA placement strategies: adapters inserted into all linear layers, adapters inserted only into the encoder-side layers, adapters inserted only into the decoder-side layers, and adapters inserted only into the output projection layer. For each target building, between one and nine available response histories were used for adaptation, and one additional response history from the same building was reserved for independent evaluation. The target of this few-shot adaptation experiment was relative displacement response prediction. The same optimizer settings and adaptation schedule were used across all LoRA placement strategies, so that the comparison reflected the effect of adapter placement rather than differences in training budget. In all LoRA adaptation experiments, the rank was r=256, the scaling factor was α_{LoRA}=512, and adaptation was performed for 5 epochs using AdamW with learning rate 5×10^{-5}.

This separation between global fine-tuning and building-specific LoRA adaptation is important. The former aligns the pretrained SeisGPT representation with the distribution of real-building finite-element responses, whereas the latter provides a lightweight mechanism for rapidly personalizing the already fine-tuned model to an individual structure when only a few response records are available.

4. SeisGPT Experiment Details and Hyperparameter Experiments

This section presents the comprehensive experimental framework supporting the development, validation, and optimization of the SeisGPT model. Three key aspects are discussed: the training of SeisGPT-Base on large-scale synthetic data, the domain adaptation to real-world structural records via fine-tuning (SeisGPT-Enhanced), and a systematic investigation into the influence of temporal sequence lengths on prediction performance.

4.1 Training and Performance of SeisGPT-Base on synthetic building structural response dataset

4.1.1 Pretraining protocol and temporal sample construction

SeisGPT-Base was pretrained on the synthetic building structural response dataset generated from nonlinear time-history analyses of code-compliant frame, frame–shear wall and shear wall buildings. The training and evaluation partitions were defined at the building level, so that finite-element models used for evaluation were not included in the pretraining set. Separate SeisGPT-Base models were trained for absolute acceleration prediction and relative displacement prediction, using the same temporal sampling protocol and optimization settings.

All response histories were sampled at $\Delta t = 0.02$ s. Training samples were generated using a sliding-window strategy. For each window, the model received the external excitation and the SDR-prior response over a sequence of $T_m + T_p$ time steps, where T_m=1000 is the context length and T_p=20 is the prediction horizon. Thus, the input sequence length used by the network was 1020 time steps. The supervised target, however, consisted only of the high-fidelity nonlinear finite-element response over the final 20 time steps. In other words, the model used known excitation information, SDR-prior response and structural descriptors to predict the nonlinear finite-element response over the target horizon.

The use of excitation and SDR-prior information over the target horizon does not introduce leakage from the high-fidelity response target. In response-history analysis, the external excitation record is an imposed input and is therefore known over the prediction interval. The SDR prior is also computed independently from the reduced stiffness–mass system under the same imposed excitation. It is a physics-derived coarse response estimate rather than a nonlinear finite-element target. The network parameters were optimized only against the high-fidelity nonlinear OpenSees response over the final T_p time steps. Consecutive samples were produced by shifting the window along each response record, allowing longduration response histories to be represented by independent windows and enabling parallel inference without fully autoregressive rollout.

Model training was conducted with the AdamW[15] optimizer, using β_1= 0.9, β_2=0.999 and a weight decay of 2×10^{-2}. The learning rate was set to 1×10^{-5} and scheduled with cosine decay after an initial linear warm-up of 4,000 optimization steps. Distributed training was implemented with DeepSpeed[16]. ZeRO-1[17] optimization was used to shard optimizer states across devices while keeping full model parameters on each device. Gradient accumulation was applied to increase the effective batch size without requiring all samples to be stored in GPU memory simultaneously. The majority of tensor operations were performed in bfloat16 precision, whereas numerically sensitive operations, including spectral-operator construction, eigendecomposition and modal-token transformations, were promoted to float32 precision for numerical stability.

The temporal configuration T_m=1000 and T_p=20 was used as the default setting for the main experiments. The sensitivity of prediction accuracy to the context length and prediction horizon is reported separately in Supplementary Tables 5 and 6.

4.1.2 Complete quantitative comparison and complementary visual validation of SeisGPT-Base

To complement the summary comparisons presented in Figure 2 of the main text, Supplementary Table 2 reports the complete numerical performance of SeisGPT-Base and all benchmark models across frame, frame–shear wall and shear wall structures. The table provides the Pearson correlation coefficient, floor-wise normalized mean absolute error and floor-wise normalized root mean squared error for both acceleration and displacement prediction. These results confirm that the performance advantages highlighted graphically in the main text are maintained consistently across all three structural categories and both response quantities.

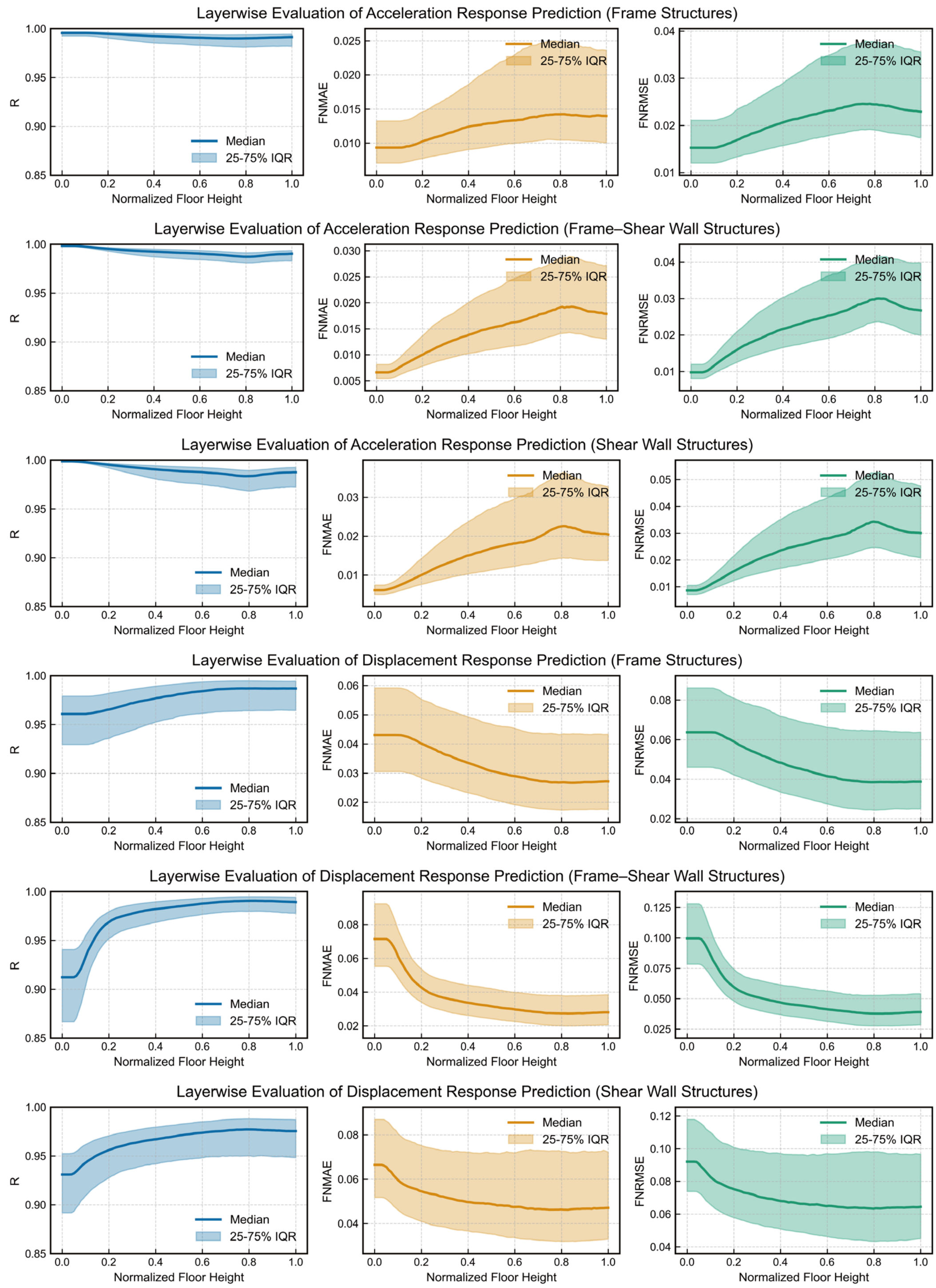
To rigorously assess the predictive performance of SeisGPT-Base across variations in building height and structural typology, a series of comprehensive post-training evaluations were performed. These analyses were designed to elucidate model behavior not only in aggregate but also with respect to vertical structural position, dynamic response type, and structural topology.

Supplementary Table 2 | Quantitative comparison of structural response prediction performance across models and structural types.

Structure Type	Models	R		FNMAE		FNRMSE	
		Acceleration	Displacement	Acceleration	Displacement	Acceleration	Displacement
Frame Structure	Phy-Seisformer	0.7554	0.8535	0.1130	0.1229	0.1542	0.1640
	SeisGRU	0.8313	0.8015	0.0624	0.0685	0.0954	0.1023
	SeisLSTM	0.8317	0.8713	0.0611	0.0771	0.0932	0.1134
	TimesNet	0.8507	0.8880	0.0575	0.0685	0.0869	0.1013
	Timer-XL	0.8281	0.8603	0.0538	0.0706	0.0924	0.1611
	N-Beats	0.7533	0.8359	0.0737	0.0856	0.1130	0.1266
	Informer	0.7080	0.7902	0.1465	0.4401	0.1902	0.4850
	TimeMixer	0.7522	0.3864	0.0752	0.1134	0.1163	0.2307
	SeisGPT-Base	**0.9840**	**0.9621**	**0.0174**	**0.0377**	**0.0287**	**0.0563**
Frame-Shear Wall Structure	Phy-Seisformer	0.7620	0.8887	0.1289	0.1270	0.1793	0.1899
	SeisGRU	0.8684	0.9154	0.0559	0.0738	0.0864	0.1127
	SeisLSTM	0.8668	0.8915	0.0558	0.0930	0.0864	0.1473
	TimesNet	0.8906	0.9018	0.0523	0.0729	0.0795	0.1121
	Timer-XL	0.8517	0.8793	0.0516	0.0764	0.0885	0.1324
	N-Beats	0.7690	0.8129	0.0729	0.1125	0.1135	0.2025
	Informer	0.6854	0.6678	0.2076	0.4142	0.3074	0.5287
	TimeMixer	0.7581	0.4080	0.0761	0.1832	0.1213	0.2712
	SeisGPT-Base	**0.9869**	**0.9684**	**0.0176**	**0.0395**	**0.0284**	**0.0592**
Shear Wall Structure	Phy-Seisformer	0.8259	0.8667	0.0628	0.1053	0.1018	0.1442
	SeisGRU	0.8765	0.8713	0.0543	0.0909	0.0842	0.1282
	SeisLSTM	0.8512	0.8573	0.0588	0.0946	0.0915	0.1371
	TimesNet	0.9187	0.8554	0.0459	0.0985	0.0712	0.1437
	Timer-XL	0.8317	0.8286	0.0550	0.1456	0.0932	0.3275
	N-Beats	0.7195	0.7667	0.0766	0.1822	0.1198	0.4602
	Informer	0.6739	0.8072	0.1388	0.1972	0.2330	0.2890
	TimeMixer	0.7356	0.3902	0.0752	0.2019	0.1219	0.3454
	SeisGPT-Base	**0.9803**	**0.9481**	**0.0202**	**0.0610**	**0.0329**	**0.0837**

First, a statistical floor-wise performance analysis was conducted using response time histories from a dataset comprising 3,000 synthetically generated buildings, equally distributed among three structural typologies: frames, frame–shear wall hybrids, and shear wall systems. For each building, complete acceleration and displacement time histories were computed for all floors. To construct the floor-wise performance profiles presented in Supplementary Figure 10, a statistical interpolation and aggregation procedure was applied to the test set. For each structure, SeisGPT-Base produced full-sequence predictions of both acceleration and displacement time histories at every storey. To enable cross-building comparison irrespective of structural height, floor indices were normalized by the total number of storeys, yielding a relative height coordinate $h = f/F \in [0,1]$. Prediction metrics were linearly interpolated to 100 uniformly spaced height levels (Δh=0.01) for each building, generating a standardized data matrix amenable to collective statistical analysis.

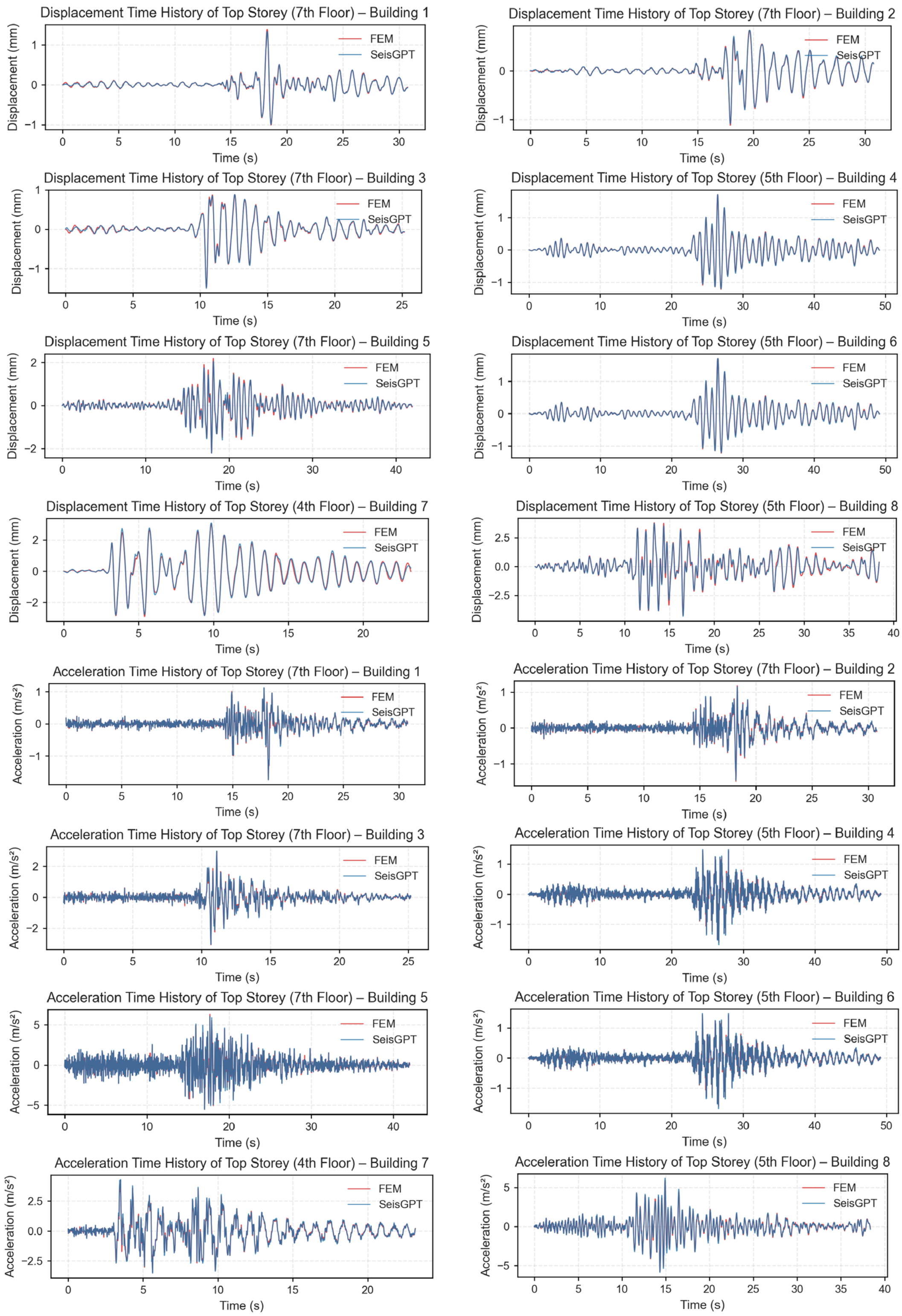
At each interpolated level, three performance metrics were computed: the Pearson correlation coefficient (R), the floor-wise normalization mean absolute error (FNMAE), and the floor-wise normalization root-mean-square error (FNRMSE). For each metric, the median, 25th percentile, and 75th percentile across all buildings were evaluated to robustly characterize central trends and variability. The resulting distributions reveal distinct vertical patterns in predictive accuracy. For acceleration responses, SeisGPT-Base achieves its highest fidelity at lower storeys, with performance gradually deteriorating toward mid-height and reaching a nadir near $h \approx$0.8, before modestly recovering at the roof level. This trend likely reflects the influence of complex boundary conditions, increased stiffness gradients, and ground input variability in the lower levels. In contrast, displacement predictions exhibit a monotonic improvement with increasing height, attaining peak accuracy in the uppermost 20% of storeys—consistent with the predominance of first-mode deformation and reduced modal complexity near the roof. These results demonstrate the model's proficiency in capturing near-ground dynamic complexity and upper-storey modal responses, while also revealing limitations in its accuracy for mid-height acceleration prediction and low-elevation displacement prediction.



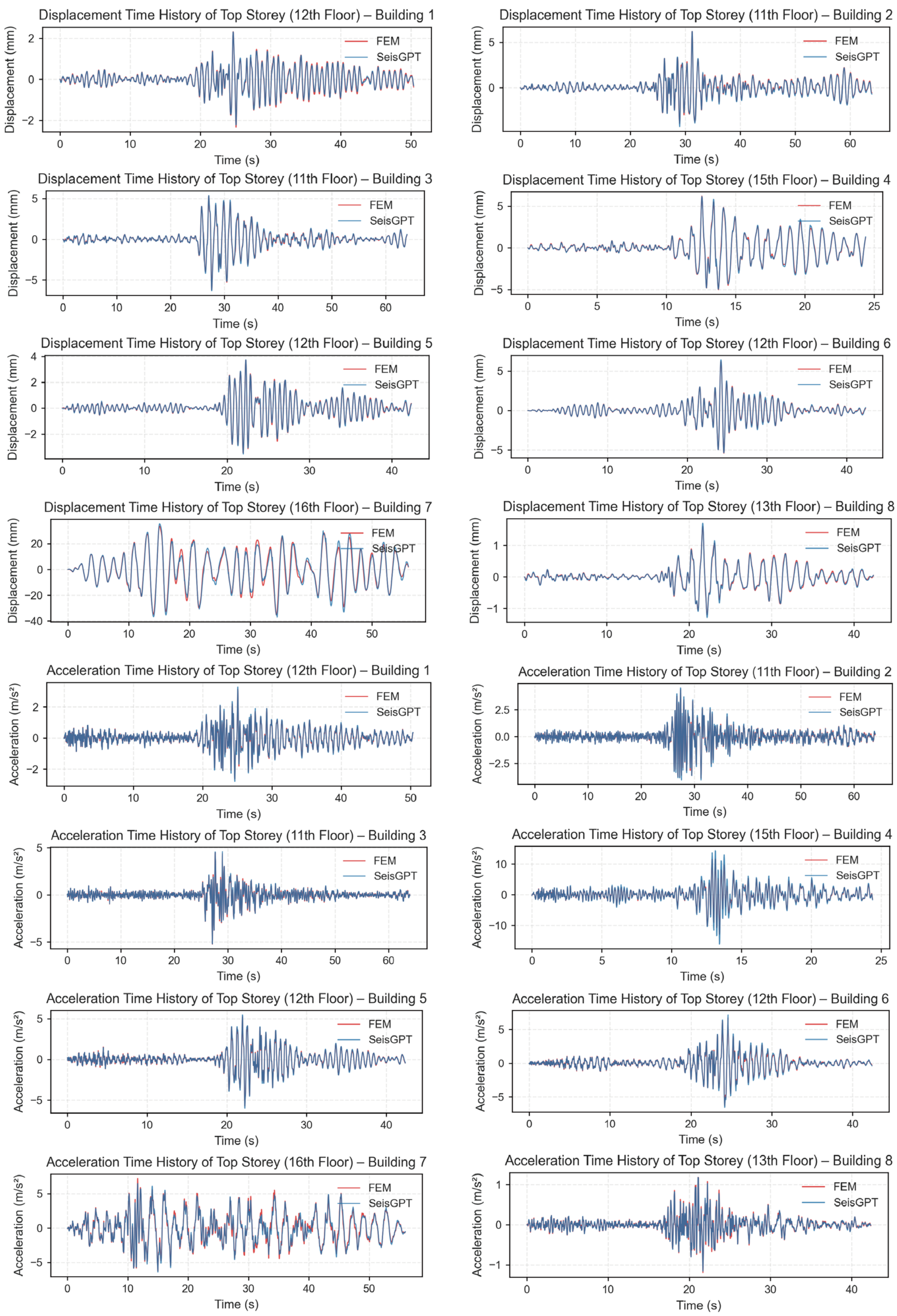
Supplementary Figure 10 | Prediction performance of SeisGPT-Base across relative floor positions in different structural types. Layer-wise prediction performance of SeisGPT-Base for acceleration and displacement responses in frame, frame–shear wall and shear wall structures. The normalized floor height ranges from the bottom storey to the roof, enabling performance trends to be compared across buildings with different numbers of floors. For each structural type and response quantity, the three columns show R, FNMAE and FNRMSE, respectively. Solid curves denote the median value at each normalized floor height, and shaded bands denote the 25th–75th percentile interquartile range across evaluated buildings. Source data are provided as a Source Data file.

Second, to examine robustness under varying dynamic inputs and structural configurations, detailed time-history comparisons were conducted for a representative subset of buildings. Ten exemplar structures were selected from each typology, encompassing a range of plan geometries, aspect ratios, and stiffness distributions. For each selected building, the predicted and reference acceleration and displacement time histories at the top floor were plotted under diverse ground motions. Visualization results are shown in Supplementary Figure 11, Supplementary Figure 12, and Supplementary Figure 13. These comparisons demonstrate that SeisGPT-Base accurately reproduces both amplitude and temporal characteristics of the structural response, including peak accelerations, phase transitions, and damping-related decay. Representative full-building response comparisons are provided in Supplementary Figure 14.

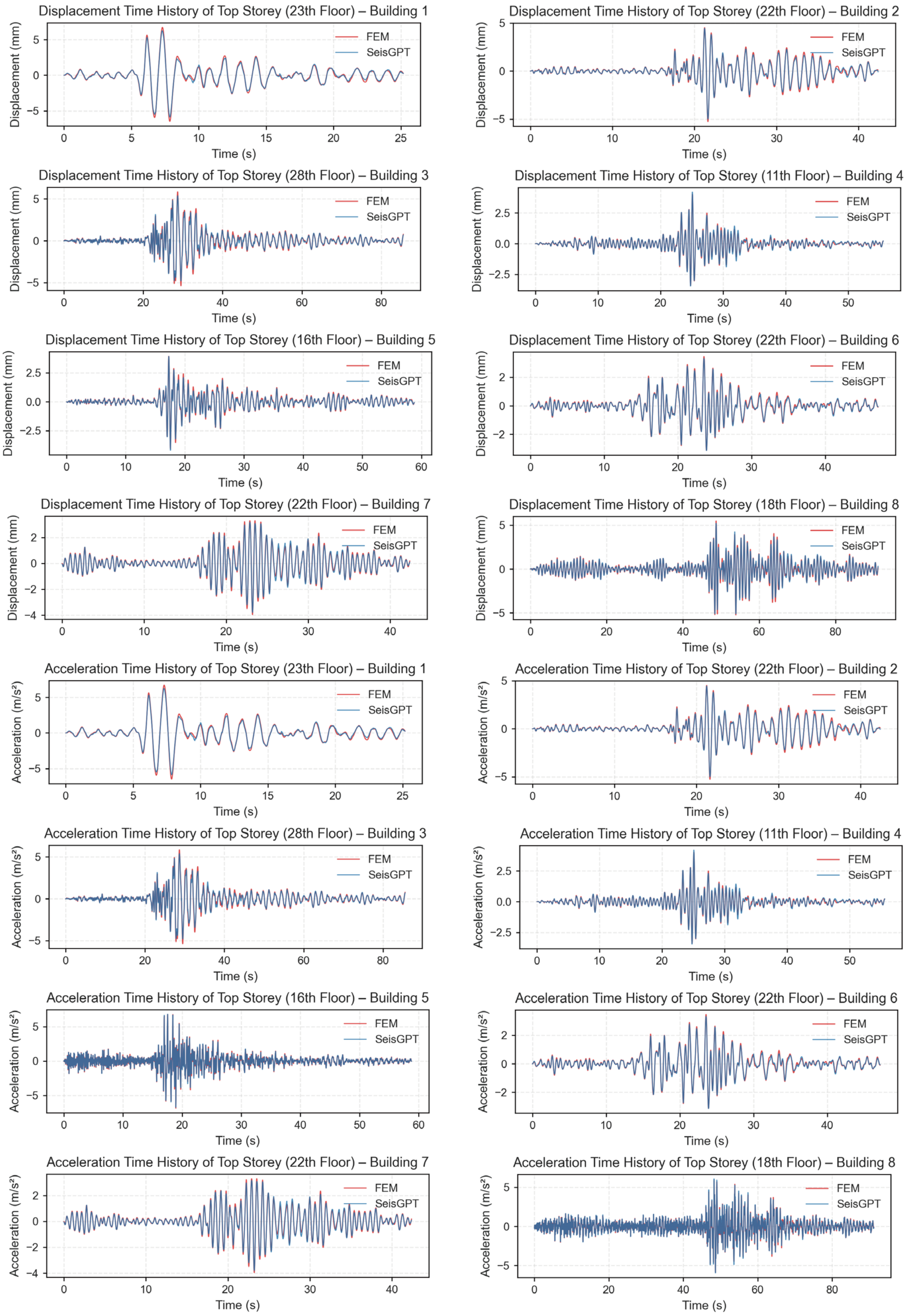
Together, these results provide compelling evidence that SeisGPT-Base achieves height-invariant accuracy, structural-type generalization, and input-motion robustness, positioning it as a reliable surrogate model for high-resolution structural response prediction under external excitation.



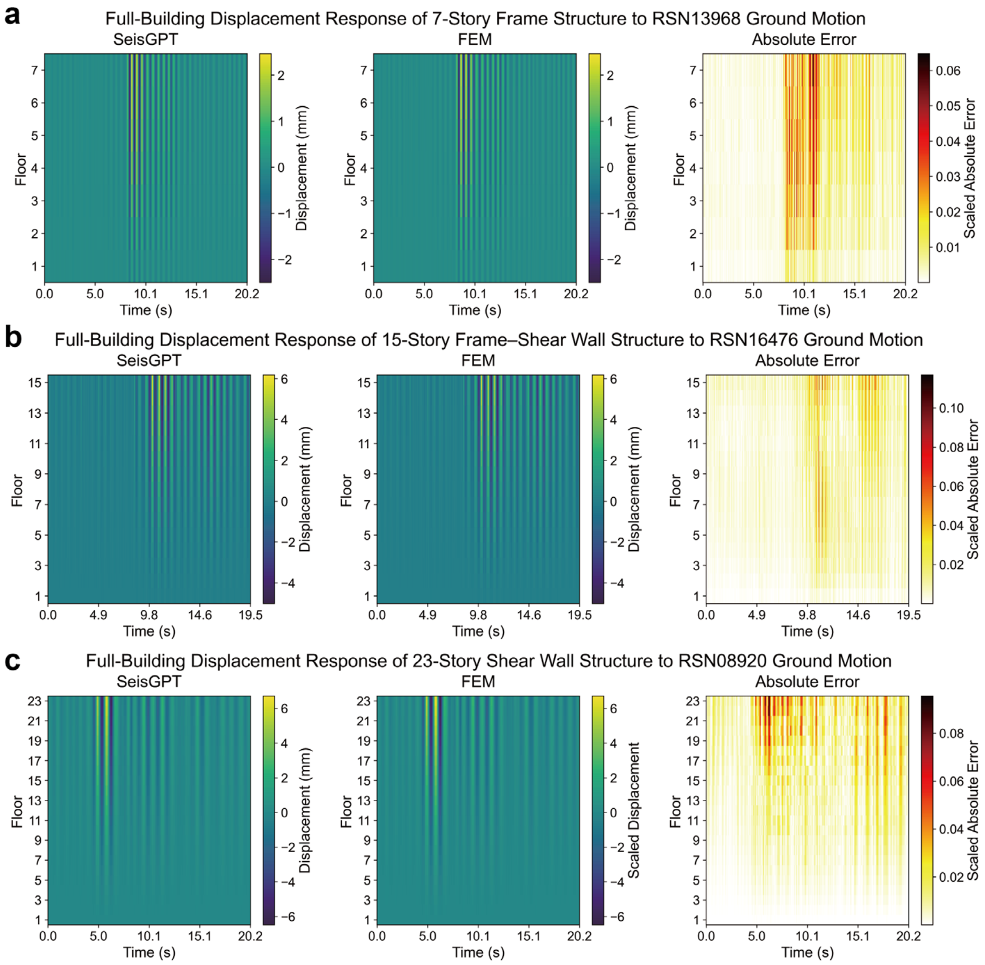
Supplementary Figure 11 | Visualization of top-storey acceleration and displacement responses in frame structures. Source data are provided as a Source Data file.



Supplementary Figure 12 | Visualization of top-storey acceleration and displacement responses in frame-shear wall structures. Source data are provided as a Source Data file.



Supplementary Figure 13 | Visualization of top-storey acceleration and displacement responses in shear wall structures. Source data are provided as a Source Data file.



Supplementary Figure 14 | Full-building structural response predictions of SeisGPT-Base compared with FEM across three representative building typologies. a, Prediction comparison for a 7-story frame structure. **b**, Results for a 15-story frame–shear wall structure. **c**, Predictions for a 23-story shear wall structure. All responses have been normalized by dividing by the peak absolute value of the ground motion, thus providing a scale-independent basis to evaluate and visualize the prediction accuracy of the model relative to FEM. Source data are provided as a Source Data file.

4.2 Fine-Tuning and Evaluation of SeisGPT-Enhanced on Real-World building structural response dataset

To promote generalization beyond synthetically generated data and enhance applicability to real-world structural monitoring scenarios, the pretrained SeisGPT-Base model was further fine-tuned using a curated dataset comprising nonlinear simulations of 694 as-built buildings. This dataset spans a broad spectrum of structural typologies, geometries, and material characteristics, thereby introducing significantly greater heterogeneity and complexity compared to the synthetic pretraining data.

Fine-tuning was conducted over 10 epochs using the same optimization pipeline as in the pretraining phase. The learning rate was initialized at 1×10^{-5} and annealed following a cosine decay schedule, with a warm-up period of 4,000 steps to ensure stable convergence during the early training iterations. Full-parameter fine-tuning was employed, with no use of parameter-efficient techniques such as Low-Rank Adaptation (LoRA), in order to retain the model's full representational capacity and enable unconstrained adaptation to the statistical properties of real structural responses.

The complete numerical results supporting the real-building evaluation in Figure 3 of the main text are provided in Supplementary Table 3 and Supplementary Table 4. Supplementary Table 3 reports the full comparison between SeisGPT-Enhanced and the baseline models on the 65-building real-world test set, whereas Supplementary Table 4 isolates the respective contributions of synthetic pretraining and real-data fine-tuning through the A1–A3 configurations. Together, these statistics show that the gain observed after fine-tuning is not merely a model-capacity effect, but arises from the combination of transferable representations learned during large-scale synthetic pretraining and their subsequent adaptation to real structural-response distributions.

Supplementary Table 3 | Quantitative evaluation of SeisGPT fine-tuning performance on real-world building response dataset.

Models	***R***		**FNMAE**		**FNRMSE**	
	Acceleration	**Displacement**	**Acceleration**	**Displacement**	**Acceleration**	**Displacement**
Phy-Seisformer	0.8091	0.8668	0.0660	0.1159	0.1042	0.1559
SeisGRU	0.8551	0.8754	0.0656	0.0901	0.0858	0.1267
SeisLSTM	0.8152	0.8544	0.0617	0.0947	0.0957	0.1371
TimesNet	0.9145	0.8606	0.0470	0.1036	0.0719	0.1453
Timer-XL	0.8062	0.8127	0.0560	0.1599	0.0960	0.3459
N-Beats	0.6721	0.7572	0.0775	0.1800	0.1217	0.4184
Informer	0.5805	0.8085	0.3300	0.3008	0.4227	0.4049
TimeMixer	0.7174	0.3686	0.0748	0.1999	0.1230	0.3395
SeisGPT-Enhanced	**0.9829**	**0.9423**	**0.0195**	**0.0639**	**0.0317**	**0.0882**

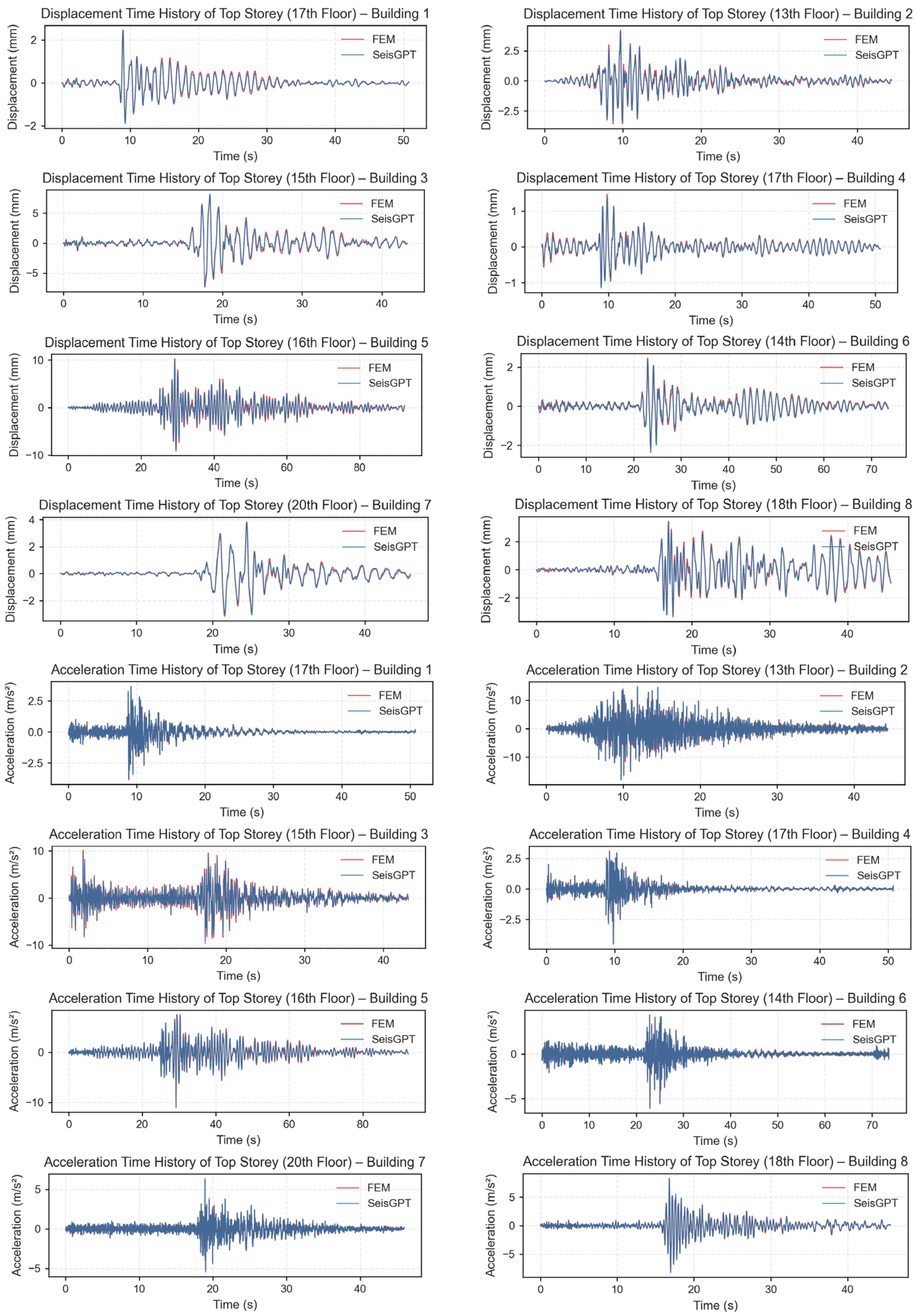
Supplementary Table 4 | Impact of pretraining and fine-tuning on SeisGPT performance for real-world building structural response prediction.

Models	***R***		**FNMAE**		**FNRMSE**	
	Acceleration	**Displacement**	**Acceleration**	**Displacement**	**Acceleration**	**Displacement**
A1	0.9079	0.8406	0.0524	0.1189	0.0747	0.2404
A2	0.9781	0.9359	0.0227	0.0666	0.0361	0.0924
A3	**0.9829**	**0.9423**	**0.0195**	**0.0639**	**0.0317**	**0.0882**

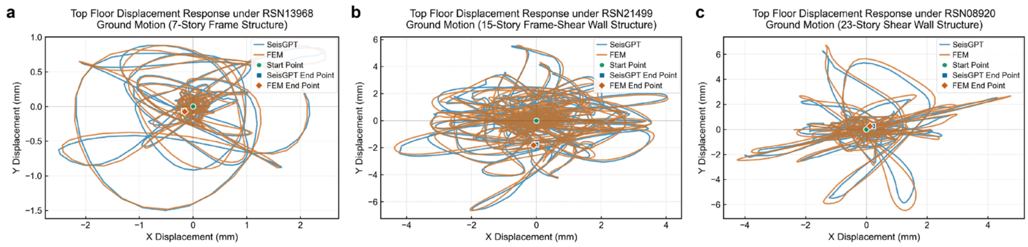
Comparison of three model configurations evaluated on 65 real buildings: A1, trained from scratch solely on the real buildings' dataset (no synthetic pretraining; A2, SeisGPT-Base pretrained on the large-scale synthetic building dataset and evaluated on real buildings without real-data fine-tuning; and A3, SeisGPT-Enhanced obtained by fine-tuning SeisGPT-Base on the real buildings dataset. Pretraining on the large-scale synthetic building dataset (A2) markedly improves accuracy over training from scratch on the limited real-world building response dataset (A1), and subsequent fine-tuning on the real-world dataset (A3) yields a further consistent gain across both acceleration and displacement metrics.

To further assess generalization under varied excitations and structural configurations, a representative subset of 9 real buildings was selected for qualitative analysis. These structures reflect diverse plan geometries, aspect ratios, and dynamic properties. Roof-level acceleration and displacement time histories were visualized and compared against ground-truth simulations. As shown in Supplementary Figure 15, the predicted responses consistently tracked key dynamic features, including peak amplitudes and transient decay patterns, with minimal phase lag—demonstrating strong agreement with the reference data. The model also demonstrated high-fidelity trajectory prediction performance in representative buildings from three structural categories, as shown in Supplementary Figure 16.

These results confirm that SeisGPT-Enhanced effectively captures complex structural dynamics and maintains high predictive accuracy under real-world variability. The model's demonstrated capability to recover fine-grained response characteristics at critical locations underscores its utility in operational scenarios involving post-earthquake damage screening, rapid prioritization of structural risks, and resilience-informed decision-making.



Supplementary Figure 15 | Visualization of top-storey acceleration and displacement responses in real-world structures. Source data are provided as a Source Data file.



Supplementary Figure 16 | Top-floor displacement trajectories predicted by SeisGPT-Enhanced compared with FEM under seismic excitations. The figure shows top-floor displacement trajectories in the x and y directions for three representative building structures, with SeisGPT-Enhanced predictions compared against FEM results. Although the trajectories are plotted in two dimensions, the predictions in each direction were computed independently and subsequently combined for visualization. **a**, 7-story frame structure. **b**, 15-story frame–shear wall structure. **c**, 23-story shear wall structure. Source data are provided as a Source Data file.

4.3 Effect of Temporal Sequence Lengths on Predictive Performance

To systematically investigate the influence of temporal hyperparameters on predictive performance, a series of controlled ablation studies was conducted, with a specific focus on the lengths of both input and output sequences. These experiments aimed to identify the optimal temporal configuration for time-series forecasting of structural responses subjected to external excitation.

Models were trained and evaluated using varying input sequence lengths L of 50, 100, 500, and 1,000 time steps, and output sequence lengths P of 1, 5, 10, 20, and 50 time steps. All configurations employed the SeisGPT architecture, utilizing the standard two-stage training framework: initial pretraining on the synthetic structural dataset, followed by fine-tuning on the real-world building dataset. The analysis was restricted to the prediction of relative displacement responses to isolate the effects of temporal windowing on structural dynamics forecasting.

The results, summarized in Supplementary Table 5 and Supplementary Table 6, reveal consistent patterns. Increasing L led to marked improvements in model accuracy, reflecting the benefit of extended temporal context for capturing structural memory effects, damping behavior, and multi-modal interactions. The relationship between P and predictive performance was non-monotonic. Shorter prediction windows require more frequent window evaluations to cover a long response record and may reduce temporal continuity across adjacent predicted segments.

Supplementary Table 5 | Influence of historical input length *L* on predictive performance

L	**50**	**100**	**500**	**1000**
R	0.9256	0.9356	**0.9428**	0.9423
FNMAE	0.0723	0.0668	0.0639	**0.0639**
FNRMSE	0.1002	0.0935	0.0888	**0.0882**

Supplementary Table 6 | Influence of output sequence length *P* on predictive performance

P	**1**	**5**	**10**	**20**	**50**
R	0.9292	0.9416	0.9420	**0.9423**	0.9422
FNMAE	0.0689	0.0640	**0.0635**	0.0639	0.0636
FNRMSE	0.0964	0.0895	0.0887	**0.0882**	0.0884

4.4 Ablation on Decoder Backbone

To directly quantify the contribution of the proposed Spectral Duhamel–Green (SDG) Mixer as the decoder backbone, we performed a controlled ablation in which all components of SeisGPT were kept identical except for the decoder. The objective of this experiment is to assess whether SDG-Mixer can achieve Transformer-level accuracy on real-world buildings after fine-tuning, while improving model compactness and deployment efficiency (FLOPs and inference speed).

4.4.1 Compared variants

We evaluated two model variants that differ exclusively in the decoder backbone.

(i) SeisGPT (SDG-Mixer decoder; proposed).

The decoder consists of a stack of SDG-Mixer blocks that implement dynamics-informed global token mixing through spectral propagation combined with bounded residual gating. All other architectural components, including the encoder, embedding layers, conditioning mechanisms, output head, and loss formulation, are kept identical to the baseline configuration.

(ii) SeisGPT (Transformer decoder; baseline).

In the baseline model, the SDG-Mixer is replaced with a standard Transformer decoder. The implementation follows contemporary large language model design practices, incorporating RMSNorm normalization, SwiGLU feed-forward networks, and rotary positional embeddings (RoPE). For computational efficiency, the attention mechanism adopts grouped-query attention or multi-query attention. The decoder depth and hidden dimensionality are configured to match the overall parameter scale of SeisGPT used in our prior implementation.

Empirical observations indicate that a Transformer-based decoder requires substantially more parameters to achieve comparable modeling capacity for long-range spatiotemporal dependencies, whereas the SDG-Mixer attains similar expressivity through a more structured mixing operator. Consequently, the Transformer baseline in this ablation contains more parameters than the SDG-Mixer variant, as reported in Supplementary Table 7. This setup renders the comparison conservative with respect to the accuracy and efficiency trade-off.

4.4.2 Training, Evaluation, and Efficiency Protocol

All model variants were trained, evaluated, and benchmarked under an identical experimental and computational protocol to ensure a controlled and unbiased comparison. Experiments were conducted on the real-world building dataset using the same train, validation, and test partitions as in the main experiments, where test buildings were strictly disjoint from those used for training.

Both variants were fine-tuned using the same two-stage training strategy and identical optimization configuration, including the AdamW optimizer with a learning rate of 1×10^{-5}, batch size of 512, bfloat16 numerical precision, and a total of 10 training epochs, as described in Section 4.1. All experiments adopted a historical input length of L=1000 and a prediction horizon of P=20. Predictive performance was evaluated using the Pearson correlation coefficient (R), FNMAE, and FNRMSE, computed for both relative displacement and absolute acceleration. Reported results are aggregated across all floors and all test sequences and presented as mean values. Deployment efficiency was assessed in terms of floating-point operations, latency, throughput, and memory usage under the fixed sequence configuration L=1000 and P=20. FLOPs per forward inference were estimated using the PyTorch profiler and are reported in GFLOPs. Latency and throughput were measured using PyTorch CUDA events with synchronization enabled after 4000 warm-up iterations. Latency is reported in milliseconds per sample at batch size 1, and throughput is reported in samples per second under the same runtime configuration. All efficiency measurements were performed on a single NVIDIA A800 GPU using bf16 precision inference, with PyTorch 2.1 and CUDA 11.8. Peak GPU memory during inference was recorded using torch.cuda.max_memory_allocated().

4.4.3 Results: Accuracy–Efficiency Trade-off on Real-World Buildings after Fine-Tuning

Supplementary Table 7 reports a decoder backbone ablation study conducted within an otherwise identical SeisGPT pipeline. Displacement forecasting performance is evaluated on four datasets: real-world buildings after fine-tuning, together with three representative structural subsets derived from synthetic building structures generated via finite element simulations, namely Frame, Frame–Shear Wall, and Shear Wall structures. Deployment-oriented efficiency metrics are also reported, including model size, FLOPs, and inference throughput.

Supplementary Table 7 | Decoder-backbone ablation on real-world buildings after fine-tuning (accuracy and efficiency)

Category	Dataset / Setting	Metric	Transformer decoder (baseline)	SDG-Mixer decoder (proposed)	Relative change
Accuracy	Frame structures	*R*	0.9495	0.9621	**+1.33%**
		FNMAE	0.0449	0.0377	**-16.04%**
		FNRMSE	0.0664	0.0563	**-15.21%**
	Frame–shear wall structures	*R*	0.9676	0.9684	**+0.08%**
		FNMAE	0.0422	0.0395	**-6.40%**
		FNRMSE	0.0616	0.0592	**-3.90%**
	Shear wall structures	*R*	0.9443	0.9481	**+0.40%**
		FNMAE	0.0624	0.0610	**-2.24%**
		FNRMSE	0.0862	0.0837	**-2.90%**
	Real-world structures	*R*	0.9292	0.9423	**+1.41%**
		FNMAE	0.0745	0.0639	**-14.23%**
		FNRMSE	0.1213	0.0882	**-27.29%**
Model size	Architecture-level; independent of dataset	Total parameters (M)	1764.08	570.49	**−67.7% (SDG is 3.09× smaller)**
		Decoder parameters (M)	1214.30	304.93	**−74.8% (SDG is 3.98× smaller)**
Compute	-	FLOPs per inference (GFLOPs)	761.07	185.12	**-75.7% (SDG is 4.11× smaller)**
Runtime	(batch size=1; same hardware/precision)	Throughput (samples/s)	54.47	127.41	**+133.9% (SDG is 2.34× faster)**

Overall, the proposed SDG-Mixer achieves Transformer-level displacement accuracy on the fine-tuned real-world test set and consistently maintains comparable or slightly improved performance across the three synthetic structural subsets. Importantly, SDG-Mixer attains this performance with substantially improved parameter efficiency, requiring far fewer total and decoder parameters than the Transformer baseline. This reduction in model size correspondingly decreases computational cost in terms of FLOPs and increases inference throughput. These findings indicate that the proposed dynamics-grounded mixing operator preserves predictive fidelity while markedly improving the accuracy–efficiency trade-off, which is particularly advantageous for time-critical structural response forecasting and scalable deployment.

For correlation R (higher is better), we report the relative improvement as $\frac{R_{SDG}-R_{Trans}}{|R_{Trans}|} \times 100\%$. For error metrics (FNMAE and FNRMSE; lower is better), we report the relative error reduction as $\frac{E_{Trans}-E_{SDG}}{|E_{Trans}|} \times 100\%$. For efficiency metrics where lower is better (parameters, FLOPs, latency), we report the relative reduction using the same form; for throughput (higher is better), we report relative improvement.

4.5 Additional quantitative analyses supporting cross-system generalization and robustness evaluation

4.5.1 Zero-shot quantitative comparison across unseen structural systems

The zero-shot generalization analysis summarized in Figure 5a–c of the main text was further documented through a complete numerical comparison across steel, masonry and base-isolated structures. Supplementary Table 8 reports the Pearson correlation coefficient, floor-wise normalized mean absolute error and floor-wise normalized root mean squared error for SeisGPT-Base and all benchmark models in each unseen structural category. The detailed statistics confirm that the performance advantage of SeisGPT-Base is consistently maintained across systems that differ substantially from the primary reinforced-concrete pretraining distribution in material behaviour, stiffness organization and energy-dissipation mechanisms.

4.5.2 Sensitivity to temporal resolution

To complement the temporal-resolution analysis shown in Figure 5e of the main text, Supplementary Table 9 reports the complete numerical performance of SeisGPT-Base across sampling intervals ranging from 0.001 s to 0.100 s. The model achieves its best performance near the training resolution of 0.020 s and remains accurate under moderate deviations from this setting, while pronounced degradation emerges at substantially coarser temporal discretizations. These results quantify the operating range over which a model trained at a fixed resolution can be applied without retraining and clarify when direct adaptation to the deployment sampling rate becomes preferable.

Supplementary Table 8 | Zero-shot prediction performance of SeisGPT and baseline models across steel, masonry, and base-isolated structures.

Structure Type	Models	R	FNMAE	FNRMSE
Steel Structure	Phy-Seisformer	0.6665	0.1716	0.2387
	SeisGRU	0.7210	0.1610	0.2274
	SeisLSTM	0.5307	0.1685	0.2376
	TimesNet	0.5268	0.1659	0.2338
	Timer-XL	0.8175	0.1045	0.1551
	N-Beats	0.4664	0.6886	0.8558
	Informer	0.3077	1.3679	1.5442
	TimeMixer	0.1730	0.1842	0.2582
	SeisGPT-Base	**0.9675**	**0.0340**	**0.0535**
Masonry Structure	Phy-Seisformer	0.0789	0.1726	0.2127
	SeisGRU	0.2899	0.1571	0.2186
	SeisLSTM	0.1456	0.1505	0.2147
	TimesNet	0.0787	0.1491	0.2140
	Timer-XL	0.5294	0.1701	0.2057
	N-Beats	0.1424	0.7972	1.3555
	Informer	0.2346	0.1566	0.2188
	TimeMixer	-0.0129	1.1805	1.8541
	SeisGPT-Base	**0.8654**	**0.1002**	**0.1471**
Base-isolated Structure	Phy-Seisformer	0.2882	0.1985	0.2542
	SeisGRU	0.5684	0.1947	0.2859
	SeisLSTM	0.4901	0.1959	0.2874
	TimesNet	0.1840	0.1961	0.2875
	Timer-XL	0.7147	0.1565	0.1649
	N-Beats	0.5347	0.2950	0.4187
	Informer	0.3102	0.1959	0.2870
	TimeMixer	0.1052	0.5478	0.7176
	SeisGPT-Base	**0.8731**	**0.0827**	**0.1217**

Supplementary Table 9 | Influence of temporal resolution (Δt) on predictive performance.

Δt (s)	0.001	0.002	0.005	0.01	0.02	0.03	0.04	0.08	0.1
R	0.8653	0.8612	0.8951	0.9313	**0.9595**	0.9519	0.9401	0.8805	0.8478
FNMAE	0.1077	0.1095	0.0920	0.0693	**0.0460**	0.0555	0.0672	0.1020	0.1145
FNRMSE	0.1461	0.1502	0.1255	0.0964	**0.0663**	0.0877	0.1131	0.1857	0.2117

4.5.3 Influence of pretraining data scale

The data-scaling trend visualized in Figure 5f of the main text is quantified in Supplementary Table 10. When the proportion of available pretraining data increases from 50% to 100%, predictive accuracy improves markedly, with the Pearson correlation coefficient rising from 0.8551 to 0.9596 and the corresponding normalized error metrics decreasing substantially. Although the results at very small data fractions do not follow a strictly monotonic trajectory, the overall pattern demonstrates that the structural-response foundation model benefits strongly from the expansion of the pretraining corpus, supporting the central role of large-scale physics-consistent data construction in the proposed framework.

Supplementary Table 10 | Influence of pretraining data scale on predictive performance.

Data Ratio	5%	10%	25%	50%	100%
R	0.3587	0.5392	0.3163	0.8551	**0.9596**
FNMAE	0.5369	0.6075	0.5942	0.1569	**0.0461**
FNRMSE	1.4363	1.9440	1.4965	0.4318	**0.0664**

4.6 IDA-based evaluation in near-collapse and damage-state regimes

4.6.1 Purpose and engineering scope of the IDA-based evaluation

In modern Performance-Based Earthquake Engineering (PBEE) frameworks[18], Incremental Dynamic Analysis (IDA) serves as the prevalent and most rigorous computational approach for quantifying structural collapse margins and evaluating damage-state fragilities[19]. By subjecting a computational model to an ensemble of ground motions scaled to progressively increasing intensity levels, IDA captures the continuous evolution of structural behavior—ranging from initial elasticity, through progressive yielding and severe cyclic degradation[20], to the ultimate onset of global dynamic instability[21]. However, from a computational mechanics perspective, conventional continuum-based finite element modeling (FEM) is inherently limited in simulating the explicit physical process of structural collapse. As a structure transitions from severe damage to global failure, the tangent stiffness matrix typically becomes singular or ill-conditioned due to extreme material degradation and geometric nonlinearities (e.g., P-Delta effects[22]). Standard FEA can only reliably simulate the structural response up to the onset of instability; tracking post-collapse kinematics—such as component separation, rigid-body motion, or debris interaction—requires distinct physical simulation paradigms like the Discrete

Element Method (DEM) or specialized explicit physics engines[23].

Because of these fundamental computational boundaries, collapse capacity in standard earthquake engineering practice is not identified through the explicit resolution of a stable post-collapse trajectory. Instead, it is operationally defined by the transition from a converged, stable response to dynamic instability, typically characterized by numerical non-convergence or a boundless accumulation of interstory drift (i.e., the "flatlining" of the IDA curve)[21,24]. Recognizing this established methodological framework, we conducted a dedicated IDA-based downstream evaluation to further assess the applicability of SeisGPT in highly demanding, failure-adjacent response regimes.

The purpose of this supplementary analysis is not to redefine the fundamental role of nonlinear FEA in collapse assessment, nor to treat SeisGPT as a replacement for the physical FEM-based collapse criterion itself. Instead, SeisGPT is evaluated as a high-fidelity surrogate for the converged-response component of the IDA workflow. Specifically, it acts as the module responsible for estimating response histories, engineering-demand parameters (EDPs), and the threshold-crossing intensities associated with successive damage states up to the point of structural instability. This evaluation design directly addresses whether SeisGPT maintains its predictive reliability not only under ordinary nonlinear response amplitudes—which are extensively validated in the main manuscript—but also under extreme conditions. By assessing the fidelity of SeisGPT in drift-based damage-state formulations derived from IDA, we validate its capacity to capture the highly nonlinear dynamic characteristics that precede structural failure. The resulting analyses substantiate the near-collapse scatter plots, the representative fragility-curve comparisons, and the population-level limit-state concordance statistics presented in Figure 7 and Table 2 of the main text.

4.6.2 IDA benchmark composition and case definition

The IDA benchmark was constructed using a representative evaluation set of 65 real-world buildings, drawn directly from the test split of the building-specific dataset employed in our main downstream experiments. To rigorously evaluate the collapse margin under the most severe dynamic demands, each building was assessed strictly along its most critical seismic loading direction (i.e., the principal axis exhibiting the highest vulnerability or dominant modal mass participation). Consequently, each IDA trajectory in this supplementary benchmark is uniquely defined by a specific triplet: (building, critical direction, ground motion). For the seismic input, we adopted the standardized suite of 22 far-field ground motion records recommended by the FEMA P-695 methodology[19]. Sourced from the Pacific Earthquake Engineering Research Center (PEER) Next Generation Attenuation (NGA) database[25], this specific ensemble is widely recognized as the established benchmark for collapse-oriented performance assessment. It provides a statistically robust representation of record-to-record variability (aleatory uncertainty) while deliberately avoiding near-fault directivity biases that could skew the generalized collapse capacity.

For each building's critical direction, these 22 ground-motion records were applied under progressively increasing intensity levels. The fundamental intensity measure (IM) used throughout the benchmark is the 5%-damped spectral acceleration evaluated at the first-mode period of the corresponding structural system in the selected direction:

$$\mathrm{IM} = S_{\mathrm{a}}(T_1, 5\%) \tag{39}$$

where T_1 denotes the first natural period obtained from the corresponding stiffness–mass system. For each unscaled ground-motion record, the reference spectral acceleration $S_{\mathrm{a},0}(T_1, 5\%)$ was first computed. The target intensity level $S_{\mathrm{a,target}}(T_1, 5\%)$ was then imposed through a scalar amplitude factor:

$$\lambda = \frac{S_{\mathrm{a,target}}(T_1, 5\%)}{S_{\mathrm{a},0}(T_1, 5\%)} \tag{40}$$

which uniformly scales the input acceleration time history. This rigorous scaling procedure ensures that both the FEM reference analysis and the corresponding SeisGPT surrogate evaluation are executed under strictly identical building, directional, and ground-motion intensity conditions.

Ultimately, the resulting benchmark yields a comprehensive collection of 1,430 individual IDA trajectories (65 buildings $\times$ 22 records). This extensive formulation allows the evaluation to rigorously probe both the highly non-linear response behavior immediately prior to dynamic instability and the consistency of drift-based damage-state intensities across a broad, real-world structural population.

4.6.3 FEM-based IDA procedure and collapse bracketing strategy

All reference IDA results were generated using the same nonlinear finite-element framework adopted for the structural response database described in Section 2.6. Specifically, gravity loads were first applied to establish the initial stress state, and nonlinear time-history analysis was then performed in OpenSees under ground-acceleration excitation. The solution algorithm utilized Rayleigh damping, constant-average Newmark time integration, Newton–Raphson equilibrium iterations, and adaptive recovery procedures for challenging nonlinear increments. This setup ensured that the IDA benchmark was fully consistent with the rigorous nonlinear dynamic analysis protocol used throughout the main study.

To efficiently and accurately identify the collapse capacity, we implemented a highly automated, multi-stage heuristic search algorithm to trace the structural response from elasticity to dynamic instability. The hunting strategy for the collapse bracket was executed in three progressive phases:

- Phase I: Coarse Initialization. For each structure–direction–ground-motion triplet, the search initiated at a relatively low target intensity of S_{a}=0.05g. If the structural response converged, the analysis advanced through a predefined, quasi-logarithmic discrete schedule to quickly capture the ordinary nonlinear behavior.
- Phase II: Dynamic Extension and Downprobing. If a structure survived the entire coarse schedule, the target intensity was exponentially extended by a constant multiplier (α=1.25) until dynamic instability was observed or a hard upper bound ($S_{\mathrm{a,max}}$=3.0g) was reached. Conversely, if a highly vulnerable structure failed at the initial 0.05g step, the algorithm triggered a downprobing sequence, recursively dividing the intensity by α until convergence was achieved or the absolute lower limit ($S_{\mathrm{a,min}}$=0.01g) was hit.
- Phase III: Geometric Bisection and Collapse Bracketing. Once the algorithm successfully identified an intensity interval bounding the transition from a stable response to a failed response, a rigorous bisection phase was engaged. The collapse bracket was continuously refined by inserting new trial intensities evaluated at the geometric mean of the bounds:

$$S_{\mathrm{a,mid}} = \exp\left(\frac{\ln\left(S_{\mathrm{a,low}}\right) + \ln\left(S_{\mathrm{a,high}}\right)}{2}\right) \tag{41}$$

  where $S_{\mathrm{a,low}}$ and $S_{\mathrm{a,high}}$ are dynamically defined as the highest converged intensity level and the lowest failed intensity level, respectively.

The interval $[S_{\mathrm{a,low}}, S_{\mathrm{a,high}}]$ constitutes the FEM collapse bracket in the operational sense of IDA-based collapse assessment. To ensure high precision without squandering computational resources, the bisection refinement was strictly terminated when either a logarithmic tolerance or an absolute tolerance was satisfied:

$$\ln\left(\frac{S_{\mathrm{a,high}}}{S_{\mathrm{a,low}}}\right) \le 0.02 \;\text{ or }\; (S_{\mathrm{a,high}} - S_{\mathrm{a,low}}) \le 0.005g \tag{42}$$

In this framework, $S_{\mathrm{a,low}}$ corresponds to the ultimate intensity at which the nonlinear structural response can still be computed successfully, whereas $S_{\mathrm{a,high}}$ marks the definitive onset of numerical instability, unbounded drift accumulation, or loss of convergence. This strictly defined bracket—rather than an explicitly resolved post-collapse trajectory—serves as the reference descriptor for collapse-adjacent behavior.

This methodological precision is critical for the interpretation of the SeisGPT evaluation. The benchmark is deliberately designed to test whether SeisGPT can faithfully reproduce the converged portion of the IDA response trajectory (up to $S_{\mathrm{a,low}}$) and the response-derived demand parameters used in structural assessment, without erroneously asserting that the surrogate model replaces the fundamental FEM-based criteria for identifying collapse onset.

4.6.4 Response prediction along the converged IDA branch and hierarchical result assembly

For each converged intensity point on the reference IDA curves, SeisGPT was evaluated under strictly identical dynamic demands—preserving the specific building configuration, loading direction, ground-motion record, and spectral intensity. The predicted spatiotemporal displacement histories were then paired one-to-one with the finite-element (FEM) ground truth, ensuring that every converged IDA state yields a perfectly matched FEM–SeisGPT response pair under equivalent excitation.

The downstream evaluation relies exclusively on these raw predicted kinematic histories, explicitly excluding any auxiliary damage classifiers or empirical collapse criteria within the SeisGPT architecture. All engineering demand parameters (EDPs) and subsequent damage-state classifications were derived via identical physical post-processing of both the predicted and reference structural responses. This design guarantees that the assessment strictly isolates the model's physical fidelity: its capacity to autonomously output spatiotemporal responses accurate enough to support standard IDA-based engineering interpretations.

To systematically parse the evaluation dataset, the response-derived quantities were organized hierarchically. At the local scale, run-level summaries were defined for every discrete converged intensity point to capture instantaneous dynamic metrics, most notably the peak interstory drift ratio (PIDR). Concurrently, curve-level summaries were aggregated for each continuous IDA trajectory to delineate the global structural evolution, encompassing the operationally defined collapse bracket, the near-collapse comparison state, and the intensity thresholds corresponding to successive damage states. This dual-resolution framework enables a comprehensive validation of both the pointwise predictive accuracy at failure-adjacent states and the trajectory-level consistency of standard engineering limit states.

4.6.5 Engineering demand parameters and near-collapse state extraction

The principal engineering demand parameter (EDP) adopted for this supplementary evaluation is the maximum interstory drift ratio (MIDR), a quintessential metric in performance-based earthquake engineering (PBEE) that directly correlates nonlinear kinematic demands with structural damage states. For a given story i, the transient interstory drift ratio is defined as:

$$\theta_i(t) = \frac{u_i(t) - u_{i-1}(t)}{h_i} \tag{43}$$

where $u_i(t)$ and $u_{i-1}(t)$ denote the lateral displacement histories of adjacent floor diaphragms, and h_i is the corresponding story height. The global MIDR over the entire dynamic excitation is subsequently extracted as:

$$\mathrm{MIDR} = \max_{i,t} |\theta_i(t)| \tag{44}$$

To rigorously evaluate predictive fidelity under extreme dynamic demands, we explicitly isolated the near-collapse state for each IDA trajectory. This critical state is deterministically anchored to the FEM collapse bracket limit, $S_{\mathrm{a,low}}$—the ultimate converged intensity level immediately preceding global dynamic instability or loss of convergence. By extracting and directly comparing the FEM- and SeisGPT-derived MIDR values precisely at $S_{\mathrm{a,low}}$, we obtain exactly one maximally demanding response point per valid building–direction–ground-motion trajectory.

The near-collapse MIDR distribution, pooled across the entire benchmark, is presented in Figure7a of the main text. By construction, this cross-sectional analysis is fundamentally distinct from the multi-intensity evaluations validated in the main manuscript; it strictly isolates structural behavior at the absolute boundary of numerical stability. Statistical concordance across this extreme-state dataset was quantified using the Pearson correlation coefficient, median absolute percentage error (MAPE), and 90$^{\text{th}}$-percentile absolute percentage error. Ultimately, this targeted extraction substantiates the robustness of the surrogate model, demonstrating that SeisGPT maintains high physical fidelity not only under ordinary inelastic excursions, but also under the most severe, failure-adjacent conditions encountered in standard IDA workflows.

4.6.6 Drift-based damage-state definition and extraction of limit-state intensities

Beyond the near-collapse point evaluation, the IDA benchmark was further used to examine whether SeisGPT can reproduce standard drift-based damage-state transitions along the converged nonlinear response branch. To place this analysis on a standardized engineering basis, four limit states, denoted here as LS1–LS4, were defined in accordance with the FEMA Hazus structural damage-state framework, corresponding to Slight, Moderate, Extensive, and Complete structural damage, respectively. In contrast to ad hoc benchmark-wide quantile thresholds, this formulation anchors the damage-state definition to an established earthquake damage classification system widely used for building-level fragility assessment across heterogeneous structural typologies.

For each building b, a Hazus-compatible structural class was assigned according to its primary lateral-force-resisting system and height category. The drift thresholds associated with the four structural damage states were then taken from the corresponding Hazus class and are denoted by $\theta_{b,k}$, with $k \in \{1,2,3,4\}$ corresponding to LS1–LS4. Accordingly,

$$\theta_{b,k} = \theta_k^{\mathrm{Hazus}}(c_b) \tag{45}$$

where c_b denotes the Hazus structural class assigned to building b. For each IDA trajectory, the structural demand was characterized by the maximum inter-storey drift ratio (MIDR) as a function of the intensity measure $\mathrm{IM} = S_{\mathrm{a}}(T_1, 5\%)$. Because local non-monotonicity may arise in raw IDA trajectories due to higher-mode effects, stiffness degradation, or period elongation, the MIDR sequence was converted into a monotone non-decreasing envelope prior to threshold-crossing extraction. For a discrete set of intensities $\mathrm{IM}_1 < \mathrm{IM}_2 < \cdots < \mathrm{IM}_N$, with corresponding raw drift demands $\theta(\mathrm{IM}_n)$, the envelope was defined as

$$\tilde{\theta}(\mathrm{IM}_n) = \max_{1 \le m \le n} \theta(\mathrm{IM}_m) \tag{46}$$

This standard envelope construction ensures that each damage-state threshold is crossed at most once along a given converged IDA branch and yields a unique, physically interpretable threshold-crossing intensity.

For a prescribed building-specific threshold $\theta_{b,k}$, the corresponding limit-state intensity $\mathrm{IM}_{b,k}$ was defined as the lowest intensity at which the monotone MIDR envelope first exceeded that threshold. If the crossing occurred between two adjacent analyzed intensities IM_j and IM_{j+1}, such that

$$\tilde{\theta}(\mathrm{IM}_j) \leq \theta_{b,k} < \tilde{\theta}(\mathrm{IM}_{j+1}) \tag{47}$$

then the threshold-crossing intensity was obtained by interpolation in logarithmic intensity space:

$$\ln(\mathrm{IM}_{b,k}) = \ln(\mathrm{IM}_j) + \frac{\theta_{b,k} - \tilde{\theta}(\mathrm{IM}_j)}{\tilde{\theta}(\mathrm{IM}_{j+1}) - \tilde{\theta}(\mathrm{IM}_j)}\left[\ln(\mathrm{IM}_{j+1}) - \ln(\mathrm{IM}_j)\right] \tag{48}$$

The use of logarithmic interpolation is consistent with the conventional log-scale representation of seismic intensity measures in fragility analysis.

If a given damage-state threshold was not reached along the converged branch of an IDA trajectory, that trajectory was classified as "not reached" for the corresponding limit state and was excluded from the paired $\mathrm{IM}_{b,k}$ concordance statistics for that state. This situation may arise, for example, when a curve terminates at the collapse bracket before a higher drift threshold is attained on the converged branch. In the present study, threshold-crossing intensities were extracted independently from the FEM-derived and SeisGPT-derived MIDR envelopes, yielding paired quantities $\mathrm{IM}_{b,k}^{\mathrm{FEM}}$ and $\mathrm{IM}_{b,k}^{\mathrm{SeisGPT}}$ wherever both were finite. Because the building, loading direction, ground-motion record, and base intensity grid were identical for the two analyses, any discrepancy in $\mathrm{IM}_{b,k}$ reflects only differences in the predicted dynamic response trajectory rather than differences in loading protocol or damage-state definition.

This response-driven formulation is important for the interpretation of the downstream benchmark. No auxiliary damage classifier was introduced, and no explicit collapse label was inferred from the prediction model. Instead, the damage-state assessment was derived entirely from the predicted displacement histories through the same MIDR-based engineering post-processing used for the FEM reference trajectories. In this sense, the benchmark evaluates whether SeisGPT can recover standardized drift-based structural damage states from its own predicted response fields, rather than whether it can reproduce a separately trained classification target.

4.6.7 Representative fragility-curve construction and population-level statistical concordance

To provide an interpretable engineering visualization of the downstream implications of the IDA predictions, a representative fragility analysis was first conducted for an individual real-world building. For this structure, threshold-crossing intensities were extracted for LS1–LS4 over the set of scaled ground motions, yielding one value $\mathrm{IM}_{b,r,k}$ for each record *r* and damage state *k*, wherever the threshold was reached on the converged branch. Following standard seismic fragility practice, the exceedance probability for each damage state was then represented using a lognormal cumulative distribution function,

$$P(\mathrm{DS} \geq k|\mathrm{IM} = x) = \Phi(\frac{\ln x - \mu_{b,k}}{\beta_{b,k}}) \tag{49}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\mu_{b,k}$ and $\beta_{b,k}$ are the location and dispersion parameters of the fragility function for building *b* and damage state *k*. In the present study, these fragility curves were used as a compact visualization of the relationship between IDA-derived response demand and damage-state exceedance probability, rather than as the sole basis for benchmark-level validation. The resulting panel shown in Figure 7b of the main text therefore serves as an illustrative case-level comparison between SeisGPT-derived and FEM-derived multi-state vulnerability trends.

Because individual fragility plots cannot establish population-level generality on their own, benchmark-wide concordance was assessed directly at the level of the threshold-crossing intensities. For each valid building–record trajectory and each damage state k, the predictive discrepancy was quantified by the logarithmic residual

$$\Delta_{b,r,k} = \ln{(\frac{\mathrm{IM}_{b,r,k}^{\mathrm{SeisGPT}}}{\mathrm{IM}_{b,r,k}^{\mathrm{FEM}}})} \tag{50}$$

This logarithmic form is consistent with the conventional lognormal representation of seismic fragility and provides a scale-invariant measure of the discrepancy in damage-state capacity. Under this definition, $\Delta_{b,r,k} = 0$ indicates perfect agreement, $\Delta_{b,r,k} > 0$ indicates that SeisGPT predicts a higher threshold-crossing intensity than FEM, and $\Delta_{b,r,k} < 0$ indicates that it predicts a lower one. Figure 7c of the main text summarizes the benchmark-wide distribution of these residuals for LS1–LS4, thereby providing a compact view of the agreement between SeisGPT-derived and FEM-derived damage-state intensities across the structural population.

Table 2 of the main text reports the principal summary statistics corresponding to this population-level comparison. These include the Pearson correlation coefficient between $\mathrm{IM}_{b,r,k}^{\mathrm{SeisGPT}}$ and $\mathrm{IM}_{b,r,k}^{\mathrm{FEM}}$, the median and 90$^{\text{th}}$ percentile of the absolute logarithmic discrepancy $|\Delta_{b,r,k}|$, and the median and 90$^{\text{th}}$ percentile of the absolute percentage error in $\mathrm{IM}_{b,r,k}$. To prevent the benchmark-level summary from being dominated by buildings associated with unusually large numbers of converged trajectories, a hierarchical aggregation metric was also reported: the across-building median of the within-building median absolute percentage error. In this statistic, the median IM-error was first evaluated within each building and then aggregated across the building population, thereby reducing statistical bias arising from unequal trajectory counts.

Together, the representative fragility curves and the benchmark-wide $\mathrm{IM}_{b,r,k}$ concordance statistics provide complementary evidence. The former illustrates, for an individual structure, that SeisGPT preserves the sequential ordering and spacing of the four Hazus-consistent drift-based damage states over increasing intensity. The latter demonstrates that this agreement is not confined to isolated case studies, but is maintained across the broader real-building benchmark when evaluated directly in terms of standardized IDA-derived limit-state intensities. Under this formulation, the central question is not whether SeisGPT can replace the FEM-based collapse criterion itself, but whether it can accurately recover the engineering quantities required for drift-based damage-state assessment within a standard IDA workflow. The present results indicate that it can do so with strong statistical consistency.

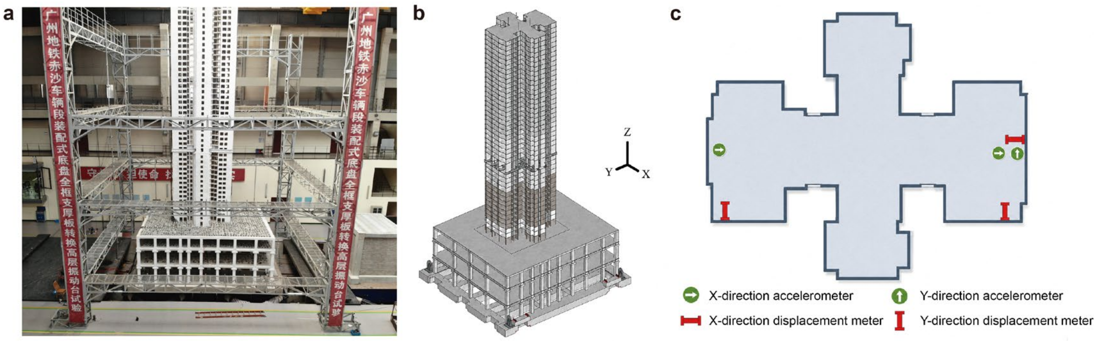
5. Details of Shaking Table Testing and Structural Response Reconstruction Using SeisGPT-R

This section presents the experimental validation of SeisGPT-R using a shake-table test dataset generated from a scaled reinforced-concrete structural specimen. The experimental setup, specimen scaling, instrumentation scheme, input motions, model inference pipeline and sensor-placement sensitivity analysis are summarized below[26]. A scaled physical replica of a real-world reinforced concrete structure was constructed and subjected to controlled seismic excitations and white noise inputs to assess the model's reconstruction capabilities under realistic dynamic conditions. The experimental setup, instrumentation scheme, input selection, model inference pipeline, and sensor-placement sensitivity analysis are detailed below.

5.1 Experimental Setup and Structural Model

The shake table experiment was conducted using a scaled physical model representing the over-track development situated above the Chisha Depot in Guangzhou, China. The prototype structure comprises a substructure—consisting of a basement and two skirt floors—and a superstructure primarily composed of shear walls, reaching a total height of 131.4 m above ground level, with the substructure accounting for approximately 29.2 m. To achieve mechanical similitude under laboratory conditions, a geometric scaling factor of 1:10 was applied, yielding a model footprint of 6.90 m×9.16 m. The structure was constructed using micro-aggregate concrete, galvanized steel wire mesh, and reinforcing steel to replicate the elastic and inelastic characteristics of the full-scale reinforced concrete system. A strength scaling factor of 1:5 was adopted to maintain consistent force–displacement behavior across scales.

The completed model was mounted on a tri-axial shake table capable of delivering complex multi-directional seismic inputs. Instrumentation was meticulously configured to capture dynamic responses at structurally significant locations: accelerometers and displacement transducers were installed on storeys exhibiting pronounced modal activity and stiffness discontinuities. The sensor arrangement and photographs of the scaled model are provided in Supplementary Figure 17.



Supplementary Figure 17 | Shaking-table test model and sensor layout used in the SeisGPT-R experiment. a, Photograph of the 1:10-scale physical model. **b**, Schematic of the 6.90 m × 9.16 m model mounted on the tri-axial shaking table. Global X, Y and Z axes are indicated. **c**, Sensor layout.

External dynamic inputs consisted of both recorded and synthetic ground motions selected to span a broad range of intensity levels and spectral characteristics. Four seismic hazard levels were considered: frequent shaking at intensity VII, design-level shaking at intensity VII, rare shaking at intensity VII, and rare shaking at intensity VIII. Three representative waveforms were used as excitation inputs: RSN1164, RSN3942, and a synthetically generated artificial wave (AW). To account for discrepancies between intended waveforms and the actual excitation delivered by the shake table, only the measured acceleration and displacement signals recorded directly at the table platform were used as ground motion inputs for SeisGPT-R inference. In total, ten distinct test scenarios were conducted to evaluate SeisGPT-R's ability to reconstruct full-structure structural responses using sparse sensor observations.

5.2 Comprehensive Assessment and Visualization of SeisGPT-R Reconstruction Performance

The reconstruction performance of SeisGPT-R was systematically evaluated against two baselines—nonlinear finite element (FEM) simulations and shake table measurements—under realistic sparse-sensing conditions, where only base excitation and mid-height floor sensor data were available. This setup reflects practical deployment scenarios in which dense instrumentation is infeasible, necessitating accurate inference from limited input. Notably, SeisGPT-R successfully inferred full-building responses from mid-height input, underscoring its capacity to model long-range spatiotemporal dependencies. As shown in Figure 6 of the main text, SeisGPT-R consistently outperformed FEM in both temporal and spatial fidelity across all test cases, achieving superior results across all three evaluation metrics, even when compared to extensively tuned FEM models. Notably, while each FEM simulation required one to two days per ground motion due to model complexity, SeisGPT-R achieved comparable or better accuracy in real time—offering prediction speeds over four orders of magnitude faster. This highlights SeisGPT-R's fundamental advantage: rapid, high-fidelity response reconstruction at a fraction of the computational cost.

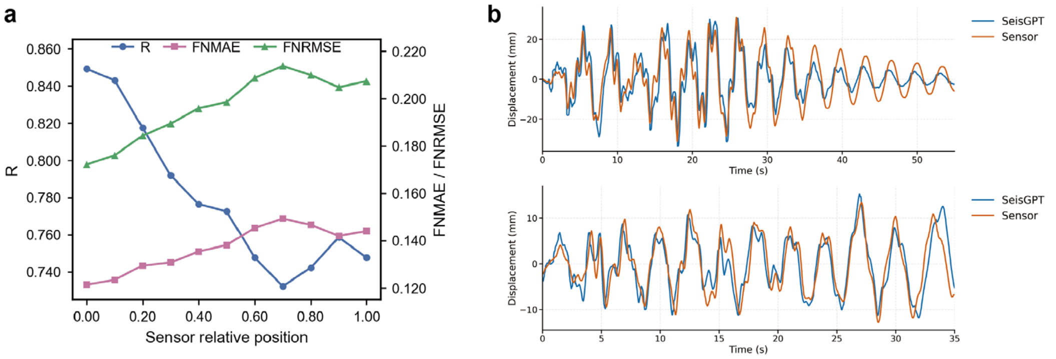
Given that the core comparative analysis is already presented in Figure 6, this section emphasizes additional visualizations. Supplementary Figure 18b shows representative top-floor displacement time histories under different ground motion inputs, further illustrating the model's robustness and effectiveness as a surrogate for full-structure response reconstruction in hybrid digital–physical testing environments.

5.3 Influence of Sensor Placement on Reconstruction Accuracy

A critical factor affecting the performance of data-driven structural response reconstruction in real-world scenarios is the vertical placement of physical sensors. To quantitatively assess the sensitivity of SeisGPT-R to sensor location, a controlled experiment was performed using the shake table test dataset. The position of a single input sensor was systematically varied from the mid-height to the roof level, while all other experimental variables were held constant. Reconstruction performance was evaluated using three metrics—Floor-Normalized Mean Absolute Error (FNMAE), Pearson correlation coefficient (R), and Floor-Normalized Root Mean Squared Error (FNRMSE). The results are presented in Supplementary Figure 18a, with the horizontal axis indicating the relative sensor relative position and the vertical axis showing the corresponding metric values.

The analysis reveals a distinct, nonlinear relationship between sensor elevation and reconstruction fidelity. Contrary to intuitive expectations, sensors placed at the upper storeys—particularly at the roof—exhibited reduced accuracy. This is likely attributable to their limited ability to capture the globally integrated dynamic behavior of the structure. In contrast, mid-height sensors provided more balanced and informative inputs, supporting improved reconstruction of both localized and system-level response characteristics. Optimal accuracy was achieved when the input sensor was placed near the structural mid-height, a location that appears to most effectively encode modal richness and dynamic coupling effects necessary for accurate full-structure inference.

These findings highlight the critical role of sensor positioning in AI-enabled structural monitoring. By exploiting SeisGPT-R's capacity to generalize from sparse but strategically located observations, high-fidelity reconstructions can be achieved even under stringent instrumentation constraints, offering a practical pathway toward scalable and cost-efficient deployment in structural response assessment.



Supplementary Figure 18 | Sensor placement sensitivity and representative reconstruction results of SeisGPT-R. a, Reconstruction performance as a function of the relative floor position of the single input sensor, evaluated using FNMAE, FNRMSE and Pearson correlation coefficient. **b**, Representative top-floor displacement histories reconstructed by SeisGPT-R under different ground-motion inputs, compared with shake-table measurements and FEM simulations. The results illustrate that mid-height sensor placement provides the most informative sparse observation for full-building response reconstruction. Source data are provided as a Source Data file.

Supplementary References

1 Ministry of Housing and Urban-Rural Development of the People's Republic of China. Code for Seismic Design of Buildings. GB 50011–2010. Beijing: 2016.

2 Ministry of Housing and Urban-Rural Development of the PRC. Technical Specification for Concrete Structures of Tall Buildings. JGJ 3-2010. Beijing: China Architecture & Building Press, 2010.

3 Liu, X., Gong, C. & Liu, Q. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003* (2022).

4 Su, J. *et al.* Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing* **568**, 127063 (2024).

5 Radford, A. *et al.* Learning transferable visual models from natural language supervision. *In Proceedings of the International conference on machine learning*, 8748-8763 (2021).

6 Raffel, C. *et al.* Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research* **21**, 1-67 (2020).

7 Shazeer, N. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202* (2020).

8 Zhang, B. & Sennrich, R. Root mean square layer normalization. *Advances in Neural Information Processing Systems* **32** (2019).

9 Zhang, L., Rao, A. & Agrawala, M. Adding conditional control to text-to-image diffusion models. *In Proceedings of the the IEEE/CVF international conference on computer vision*, 3836-3847 (2023).

10 Ministry of Housing and Urban-Rural Development of the People's Republic of China. Standard for design of steel structures: GB 50017-2017. Beijing: China Architecture & Building Press, 2018.

11 Martinelli, L., Martinelli, P. & Mulas, M. G. Performance of fiber beam–column elements in the seismic analysis of a lightly reinforced shear wall. *Engineering structures* **49**, 345-359 (2013).

12 Kolozvari, K., López, C. N. & Massone, L. M. Efficient three-dimensional shear-flexure interaction model for reinforced concrete walls. *Engineering Structures* **294**, 116700 (2023).

13 Charney, F. A. Unintended consequences of modeling damping in structures. *Journal of structural engineering* **134**, 581-592 (2008).

14 Sharma, S., Silva, L., Graziotti, F., Magenes, G. & Milani, G. Modelling the experimental seismic out-of-plane two-way bending response of unreinforced periodic masonry panels using a non-linear discrete homogenized strategy. *Engineering Structures* **242**, 112524 (2021).

15 Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

16 Rasley, J., Rajbhandari, S., Ruwase, O. & He, Y. Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters. *In Proceedings of the Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 3505-3506 (2020).

17 Rajbhandari, S., Rasley, J., Ruwase, O. & He, Y. Zero: Memory optimizations toward training trillion parameter models. *In Proceedings of the SC20: International*

Conference for High Performance Computing, Networking, Storage and Analysis, 1-16 (2020).

18 Moehle, J. & Deierlein, G. G. A framework methodology for performance-based earthquake engineering. *In Proceedings of the 13th world conference on earthquake engineering*, 12 (2004).

19 Council, A. T. *Quantification of building seismic performance factors*. (US Department of Homeland Security, FEMA, 2009).

20 Lignos, D. G. & Krawinkler, H. Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading. *Journal of Structural Engineering* **137**, 1291-1302 (2011).

21 Vamvatsikos, D. & Cornell, C. A. Incremental dynamic analysis. *Earthquake engineering & structural dynamics* **31**, 491-514 (2002).

22 Bernal, D. Instability of buildings subjected to earthquakes. *Journal of Structural Engineering* **118**, 2239-2260 (1992).

23 Lu, X., Lu, X., Guan, H. & Ye, L. Collapse simulation of reinforced concrete high‐rise building induced by extreme earthquakes. *Earthquake Engineering & Structural Dynamics* **42**, 705-723 (2013).

24 Zareian, F., Krawinkler, H., Ibarra, L. & Lignos, D. Basic concepts and performance measures in prediction of collapse of buildings under earthquake ground motions. *The Structural Design of Tall and Special Buildings* **19**, 167-181 (2010).

25 Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S. J., Donahue, J. L.,. NGA-West2 Database. PEER Report 2013/03. Pacific Earthquake Engineering Research Center, University of California, Berkeley, 2014.

26 Zhou, Y. *et al.* Seismic performance of a frame‐supported shear wall over‐track building through shaking table test. *The Structural Design of Tall and Special Buildings* **33**, e2098 (2024).