ENHANCING STRUCTURAL HEALTH MONITORING THROUGH ADVANCED TIME SERIES-TARGETED NEURAL NETWORKS: FROM SENSOR PLACEMENT TO DATA FILTERING, INFERENCE, AND DIAGNOSIS

TAN YANKE

PhD

The Hong Kong Polytechnic University

This programme is jointly offered by The Hong Kong Polytechnic University and Tongji University

2025

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

Tongji University

College of Civil Engineering

Enhancing structural health monitoring through advanced time series-targeted neural networks: From sensor placement to data filtering, inference, and diagnosis

TAN Yanke

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

May 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

TAN Yanke (Name of student)

Dedicated to my beloved family,

for their love and support

ABSTRACT

With the accelerated advancement of global urbanization, the coexistence of large-scale infrastructure construction projects and aging existing structures has imposed more severe challenges on structural health monitoring (SHM). The inherent conflict between continuous investment in infrastructure and the demand for rational lifecycle maintenance further highlights the necessity to develop advanced SHM techniques. Current SHM systems still face multiple technical bottlenecks in sensor placement optimization, data denoising, damage identification, and decision-making: sensor placement relies on empirical experience without comprehensive optimization criteria balancing global and local requirements; conventional signal preprocessing methods employing simple frequency-domain analysis struggle with data loss and extreme noise contamination; most damage diagnosis approaches focusing on macro-scale dynamic characteristics like modal parameters exhibit limited sensitivity to early-stage damage. Even machine learning (ML)-based methods suffer from insufficient accuracy, weak generalization capability, and incompatibility with massive monitoring data. This study proposes to reframe SHM data mining as time series analysis tasks, systematically optimizing key SHM technologies through proposing and improving neural networks (NNs) specialized for sequential data, thereby enhancing automation and intelligence in structural safety assessment. The specific research contributions are outlined as follows.

For sensor optimization, a novel optimization criterion integrating local signal representativeness and global modal independence is proposed, with a multi-phase heuristic optimization algorithm addressing combinatorial optimization problem. An automatic candidate selection method based on structural characteristic matrices and graph data theory is developed. In the accelerometer deployment for Rafael Cloud Corridor, the proposed criterion successfully achieved most accurate reconstruction of structural acceleration fields. The improved search algorithm demonstrated faster convergence and higher-quality local optima compared to conventional evolutionary

algorithms. By simplifying structures as graphs, the node ranking algorithm enables interpretable automated generation of candidate sets with arbitrary quantities.

In data preprocessing, an integrated solution incorporating dynamic adaptive denoising, missing data reconstruction, and outlier detection is established. An improved Kalman filter with a state transition model dynamically adjusts filtering parameters based on sensor performance, effectively balancing frequency retention and noise suppression while avoiding rigid manual thresholds. The bidirectional echo state network (Bi-ESN) achieves high-precision reconstruction of extreme missing scenarios (including long-term and multi-channel losses) in Canton Tower monitoring data by simultaneously leveraging historical-future information and cross-channel correlations. A frequency-decomposition anomaly detection framework combining generative models and local outlier factor algorithms precisely identifies various outliers in Shanghai Tower monitoring data with distinct types, quantities, distributions, and abnormality levels, while enabling reasonable signal recovery.

Data inferencing tasks focus on three application scenarios: prediction, synchronized mapping, and physical system solving. For time series forecasting, a memory-enhanced network with dynamic feature extractors achieves accurate wake flow prediction above tracks for wind barriers along the Lanzhou-Xinjiang high-speed railway. Addressing synchronized mapping between homologous variables, innovative random reservoir-based positional embedding breaks network input constraints, realizing vibration-based far-field noise substitution measurement in Shenzhen Metro. For differential equation-governed physical systems with environment-dependent correlations, a general generalization performance expanding structure successfully solves chaotic particle damper (PD) excitation-reaction mapping, establishing new paradigms for complex system calibration.

The assessment phase encompasses damage detection and condition evaluation. A supervised multi-dimensional multi-scale convolutional classifier achieves precise multi-position combined damage localization and severity quantification on a three-span model

bridge, complemented by a redundant data screening module enhancing noise robustness while reducing dependency on complete damage datasets. An updatable digital twin (DT) framework utilizing long-term operational data enables unsupervised rapid structural assessment, guiding post-disaster rehabilitation. Transfer learning facilitates convenient model adaptation to structural evolution over extended service periods.

LIST OF PUBLICATIONS

Journal articles

Tan, Y. K., Wang, Y. L., Ni, Y. Q., Zhang, Q. L., & Wang, Y. W. Improved bidirectional echo state network-based time series reconstruction and prediction for structural response. Structural Health Monitoring, 2024: 14759217241253082.

Tan, Y. K., Wang, Y. W., Ni, Y. Q., & Zhang, Q. L. Parallel reservoir computing based signal outlier detection and recovery method for structural health monitoring. Developments in the Built Environment, 2024, 18: 100463.

Tan, Y. K., Wang, Y. L., Deng, E., Ye, X., Zhang, Y., & Ni, Y. Q. Automatic damage detection and data completion for operational bridge safety using convolutional echo state networks. Automation in Construction, 2024, 166: 105606.

Tan, Y. K., Ni, Y. Q., Zhang, S. X., Zhang, Q. L., & Wang, Y. W. A fast, information-interactive, and reservoir computing-based digital twin for high-rise building operation. Expert Systems with Applications, 2025, 269: 126390.

Tan, Y. K., Ouyang, D. H., Deng, E, Yue, H., & Ni, Y. Q. CFD-guided memory-enhanced LSTM predicts leeward flow of railway windproof structures. Advanced Engineering Informatics, 2025, 65: 103253.

Ye, X., **Tan, Y. K.**, & Ni, Y. Q. Echoformer: An echo state-embedded transformer for robust reconstruction of railway trackside noise on urban metro lines. Mechanical Systems and Signal Processing, 2025, 229: 112491.

Zhang, S. Q., **Tan, Y. K.**, Ge, H. B., & Zhang, Q., L. An ultra-thin bolt tension sensor and online monitoring system: For application in hydropower plant unit. Frontiers of Structural and Civil Engineering, 2024, 18(9): 1388-1400.

Zhang, S. Q., **Tan, Y. K.**, Ge, H. B., & Zhang, Q., L. Safety-Function-Environment

Evaluation System for Large-Span Cable-Supported Bridges: Theory and Case Studies. Sustainability, 2024, 16(4): 1414.

Tan, Y. K., Ye, X., Ni, Y. Q., Zhang, Q. L., Wang, Y. W., & Seyed M., S. A. Generalization performance expansion of Seq2Seq models: For dynamic systems with inconstant mapping relations. Structural Control and Health Monitoring (Under review).

Conference papers

Tan, Y. K., Wang, Y. L., Ni, Y. Q., Zhang, Q. L., & Wang, Y. W. Large-Span Bridge Strain Reconstruction Based on Bidirectional LSTM and ESN. Proceedings of the 14th International Workshop on Structural Health Monitoring, 12-14 September 2023, Stanford University, California, USA.

Awards

Deng, E, Zhang, Y., **Tan, Y. K.**, Wang, Y. L., Ye, X. The Third Prize in the 3rd International Competition for Structural Health Monitoring (IC-SHM, 2022).

ACKNOWLEDGEMENTS

Looking back on my doctoral journey, I wish to express deepest gratitude to those who have illuminated my path with their wisdom, support, and warmth.

Foremost, I extend profound gratitude to my supervisor, Professor Ni Yiqing, for accepting me as your doctoral student three years ago and guiding me with unwavering dedication. Your encyclopedic knowledge and rigorous scholarship laid the foundation for my academic development. Your philosophy that "technological innovation must stem from practical engineering needs" shaped my research direction, while your humility and magnanimity taught me to maintain perseverance in scientific inquiry.

During my study period at Tongji University, I was privileged to receive guidance from Professor Zhang Qilin. Your global vision and keen insights into cutting-edge technologies revealed new dimensions of academic research.

I gratefully acknowledge Dr. Wang Youwu and Dr. Chen Zhengwei at the Hong Kong Polytechnic University for your generous support and invaluable technical advice during this study, which significantly broadened my academic perspectives. Special thanks to Dr. Luo Xiaoqun, Dr. Yang Huizhu, and Mrs. Tang Zhaohui at Tongji University for their mentorship. Your rigorous work ethic and profound engineering expertise set exemplary models for my future research.

My sincere appreciation goes to Dr. Deng E, Dr. Ye Xin, and Dr. Chen Siyi for sharing invaluable advice during my formative years. I am indebted to Dr. Pan Licheng, Dr. Yuan Ye, Dr. Huang Yanan and Dr. Zhu Peihua for their intellectual inspiration and collaborative opportunities. To my peers Mr. Dang Dazhi, Mr. Wang Yuling, Mr. Yuan Lei, Mr. Rui Enze, Mr. Zhang Weijia, Mr. Lu Jiahao; Ms. Li Xinye, Mr. Zhang Shuxiang, Mr. Yu Qihang, Ms. Liu Yiyin, Mr. Zeng Zixuan; Mr. Zhang Hanghua, Mr. Yu Jie; Mr. Xia Tian, Mr. Xia Haiyang, Mr. Fang Zhihong, Mr. Li Bingze, Mr. Song Guanqing, and Mr. Meng Kaiqiang, your companionship transformed this solitary journey into a shared adventure of discovery.

Special tribute to my partner Wang Yuxiang. For nine years, your unwavering support and gentle understanding have been a constant source of strength during moments of doubt and exhaustion. Our shared journey proves that the most profound academic perseverance grows not just in laboratories, but in the quiet constancy of mutual devotion.

Finally, my deepest love and gratitude belong to my family. My mother, with her resilience, has shouldered immense responsibilities to support my journey. My grandparents, through their advanced years, continue to sustain me with simple yet profound wisdom. Your silent sacrifices and steadfast faith have become the constellation guiding my pursuit of knowledge.

Words cannot fully capture the depth of my gratitude to all who have contributed, directly or indirectly, to this academic odyssey. These collective kindnesses have become indelible marks in my life's journey. As I continue exploring the frontiers of knowledge, I shall carry these gifts forward, remaining ever mindful that true scholarship thrives not in isolation, but through human connection and shared as privation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full-name
SHM	Structural health monitoring
FBG	Fiber Bragg grating
HSR	High-speed railway
WSN	Wireless sensor network
ML	Machine learning
DL	Deep learning
DT	Digital twin
5G	Fifth generation
WT	Wavelet transform
NN	Neural network
FEM	Finite element model
LDV	Laser Doppler vibrometer
MEMS	Micro-electro-mechanical system
UAV	Unmanned aerial vehicle
LiDAR	Laser detection and ranging
BIM	Building information model

GPS	Global positioning system
IMU	Inertial measurement unit
IoT	Internet of things
OSP	Optimal sensor placement
MAC	Modal assurance criterion
EI	Effective independence
MKE	Modal kinetic energy
FFT	Fast Fourier transform
EMD	Empirical mode decomposition
RNN	Recurrent neural network
LSTM	Long short-term memory
GRU	Gated recurrent unit
ESN	Echo state network
PINN	Physics-informed neural network
PDE	Partial differential equation
NODE	Neural ordinary differential equation
ODE	Ordinary differential equation
DeepONet	Deep operator network
LNN	Liquid neural network

TCN	Temporal convolutional network
CNN	Convolutional neural network
GCN	Graph convolutional network
NLP	Natural language processing
LLM	Large language model
HEOA	Human evolutionary optimization algorithm
FIA	Field inference accuracy
DoF	Degree of freedom
FFM	Feature-forcing mechanism
MLP	Multi-layer perceptron
MSE	Mean square error
SGD	Stochastic gradient descent
GA	Genetic algorithm
PSO	Particle swarm optimization
ARIMAX	Auto-regressive integrated moving average model with exogenous variables
KF	Kalman filter
LOF	Local outlier factor
IKF	Improved Kalman filter

ARIMA	Auto-regressive integrated moving average model
ACF	Auto-correlation function
PACF	Partial auto-correlation function
BIC	Bayesian information criterion
Bi-ESN-R	Dual-side bidirectional echo state network reconstructor
Bi-ESN	Bidirectional echo state network
SNR	Signal-to-noise ratio
FD-ESN	Frequency-decomposed echo state network
CWT	Continuous wavelet transform
ICWT	Inverse continuous wavelet transform
LRD	Local reachability density
eLSTM	Memory-enhanced long short-term memory
dESN	Dynamic echo state network
HST	High-speed trains
WS	Windproof structure
CFD	Computational fluid dynamics
UDF	User-defined function
DDES	Delayed detached eddy simulation

SST	Shear stress transport
TVD	Total variation diminishing
PSD	Power spectral density
Seq2Seq	Sequence-to-sequence
FFNN	Feed-forward neural network
SPL	Sound pressure level
ESDN	Echo state deep network
GPES	Generalization performance expansion structure
Ex-ESN	Expanded echo state network
PD	Particle damper
MDSC	Multi-dimensional and multi-scale convolution
PCA	Principal component analysis
ResNet	Residual network
VGG	Visual geometry group network
YOLO	You only look once
MFRC	Multi-head full-rank reservoir computing
WD	Wasserstein distance
SQL	Smooth quadratic loss

INTRODUCTION

1.1 Research background

The accelerated pace of global urbanization has driven the intensive construction of large-scale civil engineering projects, including long-span bridges, super high-rise buildings, and undersea tunnels. Taking China as an example, the total value of newly signed construction contracts in 2024 reached RMB 4.5 trillion, with infrastructure project contracts showing a 21.1% year-on-year increase, demonstrating sustained investment in mega-engineering projects (Farrar, Dervilis & Worden 2025; Hao et al., 2023; Yan et al., 2024). Meanwhile, the aging problem of existing structures has become increasingly prominent. As statistics indicate, approximately 30% of China's current civil engineering structures (particularly bridges) have been in service for over 30 years, while some dams and industrial buildings face performance degradation risks due to material fatigue and environmental corrosion (Zhang et al., 2022; Liao et al., 2010). Against this paradoxical backdrop of new construction demands and aging infrastructure, ensuring the safety of emerging engineering projects while extending the service life of existing structures has emerged as a core challenge in civil engineering. This dual imperative has positioned structural health monitoring (SHM) technology as a critical solution framework (Farrar, Dervilis & Worden 2025; Farrar & Worden 2007).

Large-scale engineering projects, due to their structural complexity and environmental sensitivity, pose heightened demands for real-time monitoring. These projects predominantly involve transportation hubs and smart city complexes (Javadinasab, Gutierrez & Adeli 2021). During both constructional and operational phases, such infrastructure must address multiple risks: vibrational mode alterations in long-span bridges (Tan et al., 2019), seismic and wind-induced responses of super high-rise buildings (Jiang et al., 2014; Zhang et al., 2020), surrounding rock stability in underground tunnels (Qiu et al., 2022), etc. Traditional manual inspections struggle to

meet real-time requirements and comprehensive coverage demands, whereas sensor-based SHM systems enable continuous acquisition of structural parameters including strain, temperature, and displacement. For instance, vibration analysis technology, which utilizes accelerometers to capture dynamic responses and combines modal analysis for damage identification, has been successfully implemented in landmark projects such as the Shanghai Tower and Hong Kong Tsing Ma Bridge (Su et al., 2013; Chan et al., 2006). Furthermore, fiber Bragg grating (FBG) sensors, leveraging their electromagnetic interference resistance and high sensitivity, have achieved micrometer-level strain detection in high-speed railway (HSR) track monitoring, delivering robust technical safeguards for infrastructure safety (Dang et al., 2022; Zhang et al., 2024a).

The aging of existing structures presents equally critical concerns. Statistical analyses reveal that China's urban residential building stock exceeds 50 billion square meters, with approximately 20% of these structures constructed in the last century exhibiting latent risks of material degradation and inadequate seismic performance (Garriga et al., 2023). The long-term compounding effects of environmental factors have accelerated this deterioration trend: coastal concrete structures demonstrate an annual 1.5%-3% increase in steel corrosion rates due to chloride ion erosion (Jin et al., 2018), while freeze-thaw cycles in northern regions cause 8%-12% annual strength loss in pavement base layers (Li et al., 2017). According to statistical data of 2024, approximately 15% of public buildings (e.g., cultural and recreational venues) require periodic closure for maintenance due to structural safety concerns, incurring direct economic losses exceeding RMB 10 billion. These cases underscore that aging structures lacking effective monitoring may trigger cascading societal risks, as exemplified by the 2019 roof collapse incident at a sports stadium caused by undetected long-term fatigue damage (Mander, Panthaki & Kasalanati 1994; Ye, Su & Han 2014).

The concurrent construction boom of large-scale infrastructure projects and the aging crisis of existing structures collectively form dual drivers propelling the advancement of SHM technologies. SHM systems are evolving from single-parameter monitoring

towards intelligent diagnostics enabled by multi-source data fusion. For example, wireless sensor networks (WSNs) achieve large-scale deployment through low-power nodes, demonstrating 40% cost reduction compared to wired systems, with their reliability validated in the corridor monitoring of the Hong Kong-Zhuhai-Macao Bridge (Chen et al., 2022). Intelligent diagnostic techniques integrating machine learning (ML), particularly deep learning (DL) algorithms, have enhanced damage identification accuracy (e.g., concrete crack width quantification) by 25-35% (Dung 2019; Prasanna et al., 2014; Rizvi, Khan & Ahmad 2017). Additionally, digital twin (DT) implementations enable dynamic interaction between physical structures and virtual models (Khajavi et al., 2019; Zhou et al., 2022), where high-rise buildings utilizing twin platforms to synchronize sensor data achieve 60% improvement in maintenance decision-making efficiency (Hu et al., 2024). Industry reports indicate the global SHM market surpassed USD 20 billion in 2024, exhibiting an 11.3% compound annual growth rate, reflecting its technological value and industrial potential (Franchi et al., 2022). Looking forward, the synergistic integration of the fifth generation (5G) communication (Peng et al., 2024), edge computing (Deng et al., 2023a), high-performance low-energy sensors, and more generalizable intelligent assessment methodologies will drive SHM systems toward real-time responsiveness and cognitive intelligence, thereby establishing robust safeguards for structural safety (Sun et al., 2025).

Concretely, for specified target structures with defined monitoring requirements and available sensor or data transmission devices, the primary challenges confronting SHM system designers, operators, and decision-makers fall into two methodological frameworks: hardware configuration schemes (Sun et al., 2025; Guratzsch & Mahadevan 2010) and data post-processing methodologies (Zhang et al., 2024b; Hassani et al., 2024; Deng et al., 2024). The former principally involves determining monitoring variables and establishing corresponding sensor deployment strategies based on structural environmental parameters and sensor specifications. The latter aims to assess structural conditions through acquired sensor data, providing informational support for long-term

safety and functional maintenance. To achieve this objective, SHM systems' data post-processing modules require not only robust evaluation models but also efficient data denoising mechanisms and imputation capabilities (Ravizza et al., 2021). However, current sensor placement optimization methods lack theoretically-grounded yet practically-feasible criteria, with their optimization processes heavily reliant on manual preliminary site selection (Dhillon & Chakrabarty 2003). Meanwhile, signal filtering algorithms in SHM post-processing predominantly remain confined to time-frequency domain transformation methods like low-pass filters and wavelet transform (WT) (Sun & Chang 2002; Silik et al., 2021), which exhibit limited effectiveness against extreme data contamination issues such as outliers (Li et al., 2018; Chen, Wang & Ni 2022) and missing segments (Lei, Sun & Xia 2021; Tan et al., 2024a; Wan & Ni 2019). Furthermore, most damage detection algorithms focus excessively on modal information (Azhar et al., 2024; Perera & Torres 2006), demonstrating insensitivity to damages occurring in structures with high design margins. Even recently proposed ML-integrated intelligent algorithms still face constraints including low prediction accuracy and scarcity of labeled training data (Avci et al., 2021; Santos et al., 2016; Gui et al., 2017).

SHM systems focus on the long-term operational management of structures. Most data transmitted from SHM systems contain temporal labels, where monitored variables can be conceptualized as time-domain functions, with actual acquired data over specific periods constituting time series (Hamilton 2020). All aforementioned tasks, including denoising, imputation, and damage detection, essentially represent preprocessing and postprocessing operations on temporal sequences (even sensor placement optimization can be abstracted as configuring spatial arrangements to maximize time series quality). The field of time series processing has witnessed revolutionary theoretical advancements in recent years, particularly through the integration of neural networks (NNs) (Hsieh 2004) specifically designed for sequential data processing. These methodologies employ recursive structures (Che et al., 2018; Guo et al., 2016) or attention mechanisms (Shih, Sun & Lee 2019; Du, Cote & Liu 2023) to model temporal evolution patterns, utilizing

numerous trainable parameters to capture latent regularities from observational data for task completion. Through over two decades of development, modern NN frameworks for time series analysis have comprehensively surpassed conventional temporal processing methods in performance metrics, though their practical implementation still requires domain-specific modifications and knowledge integration (Nguyen, Dang & Pham 2024). Currently, SHM technology primarily faces issues such as unclear sensor placement guidelines, difficulty in removing complex and extreme data contamination, insufficient model memory capacity and generalization performance, and heavy reliance on labeled damaged data. This study therefore proposes to adapt and optimize state-of-the-art time series NN algorithms for SHM applications, enhancing functional capabilities and improving task-specific implementation accuracy across the monitoring workflow.

1.2 Research objectives

To address the deficiencies in sensor placement optimization objectives and the mismatch between SHM post-processing algorithms and advanced temporal NNs, this study employs architectural refinement and optimization of NNs as technical means to deliver enhanced solutions for these critical SHM tasks. The research comprises four interconnected components:

(1) A novel sensor placement criterion is proposed to optimize sensor configurations by simultaneously considering global structural characteristics and local signal quality metrics, with verifiability embedded in the evaluation framework. This criterion is synergistically integrated with advanced optimization algorithms through adaptation for combinatorial optimization problems, ensuring rapid convergence. To resolve the conflict between local optimum quality dependence on initial point sets and convergence difficulties caused by excessive candidate points, an automatic initial point set generation algorithm is developed. (Chapter 3)

(2) A composite denoising strategy is established to address diverse data anomalies. For regular noise, adaptive filtering standards are dynamically configured according to signal fluctuation patterns and sensor specifications, replacing rigid “one-size-fits-

all" approaches. Specialized reconstruction algorithms are created to handle signal loss and outliers caused by sensor failures, component loosening, or transmission interruptions, demonstrating robustness against synchronous or long-term missing data scenarios while achieving comprehensive outlier detection and recovery. (Chapter 4)

(3) NN architectures specifically tailored for temporal inference tasks (including prediction and cross-variable mapping) are developed to realize higher accuracy than traditional methods. These architectures should exhibit enhanced memory capacity for long-term prediction and superior generalization performances for variable transformation tasks in physical systems. (Chapter 5)

(4) Leveraging purified monitoring data and high-precision inference results, damage diagnosis algorithms with heightened sensitivity to early-stage and localized defects are formulated, enabling simultaneous damage localization and severity quantification. Given the scarcity of labeled damage data, these algorithms incorporate advanced interpolation and extrapolation mechanisms and exceptional spatiotemporal correlation extraction capacities. Concurrently, rapid-update DT models are constructed to enable real-time structural state visualization and maintenance decision support. (Chapter 6)

The mentioned research objectives are systematically presented in Fig. 1.1.

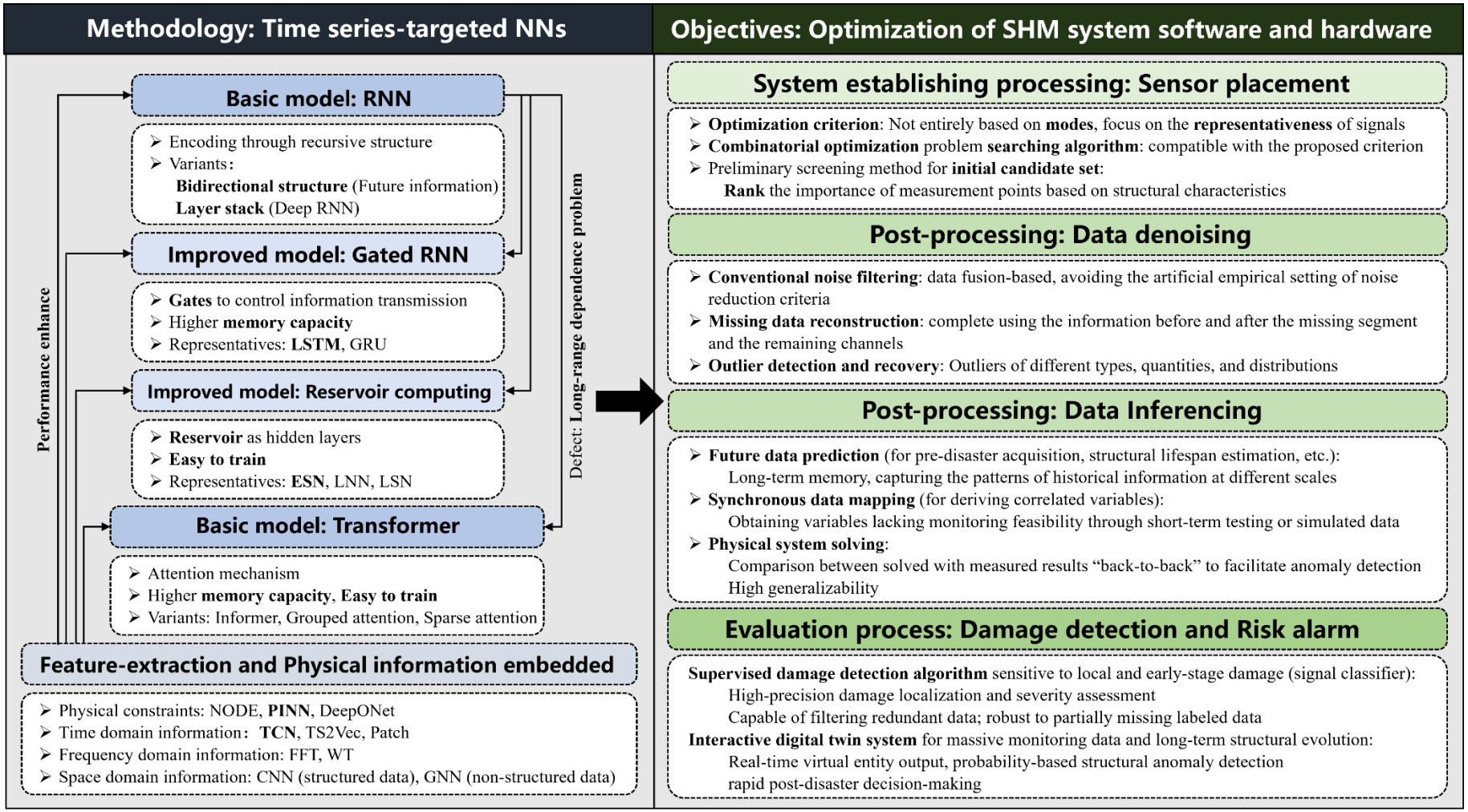


Fig. 1.1 Research objectives

LITERATURE REVIEW

2.1 SHM and its applications

SHM, as an interdisciplinary technological domain, systematically integrates real-time or periodic data acquisition (including acceleration, displacement, temperature and humidity, wind speed and direction, seismic motions, etc.) with analytical evaluation to diagnose structural integrity across engineered systems such as bridges (Zhang et al., 2024c; Li et al., 2008), buildings (Liu, Yang & Zhang 2017; Ni et al., 2009), railway tracks (Wang, Ni & Wang 2020), tunnels (Ding et al., 2013), mechanical systems (Zhao et al., 2019), and aerospace equipment (Diamanti & Soutis 2010). Its core objectives encompass identifying latent damage or degradation (Tibaduiza et al., 2018; Tan et al., 2024b), predicting residual service life (Hajializadeh, OBrien & O'Connor 2017), and optimizing maintenance strategies to reduce operational costs while ensuring safety (Tan et al., 2025). SHM is fundamentally defined as a systematic process combining sensor networks, data acquisition and processing technologies, and physical or data-driven models to continuously monitor and interpret structural dynamic responses (Sohn et al., 2003).

Initially applied in aerospace engineering for aircraft fatigue crack detection, SHM has expanded into civil and mechanical engineering and energy infrastructure, becoming indispensable for modern industrial asset management. Within civil engineering, SHM implementation has achieved widespread adoption. For material-level monitoring, Behnia, Chai & Shiotani (2014) implemented a remote real-time system tracking concrete strength evolution during construction of a 250-meter super-tall building, while Perry et al. (2017) employed thermocouple sensor networks with the maturity method to predict concrete strength degradation patterns in wind turbine foundations. In structural applications, Ibrahim, Han & Liu (2008) developed distributed sensor networks capturing bridge displacement ratios for post-earthquake condition assessment. Xie et al. (2018) created a real-time tunnel settlement risk monitoring system integrated with mobile

applications. The P-block at Queensland University of Technology, Australia, utilizes distributed sensors coupled with finite element model (FEM) updating techniques to monitor wind and seismic responses and optimize seismic designs (Nguyen et al., 2015). Bezas et al. (2020) devised an SHM system for historical structures, employing accelerometers and displacement sensors to track deformation in complex fragile components.

Monitoring data enable the realization of SHM's functionalities, with their effective and multivariate acquisition relying on advancements in sensing technologies. Conversely, SHM innovations have significantly enhanced monitoring precision and expanded application scenarios. Traditional vibration monitoring depends on contact-based sensors such as piezoelectric accelerometers, displacement sensors (e.g., eddy current sensors), and velocity sensors (e.g., electromagnetic vibrometers). Among these, accelerometers are widely adopted in dynamic system vibration analysis due to their high-frequency response and compact design, though their performance is constrained by sensitivity to installation positions and environmental noise (Zhu et al., 2018). Eddy current sensors excel in shaft displacement monitoring for mechanical systems but exhibit limited low-frequency capabilities.

The advent of non-contact measurement technologies has introduced laser Doppler vibrometers (LDVs), which offer high resolution and non-invasive operation, making them ideal for complex structural vibration monitoring (Staszewski et al., 2012; Yu, Tang & Vinayaka 2024). However, their high cost and operational complexity restrict large-scale deployment. Recent progress in fiber-optic sensing technologies, exemplified by FBGs, demonstrates advantages including electromagnetic interference immunity, distributed measurement capacity, and long-term stability. These systems enable simultaneous monitoring of strain and temperature variations in civil structures like bridges and dams (Majumder et al., 2008; Kinet et al., 2014; Soman et al., 2025). Furthermore, miniaturization and cost reduction of micro-electro-mechanical systems (MEMS) (Sivasuriyan et al., 2024), piezoelectric materials (Zhang et al., 2024d),

magnetorheological materials (Chong, Kim & Chon 2014), and metamaterials (Ozbey et al., 2014) have pioneered new approaches for embedded monitoring.

Advancements in imaging technologies have significantly enhanced non-contact and remote monitoring capabilities. Conventional three-channel optical imaging systems, when integrated with computer vision techniques, can extend their application scope, for instance, identifying concrete cracks or steel corrosion through high-resolution image analysis (Ali et al., 2023; Dong et al., 2022), and extracting subtle vibrational modes from videos using phase-based motion magnification techniques (Gao et al., 2023). Mobile platforms such as unmanned aerial vehicles (UAVs) (Sreenath et al., 2020; Akbar, Qidwai, & Jahanshahi 2019), submersible robots (Jiao et al., 2024), and cable-climbing or rail-mounted devices (Shin et al., 2013; Myeong et al., 2015) further enable efficient surface damage inspection for large-scale linear infrastructure. Hyperspectral cameras expand spectral channels to leverage material-specific reflectance characteristics across wavelengths, facilitating subsurface defect detection including material identification and corrosion product analysis (Shaban 2013). Depth-sensing cameras utilizing structured light or time-of-flight principles generate 3D depth maps, supplementing spatial detail acquisition in confined spaces or dynamic environments (Ye, Dong & Liu 2016).

More precise three-dimensional imaging is achievable via 3D scanning technologies with superior data acquisition efficiency and scenario adaptability. As a core technology, laser detection and ranging (LiDAR) (Kaartinen, Dunphy & Sadhu 2022; Park et al., 2007; Ochieng et al., 2018) employs pulsed laser emissions and measures reflection time or phase differences to generate high-density 3D point clouds, widely applied in topographic mapping, structural monitoring, and industrial inspection. For example, dense LiDAR-derived point clouds can be imported into building information modeling (BIM) platforms during construction acceptance, enabling precise quantification of crack propagation or settlement by comparing as-built and in-service structural deformation data (Riveiro, DeJong & Conde 2016; Jafari, Khaloo & Lattanzi 2016). For dynamic targets, terrestrial laser scanning systems integrated with global positioning systems (GPS)

(Wong, Man & Chan 2001) and inertial measurement units (IMUs) (Hester et al., 2018) achieve millimeter-level 3D spatial modeling and real-time displacement field reconstruction, delivering high-precision spatial data for heritage restoration, vibration control, and autonomous navigation (Sofi et al., 2022). Fig. 2.1 illustrates the sensor technologies applied in SHM systems.

Advancements in sensors and the Internet of Things (IoT) have enriched SHM with broader measurable variables and enhanced precision and resolution (Mahmud et al., 2018), while simultaneously introducing the first critical technical challenge: optimal sensor placement (OSP). Ideally, increasing the number of sensors improves data completeness, but practical constraints limit sensor quantity due to economic costs, data transmission and storage bottlenecks, and structural load impacts from excessive sensor deployment. For a fixed sensor count, data quality varies significantly with placement locations, as structural responses exhibit interdependencies. Thus, OSP aims to maximize information independence while minimizing redundancy (Chmielewski, Palmer & Manousiouthakis 2002). The process of defining evaluation metrics and applying optimization algorithms to determine optimal sensor positions is termed OSP (Tan & Zhang 2019; Liu et al., 2021; Mehrjoo et al., 2022; Yang et al., 2022).

Since the late 20th century, numerous OSP methods have emerged. Carne et al. proposed the modal assurance criterion (MAC) (He et al., 2015) and its corresponding minimization algorithm, iteratively selecting measurement points that minimize the squared cosine of modal vector angles to ensure modal independence. Kammer and Yao (1994) developed the effective independence (EI) method, optimizing sensor placement by maximizing specific norms of the Fisher information matrix to evaluate modal contributions. Ercan and Papadimitriou (2023) investigated entropy minimization-based strategies to reduce information uncertainty. Cobb and Liebst (1997) and Shi, Luo & Zhang (2000) introduced damage indices based on eigenvalue and eigenvector errors and mode shape variations, prioritizing measurement points by their sensitivity to damage indicators. For displacement field reconstruction, researchers employed inclinometers and nonlinear

least-squares fitting to optimize sensor layouts based on rotational data accuracy (Cao et al., 2023). Other approaches include modal kinetic energy (MKE) (Pooya & Massumi 2022) and spatial sampling methods (Stubbs & Park 1996), which rank measurement points by their contribution to monitoring objectives.

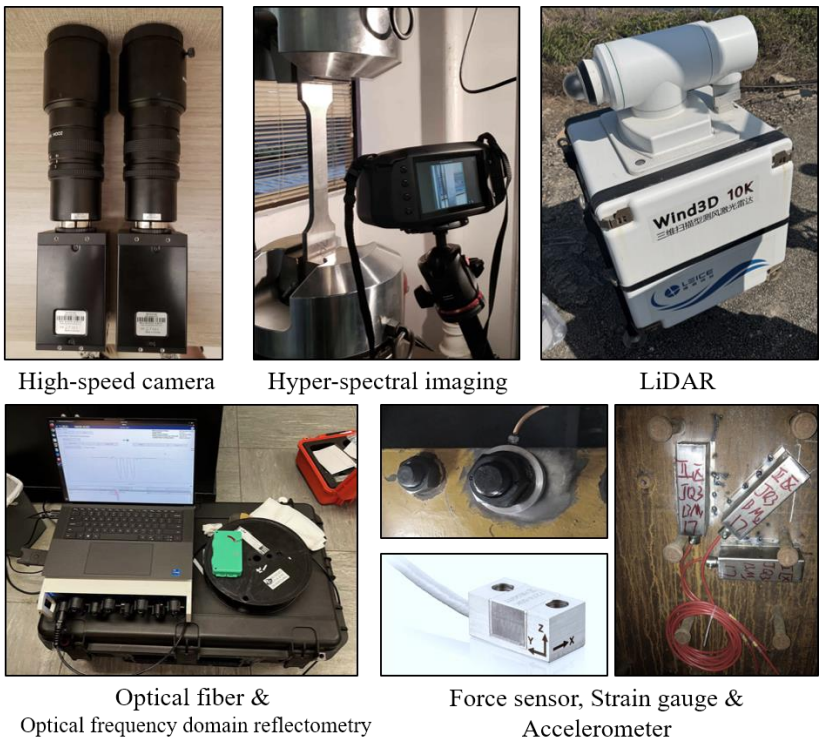


Fig. 2.1 Sensors used in SHM systems

However, most important or large-scale structures possess high design margins. The sensor placement fully determined by modal information may fail to detect early-stage or local damage under temperature effects. More importantly, in order to capture more modal information, measurement points tend to be concentrated in areas with significant vibrations, while areas such as beam-column joints, which have high stiffness (small responses) and are prone to damage, are easily overlooked. Moreover, validating algorithm efficacy requires impractical large-scale sensor deployment on actual structures, leaving many methods unproven in real-world scenarios. Consequently, the OSP field demands novel optimization objectives that prioritize local information accuracy while ensuring verifiability. Concurrently, emerging optimization algorithms, such as evolutionary algorithms (Civera et al., 2021) and heuristic algorithms (Jia, Feng & Liu

2015; Yi et al., 2017; Gomes & Pereira 2020; Deif & Gadallah 2017), also require domain-specific adaptations to address OSP challenges effectively.

Following the implementation of OSP solutions on target structures, continuous SHM data streams necessitate subsequent processes including damage diagnosis, condition assessment, residual life prediction, and risk warning based on time-varying data. Practically, raw acquired data require essential denoising, and often demand data inferencing (prediction, synchronized mapping, variable transformation) to fulfill final evaluation objectives. These tasks fundamentally constitute mathematical challenges in temporal sequence processing, such as reconstruction, forecasting, and anomaly detection. The evolution of data processing methodologies represents another core driver of SHM technological advancement.

Early signal processing relied on time-domain analysis (amplitude, mean, variance, gradient) and frequency-domain transformations like fast Fourier transform (FFT). However, standard FFT fails to capture transient features in non-stationary signals, prompting the emergence of time-frequency techniques including windowed FFT (Giurgiutiu & Yu 2003) and WT. WT's multi-scale analysis capability excels in crack localization for steel and concrete components. Researchers further proposed empirical mode decomposition (EMD) to handle nonlinear and non-stationary signals, successfully identifying delamination damage in composite structures (Pines & Salvino 2006; Zheng et al., 2021).

Damage identification methods based solely on frequency and modal changes once dominated structural diagnostics (Azhar et al., 2024; Bisheh & Amiri 2023), yet these approaches only detect severe damage affecting global dynamic characteristics, lacking sensitivity to early-stage localized defects. Recent data-driven paradigms have revolutionized SHM post-processing. Statistical time-series models like auto-regressive integrated moving average (ARIMA) (Peter & Silvia 2012; Chen et al., 2022; Zhang 2003) analyze residual fluctuations to identify structural stiffness changes (Chen, Dai & Zheng 2022). Early ML algorithms (support vector machines (Sapankevych & Sankar 2009),

random forests (Lin et al., 2017)) leveraged extracted modal parameters (frequencies, mode shapes) for damage classification. Probabilistic approaches integrating Bayesian theory (Du et al., 2022; Dezhkam et al., 2023) and Monte Carlo simulations (Wang et al., 2023) enhanced robustness through uncertainty quantification, often serving as correction modules for primary algorithms. Nevertheless, real-world data containing noise and sensor faults challenge the precision of these methods, failing to meet escalating demands from structural operators. High-precision, robust, and generalizable SHM post-processing algorithms remain imperative for development.

2.2 Time series-targeted NNs and their applications

Sequence data processing has long been a central challenge in artificial intelligence. Early recurrent neural networks (RNNs) addressed temporal dependencies by introducing delayed loop connections that encode historical sequence data into hidden layer neuron states, enabling information transfer across time steps and overcoming the limitations of traditional feedforward networks in modeling temporal relationships (Schmidt 2019; Salehinejad et al., 2017). However, standard RNNs suffer from vanishing and exploding gradient issues (Al-Selwi et al., 2023), hindering their capacity to capture long-range dependencies. This limitation catalyzed the development of long short-term memory (LSTM) networks, which employ gated mechanisms (input gates, forget gates, and output gates) to maintain parallel cell states. These states enable selective retention of critical information and dynamic filtering of irrelevant data over extended periods (Hochreiter & Schmidhuber 1997; Li et al., 2020; Yu et al., 2019). The gated recurrent unit (GRU), a simplified LSTM variant, consolidates three gating structures into update and reset gates while maintaining comparable performance with reduced computational complexity (Cho et al., 2014; Chung et al., 2014; Ravanelli et al., 2018). Gated RNN architectures, particularly LSTM and its derivatives, have dominated time series processing benchmarks and been widely adopted in SHM applications. Notable implementations include damage identification (Choe, Kim & Kim 2021), wind speed prediction (Jaseena & Kovoor 2021), and occupant comfort assessment (Yelisetti et al., 2023), demonstrating

their versatility in structural monitoring tasks.

To further enhance the performance of RNN-family models, researchers typically employ three principal strategies. First, bidirectional architectures enable dual-directional context encoding: forward layers process historical sequences while backward layers analyze future contexts, allowing comprehensive temporal feature extraction (Schuster & Paliwal 1997; Graves, Jaitly & Mohamed 2013). This approach significantly strengthens contextual understanding but remains incompatible with forecasting tasks requiring causality. Second, deep stacking through multi-layer RNN configurations (Pascanu et al., 2013; Sagheer & Kotb 2019) enhances the upper bound of mappable functional complexity via network scaling. Third, auxiliary techniques like spectral feature integration (Xu, Zhang & Xiao 2019) and temporal patch aggregation (Gong, Tang & Liang 2023) augment learning capacity by introducing frequency-domain characteristics or segment-wise pattern recognition. Notably, RNN variants still encounter computational inefficiency when handling ultra-long sequences, driving ongoing exploration of alternative architectures.

Emerging architectures improve upon RNN foundations through innovative modifications. Reservoir Computing circumvents recurrent structure training by introducing large-scale reservoirs to mitigate long-term dependency issues. A representative implementation, the echo state network (ESN) proposed by Jaeger and Haas (2004), achieves enhanced accuracy while simplifying training to linear regression through optimized reservoir configurations (He et al., 2023; Larger et al., 2017; Antonelo, Camponogara & Foss 2017). Meanwhile, leveraging RNN's inherent alignment with physical system dynamics, researchers have developed physics-informed neural networks (PINNs) (Raissi, Perdikaris & Karniadakis 2019; Yuan et al., 2022). These integrate partial differential equations (PDEs) into loss functions, synergizing physical laws with data-driven approaches. The partial derivative terms in PDEs can be automatically obtained through the error backpropagation algorithm. Ultimately, PINNs demonstrate unique advantages in spatiotemporal forecasting tasks like fluid dynamics (Rui et al.,

2024) and heat transfer (Bararnia & Esmaeilpour 2022). Neural ordinary differential equations (NODEs) employ continuous-depth models to replace discrete network layers, defining the hidden state dynamics via a neural network-parameterized ordinary differential equation (ODE) and treating depth as a continuous time variable. Utilizing ODE solvers for gradient backpropagation via the adjoint method, they establish mathematically rigorous frameworks particularly well-suited for modeling irregularly sampled time series due to their inherent continuous-time formulation (Chen et al., 2018; Yi 2023). Deep operator networks (DeepONets) pioneer operator learning paradigms through separate branch and trunk networks that map input functions to spatial coordinates; the branch network encodes the input function, the trunk network encodes the output location, and their dot product predicts the output function value, enabling learning of mappings between function spaces (Lu et al., 2021). Combining reservoir computing with physics-informed principles, Liquid neural networks (LNNs) (Hasani et al., 2021) incorporate dynamically adjustable time constants to enhance network kinetics, enabling adaptive processing speeds aligned with signal temporal variations. However, physics-informed networks face limited real-world adoption due to unresolved conflicts between physical constraints and data-driven training objectives, with most applications remaining confined to numerical case studies.

Alternative architectures transcend recurrent structure limitations through innovative approaches. temporal convolutional networks (TCNs) employ dilated causal convolutions with hierarchical expansion coefficients to achieve parallel processing and extended receptive fields, ensuring temporal causality while accelerating training (Bai, Kolte & Koltun 2018; Oh et al., 2020; Lea et al., 2017). Despite their proliferation of variants, these convolutional neural network (CNN)-inspired architectures completely circumvent long-term dependency issues. However, their reliance on deep stacking for temporal pattern extraction results in performance parity rather than superiority compared to advanced recurrent models like LSTM and ESN. Nevertheless, TCNs' training efficiency has garnered significant adoption in SHM practice, with recent literature documenting

successful applications in sequence classification and anomaly detection (He et al., 2016; Szegedy et al., 2015; Redmon et al., 2016). While excelling in temporal feature extraction, TCNs demonstrate greater aptitude for spatial domain information processing, particularly inter-channel relationship learning in multivariate time series. For complex unstructured data beyond grid-aligned formats, graph convolutional networks (GCNs) (Yue et al., 2022; Yu, Yin & Zhu 2017; Wu et al., 2018) extend CNN principles to extract spatiotemporal correlations from multi-channel signals. GCN-embedded temporal neural architectures have been successfully deployed in urban wind and temperature field interpolation (Khodayar & Wang 2018) and transportation load forecasting (Wang et al., 2020), demonstrating versatility in infrastructure monitoring tasks.

The advent of Transformer architectures has revolutionized sequence data processing by fundamentally altering computational paradigms (Vaswani et al., 2017). Its self-attention mechanism computes global correlations among sequence points, overcoming the parallelization limitations inherent in RNN-based models. Positional encoding effectively preserves sequential order information, while multi-head attention enables multi-dimensional feature extraction (Zhou et al., 2021). This breakthrough ushered in the era of large-scale models for temporal data processing. In SHM applications, Wang, Shang & Song (2023) demonstrated that attention mechanisms significantly enhance damage feature extraction by capturing multi-scale temporal dependencies. Meanwhile, Nong et al. (2023) explored Transformer-based multimodal fusion, integrating visual and sensor data for comprehensive structural assessment. Yuan et al. (2024) developed physics-guided Transformers that synergize mechanical principles with DL to improve damage localization accuracy.

The performance leap of Transformer architectures stems from their exponentially increased parameter number and heightened demands for training data volume and computational resources. Subsequent improvements introduced computational efficiency enhancements through sparse attention (restricting attention scope) (Roy et al., 2021), grouped attention (parallel computation) (Burchi & Vielzeuf 2021), and linear attention

using kernel approximations (Shen et al., 2021). Originally developed for natural language processing (NLP), Transformers excel in handling sparse word-vector sequences where similarity-based attention mechanisms align with linguistic semantics. However, SHM datasets exhibit fundamentally different characteristics: engineering sensor data lacks inherent sparsity, and attention scores derived from temporal similarity calculations lack clear physical interpretations. Empirical studies reveal limited success in directly applying large language models (LLMs) (Achiam et al., 2023; Touvron et al., 2023; Guo et al., 2024; Zhao et al., 2023) to SHM tasks, with performance often inferior to specialized compact models. Consequently, strategic algorithm development should focus on two complementary directions: enhancing small-model capabilities through knowledge distillation (Gou et al., 2021) and meta-learning (Hospedales et al., 2021), and customizing large models via domain-specific lightweight modifications. These approaches aim to optimize SHM system outputs for precise condition evaluation and safety-critical decision-making.

SENSOR PLACEMENT BASED ON TIME SERIES-TARGETED NEURAL NETWORKS

3.1 Introduction

The determination of sensor placement configurations constitutes the primary technical challenge in SHM system implementation, known as OSP problem as discussed in Section 2.1. Logically, since no measured data exists during sensor placement design, all methodologies in this chapter are developed based on structural numerical models (primarily FEM). Theoretically, every FEM node could serve as a potential sensor location except those restricted by physical constraints. However, information redundancy among different structural positions and practical limitations (e.g., cost) necessitate systematic strategies to design evaluation metrics that quantify the effectiveness of sensor configurations relative to monitoring objectives. Thus, OSP problem-solving can be formulated as: establishing optimization criteria based on these metrics, then iteratively identifying optimal configurations from all feasible candidate positions through optimization algorithms to maximize criterion fulfillment.

Three critical technical components emerge in this framework. Firstly, traditional criteria like MAC exhibit limitations in practical verification and lack capacity to evaluate local information acquisition quality. This chapter proposes time series-targeted NN inference accuracy-based metrics to assess measurement point representativeness. Secondly, the author modifies the state-of-the-art human evolutionary optimization algorithm (HEOA) (Lian & Hui, 2024) for combinatorial optimization problems, developing a novel variant that balances convergence speed with near-global optimum attainment. Thirdly, though not mandatory for all OSP tasks, the initial candidate set screening critically impacts searching efficiency. To address the challenges of manual selection from excessive FEM nodes, we propose a PageRank-inspired candidate ranking algorithm (Berkhin, 2005) based on the graph data theory that adaptively generates initial point sets of varying scales.

3.2 OSP problem-solving

For a given structure, omitting initial candidate set screening, OSP problem-solving comprises two essential components: defining and computing optimization objectives, and iterative search via optimization algorithms. This section proposes localized information-focused optimization criteria to address deficiencies in conventional approaches. Structural sensors essentially sample specific physical fields (e.g., accelerometers sample vibration acceleration fields). The OSP problem fundamentally seeks positions that optimally represent the complete physical field where representativeness implies maximal information coverage about unsampled regions. Time series-targeted NNs enable mapping between correlated sequences, allowing the most representative measurement subset to be defined as configurations that minimize the error in inferring full-field data from subset measurements. This criterion is designated as the field inference accuracy (FIA). Regarding optimization algorithms, this section enhances HEOA for combinatorial optimization. HEOA, a metaheuristic optimization algorithm inspired by human societal evolution mechanisms, divides global search into human exploration and human development phases, simulating collective human collaboration for efficient optimization (Cheng, Zhang & Cao, 2024).

3.2.1 FIA and its inferencing model

The proposed OSP solution in this chapter introduces spatial domain sampling concepts. Selected sensor positions serve as sampling points for the target physical field, requiring optimal configurations to enable high-accuracy inference (inversion) of the original field from their acquired data. After feasibility analysis, suppose there exist n candidate sensor positions on the target structure, forming set Ω_n. Given economic and structural constraints allowing a maximum of m sensors, let subset $\Omega_m \subseteq \Omega_n$ denote the chosen positions, with the remaining $(n-m)$ positions forming unselected set Ω_u. The physical field monitored by these sensors (e.g., acceleration field for accelerometers) is defined as $\boldsymbol{Z} \in \mathbb{R}^{n\times T}$ (assuming single vibration degree of freedom (DoF) per position,

while multi-DoF cases follow analogous logic; T is the time length of the measurement). Accordingly, the monitored and unmonitored time series under configuration Ω_m are denoted as $\boldsymbol{X} \in \mathbb{R}^{m\times T}$ and $\boldsymbol{Y} \in \mathbb{R}^{(n-m)\times T}$, respectively (throughout this study, these definitions are maintained, that $\boldsymbol{Z}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ stands for whole dataset, input, and output, respectively). Considering spatiotemporal correlations across structural positions, there must be inherent mapping relationships between $\boldsymbol{X}$ and $\boldsymbol{Y}$, which can be learnt with the utilization of time series-targeted NNs. Let $\mathcal{G}(\cdot)$ represent the NN learning this mapping to infer $\boldsymbol{Y}$ from $\boldsymbol{X}$, as,

$$\boldsymbol{Y} \approx \widehat{\boldsymbol{Y}} = \mathcal{G}(\boldsymbol{X}) \tag{3.1}$$

Thus, for a given $\mathcal{G}(\cdot)$, the FIA criterion can be expressed as,

$$\underset{\Omega_m \subseteq \Omega_n}{argmin}\, FIA(\Omega_m) = \sum_{t=1}^{T}\sum_{k=1}^{n-m} \beta_k \big(Y_k(t) - \mathcal{G}(\boldsymbol{X})_k(t)\big)^2 \tag{3.2}$$

where $FIA(\Omega_m)$ denotes the score of sensor configuration Ω_m under the FIA criterion, which correlates with the precision of $\widehat{\boldsymbol{Y}}$. A smaller value indicates higher inference accuracy and better configuration. The term β_k represents the inference error amplification factor for each channel, primarily determined by the ratio of the channel's variance to the maximum variance across all channels. This factor prevents the FIA criterion from disproportionately favoring channels with larger vibration amplitudes.

In most cases, the OSP problem is a priori because field measurement data is unavailable during the sensor placement phase. Thus, all subsequent data are derived from structural FEM simulations. Since most damage diagnosis and modal identification methods focus on free-vibration time series (response signals under seismic or strong wind loads are often contaminated by excitation characteristics), the physical field here typically represents the FEM response field under white noise excitation.

Fig. 3.1 illustrates the iterative solution framework for the proposed OSP based on the FIA definition. At the κ th iteration, the search algorithm proposes a sensor configuration $\Omega_m^{(\kappa)}$, with its complement in Ω_n denoted as $\Omega_u^{(\kappa)}$. A new time series-

targeted NN with identical architecture and hyperparameters is constructed. A subset of $\boldsymbol{X}^{(\kappa)}$ and $\boldsymbol{Y}^{(\kappa)}$ trains the model (using consistent training strategies across iterations), while the remaining data tests performance. By comparing the network output $\widehat{\boldsymbol{Y}}^{(\kappa)}$ with ground truth $\boldsymbol{Y}^{(\kappa)}$, the score $FIA\left(\Omega_m^{(\kappa)}\right)$ is computed as the fitness value for the searching algorithm. The algorithm generates a new configuration $\Omega_m^{(\kappa+1)}$ by synthesizing current and historical FIA scores for the $(\kappa+1)$th iteration. The final Ω_m is selected upon FIA score convergence after sufficient iterations.

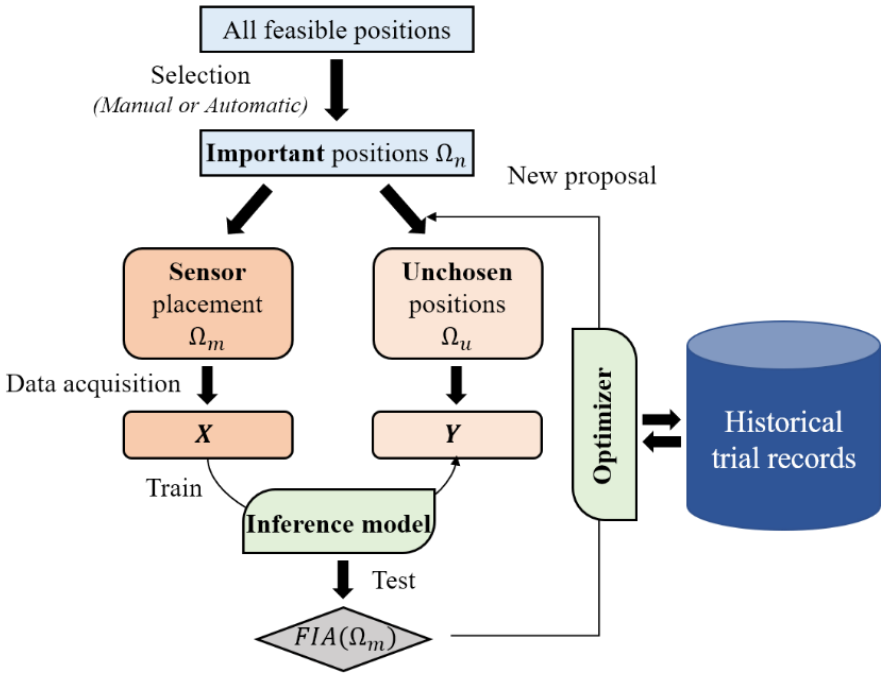


Fig. 3.1 FIA-guided OSP problem-solving

Theoretically, all architectures mentioned in Section 2.2 are applicable to this mapping task. However, considering subsequent iterative processes, the selected model should prioritize computational efficiency. This section employs a RNN with a feature-forcing mechanism (FFM) to implement field inference.

The standard RNN represents the most fundamental neural architecture for sequential data, utilizing delay loops to encode time series. The encoded information, manifested as sequential signals, is stored in hidden layer neurons and termed "states", denoted as $\boldsymbol{H} \in \mathbb{R}^{N_h \times t}$, where N_h indicates the number of hidden neurons. States are acquired through

recursive computation: the current state is derived by mapping both the current input and the preceding state through a linear layer, as formalized in Eq. (3.3).

$$\boldsymbol{H}(t) = f(\boldsymbol{W_h}[\boldsymbol{H}(t-1); \boldsymbol{X}(t)] + \boldsymbol{b_h}) \tag{3.3}$$

where $\boldsymbol{W_h} \in \mathbb{R}^{N_h \times m}$ represents the trainable weight matrix of the linear layer, $\boldsymbol{b_h}$ denotes the bias term, and $f(\cdot)$ signifies the nonlinear activation function (typically $tanh(\cdot)$). The notation $[\cdot;\cdot]$ indicates vector concatenation. As shown, given an initial state $\boldsymbol{H}(0)$, the RNN progressively encodes the entire sequence state $\boldsymbol{H}$ based on input $\boldsymbol{X}$, with parameter $\boldsymbol{W_h}$ shared across all encoding steps. At any timestep t, $\boldsymbol{H}(t)$ incorporates not only the contribution from the current input $\boldsymbol{X}(t)$ but also partially retained historical information from the sequence $\boldsymbol{X}(1)\sim\boldsymbol{X}(t-1)$, enabling downstream sequence processing tasks using $\boldsymbol{H}(t)$.

However, the state quality obtained from single-layer RNN encoding remains suboptimal. A common enhancement involves layer stacking, such as extending ESN to deep ESN as discussed later in Section 5.4.2. While this approach improves representational capacity, it exacerbates gradient chain complexity and amplifies long-term dependency issues (a problem inherently mitigated by ESN architectures but critical for standard RNNs). An alternative extension replaces the single layer mapping $[\boldsymbol{H}(t-1); \boldsymbol{X}(t)]$ to $\boldsymbol{H}(t)$ in Eq. (3.3) with a multi-layer perceptron (MLP) incorporating nonlinear activations. By substituting $\boldsymbol{W_h}$ with an $MLP(\cdot)$ function, the model gains enhanced spatiotemporal feature extraction capabilities. The architecture described above forms the inference framework under the FIA criterion, as illustrated in Fig. 3.2. Since synchronized mapping imposes lower memory demands, the performance impact of RNNs' long-term dependency limitations remains manageable.

To enhance model performance without increasing architectural complexity, this study proposes the FFM, which enforces temporal characteristics of target outputs by modifying the loss function. This mechanism draws from the PINN framework where variables are governed by PDEs. In PINNs, automatic differentiation through the computational graph enables the calculation of partial derivatives in PDEs. By incorporating deviations of

these derivatives into the loss function as penalty terms, the model learns physical laws encoded by the PDEs during training. A key contribution of PINNs lies in their systematic integration of physical constraints into loss functions. However, many tasks lack explicit PDE constraints, limiting PINNs' practical applicability.

Analysis of standard RNN predictions in this study reveals that errors in SHM data inference (e.g., acceleration signals) predominantly concentrate at signal peaks, often manifesting as premature descent phases in predictions compared to ground truth. This discrepancy stems from inaccurate slope estimations during the ascending phase preceding peaks, resulting in underestimated amplitude growth. To address this, we propose incorporating signal gradients into the loss function. Specifically, both 1st and 2nd order gradients are integrated into the optimization objective. To enable gradient computation, the output layer's single-step mapping is extended to five steps, i.e.,

$$\begin{aligned}&\left[\widehat{\boldsymbol{Y}}(t-2);\widehat{\boldsymbol{Y}}(t-1);\widehat{\boldsymbol{Y}}(t);\widehat{\boldsymbol{Y}}(t+1);\widehat{\boldsymbol{Y}}(t+2)\right]=\\&\boldsymbol{W_{out}}[\boldsymbol{H}(t-2);\boldsymbol{H}(t-1);\boldsymbol{H}(t);\boldsymbol{H}(t+1);\boldsymbol{H}(t+2)]+\boldsymbol{b_y}\end{aligned}\tag{3.4}$$

where $\boldsymbol{W_{out}} \in \mathbb{R}^{5(n-m)\times 5N_h}$ parameters to be trained in the output layer and $\boldsymbol{b_y}$ denotes the bias term. Besides, the 1st and 2nd order gradients at step t of the ground truth sequence can be computed using the central difference method, as shown in Eq. (3.5) and Eq. (3.6).

$$\nabla\boldsymbol{Y}(t)=\frac{\boldsymbol{Y}(t+1)-\boldsymbol{Y}(t-1)}{2}\tag{3.5}$$

$$\nabla^2\boldsymbol{Y}(t)=\frac{\nabla\boldsymbol{Y}(t+1)-\nabla\boldsymbol{Y}(t-1)}{2}=\frac{\boldsymbol{Y}(t+2)+\boldsymbol{Y}(t-2)-2\boldsymbol{Y}(t)}{4}\tag{3.6}$$

The same method can be applied to calculate the 1st and 2nd order gradients of the predicted sequence at step t, denoted as $\nabla\widehat{\boldsymbol{Y}}(t)$ and $\nabla^2\widehat{\boldsymbol{Y}}(t)$. Consequently, the loss function for the FFM-RNN model can be formulated with a base structure of mean square error (MSE). In addition to the MSE of the predicted values themselves, the loss function incorporates MSE terms for both the error of determining the 1st and 2nd order gradients, as specified in Eq. (3.7).

$$L(\boldsymbol{Y},\widehat{\boldsymbol{Y}})=\sum_{t=1}^{T}\left(\begin{array}{c}\beta_p\left(\boldsymbol{Y}(t)-\widehat{\boldsymbol{Y}}(t)\right)^2+\\ \left(\nabla\boldsymbol{Y}(t)-\nabla\widehat{\boldsymbol{Y}}(t)\right)^2+\left(\nabla^2\boldsymbol{Y}(t)-\nabla^2\widehat{\boldsymbol{Y}}(t)\right)^2\end{array}\right) \tag{3.7}$$

Here, β_p serves as a hyperparameter that regulates the contribution of each term in the loss function. Additionally, β_p amplifies the weight of peak points. Since inaccuracies at peak points tend to induce larger errors in both the peaks themselves and neighboring data points, β_p increases the contribution weight of peak points to the overall loss. The peaks are those points meet $\widehat{\boldsymbol{Y}}(t)=\max\left(\widehat{\boldsymbol{Y}}(t-2),\widehat{\boldsymbol{Y}}(t-1),\widehat{\boldsymbol{Y}}(t),\widehat{\boldsymbol{Y}}(t+1),\widehat{\boldsymbol{Y}}(t+2)\right)$. Consequently, errors occurring at peak locations are penalized more severely. The complete architecture of the FFM-RNN is also illustrated in Fig. 3.2.

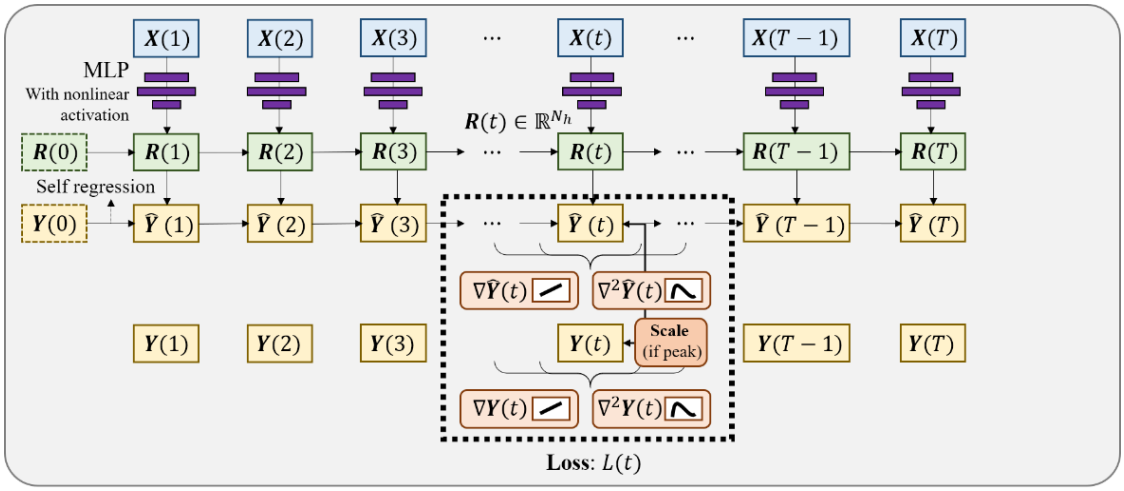


Fig. 3.2 FFM-RNN

3.2.2 Combinatorial HEOA

To collaboratively solve the OSP problem with the FIA criterion, this section improves the HEOA search algorithm to meet the requirements of combinatorial optimization problems, with the enhanced method termed combinatorial HEOA. The original HEOA draws inspiration from human evolutionary processes in complex environments (Lian & Hui, 2024). In HEOA, logistic chaos mapping serves as the initialization technique to simulate chaotic processes during early evolutionary stages. By assigning indices to all potential sensor placement locations, Ω_n can be represented as the set of the

first n positive integers. Then, an initial sensor configuration $\Omega_m^{(\kappa)} = [\mathcal{S}_1^{(0)}; \mathcal{S}_2^{(0)}; \dots; \mathcal{S}_m^{(0)}]$ can be determined via Eq. (3.8).

$$\mathcal{S}_i^{(0)} = \left\lfloor \epsilon \frac{\mathcal{S}_{i-1}^{(0)}}{n} \left(1 - \frac{\mathcal{S}_{i-1}^{(0)}}{n}\right)(n-1) + 1 \right\rfloor, i = 1, 2, \dots, m \tag{3.8}$$

where, $\mathcal{S}_0^{(0)}$ is a number randomly sampled from Ω_n, ϵ is a tunable hyperparameter, and $\lfloor \cdot \rfloor$ denotes the floor operation. As HEOA is a population-based search algorithm involving simultaneous competition and evolution among multiple individuals, $\mathcal{H}$ initial configurations are generated using the above method to form the initial population, denoted as $\boldsymbol{\mathcal{Z}^{(0)}} = [\Omega_{m,1}^{(0)}; \Omega_{m,2}^{(0)}; \dots; \Omega_{m,\mathcal{H}}^{(0)}]$.

Subsequently, the population enters the human exploration phase, which typically persists until the first quarter of the maximum iteration count. During this phase, all individuals adopt a unified search strategy. At the start of each iteration, the individual $\Omega_{m,best}$ achieving the best fitness across all prior iterations is recorded. In Combinatorial HEOA, the concept of "average position" ($\Omega_{m,mean}^{(\kappa)}$) from the original HEOA is replaced with the "most frequently selected positions" ($\Omega_{m,most}^{(\kappa)}$) from the previous iteration. $\Omega_{m,most}^{(\kappa)}$ comprises the m positions with the highest occurrence frequency across all configurations in $\boldsymbol{\mathcal{Z}^{(\kappa)}}$.

During the update process, the $\hbar$th individual ($\Omega_{m,\hbar}^{(\kappa+1)}$) in $(\kappa+1)$th iteration first assimilates a subset of positions from $\Omega_{m,best}$ proportionally. Subsequently, a portion of positions included in $\Omega_{m,most}^{(\kappa)}$ but excluded from $\Omega_{m,best}$ is selected. Finally, any remaining vacancies are filled by randomly choosing unselected positions from the individual's previous configuration $\Omega_{m,\hbar}^{(\kappa)}$. This update procedure is shown in Eq. (3.9).

$$\begin{aligned}\Omega_{m,\hbar}^{(\kappa+1)} = \mathcal{C}\left(\Omega_{m,best}, \mathcal{P}_1(\kappa)\right) \cup \mathcal{C}\left(\left(\Omega_{m,most}^{(\kappa)} \backslash \Omega_{m,best}\right), \mathcal{P}_2(\kappa)\right) \cup \\ \mathcal{C}\left(\left(\Omega_{m,\hbar}^{(\kappa)}\left(\Omega_{m,best} \cup \Omega_{m,most}^{(\kappa)}\right)\right), \mathcal{P}_3(\mathcal{P}_1, \mathcal{P}_2)\right)\end{aligned} \tag{3.9}$$

where, $(\Omega_1 \backslash \Omega_2)$ denotes the set difference operation, and $\mathcal{C}(\Omega, \mathscr{p})$ represents randomly selecting elements from set Ω according to proportion $\mathscr{p}$. The parameters $\mathscr{p}_1$, $\mathscr{p}_2$, and $\mathscr{p}_3$ correspond to the sampling ratios from the three sources. Specifically, $\mathscr{p}_1$ and $\mathscr{p}_2$ decay as the iteration count κ increases, while $\mathscr{p}_3$ is determined by the number of remaining vacancies in $\Omega_{m,\hbar}^{(\kappa+1)}$. To introduce stochastic movement and maintain population diversity, elements in $\Omega_{m,\hbar}^{(\kappa+1)}$ are partially replaced (with probability $\mathscr{p}_r$) by positions not included in any individual of the current iteration, as formalized in Eq. (3.10).

$$\Omega_{m,\hbar}^{(\kappa+1)} \leftarrow \mathcal{B}(\mathscr{p}_r)\mathcal{R}\left(\Omega_{m,\hbar}^{(\kappa+1)}, \mathcal{C}\left(\left(\Omega_n \backslash \Omega_{m,\hbar}^{(\kappa+1)}\right), \mathscr{p}_4(k)\right)\right) \tag{3.10}$$

where, $\mathcal{B}(\mathscr{p}_r)$ denotes a Bernoulli trial, and $\mathcal{R}(\Omega_1, \Omega_2, \mathscr{p})$ represents replacing a random subset of elements in Ω_1 with elements selected from Ω_2 according to proportion $\mathscr{p}$. The parameter $\mathscr{p}_4(k)$, which also decays with iteration count, governs this replacement ratio.

Following the egalitarian search phase, individuals in the population gradually develop distinct capabilities and assume specialized roles during the subsequent human development phase. These roles are divided into four categories based on fitness rankings: leaders (top 40% in fitness), explorers (40th to 80th percentile), followers (80th to 90th percentile), and stragglers (bottom 10% in fitness). Leaders primarily occupy the most optimal regions within the explored space and conduct thorough local searches, as global optima are likely to reside in these areas. Explorers are assigned to investigate uncharted regions, while followers focus on approaching high-quality regions identified by leaders and explorers to bridge gaps between different zones. The remaining individuals are classified as stragglers, which are eliminated and regenerated based on current optimal positions to refresh population diversity. The update mechanisms for these four roles during the human development phase are formalized as follows:

$$\Omega_{m,\hbar}^{(\kappa+1)} = \mathcal{B}(\mathscr{p}_r)\mathcal{P}(\Omega_{m,\hbar}^{(\kappa)}, \Omega_n, \mathscr{p}_{le}) \tag{3.11}$$

$$\Omega_{m,\hbar}^{(\kappa+1)} = \mathcal{C}\left(\Omega_{m,worst}^{(\kappa)}, \mathscr{p}_e\right) \cup \mathcal{C}\left(\left(\Omega_{m,\hbar}^{(\kappa)} \backslash \Omega_{m,worst}^{(\kappa)}\right), (1-\mathscr{p}_e)\right) \tag{3.12}$$

$$\Omega_{m,\hbar}^{(\kappa+1)} = \mathcal{C}\left(\Omega_{m,best}^{(\kappa)}, \mathscr{p}_f\right) \cup \mathcal{C}\left(\left(\Omega_{m,\hbar}^{(\kappa)} \backslash \Omega_{m,best}^{(\kappa)}\right), \left(1 - \mathscr{p}_f\right)\right) \tag{3.13}$$

$$\Omega_{m,\hbar}^{(\kappa+1)} = \mathcal{C}\left(\Omega_{m,best}, \mathscr{p}_{lo}\right) \cup \mathcal{C}\left(\left(\Omega_{m,\hbar}^{(\kappa)} \backslash \Omega_{m,best}\right), \left(1 - \mathscr{p}_{lo}\right)\right) \tag{3.14}$$

Here, $\mathcal{P}(\Omega_1, \Omega_2, \mathscr{p})$ denotes an operation that slightly perturbs Ω_1 by replacing a subset of its elements with values selected from Ω_2 according proportion $\mathscr{p}$. $\Omega_{m,worst}^{(\kappa)}$ represents the sensor locations least frequently selected across all individuals in the previous iteration, while $\Omega_{m,best}^{(\kappa)}$ corresponds to the configuration provided by the individual with the highest fitness in the prior iteration. The constants $\mathscr{p}_{le}$, $\mathscr{p}_e$, $\mathscr{p}_f$, and $\mathscr{p}_{lo}$ denote the extraction ratios applied during the update processes for leaders, explorers, followers, and stragglers, respectively. The human development phase continues until either the maximum iteration count is reached or fitness convergence is achieved (defined as consecutive iterations with fitness improvement below a predefined threshold).

Validation of the FIA criterion and HEOA optimization algorithm for solving the OSP problem in SHM will be presented in Section 3.3, following the introduction of the initial candidate set generation method.

3.3 Initial candidate set generation method

The rationale for introducing the combinatorial HEOA and other search algorithms in Section 3.2 stems from the computational infeasibility of exhaustively evaluating all $\mathcal{C}_n^m$ possible configurations to identify the optimal solution, even with computationally tractable optimization criteria. For instance, selecting 10 sensor locations from 100 candidate positions yields 1.7×10^{13} possibilities, while real-world FEMs often exceed millions of nodes. Consequently, we resort to search algorithms to approximate local optima. The convergence behavior and solution quality of these algorithms critically depend on the candidate set Ω_n. In Section 3.2, Ω_n encompasses all feasible installation locations, but excessively large candidate sets drastically reduce search coverage, leading to slow convergence or suboptimal local solutions.

Traditional SHM configuration tasks often relied on expert engineers to manually curate a reduced FEM node subset as Ω_n or directly as Ω_n. However, such empirical, intuition-driven approaches are labor-intensive and lack reliability. To address this, this section proposes a sensor placement prioritization algorithm for efficiently generating Ω_n at scale. For geometrically simple structures, the top m ranked positions may even serve directly as Ω_m. Inspired by ranking methods in graph data for web pages, products, social network entities, and knowledge graph nodes (Wu et al., 2022; Huang et al., 2021), this algorithm adapts structural feature matrices (e.g., mass, stiffness, damping) as nodal adjacency matrices. After processing graph connectivity characteristics, nodes are ranked using a modified PageRank algorithm (Berkhin 2005), enhanced to account for asymmetric nodal connections.

3.3.1 Construction of structural graph

The structural FEM simulates the behavior of an entire structure by discretizing it and assigning specific deformation modes to each discrete element. Taking mechanical analysis as an example, the mechanical response (stress, strain) of the structure can be characterized through the displacements of discretized nodes. The correlations between nodal displacements are governed by the mass ($\mathbb{M}$), flexibility ($\mathbb{F}$), and damping ($\mathbb{C}$) matrices, as well as the structural vibration equation (e.g., Eq. (3.15) representing free vibration). Once these characteristic matrices are determined, the structural properties become fully defined.

$$\mathbb{F}\mathbb{M}\ddot{\boldsymbol{Z}} + \mathbb{F}\mathbb{C}\dot{\boldsymbol{Z}} + \boldsymbol{Z} = \boldsymbol{0} \tag{3.15}$$

All three matrices have dimensions of $\mathbb{R}^{N_n \times N_n}$ (where N_n denotes the number of structural DoFs) and are sparse symmetric matrices. An element in the ith row and jth column may be non-zero only if there exists a direct material connection between the ith and jth DoFs. Consequently, the structure can be abstracted as a graph $\mathbb{G}$. Due to the symmetry of the characteristic matrices, the connections (edges) between nodes in $\mathbb{G}$ are undirected. Let the node set $\mathcal{A}$ contain n elements and the edge set $\mathcal{V}$. However, since

each node typically possesses multiple DoFs, the characteristic matrices require preprocessing to describe nodal connectivity.

As illustrated in Fig. 3.3 (using a spatial truss structure for intuitive visualization, though applicable to other structures), each node has three translational DoFs in space, corresponding to three rows and columns in the characteristic matrices. To define the connection $\mathcal{V}_{ij}$ between nodes $\mathcal{A}_i$ and $\mathcal{A}_j$, consider the flexibility matrix: a force applied to any DoF at node j may induce displacements across all three translational DoFs at node i. This interaction is represented by a 3×3 block matrix$\mathbb{f}_{ij}$ in $\mathbb{F}$, formed by the intersection of the three columns corresponding to node j and the three rows corresponding to node i. This block quantifies the influence of forces at node j on the vibrational response of node i.

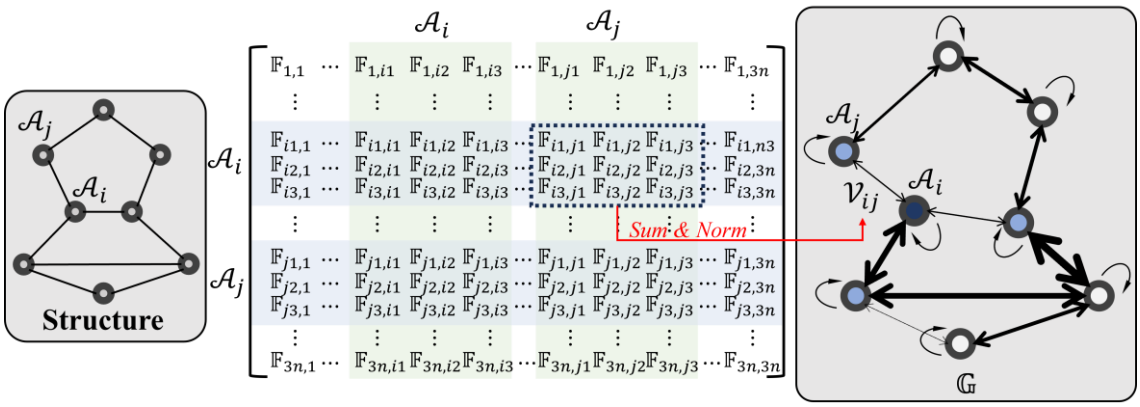


Fig. 3.3 Graph of a structure (FEM)

Theoretically, all nine elements in $\mathbb{f}_{ij}$ hold physical significance, and their collective influence can be quantified using the Frobenius norm $\|\mathbb{f}_{ij}\|_F$ to represent $\mathcal{V}_{ij}$. However, vibration monitoring in real-world structures often exhibits directional sensitivity. For instance, vertical displacements are prioritized in spatial structures, while lateral displacements are critical for super-tall buildings. To account for this anisotropy, the calculation of $\mathbb{f}_{ij}$ can incorporate directional weighting by amplifying matrix rows corresponding to critical displacement components. It is essential to note that non-truss elements may introduce rotational DoFs, resulting in mixed dimensional units and disparate magnitude scales between translational and rotational components within the

characteristic matrices. To address this, the matrix should be partitioned appropriately. When deploying accelerometers focused on translational motion, rows and columns associated with rotational DoFs in $\mathbb{f}_{ij}$ must be removed. Conversely, for inclinometers monitoring rotational behavior, translational components should be eliminated.

The same methodology applies to computing the Frobenius norms $\left\|\mathbb{m}^{-1}{}_{ij}\right\|_F$ and $\left\|\mathbb{c}^{-1}{}_{ij}\right\|_F$. The rationale for using inverse matrices stems from the monitoring objective: $\mathcal{V}_{ij}$ quantifies the susceptibility of nodal vibrations to external influences rather than the structure's resistance to excitation. The final aggregated $\mathcal{V}_{ij}$ is formulated through the following procedure, ensuring alignment with directional priorities and dimensional consistency.

$$\mathcal{V}_{ij} = \beta_{\mathbb{f}}\left\|\mathbb{f}_{ij}\right\|_F + \beta_{\mathbb{m}}\left\|\mathbb{m}^{-1}{}_{ij}\right\|_F + \beta_{\mathbb{c}}\left\|\mathbb{c}^{-1}{}_{ij}\right\|_F \tag{3.16}$$

Here, $\beta_{\mathbb{f}}$, $\beta_{\mathbb{m}}$, and $\beta_{\mathbb{c}}$ are weighting coefficients whose determination must account for both their relative importance and the need for magnitude normalization. Additionally, the 3×3 block formed by the intersection of the three rows and three columns corresponding to each node itself represents its intrinsic characteristics. This block is processed using the aforementioned method to compute $\mathcal{V}_{ii}$, which establishes a self-referential connection for node $\mathcal{A}_i$.

3.3.2 Candidate ranking based on PageRank

The PageRank algorithm, originally developed by Google, is a graph-based analytical method designed to assess node importance through modeling user navigation behavior as a random walk process to compute steady-state node probabilities (Berkhin 2005). This algorithm conceptualizes webpages as nodes and hyperlinks between them as edges within a massive graph structure. PageRank calculates an importance score ($PR(\mathcal{A}_i)$) for each node to establish rankings. Its foundational premises are: first, a node's importance increases with the number of in-links (connections from other nodes $\mathcal{A}_j$ to $\mathcal{A}_i$); second, in-links originating from highly important nodes carry greater weight. For any node $\mathcal{A}_i$,

let $\Omega_{in,i}$ denote the subset of nodes containing its in-link neighbors, with cardinality $|\Omega_{in,i}|$, and let $\Omega_{in,j}$ represent the out-link neighbor set of nodes $\mathcal{A}_j$ with cardinality $|\Omega_{out,j}|$. The PageRank score of $\mathcal{A}_i$ is then defined by:

$$PR(\mathcal{A}_i) = \sum_{j=1}^{|\Omega_{in,i}|} \frac{1}{|\Omega_{out,j}|} PR(\mathcal{A}_j)\Big|_{\mathcal{A}_j \in \Omega_{in,i}} \tag{3.17}$$

In the computational method presented in Eq. (3.17), node $\mathcal{A}_j$ equally distributes its importance to all outgoing links ($|\Omega_{out,j}|$), resulting in a transmission coefficient of $1/|\Omega_{out,j}|$ for each receiver node $\mathcal{A}_i$. Notably, Eq. (3.17) exhibits a recursive structure, which can be addressed through two primary approaches. The first involves reformulating it into a homogeneous linear system of N_n equations via eigenvalue decomposition, though this method becomes computationally prohibitive for large-scale networks. The second approach leverages stochastic modeling by simulating user navigation as a Markov chain, where the steady-state visitation frequency of nodes is normalized to approximate $PR(\mathcal{A}_i)$.

The original PageRank algorithm suffers from limitations beyond convergence and computational complexity: it neglects edge importance. Connections are treated as homogeneous, where contributions depend solely on the source node's importance rather than the edge's intrinsic properties. In other words, nodes distribute importance uniformly to all members of $\Omega_{out,j}$, disregarding connection strength. However, edge heterogeneity is fundamental to many graph structures. For instance, in the simplified truss-derived graph discussed in Section 3.3.3, connection strength reflects critical mechanical properties such as member stiffness, material, and cross-sectional characteristics.

To address this, this section proposes an enhanced algorithm incorporating edge strength $\mathcal{V}_{ij}$, where node $\mathcal{A}_j$ distributes its importance proportionally to $\mathcal{V}_{ij}$. For undirected structural graphs, we define $\Omega_{out,j} = \Omega_{in,j} = \Omega_j$, leading to the modified formulation in Eq. (3.18). This adaptation preserves the equivalence between $PR(\mathcal{A}_i)$ calculation and homogeneous linear system solutions, allowing

existing PageRank solvers to remain applicable. Post-calculation, nodes are ranked by $PR(\mathcal{A}_i)$ values, enabling efficient selection of top-ranked nodes as the candidate set Ω_n based on computational constraints.

$$PR(\mathcal{A}_i) = \sum_{j=1}^{|\Omega_i|} \frac{\mathcal{V}_{ij}}{\sum_{k=1}^{|\Omega_j|} \mathcal{V}_{ik}} PR(\mathcal{A}_j) \Bigg|_{\mathcal{A}_j \in \Omega_i} \tag{3.18}$$

3.3.3 Application: Accelerometer placement of Raffles Cloud Corridor

The Raffles Cloud Corridor, also known as the G60 Sci-Tech Innovation Cloud Corridor, is a mixed-use commercial complex located in Shanghai Songjiang, integrating office spaces, hotels, dining, entertainment, and lifestyle amenities. The complex comprises 23 high-rise buildings arranged in two staggered rows. A monumental aluminum alloy canopy spans the rooftops of all structures, featuring a unidirectional length exceeding 700 m, and a coverage area surpassing $80000\ m^2$, as illustrated in Fig. 3.4(a).

This free-form grid shell canopy exhibits an undulating wave-like profile. Supported by tree-shaped columns atop 11 office buildings, its geometry alternates between peaks (above support points) and valleys (between buildings). The system incorporates over ten thousand aluminum alloy members, predominantly with cross-sectional dimensions of 250 mm (width) × 550 mm (height).

This large-span canopy exhibits complex structural characteristics, challenges in comprehensive maintenance, and non-uniform support stiffness, making it a critical candidate for SHM systems. Unlike super-tall buildings that can be simplified as “gourd-chain” models with strengthened floors or long-span bridges with uniformly varying cross-sections where weak zones typically occur at mid-span, such spatial structures, with featuring numerous components and complex load-deformation interactions, defy intuitive sensor placement based solely on engineering expertise. This scenario justifies the proposed two-stage FIA criterion integrated with PageRank and HEOA algorithms to address its OSP problem.

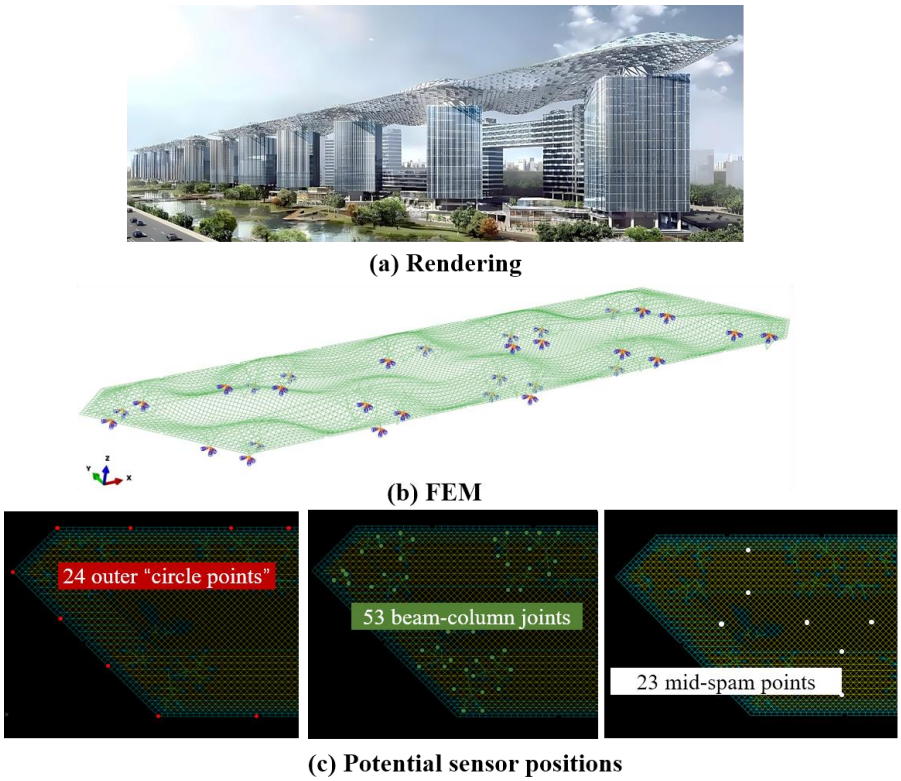


Fig. 3.4 Roof of Raffles Cloud Corridor

Fig. 3.4(b) displays the FEM of the Raffles Cloud Corridor roof constructed in ABAQUS, with the tree-shaped columns constrained by fixed supports. All elements employ two-node B33 beam elements, resulting in a total of 21,094 elements and 9,040 nodes. For acceleration monitoring, assuming the use of triaxial accelerometers (with 10 sensors ignoring rotational degrees of freedom, yielding $m = 30$), the FIA criterion posits that the most representative measurement points should optimally reconstruct the complete acceleration field. However, given the excessive number of FEM nodes, preliminary filtering of redundant or insignificant nodes is necessary. Here, multiple structural engineers conservatively selected 100 candidate measurement locations ($n = 300$) deemed structurally significant, including 24 outer truss nodes spaced every 15 to 17 members, 53 beam-column joints, and 23 mid-span nodes.

White-noise seismic excitation was applied to simulate free vibration (commonly used to infer modal properties and diagnose damage from vibration responses). The combinatorial HEOA algorithm iteratively selects Ω_m to train the FFM-RNN for $30 \rightarrow 270$ sequence inference tasks, evaluating the $FIA(\Omega_m)$ for different measurement

configurations, with the lowest score determining the final solution. To reduce subjectivity in manual selection, subsequent analyses will explore replacing the 100-point preselection process with the PageRank algorithm.

To match sensor sampling rates, triaxial white-noise seismic excitation with a 0.01s time step and 100s duration was applied simultaneously to all 11 supports. The FEM output provides triaxial accelerations for all nodes, recorded as $\boldsymbol{Z} \in \mathbb{R}^{300\times10000}$, normalized according to Eq. (3.19). During each iteration, the newly selected Ω_m partitions the dataset into $\boldsymbol{X} \in \mathbb{R}^{30\times10000}$ and $\boldsymbol{Y} \in \mathbb{R}^{270\times10000}$, split in an $8{:}2$ ratio to form training and testing sets. The $FIA(\Omega_m)$ is evaluated exclusively on the testing set samples.

$$Z_k(t) \leftarrow \frac{Z_k(t) - mean(\boldsymbol{Z})}{std(\boldsymbol{Z})} \tag{3.19}$$

where, $mean(\cdot)$ and $std(\cdot)$ denote the expectation and standard deviation operations, respectively. The normalization employs the global mean and standard deviation of the entire dataset to preserve inter-channel numerical relationships. Unless otherwise specified, all subsequent data normalization follows this method.

3.3.4 OSP problem-solving results and discussion

This section presents a comprehensive evaluation of the proposed FIA criterion and combinatorial HEOA algorithm for solving the OSP problem in the Raffles Cloud Corridor structure. The analysis begins with a comparative study between the Ω_m configurations obtained through manual selection of Ω_n as well as the comparison between HEOA and alternative optimization algorithms. The performance of the FFM-RNN is then examined in terms of its accuracy in inferencing the structural acceleration field and computational efficiency during training, with direct comparisons made against conventional approaches. The potential for replacing manual node selection with the PageRank-generated Ω_n is subsequently investigated.

The implementation builds upon the OSP framework established in Section 3.3.3, with specific parameter configurations for each algorithmic component. The FFM-RNN architecture incorporates a single hidden layer with $N_h = 200$ neurons and utilizes a

two-layer MLP for input-state mapping, where the first layer contains 100 neurons. The loss function employs distinct β_p values of 10 for non-peak points and 100 for peak points. Training is conducted via stochastic gradient descent (SGD) with a learning rate of 0.005 over 50 epochs. The hybrid HEOA algorithm operates with a population size of 10 individuals and a maximum of 1000 iterations, utilizing the proportional parameters detailed in Table 3.1. For the PageRank implementation, the graph construction process applies weighting coefficients of $\beta_{\mathbb{f}} = 0.8$ for flexibility and $\beta_{\mathbb{c}} = \beta_{\mathbb{m}} = 0.1$ for mass and damping matrices following normalization. All computational procedures mentioned in this thesis are conducted on a standardized hardware comprising an NVIDIA RTX 3060 GPU and Intel i7-12700F CPU to ensure consistent evaluation conditions.

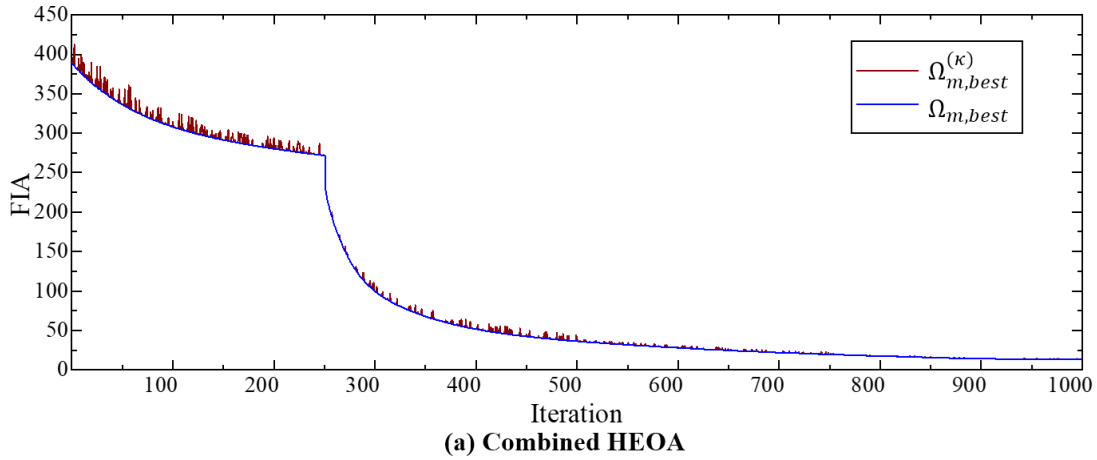
Table 3.1 Hyperparameters of combinatorial HEOA

$\boldsymbol{p_1}$	$\boldsymbol{p_2}$	$\boldsymbol{p_3}$	$\boldsymbol{p_4}$	$\boldsymbol{p_r}$	$\boldsymbol{p_{le}}$	$\boldsymbol{p_e}$	$\boldsymbol{p_f}$	$\boldsymbol{p_{lo}}$
0.3	0.3	0.4	0.1	0.1	0.2	0.2	0.3	0.3

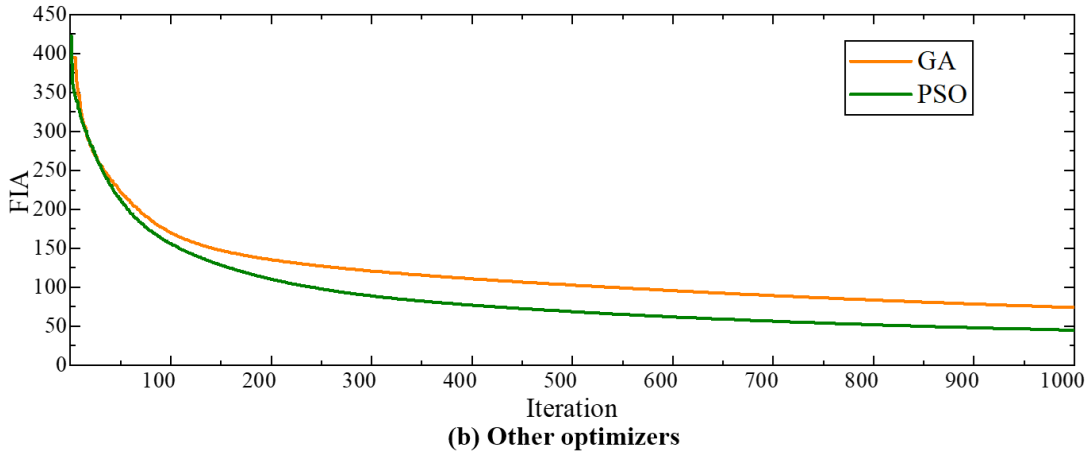
3.3.4.1 Combinatorial HEOA searching results under FIA criterion

Fig. 3.5(a) illustrates the variation process of the contemporary and historical best fitness values ($FIA\left(\Omega_{m,best}^{(\kappa)}\right)$ and $FIA\left(\Omega_{m,best}\right)$) during the search for measurement point configurations using the combinatorial HEOA algorithm. The $FIA\left(\Omega_{m,best}^{(\kappa)}\right)$ curve exhibits an overall downward trend with significant fluctuations, particularly during the initial iterations (human exploration phase). During this phase, the division of labor among population individuals remains undefined, and an increased proportion of "unselected measurement points" is introduced during individual updates to ensure comprehensive exploration. This results in substantial dispersion in the quality of selected $\Omega_m^{(\kappa)}$, with $FIA\left(\Omega_{m,best}\right)$ demonstrating a stepwise decline. However, this extensive exploration phase enables the algorithm to probe diverse search regions, subsequently facilitating rapid $FIA\left(\Omega_{m,best}\right)$ reduction during the early human development phase.

Verification reveals that over 70% of the rapid optimization metric decline in this early development phase is achieved by leader individuals, approximately 20% by explorers, and merely 5% by followers. Notably, only three $FIA(\Omega_{m,best})$ reductions are attributable to underperformers throughout the human development phase. These observations validate the effectiveness of HEOA's strategy in implementing differentiated search mechanisms based on population status. The fitness values essentially converge when κ reaches approximately 75% of the maximum iteration limit, with the minimum $FIA(\Omega_{m,best})$ occurring in the 998th generation. However, the optimization gains during the final 100 iterations prove negligible.



(a) Combined HEOA



(b) Other optimizers

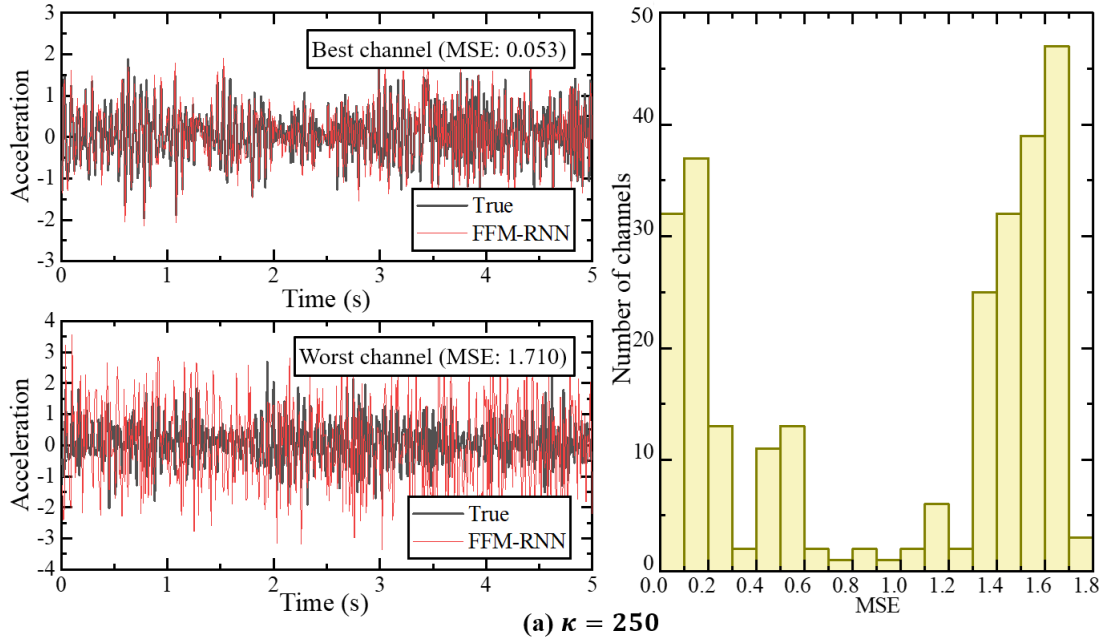
Fig. 3.5 FIA-iteration curves

To demonstrate the advantages of the combinatorial HEOA, Fig. 3.5(b) compares its performance with genetic algorithm (GA) and particle swarm optimization (PSO) using equivalent population sizes for OSP solving. GA simulates biological evolution

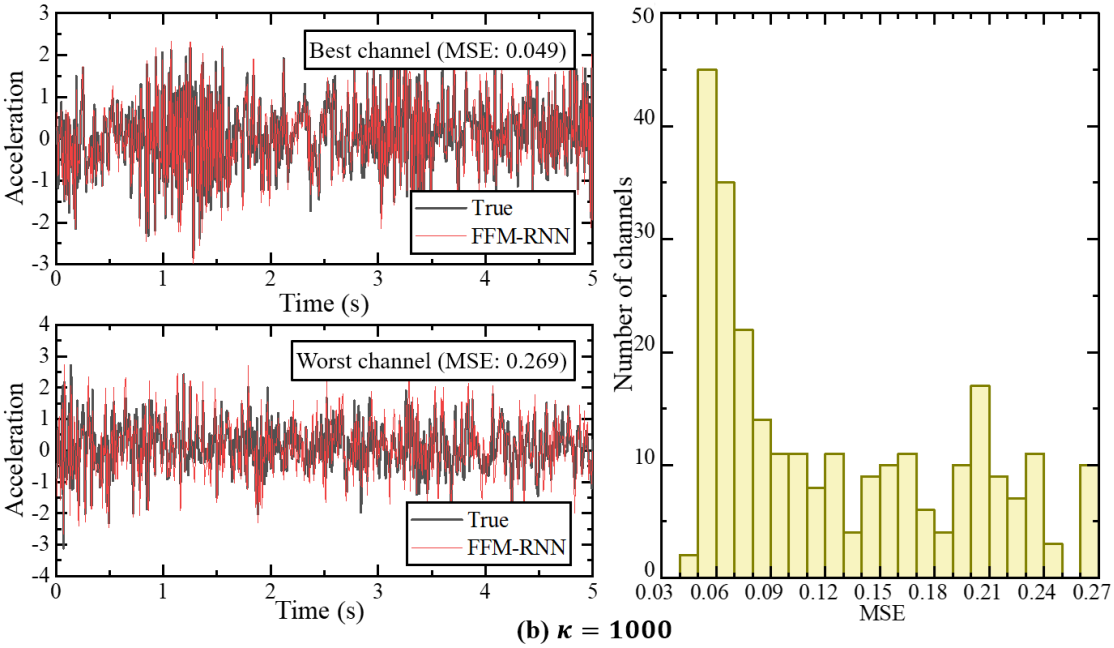
mechanisms through binary chromosome encoding, where solutions are represented as bit strings. Fitness evaluation guides parent selection, followed by crossover rate of 0.6 and mutation rate of 0.01 operations to generate offspring while maintaining population diversity. PSO models collective intelligence through particle swarms, where each particle updates its velocity by balancing inertia (linearly decreasing from 0.9 to 0.4), individual cognition of 2.0, and social influence of 2.0. Experimental results indicate that both GA and PSO obtain locally optimal solutions with higher $FIA(\Omega_{m,best})$ values compared to the composite HEOA, confirming HEOA's superior field inference capability. Furthermore, GA demonstrates inferior search efficiency relative to PSO and HEOA, with its $FIA(\Omega_{m,best})$ showing notable reductions until the 950th iteration, while the latter two algorithms achieve convergence within 700-800 iterations.

Fig. 3.6(a) presents the acceleration field inference results of the initial population ($\kappa = 250$) during the human development phase in the composite HEOA search process, specifically for the optimal solution $\Omega_{m,best}^{(250)}$. The visualization compares the normalized error distributions of the testing sets for the 270 target channels, highlighting the highest- and lowest-error channels, with the overall inference accuracy distribution of all channels displayed on the right. The optimized inference results from the final iteration $\Omega_{m,best}$ are shown in Fig. 3.6(b), where only the first 500 steps of the testing set are visualized for clarity. Significant differences emerge between these stages: during the early search phase, the measurement points exhibit substantial spatial randomness, and $\Omega_m^{(\kappa)}$ lacks essential information required for reconstructing the entire acceleration field. Consequently, most output channels demonstrate MSE exceeding 1.0, indicating the inadequacy of the initial measurement configuration. Notably, however, certain channels achieve exceptionally low MSE values. The observed high precision arises from the proximity of these channels to the measurement points selected in $\Omega_m^{(\kappa)}$, such as adjacent peripheral sensors or branched nodes on the same tree-shaped column. The inherent spatial correlation between neighboring nodes ensures that the acceleration signals for

these positions can be effectively reconstructed using the information contained in $\Omega_m^{(\kappa)}$.



(a) $\kappa = 250$



(b) $\kappa = 1000$

Fig. 3.6 FFM-RNN inferencing results

Upon convergence, the optimized sensor configuration $\Omega_m^{(\kappa)}$ achieves peak comprehensive accuracy in reconstructing the full acceleration field, attaining approximately 83% inference precision even at the most challenging reconstruction locations. This demonstrates that $\Omega_m^{(\kappa)}$ identifies an optimal sensor arrangement under

constrained computational resources and optimization time, effectively representing the acceleration field characteristics. Notably, nearly 50% of the target positions exhibit MSE below 0.1 (equivalent to over 90% accuracy). The error distribution across all position's forms overlapping Gaussian profiles, indicating strong spatial correlations between inferred locations and measurement points, though with variable predictive capabilities among different regions.

To evaluate the FFM-RNN model's performance using $\Omega_{m,best}$, three baseline architectures with equivalent hidden layers and neuron number were compared: standard RNN, LSTM (as detailed in Section 5.2.1), and bidirectional LSTM, as listed in Table 3.2. The RNN is chosen because it is the original form of FFM-RNN; while the other two used to be the mainstream architecture for Seq2Seq tasks. The FFM-RNN outperforms all counterparts by achieving superior gradient alignment and peak matching accuracy, as detailed in the comparative table below. Notably, the FFM-RNN features a lightweight network architecture, requiring only approximately 85 s for complete population training per iteration (a parallel training implementation could further accelerate this process), enabling full OSP problem resolution within one day. In contrast, while the bidirectional LSTM achieves comparable accuracy, its training duration extends to nearly sixfold that of FFM-RNN (approximately one week per solution), rendering it computationally prohibitive when accounting for hyperparameter tuning and error correction cycles.

Fig. 3.7(a) displays the sensor locations determined by the FIA criterion, demonstrating relatively uniform spatial distribution. This indicates that optimizing for global vibration field inference accuracy inherently prevents sensor clustering issues. To further validate the FIA criterion's effectiveness, we apply the MAC (He et al., 2015) to solve the Rafael Cloud Corridor OSP problem. MAC quantifies modal shape similarity by computing the correlation coefficient (cosine of the angle) between two modal vectors, as expressed in Eq. (3.20).

$$MAC_{ij} = \frac{\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j}{\sqrt{\left(\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j\right)\left(\boldsymbol{\phi}_j^T \boldsymbol{\phi}_i\right)}} \tag{3.20}$$

Table 3.2 Inference accuracy comparisons among NN models

Models	Average MSE	Highest MSE	Training time
FFM-RNN	0.1265	0.2690	0.17 s/epoch
RNN	0.2857	0.7645	0.15 s/epoch
LSTM	0.1858	0.3251	0.54 s/epoch
Bi-LSTM	0.1569	0.2951	0.95 s/epoch

where, $\boldsymbol{\phi_i} \in \mathbb{R}^n$ represents the ith modal vector of the structure. For an intact structure, the MAC matrix becomes an identity matrix due to modal orthogonality. However, when partial structural information is unavailable (i.e., no sensors are deployed at specific locations), the reduced modal vectors $\boldsymbol{\phi_i} \in \mathbb{R}^m$, formed by sensor-equipped degrees of freedom, lose their orthogonality. This results in non-zero off-diagonal elements in the MAC matrix. Larger magnitudes of these off-diagonal elements indicate higher similarity between modal vectors (smaller angular separation), implying redundant monitoring at corresponding positions. MAC-based sensor placement methods iteratively expand a manually defined subset of measurement points (with initial size smaller than m) by sequentially adding candidate positions that minimize the off-diagonal elements of the updated MAC matrix. As shown in Fig. 3.7(b), the MAC-optimized sensor configuration predominantly concentrates on peripheral roof positions. For a space structure anchored internally to columns, these peripheral locations exhibit larger amplitudes across mode shapes, and their exclusion would exacerbate orthogonality degradation in reduced modal vectors. However, from a mechanical perspective, column-to-shell connection points which are analogous to cantilever beam roots, experience stress concentration and higher vulnerability to damage. Additionally, uneven displacements between supporting structures render mid-span positions between grid modules as critical weak points. While such locations demand prioritized monitoring, modal-based OSP methods exhibit inherent limitations due to their bias toward high-amplitude regions. In contrast, the FIA-

optimized configuration $\Omega_{m,best}$ demonstrates balanced coverage across structurally diverse locations. Notably, the maximum off-diagonal element in the MAC matrix of $\Omega_{m,best}$ is 0.496, only 15% higher than that of the MAC-based configuration (0.413). This confirms that the FIA criterion inherently addresses modal independence through its global reconstruction accuracy framework while maintaining comparable modal distinguishability.

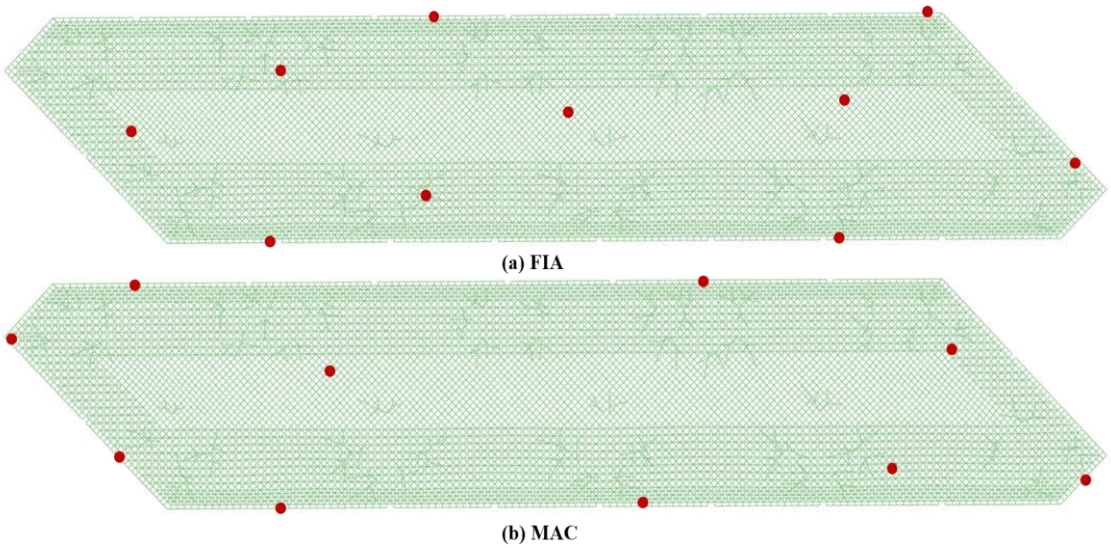


Fig. 3.7 Optimized sensor placement

3.3.4.2 Effectiveness of PageRank

The construction process of Ω_n illustrated in Fig. 3.4(c) remains challenging to fully rationalize, as even experienced structural engineers struggle to definitively select the 100 most critical degrees of freedom from tens of thousands and provide rigorous justification. The author participated in the selection of Ω_n, yet different spatiotemporal conditions led to varying selections, making the actual performance hierarchy of Ω_n inherently ambiguous. Fig. 3.8 displays the top 100 and 150 positions ranked by the PageRank algorithm.

It is observed that nodes within the annular region (characterized by dense connections in the internal triangular grids of the grid structure and connections between the grid and column top bifurcations) achieve higher scores. Conversely, peripheral nodes with larger vibration amplitudes, located at the network edges, rarely appear in the top 100 sequence.

When expanding the selection to the top 150, a limited number of mid-span nodes and central column-top nodes emerge in the candidate set. Assuming that nodes within 3rd order neighborhoods of PageRank-selected nodes represent redundant selections, the top 150 sequence covers approximately 60% of manually selected nodes. Fig. 3.8 further compares sensor placement schemes derived from PageRank's top 100 and 150 preselected nodes, with evaluation metrics summarized in Table 3.3. Results indicate that the PageRank top 100 set, overly concentrated in peripheral regions, exhibits poor inference accuracy due to incomplete coverage of acceleration field information. However, its MAC consistency benefits from the larger modal amplitudes at peripheral nodes. The PageRank top 150 set achieves inference accuracy and maximum off-diagonal element of the MAC matrix comparable to manual selection, demonstrating potential for automated substitution.

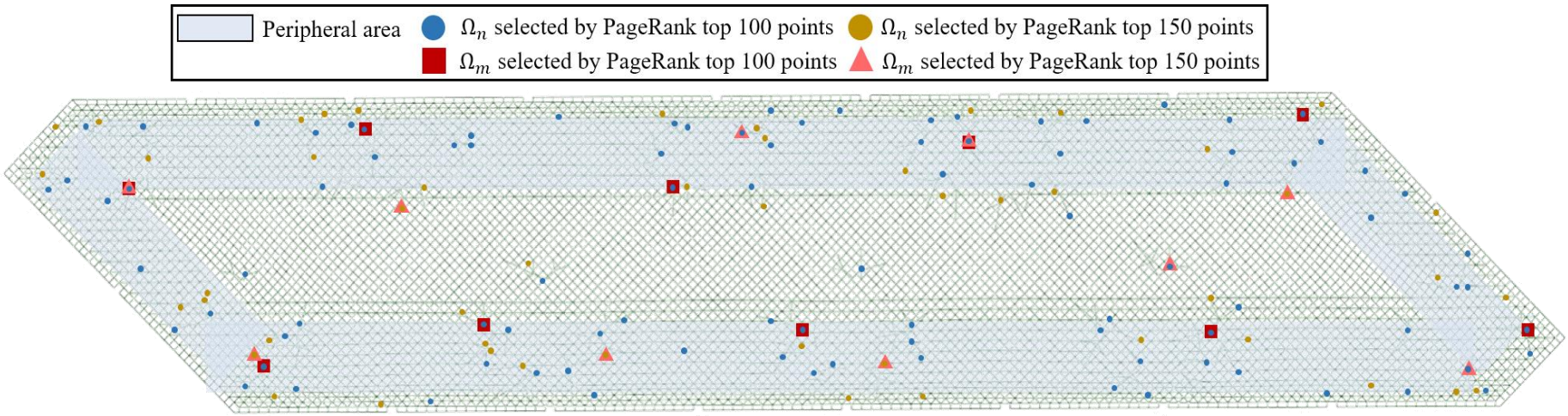


Fig. 3.8 PageRank ranking results

Table 3.3 OSP solutions based on different initial candidate set

Initial candidate set	Average MSE	Highest MSE	Maximum off-diagonal element of MAC
Manual	0.1265	0.2690	0.496
PageRank top 100	0.1856	0.4723	0.458
PageRank top 150	0.1620	0.3317	0.501

3.3.4.3 Sensor number determination

Another important parameter to be determined in the OSP problem is the number of sensors. Although this parameter is primarily determined by factors such as budget in practical engineering, technical recommendations remain important. Under the FIA framework proposed in this study, the number of sensors can be determined using Fig. 3.9. The figure shows the lowest FIA scores achievable by systems with different numbers of sensors after the search process. Based on the aforementioned dataset, we added numerical experiments for one, five, twenty, and fifty sensors in addition to the initial ten sensors. Clearly, the greater the number of sensors, the lower the achievable minimum FIA score. However, this trend gradually slows down as the number of measurement points reaches 20% or even 50% of the full set, indicating that the correlation between response data from different locations reduces the need for an excessive number of sensors. Conversely, when the number of sensors is reduced from ten to five, a significant increase in the FIA score is observed, suggesting that this reduction leads to a permanent loss of substantial effective information (which cannot be captured from the remaining five measurement points). This reveals that a reasonable number of sensors likely lies between five and ten. To determine the optimal number more precisely, numerical experiments should be conducted for configurations with six to nine sensors, provided the computational time is acceptable.

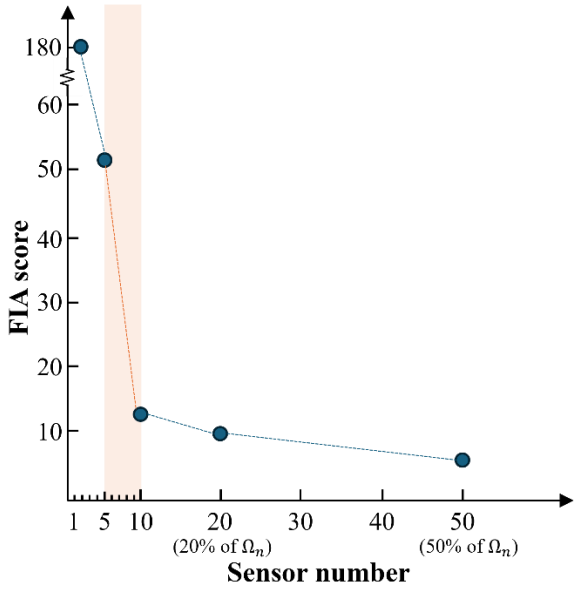


Fig. 3.9 FIAs influenced by different sensor numbers

3.4 Conclusions

This chapter proposes a NN-based solution integrating temporal data and graph data for sensor placement in SHM systems. First, the FIA criterion is introduced to evaluate the reconstruction accuracy of monitored variable fields, accompanied by a lightweight inference network, FFM-RNN, specifically designed for this criterion. Subsequently, the HEOA is enhanced to address combinatorial optimization challenges. Finally, to resolve issues of poor interpretability and weak quality assurance in initial candidate set generation, a method is developed to transform structural feature matrices into graph data, enabling rapid automated generation of preselected node sets with arbitrary sizes via PageRank ranking. Key conclusions are summarized as follows:

(1) The Ω_m derived from FIA-guided optimization achieves over 90% accuracy in inferring critical acceleration signals across the structure, aligning with the objective of deploying sensors to sample vibration states. This high accuracy stems from the inference performance of FFM-RNN, which incorporates gradient and peak penalty terms in its loss function. While retaining the lightweight architecture and rapid training advantages of conventional RNNs, FFM-RNN marginally surpasses bidirectional LSTM in accuracy. Additionally,

the proposed HEOA algorithm demonstrates faster convergence rates and identifies higher-quality local optima compared to PSO and GA. These local optima inherently satisfy modal independence criteria emphasized by traditional modal-based OSP methods like MAC, indicating that the FIA criterion inherently incorporates non-maximum suppression functionality to prevent excessive clustering of measurement points.

(2) The graph abstraction of structural features combined with PageRank-based node ranking provides an interpretable approach for automated generation of preselected node sets Ω_n. However, this method tends to prioritize areas with dense structural members. In homogeneous regions (e.g., grid zones with identical material and dimensions), it may produce batch identical scores for measurement points. This limitation can be mitigated by moderately increasing the size of Ω_n. For instance, in the Raphael Cloud Corridor OSP case study, expanding Ω_n from 100 to 150 elements achieves approximately 60% coverage of manually selected nodes while maintaining comparable inference accuracy.

It is worthy to be discussed that, as a prior problem, the placement determined by simulated data and FIA will be inevitably influenced by measuring noise. However, the conventional OSP methods (e.g., MAC, EI, MKE) are additionally affected by the model error, which makes them even harder to be applied to engineering projects. For the inaccurate boundary condition assumption, to avoid OSP with low quality, scaled or even full-scale tests are needed to determine the real stiffness curves of the constraints and joints. Moreover, in FIA, there is no requirement for candidate set Ω_n to include all DoFs. Instead, a preselection process is recommended both manually and automatically (using the method introduced in Section 3.3).

DATA DENOISING BASED ON TIME SERIES-TARGETED NEURAL NETWORKS

4.1 Introduction

The deployed SHM system will operate continuously throughout the structure's service life, transmitting sensor-acquired data persistently. The reliability of downstream tasks (e.g., damage diagnosis, condition assessment) critically depends on the completeness and accuracy of the collected datasets, while computational constraints additionally necessitate minimizing data redundancy. Given the system's prolonged operation, diverse data types, and complex transmission chains, raw data inevitably contains low-value or invalid components, which can be categorized as:

(1) Conventional noise: low-magnitude, unbiased fluctuations around true values caused by sensor hardware limitations or data transmission and storage devices, typically modeled as white noise.

(2) Missing data: partial or complete signal loss due to sensor failure, temporary power outages, or transmission interruptions.

(3) Outliers: anomalous signal values (e.g., spikes, baseline drift, or sustained deviations) induced by sensor malfunctions, external interference, or unstable installations.

(4) Redundant information: repetitive data patterns arising from environmental periodicity or structural vibration cycles.

Among these, the first three categories represent non-structural artifacts that interfere with condition evaluation, while redundancy primarily burdens computational resources. This chapter focuses on NN-based solutions for the first three issues, including: a novel filtering algorithm combining auto-regressive integrated moving average model with exogenous variables (ARIMAX) with a modified Kalman filter (KF) (Khodarahmi & Maihami 2023); a data imputation algorithm using bidirectional ESN for missing data

reconstruction; and an outlier detection and recovery method integrating frequency-domain RNN and local outlier factor (LOF) algorithms. All proposed methods will be validated on practical SHM datasets and benchmarked against conventional approaches.

4.2 Data filtering

The raw signals collected from SHM systems inevitably contain noise. Previous studies typically employ low-pass filters with predefined cutoff frequencies for noise reduction, which may introduce biased data extraction. Given that conventional noise is often assumed to be white noise encompassing diverse frequency components, manually specifying a single cutoff frequency risk filtering out valuable high-frequency components while retaining low-frequency noise. This section proposes an improved Kalman filter (IKF) as a filtering algorithm, capable of dynamically adjusting its parameters to adapt to varying noise levels. The IKF incorporates a data fusion concept to derive optimal signal estimates by combining predictions from a state transition model with sensor measurements. Here, the state transition model is constructed using the ARIMAX framework, which requires input signals to be stationary (statistically invariant in mean and covariance). Additionally, the feasibility of employing NNs as alternative state transition models is explored.

4.2.1 Data stabilization

Let the raw signal be $\boldsymbol{Z_{ori}}(t) \in \mathbb{R}^{N_z}$, where N_z denotes the total number of signal acquisition channels, and $t = 1, 2, \dots, T$ represents discrete sampling time points. Given the potential non-stationarity of the raw signal, stabilization must be performed prior to noise reduction. If the signal exhibits an exponential trend, a sign-preserving logarithmic stabilization should first be applied, as shown in Eq. (4.1).

$$\boldsymbol{Z_{log}}(t) = \begin{cases} -log\left(\boldsymbol{Z_{ori}}(t)\right), & \boldsymbol{Z_{ori}}(t) > 0 \\ log\left(-\boldsymbol{Z_{ori}}(t)\right), & \boldsymbol{Z_{ori}}(t) \leq 0 \end{cases} \tag{4.1}$$

For non-logarithmic trends, differencing is applied to achieve stationarity. Let the 1st order difference be denoted as $\nabla \boldsymbol{Z_{log}}(t) = \boldsymbol{Z_{log}}(t) - \boldsymbol{Z_{log}}(t-1)$. If $\boldsymbol{Z_{log}}(t)$ becomes

stationary after dth order differencing, the final stationary signal is expressed as:

$$\boldsymbol{Z}(t) = \nabla^d \boldsymbol{Z}_{\boldsymbol{log}}(t) \tag{4.2}$$

All subsequent denoising procedures operate on the stationary series $\boldsymbol{Z}(t)$. Upon completing the denoising process, the original signal must be reconstructed through the inverse transformations of Eq. (4.1) and (4.2).

4.2.2 IKF based on ARIMAX

In the standard KF algorithm (Revach 2022), the objective is to obtain the optimal posterior estimate $\widehat{\boldsymbol{Z}}(t)$ of the true signal value $\boldsymbol{Z}(t)$ at step t. The measured value $\boldsymbol{Z}_{\boldsymbol{me}}(t)$ is assumed to contain measurement noise that requires filtering. The algorithm operates under the following assumptions:

$$\boldsymbol{Z}_{\boldsymbol{me}}(t) = \boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{Z}(t) + \boldsymbol{\nu}(t) \tag{4.3}$$

where, $\boldsymbol{A}_{\boldsymbol{m}}$ represents the measurement matrix: when all variables in $\boldsymbol{Z}(t)$ are measured, $\boldsymbol{A}_{\boldsymbol{m}}$ is an identity matrix; when measurements are incomplete, the rows corresponding to unobserved variables are set as zero vectors. The measurement error vector $\boldsymbol{\nu}(t)$ is typically assumed to be unbiased with zero mean and a covariance matrix $\boldsymbol{Q}_{\boldsymbol{\nu}} \in \mathbb{R}^{N_z \times N_z}$. The estimation of the true signal value based on measurements is given by:

$$\widehat{\boldsymbol{Z}}^{+}(t) = \boldsymbol{A}_{\boldsymbol{m}}{}^{-1}\boldsymbol{Z}_{\boldsymbol{me}}(t) \tag{4.4}$$

The KF algorithm incorporates additional information beyond measurements to estimate the true signal value. Specifically, its originators postulated that historical values of the target sequence contain inherent evolutionary patterns. This allows constructing a model to infer the current signal value based on the signal's previous state and the current external inputs This inference also includes model error, expressed as:

$$\boldsymbol{Z}(t) = \boldsymbol{A}_{\boldsymbol{s}}[\boldsymbol{Z}(t-1); \boldsymbol{Z}_{\boldsymbol{ex}}(t)] + \boldsymbol{\omega}(t) \tag{4.5}$$

where, $\boldsymbol{A}_{\boldsymbol{s}}$ represents the state transition matrix, characterizing the evolutionary pattern of the signal between consecutive time steps, $\boldsymbol{Z}_{\boldsymbol{ex}}(t)$ denotes external inputs, typically

environmental information related to the $\boldsymbol{Z}(t)$ sequence. The model error $\boldsymbol{\omega}(t)$, arising from modeling assumptions and simplifications, is similarly assumed to have zero mean and a covariance matrix $\boldsymbol{Q_\omega} \in \mathbb{R}^{N_z \times N_z}$. Consequently, the state transition model yields an estimate of the true signal value based on: the model's previous output and the current environmental information, as:

$$\widehat{\boldsymbol{Z}}^{-}(t) = \boldsymbol{A_s}\left[\widehat{\boldsymbol{Z}}(t-1); \boldsymbol{Z_{ex}}(t)\right] \tag{4.6}$$

Both estimates of $\boldsymbol{Z}(t)$, denoted as $\widehat{\boldsymbol{Z}}^{+}(t)$ (measurement-based) and $\widehat{\boldsymbol{Z}}^{-}(t)$ (model-based), are inherently inaccurate, with their uncertainties governed by $\boldsymbol{Q_v}$ and $\boldsymbol{Q_\omega}$, respectively. The KF adopts a data fusion framework, treating $\widehat{\boldsymbol{Z}}^{+}(t)$ and $\widehat{\boldsymbol{Z}}^{-}(t)$ as samples from distinct distributions. It introduces a time-varying gain matrix $\boldsymbol{G}(t) \in \mathbb{R}^{N_z \times N_z}$ to optimally combine the two estimates, yielding the posterior estimate $\boldsymbol{Z}(t)$ as specified in Eq. (4.7).

$$\widehat{\boldsymbol{Z}}(t) = \widehat{\boldsymbol{Z}}^{-}(t) + \boldsymbol{G}(t)\left(\widehat{\boldsymbol{Z}}^{+}(t) - \widehat{\boldsymbol{Z}}^{-}(t)\right) \tag{4.7}$$

Let the Kalman gain be $\boldsymbol{G_K}(t) = \boldsymbol{G}(t){\boldsymbol{A_m}}^{-1}$, thus the above equation can be written as:

$$\widehat{\boldsymbol{Z}}(t) = \widehat{\boldsymbol{Z}}^{-}(t) + \boldsymbol{G_K}(t)\left(\boldsymbol{Z_{me}}(t) - \boldsymbol{A_m}\widehat{\boldsymbol{Z}}^{-}(t)\right) \tag{4.8}$$

The output $\widehat{\boldsymbol{Z}}(t)$ represents the final filtered result of the KF algorithm. Defining the estimation error as $\boldsymbol{e_z}(t) = \widehat{\boldsymbol{Z}}(t) - \boldsymbol{Z}(t)$, it follows that $\boldsymbol{e_z}(t) \sim \mathcal{N}[\boldsymbol{0}, \boldsymbol{\Sigma}(t)]$, where $\boldsymbol{\Sigma}(t) \in \mathbb{R}^{N_z \times N_z}$ is the error covariance matrix. The core objective of applying KF is to determine the optimal Kalman gain $\boldsymbol{G_K}(t)$ that minimizes the discrepancy between $\widehat{\boldsymbol{Z}}(t)$ and $\boldsymbol{Z}(t)$.

The diagonal elements $\sigma_{e_i}^2(t)$ $(i = 1, 2, \ldots, N_z)$ of $\boldsymbol{\Sigma}(t)$ represent the error variances for each channel of $\boldsymbol{Z}(t)$ after filtering. Their summation equals the trace of $\boldsymbol{\Sigma}(t)$, denoted as $tr\left(\boldsymbol{\Sigma}(t)\right)$, which quantifies the overall estimation error when using $\widehat{\boldsymbol{Z}}(t)$ to approximate $\boldsymbol{Z}(t)$. The optimal $\boldsymbol{G_K}(t)$ should minimize $tr\left(\boldsymbol{\Sigma}(t)\right)$, i.e.,

$$\frac{\partial tr\left(\boldsymbol{\Sigma}(t)\right)}{\partial \boldsymbol{G_K}(t)} = 0 \tag{4.9}$$

Thus, the expression of $\boldsymbol{G_K}(t)$ can be derived as:

$$\boldsymbol{G_K}(t) = \frac{\boldsymbol{\Sigma}^{-}(t)\boldsymbol{A_m}}{\boldsymbol{A_m}\boldsymbol{\Sigma}^{-}(t){\boldsymbol{A_m}}^{T} + \boldsymbol{Q_v}} \tag{4.10}$$

where, $\boldsymbol{\Sigma}^{-}(t) \in \mathbb{R}^{N_z \times N_z}$ denotes the error covariance matrix between the model-based estimate $\widehat{\boldsymbol{Z}}^{-}(t)$ and the true signal $\boldsymbol{Z}(t)$, which can be computed according to Eq. (4.11).

$$\boldsymbol{\Sigma}^{-}(t) = \boldsymbol{A_s}\boldsymbol{\Sigma}(t-1){\boldsymbol{A_s}}^{T} + \boldsymbol{Q_\omega} \tag{4.11}$$

In summary, given the current measurement, the KF output from the previous time step, and the corresponding error covariance matrices, the KF output at the current time step can be obtained through Eqs. (4.6), (4.8), (4.10), and (4.11). Subsequently, the covariance matrix $\boldsymbol{\Sigma}(t)$ must be updated via Eq. (4.12) to initiate the iterative process for the next time step.

$$\boldsymbol{\Sigma}(t) = (\boldsymbol{I} - \boldsymbol{G_K}(t)\boldsymbol{A_m}) \cdot \boldsymbol{\Sigma}^{-}(t) \tag{4.12}$$

The iterative filtering framework described above lacks a defined method for establishing the state transition model. Here, this study employs ARIMAX (Peter & Silvia 2012) for signal modeling. This method extends the traditional auto-regressive integrated moving average model (ARIMA) framework by incorporating external variables, aligning with the format of Eq. (4.5). ARIMAX posits that the current signal value can be decomposed into three components: a linear combination of the signal's past p steps (autoregressive component), accumulated errors from the past q steps (moving average component), and the current external inputs and error terms, as expressed by Eq. (4.13).

$$\boldsymbol{Z}(t) = \sum_{i=1}^{p} \boldsymbol{\varphi_{ar,i}} \odot \boldsymbol{Z}(t-i) + \sum_{j=1}^{q} \boldsymbol{\varphi_{ma,j}} \odot \boldsymbol{\omega}(t-j) + \boldsymbol{\omega}(t) + {\boldsymbol{\varphi_x}}^{T}\boldsymbol{Z_{ex}}(t) \tag{4.13}$$

where, $\boldsymbol{\varphi_{ar,i}}$, $\boldsymbol{\varphi_{ma,j}}$, and $\boldsymbol{\varphi_x}$ denote the combination coefficients for the autoregressive, moving average, and external factor terms, respectively, and $\odot$ represents element-wise multiplication. The orders p and q determine the model's complexity and fitting capability, typically selected based on the auto-correlation function (ACF) and partial auto-correlation function (PACF). The standard selection criteria are: the proper p makes the largest integer before $\boldsymbol{PACF}(p)$ falls within the 95% confidence interval; the

proper q makes the largest integer before $\boldsymbol{ACF}(q)$ falls within the 95% confidence interval (Peter & Silvia 2012). The ACF and PACF are computed as follows:

$$\boldsymbol{ACF}(q) = \frac{cov\big(\boldsymbol{Z}(t), \boldsymbol{Z}(t-q)\big)}{\sqrt{var\big(\boldsymbol{Z}(t)\big)var\big(\boldsymbol{Z}(t-q)\big)}} \tag{4.14}$$

$$\boldsymbol{PACF}(p) = \frac{cov(\boldsymbol{Z}(t), \boldsymbol{Z}(t-p)|\boldsymbol{Z}(t), \dots, \boldsymbol{Z}(t-p+1))}{\sqrt{\Big(var\big(\boldsymbol{Z}(t)\big)var\big(\boldsymbol{Z}(t-p)\big)\Big)|\boldsymbol{Z}(t), \dots, \boldsymbol{Z}(t-p+1)}} \tag{4.15}$$

where, $cov(\cdot)$ and $var(\cdot)$ denote the covariance and variance operations, respectively. However, for real-world signals, the ACF and PACF may exhibit oscillatory behavior, resulting in multiple valid combinations of p and q. In such cases, the Bayesian information criterion (BIC) (Chen et al., 2022) is introduced to identify the optimal combination, specifically, the pair (p, q) that minimizes the following expression among all qualified candidates:

$$BIC = (p + q + 1)\ln(T \times N_z) - 2\ln(\mathcal{L}) \tag{4.16}$$

where, $\mathcal{L}$ represents the maximum likelihood function. Once p and q are determined, the combination coefficients $\boldsymbol{\varphi_{ar,i}}$, $\boldsymbol{\varphi_{ma,j}}$, and $\boldsymbol{\varphi_x}$ can be identified using the least squares method based on existing signal segments.

By applying the ARIMAX model to the KF, a new state transition model is formed to accommodate multi-step backtracking scenarios, thereby upgrading the KF into an IKF. In the IKF, a sub-filter is individually established for each channel $Z_k(t)$, where signals from other channels are temporarily incorporated into $\boldsymbol{Z_{ex,k}}(t)$ for that sub-filter. The ARIMAX model coefficients for $Z_k(t)$ are then determined based on the predefined p and q. Furthermore, within this sub-KF, the p-step backtracked signal values $Z_k(t)$ to $Z_k(t-p+1)$ are treated as p distinct channels of the signal to be processed at any step t. Since the previous $(p-1)$ steps have already been resolved by step t, only the first variable $Z_k(t)$ remains a measured variable. Consequently, in the sub-KF corresponding to $Z_k(t)$, Eqs. (4.4) and (4.6) at step t should be rewritten as Eqs. (4.17) and (4.18). Based on the redefined $\boldsymbol{A_{m,k}}$ and $\boldsymbol{A_{s,k}}$ from Eqs. (3.17) and (3.18), the filtered signal

values can be computed stepwise using the aforementioned iterative method. Correspondingly, when calculating errors, the steps from $(t-1)$ to $(t-p+1)$ in the delay sequence directly utilize the already obtained filtered values, and their respective error vectors are calculated by Eqs. (4.19) and (4.20).

$$\begin{bmatrix} \hat{Z}_k^-(t) \\ \hat{Z}_k^-(t-1) \\ \vdots \hat{Z}_k^-(t-2) \\ \vdots \\ \hat{Z}_k^-(t-p+1) \end{bmatrix} = \begin{bmatrix} \varphi_{ar,1,k} & \varphi_{ar,2,k} & \cdots & \varphi_{ar,p,k} & \boldsymbol{\varphi_{x,k}}^T \\ 1 & 0 & \cdots & 0 & \mathbf{0} \\ 0 & 1 & \cdots & 0 & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{Z}_k(t-1) \\ \hat{Z}_k(t-2) \\ \vdots \\ \hat{Z}_k(t-p) \\ \boldsymbol{Z_{ex,k}}(t) \end{bmatrix} \tag{4.17}$$

$$\begin{bmatrix} \hat{Z}_k^+(t) \\ \hat{Z}_k^+(t-1) \\ \hat{Z}_k^+(t-2) \\ \vdots \\ \hat{Z}_k^+(t-p+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Z_k(t) \\ \hat{Z}_k(t-1) \\ \hat{Z}_k(t-2) \\ \vdots \\ \hat{Z}_k(t-p+1) \end{bmatrix} \tag{4.18}$$

$$\boldsymbol{\omega_k}(t) = \begin{bmatrix} \omega_k(t) + \sum_{j=1}^{q} \varphi_{ma,j,k}\omega_k(t-j) & 0 & \cdots & 0 \end{bmatrix}^T \tag{4.19}$$

$$\boldsymbol{\nu_k}(t) = [\nu_k(t) \quad 0 \quad \cdots \quad 0]^T \tag{4.20}$$

If the model errors are assumed to be uniformly distributed at each step, $\boldsymbol{Q_{\omega,k}}$ becomes a matrix with all elements zero except the top-left element, which is $\sigma_{\omega_k}^2\left(1+\sum_{j=1}^{q}\varphi_{ma,j,k}\right)$, while $\boldsymbol{Q_{\nu,k}}$ is a matrix with all elements zero except the top-left element being $\sigma_{\nu_k}^2$. Within the framework of signal denoising for the SHM system discussed in this study, σ_{ω_k} and σ_{ν_k} are determined by the final fitting accuracy of the least squares method in ARIMAX and the sensor resolution, respectively. The flowchart for the entire data denoising stage is illustrated in Fig. 4.1.

Furthermore, ARIMAX is merely one approach for time series modeling, especially with the rise of DL, where NN-based sequence models have achieved higher accuracy and performance. The state transition model required by the IKF method inherently leverages historical signal information and current external data to infer the current signal value, a functionality that aligns with RNN and its variants. Consequently, the combination of RNN+IKF can be derived using the same methodology as ARIMAX+IKF.

The experimental validation of the denoising effectiveness of the proposed IKF algorithm employs the same dataset as the missing data reconstruction validation in Section 4.3. The dataset details and verification results will be presented and discussed collectively in Section 4.3.

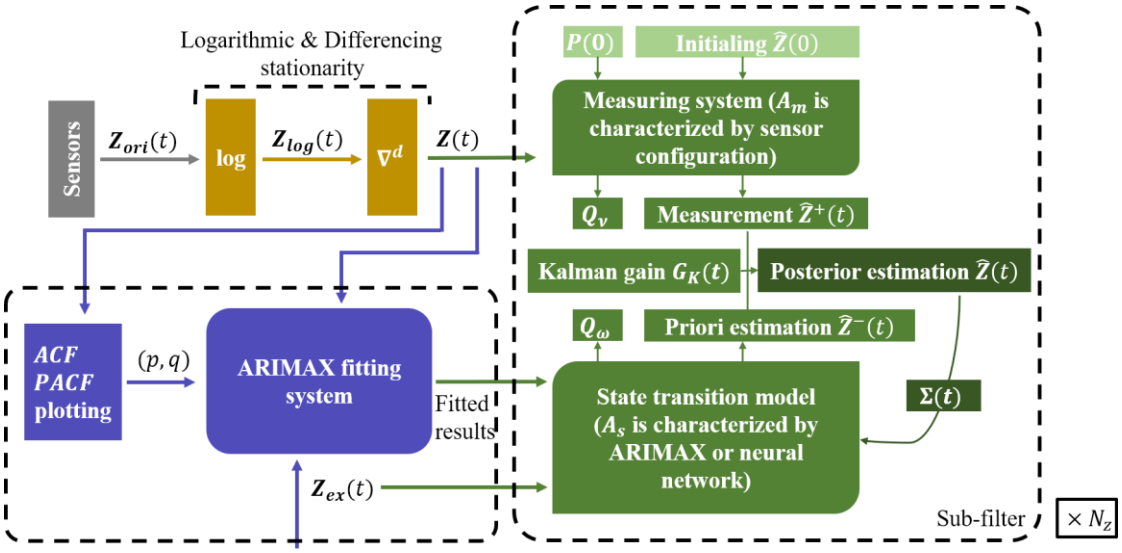


Fig. 4.1 Flowchart of IKF

4.3 Lost data reconstruction

Data missing occurs when signal values from one or multiple channels are lost (i.e., replaced with zeros) over a certain time period. To reconstruct these missing segments, we utilize available data including the non-missing data before and after the missing segment within the same channel, and complete time-series data from other channels (assuming single-channel missingness). For time-series data collected from real physical systems, inherent correlations exist both between different channels and across time within each channel. These correlations form the theoretical foundation for reconstructing missing data using available information. As discussed in Section 2.2, existing methods typically only utilize part of the available data, leaving potential for further performance improvement.

This section presents a dual-side bidirectional echo state network reconstructor (Bi-ESN-R) for missing data imputation. The “bidirectional” aspect refers to the network

architecture that extends standard ESN by processing data in both forward and backward directions to enhance information richness. The "dual-side" approach simultaneously utilizes both temporal information (adjacent data before and after the missing segment) and spatial information (complete data from other channels). The performance, robustness, and hyperparameter determination of Bi-ESN-R will be evaluated using datasets from an actual engineering SHM system.

4.3.1 Standard ESN in reconstruction task

All reconstruction tasks are performed on the filtered signal $\boldsymbol{Z}(t)$ obtained in Section 4.2. The signal from the channel with missing data serves as the network output, denoted as $\boldsymbol{Y}(t) \in \mathbb{R}^{N_{out}}$, while data from other complete channels constitute the network input $\boldsymbol{X}(t) \in \mathbb{R}^{N_{in}}$. Here, N_{in} and N_{out} represent the number of input and output channels (i.e., the neuron number in the ESN's input and output layers), with $N_{out} = 1$ for the most common single-channel missingness scenario.

Similar to the RNN described in Section 3.2.1, the ESN's hidden layer maintains a state sequence that encodes historical information through recursive processing of input and output sequences. However, unlike RNNs that require trained encoding schemes, ESNs employ numerous simple encoding patterns (Jaeger & Haas 2004). Each state sequence captures only minimal features of the original signal, but the collective effect becomes powerful when the reservoir size is sufficiently large ($N_{res} \gg max\ (N_{in},\ N_{out})$). This permits the use of randomly generated (rather than trained) weight matrices for state encoding, as formalized in Eq. (4.21).

$$\boldsymbol{R}(t) = f(\boldsymbol{W_{in}}\boldsymbol{X}(t) + \boldsymbol{W_{res}}\boldsymbol{R}(t-1) + \boldsymbol{W_{back}}\boldsymbol{Y}(t-1) + \boldsymbol{b_r}) \tag{4.21}$$

Here, $\boldsymbol{W_{in}} \in \mathbb{R}^{N_{res} \times N_{in}}$, $\boldsymbol{W_{res}} \in \mathbb{R}^{N_{res} \times N_{res}}$, and $\boldsymbol{W_{back}} \in \mathbb{R}^{N_{res} \times N_{out}}$ t represent the mapping matrices from the current input, previous state, and previous output to the current state, respectively, $\boldsymbol{b_r}$ denotes the bias term. To enhance the network's memory capacity, prior research developed leaky integrator units (Lun et al., 2023), which update the state according to the following equation:

$$\boldsymbol{R}(t) \leftarrow (1-\alpha)\boldsymbol{R}(t-1) + \alpha\boldsymbol{R}(t) \tag{4.22}$$

where, $\alpha \in [0,1]$ represents the leak rate. Previous researchers also proposed mapping matrix scaling methods to improve performance (Sun et al., 2022), introducing scaling factors β_w and β_b for $\boldsymbol{W_{in}}$ and $\boldsymbol{b_r}$, respectively. The hidden layer, which contains numerous simple state sequences, is termed the reservoir. To maintain its training-free property, the associated mapping matrices must be sufficiently large and sparse. Additionally, to ensure that historical information neither decays too rapidly nor accumulates explosively during encoding, the elements of the randomly generated $\boldsymbol{W_{res}}$ must adhere to specific constraints. Specifically, research has shown that the spectral radius $\rho(\boldsymbol{W_{res}})$ (i.e., the largest eigenvalue of W_{res}) should lie within a reasonable range. This ensures stable and sustained propagation of internal system information and external excitations over time, mimicking an "echo state" within the reservoir which is the foundational principle enabling ESN functionality. The spectral radius $\rho(\boldsymbol{W_{res}})$ and matrix sparsity $spar(\boldsymbol{W_{res}})$ (the fraction of non-zero elements) emerge as the two most critical hyperparameters (Jiang & Lai 2019). However, prior studies suggest varying optimal ranges for $\rho(\boldsymbol{W_{res}})$ across different problems. A detailed discussion on its determination is provided in Section 4.3.4. The workflow of a standard ESN is illustrated in Fig. 4.2.

Once the current state is obtained, the output for this time step can be generated through a linear combination of these states, where the combination coefficients serve as the trainable parameters of the ESN. For signals strongly influenced by values from previous time steps, delayed sequences can be incorporated into the output layer. The estimated target signal value at the current time step thus includes both contributions from the state and autoregressive characteristics, expressed as:

$$\widehat{\boldsymbol{Y}}(t) = \boldsymbol{W_{out}}\big[\boldsymbol{R}(t); \boldsymbol{Y}(t-1); \dots; \boldsymbol{Y}\big(t-t_y\big)\big] + \boldsymbol{b_y} \tag{4.23}$$

where, $\boldsymbol{W_{out}} \in \mathbb{R}^{N_{out} \times (N_{in}+N_{res}+t_Y N_{out})}$ represents the sole trainable weight parameter in this linear regression problem, where t_y denotes the number of autoregressive lookback

steps. The model's loss function can still be characterized by the MSE, defined as:

$$L(\boldsymbol{Y}, \widehat{\boldsymbol{Y}}) = \frac{1}{N_{out}} \sum_{k=1}^{N_{out}} \left(\sum_{t=T_a}^{T} \left(Y_k(t) - \widehat{Y}_k(t) \right)^2 + \eta \left\| \boldsymbol{W_{out,k}} \right\|^2 \right) \tag{4.24}$$

where, T_a denotes the initial segment length (typically set to $0.1T$) to be discarded due to encoding instability caused by random initial states, $\|\cdot\|$ represents the vector norm, and η is the regularization parameter. This constitutes a supervised training process where the network loss is determined by comparing ground truth values from the training set with current network estimates. Through optimization algorithms (e.g., SGD), the weight parameters are iteratively adjusted to progressively reduce their contribution to the loss.

Once trained, the network can autoregressively compute future outputs step-by-step, using initial output values and input sequences. In practice, the signal segment immediately preceding the missing data is typically used for network training. The last t_y points before the missing segment then serve as initial values for $\boldsymbol{Y}$, which—combined with $\boldsymbol{X}$, which initiates the autoregressive computation to reconstruct the entire missing segment.

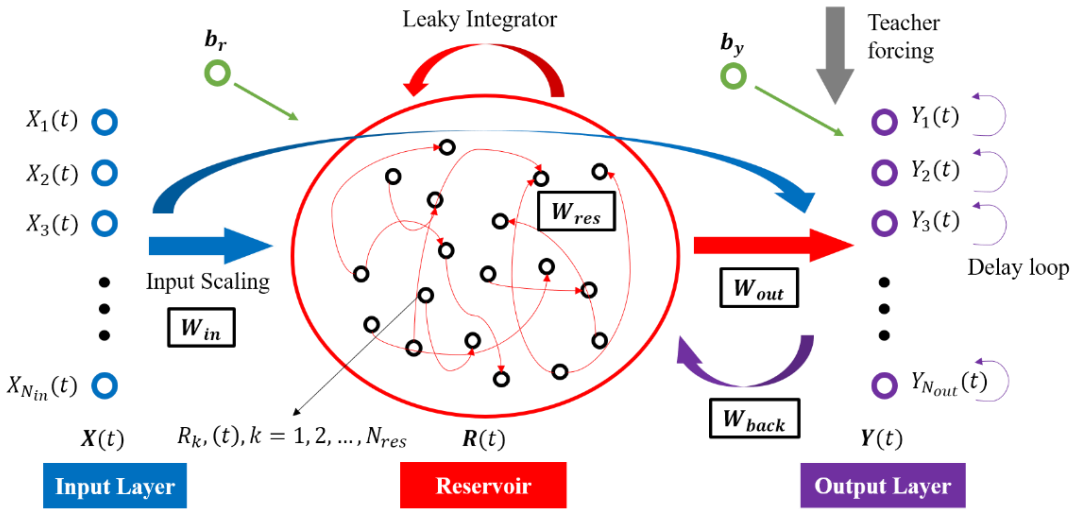


Fig. 4.2 Standard ESN

4.3.2 Bi-ESN-R in reconstruction task

The bidirectional architecture is a generalized extension of RNNs that incorporates a

parallel hidden layer processing the time series in reverse. While traditional time-series analysis assumes causality (current signals depend only on historical data), in practice, future signals, especially near-future values, can provide valuable references for inferring current states. Thus, in bidirectional ESN (Bi-ESN), at each time step t, the original hidden neurons encode historical information through progressive forward computation, starting from an initial state $\overrightarrow{\boldsymbol{R}}(0)$ to derive $\overrightarrow{\boldsymbol{R}}(t)$; while the additional hidden neurons encode future information via reverse computation from a terminal state $\overleftarrow{\boldsymbol{R}}(T+1)$ to obtain $\overleftarrow{\boldsymbol{R}}(t)$. This dual integration of past and future contexts significantly enhances the accuracy of current signal estimation. However, since outputs must still be generated sequentially, autoregressive structures in the backward pass are disabled (as $\boldsymbol{Y}(t+1)$ remains unknown when backward encoding). The encoding and output processes for Bi-ESN are formalized in Eqs. (3.25), (3.26), and (3.27), with the workflow illustrated in Fig. 4.3.

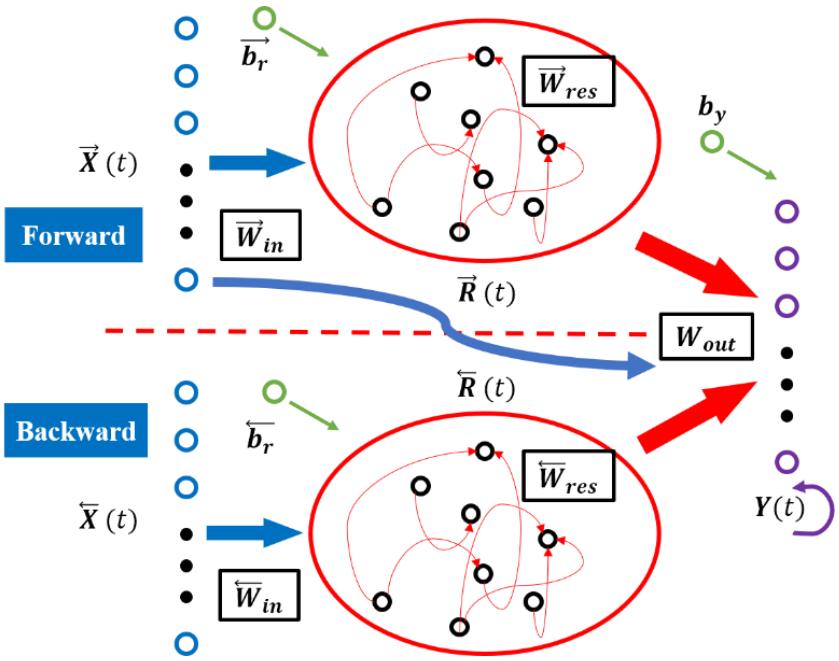


Fig. 4.3 Bi-ESN

$$\overrightarrow{\boldsymbol{R}}(t) = (1-\alpha)\overrightarrow{\boldsymbol{R}}(t-1) + \alpha f\left(\overrightarrow{\boldsymbol{W}}_{\boldsymbol{in}}\boldsymbol{X}(t) + \overrightarrow{\boldsymbol{W}}_{\boldsymbol{res}}\overrightarrow{\boldsymbol{R}}(t-1) + \overrightarrow{\boldsymbol{W}}_{\boldsymbol{back}}\boldsymbol{Y}(t-1) + \overrightarrow{\boldsymbol{b}}_{\boldsymbol{r}}\right) \tag{4.25}$$

$$\overleftarrow{\boldsymbol{R}}(t) = (1-\alpha)\overleftarrow{\boldsymbol{R}}(t-1) + \alpha f\left(\overleftarrow{\boldsymbol{W}}_{\boldsymbol{in}}\boldsymbol{X}(t) + \overleftarrow{\boldsymbol{W}}_{\boldsymbol{res}}\overleftarrow{\boldsymbol{R}}(t+1) + \overleftarrow{\boldsymbol{b}}_{\boldsymbol{r}}\right) \tag{4.26}$$

$$\widehat{\boldsymbol{Y}}(t) = \boldsymbol{W_{out}}\left[\overrightarrow{\boldsymbol{R}}(t); \overleftarrow{\boldsymbol{R}}(t); \boldsymbol{Y}(t-1); \ldots; \boldsymbol{Y}\left(t-t_y\right)\right] + \boldsymbol{b_y} \tag{4.27}$$

In the Bi-ESN architecture, two parallel reservoirs exist, each with independently sampled encoding matrices drawn from the same distribution. The forward reservoir retains the three standard ESN mapping matrices, while the backward reservoir utilizes only two mapping matrices. Due to the incorporation of backward states, the dimensionality of the trainable output weight matrix expands to $\mathbb{R}^{N_{out}\times(N_{in}+2N_{res}+t_yN_{out})}$. While Bi-ESN's training and reconstruction procedures mirror those of standard ESN, which is relying solely on unilateral (pre-missing segment) data and cross-channel correlations, its autoregressive nature causes error accumulation, degrading the accuracy of reconstructions farther from the gap's starting point. Conversely, post-gap signals remain unused despite their potential to improve distal reconstructions. To address this, this study proposes Bi-ESN-R, which employs dual Bi-ESNs: one is responsible for the forward reconstruction, starting from pre-missing segment signals and yielding $\widehat{\boldsymbol{Y}}_{\boldsymbol{form}}(t)$; and the other is responsible for the backward reconstruction, initiating from post-missing segment signals, producing $\widehat{\boldsymbol{Y}}_{\boldsymbol{latt}}(t)$. The final output combines both reconstructions via a group of adaptive parameters:

$$\widehat{\boldsymbol{Y}}(t) = \xi_1(t)\widehat{\boldsymbol{Y}}_{\boldsymbol{form}}(t) + \xi_2(t)flip\left(\widehat{\boldsymbol{Y}}_{\boldsymbol{latt}}(t)\right) \tag{4.28}$$

where, $\xi_1(t) + \xi_2(t) = 1$ defines time-varying blending parameters, and $flip(\cdot)$ denotes the vector reversal operation. The blending parameters should be determined based on the temporal proximity to the start and end of the missing segment. Specifically, reconstruction values closer to one side should receive higher weights in the blending process, i.e.:

$$\xi_1(t) = \frac{\xi T - t}{\xi T} \tag{4.29}$$

where, ξ is a hyperparameter where larger values prioritize $\widehat{\boldsymbol{Y}}_{\boldsymbol{form}}(t)$ (forward reconstruction), while smaller values favor $\widehat{\boldsymbol{Y}}_{\boldsymbol{latt}}(t)$ (backward reconstruction). The

architecture of Bi-ESN-R is illustrated in Fig. 4.4.

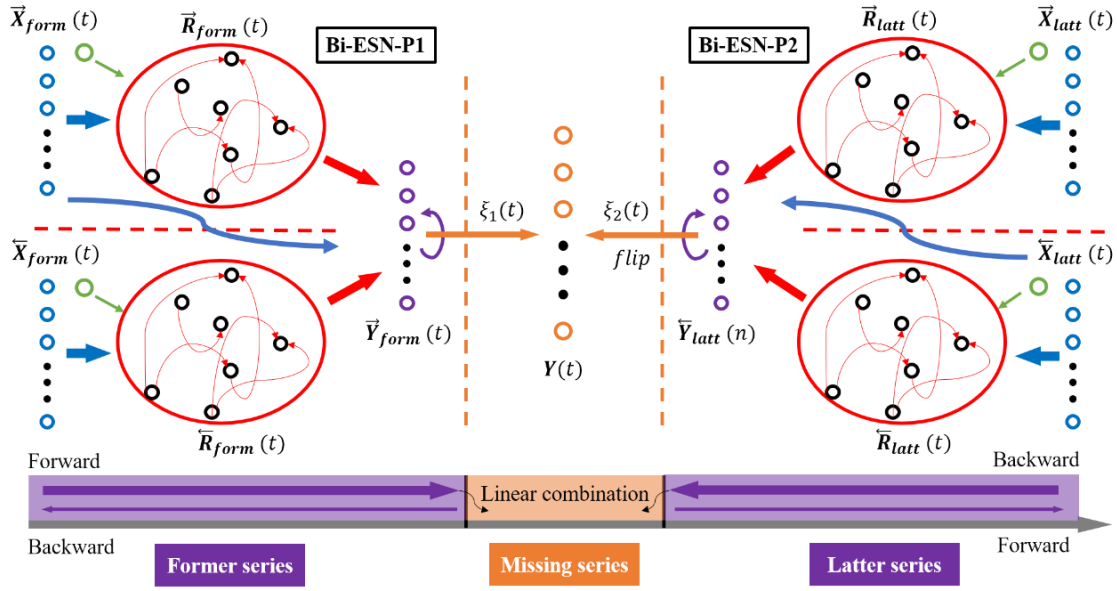


Fig. 4.4 Bi-ESN-R

4.3.3 Application: Acceleration reconstruction of Canton Tower

The Canton Tower, located in Guangzhou, China, is a super-tall structure employing a tube-in-tube structural system, with a building height of approximately 600 m. As of 2024, it ranks as the world's second-tallest tower while serving dual functions in broadcasting and telecommunications. Its inner tube, constructed of reinforced concrete, is interconnected with an outer steel frame at 37 floor levels to ensure collaborative load-bearing behavior. The innovative hyperbolic design not only creates a visually striking profile but also introduces complex structural dynamics. Situated in a typhoon-prone region, the tower faces inevitable long-term degradation, including steel fatigue, concrete cracking, and steel-concrete interface delamination, over its 100+ year design lifespan, necessitating rigorous monitoring of operational states and environmental conditions.

A long-term SHM system operated by the Hong Kong Polytechnic University integrates 16 sensor types to measure variables such as wind pressure, tilt, temperature, and acceleration (Ni et al., 2009), enabling continuous data acquisition for operational insights.

Acceleration signals are particularly critical for SHM post-processing, as they reflect

vibration characteristics, structural damping, and stiffness degradation with higher practicality compared to displacement monitoring (eliminating the need for reference points). This study validates the Bi-ESN-R model using acceleration data from the Canton Tower SHM system.

The tower features eight acceleration monitoring segments at elevations of 30.63 m, 119.30 m, 171.10 m, 228.50 m, 275.30 m, 332.15 m, 384.24 m, and 446.80 m. Each segment is equipped with a pair of uniaxial accelerometers (AS-2000C) measuring orthogonal directional accelerations. The fourth and eighth segments include an additional accelerometer pair, totaling 20 accelerometers. A triaxial seismometer (SPC-51C) is installed at ground level to monitor seismic accelerations. All sensors operate at a 50 Hz sampling frequency, with their layout detailed in Fig. 4.5.

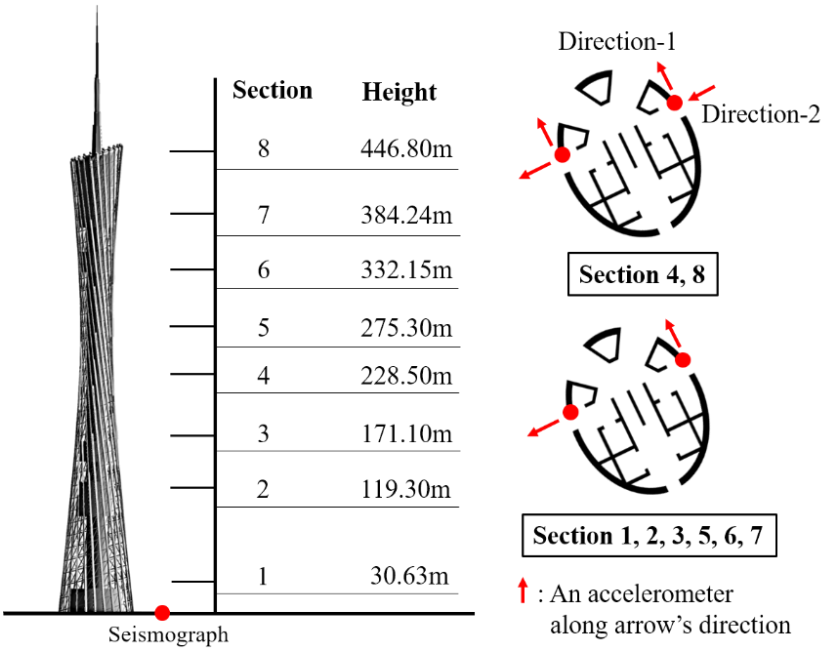


Fig. 4.5 Accelerometer layout of Canton Tower SHM system

Considering the higher probability of sensor failures during disasters and the predominantly non-stationary nature of structural acceleration data under such conditions, this chapter aims to validate the superior performance of the proposed model using two datasets: earthquake-induced vibrations and free vibrations. The free vibration acceleration signals are nearly stationary but exhibit a lower signal-to-noise ratio (SNR), whereas earthquake-induced accelerations are distinctly non-stationary but have a higher

SNR. Here, the non-stationary data has changing variance and covariance, but no seasonality. Through data screening, the acceleration signal along Direction-1 (as shown in Fig. 4.5) of the Canton Tower during the Huizhou earthquake on March 14, 2022, was selected as the non-stationary dataset, while acceleration data recorded one hour prior to the earthquake in the same direction served as the stationary dataset. The epicenter of this earthquake was located at 115.04°E, 22.51°N (Pinghai, Huidong, which is 167 km from the structure), with a focal depth of 25 km and a magnitude of M_W4.1.

Table 4.1 Reconstruction conditions

Condition No.	Dataset	Missing channel	Missing rate	Missing location
S-1-1/2/3	$\boldsymbol{ST}$	$\boldsymbol{ST_5}$	0.04/0.08/0.12	/
S-5-2		$\boldsymbol{ST_{1,3,5,7,9}}$	0.08	/
S-9-2		$\boldsymbol{ST_{1\sim3}, ST_{5\sim10}}$	0.08	/
E-1-1/2/3-P	$\boldsymbol{EQ}$	$\boldsymbol{EQ_5}$	0.04/0.08/0.12	Peak phase
E-1-1/2/3-I		$\boldsymbol{EQ_5}$	0.04/0.08/0.12	Starting phase
E-1-1/2/3-E		$\boldsymbol{EQ_5}$	0.04/0.08/0.12	Ending phase
E-1-4/5/6-P		$\boldsymbol{EQ_5}$	0.20/0.35/0.50	Peak phase
E-5-2-P		$\boldsymbol{ST_{1,3,5,7,9}}$	0.08	Peak phase
E-9-2-P		$\boldsymbol{ST_{1\sim3}, ST_{5\sim10}}$	0.08	Peak phase

The earthquake and stationary datasets are designated as $\boldsymbol{EQ}$ and $\boldsymbol{ST}$, respectively, each comprising 11 channels (10 accelerometers and 1 seismograph): $\boldsymbol{EQ_g, EQ_1, \dots, EQ_{10}}$, and $\boldsymbol{ST_g, ST_1, \dots, ST_{10}}$. The $\boldsymbol{EQ}$ dataset spans 50 s (2500 data points), while the $\boldsymbol{ST}$ dataset covers 200 s (10000 data points). As outlined in Table 4.1, a series of data loss scenarios were established, including single-channel losses of varying durations and

multi-channel losses of different lengths. For the $\boldsymbol{EQ}$ dataset, scenarios with identical loss durations further considered the positioning of missing segments within distinct earthquake phases: the starting phase, peak phase, and ending phase.

4.3.4 Lost reconstruction results and discussion

This section will first demonstrate the reconstruction performance of Bi-ESN-R for regular data loss and compare it with common NN-based methods. Since all performance evaluations are based on denoised datasets, the impact of the IKF preprocessing method on reconstruction will also be discussed. Subsequently, the robustness of Bi-ESN-R under extreme scenarios, such as long-term loss, limited adjacent data availability, and multi-channel loss, will be explored. Finally, the influence of three critical hyperparameters on the model will be tested to provide selection recommendations.

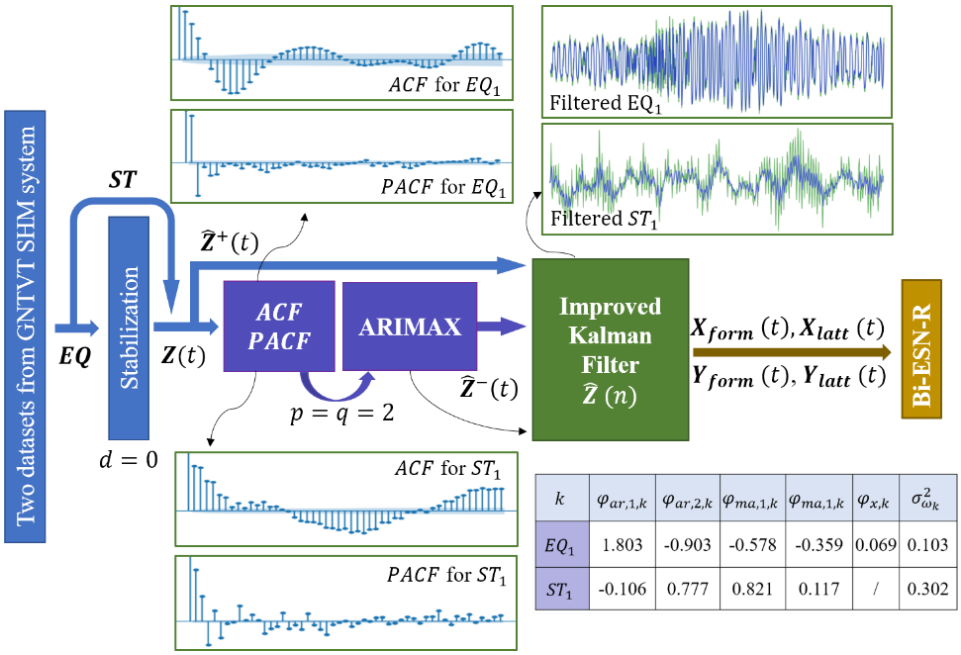
Based on preliminary data analysis, the IKF method grounded in ARIMAX was configured with $d = 0$ (for the non-stationary $\boldsymbol{EQ}$ dataset, only logarithmic stabilization was applied), and $p = q = 2$. Taking channels $\boldsymbol{EQ_1}$ and $\boldsymbol{ST_1}$ as examples, Fig. 4.6 illustrates the ACF, PACF, and time-domain comparisons of signals before and after denoising during the IKF process. In Bi-ESN-R, the hyperparameter t_y is set equal to p to ensure consistent signal autoregressive characteristics throughout the process. Other hyperparameters related to the network structure follow the values specified in Table 3.2.

Table 4.2 Hyperparameters of Bi-ESN-R

$\boldsymbol{spar(W_{res})}$	$\boldsymbol{\rho(W_{res})}$	$\boldsymbol{N_{res}}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta_w}$	$\boldsymbol{\beta_b}$	$\boldsymbol{\xi}$
$\boldsymbol{2.0N_{res}}$	5.2	$\boldsymbol{0.6T}$	0.7	2.0	2.0	1.0

The model training optimization method remains SGD, with a learning rate $lr = 0.001$, $\eta = 0.05$, and training cycles lasting 100 epochs. All hyperparameters above were determined through grid search, with an analysis of the search results to be completed in Section 4.3.4.3. To facilitate horizontal comparisons of reconstruction

performance under data loss, the filtered raw data were normalized as illustrated before. The accuracy of reconstruction results is evaluated using the MSE between the reconstructed signals and the ground truth.



k	$\varphi_{ar,1,k}$	$\varphi_{ar,2,k}$	$\varphi_{ma,1,k}$	$\varphi_{ma,1,k}$	$\varphi_{x,k}$	$\sigma^2_{\omega_k}$
EQ_1	1.803	-0.903	-0.578	-0.359	0.069	0.103
ST_1	-0.106	0.777	0.821	0.117	/	0.302

Fig. 4.6 Denoising process of IKF

4.3.4.1 Basic reconstruction performance of Bi-ESN-R

The single-channel loss scenarios S-1-2 and E-1-2-P were selected as the baseline cases for reconstructing stationary and non-stationary data, respectively. The missing segment lengths were 200- and 800-time steps, with the missing segment in the non-stationary dataset located at the peak phase of the seismic wave. The reconstruction results are compared with traditional time-series processing methods in Figs. 4.7 and 4.8. The traditional algorithms used for comparison include ARIMAX, KF, and a standard RNN (details in Section 3.2.1). The RNN here consists of two hidden layers, each containing 100 neurons.

For the S-1-2 scenario, the MSE values of Bi-ESN-R, standard RNN, ARIMAX, and KF were 0.069, 0.174, 0.366, and 0.353, respectively. For E-1-2-P, the MSE values were 0.038, 0.081, 1.02, and 0.98. Clearly, the reconstruction accuracy of the Bi-ESN-R model far exceeded that of traditional methods like KF and ARIMAX, and even surpassed the

standard NN-based approach. Specifically, the Bi-ESN-R model demonstrated robust reconstruction capabilities across the entire missing segment, whereas classical time-series methods maintained high accuracy only in approximately the first 20% of the reconstructed segment. Beyond this range, these methods could only roughly preserve the fundamental frequency alignment between the reconstructed and true signals, failing to recover high-frequency components, resulting in reconstructed amplitudes smaller than the true values. In the final 15% of the reconstructed segment, even the consistency in the fundamental frequency was largely lost. This discrepancy arises because KF and ARIMAX rely on a limited number of parameters to describe signal patterns and over-depend on autoregressive structures, making their predictive capabilities vulnerable to error accumulation. In contrast, Bi-ESN-R employs four large reservoirs to explore and fit signal relationships, while only a small fraction of its output relies on autoregressive contributions, thereby mitigating error accumulation.

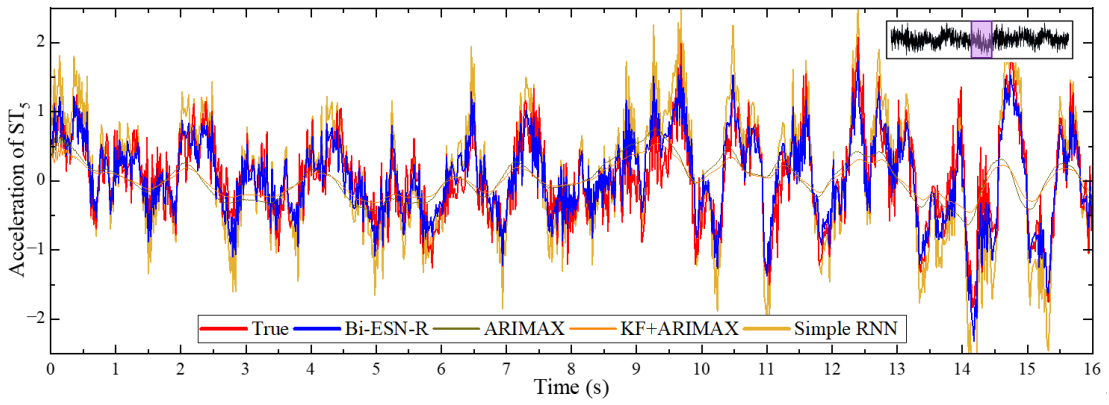


Fig. 4.7 S-1-2 reconstruction results and model comparison

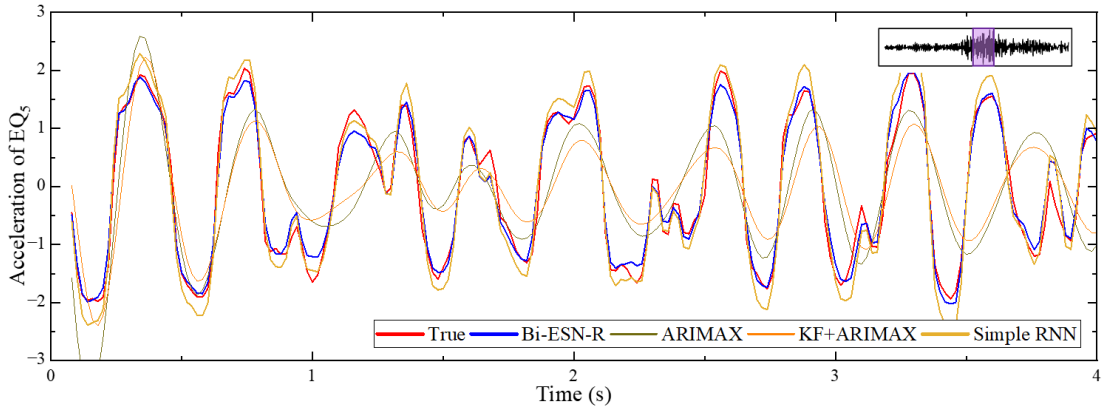


Fig. 4.8 E-1-2-P reconstruction results and model comparison

To further demonstrate the effectiveness of the bidirectional and bilateral improvements in Bi-ESN-R, the reconstruction results of unidirectional ESN (standard ESN) and single-side bidirectional ESN are presented in Figs. 4.9 and 4.10. The suffix "F" denotes a single-side model that uses only pre-missing segment information and external signals for reconstruction (i.e., solely $\widehat{\boldsymbol{Y}}_{\boldsymbol{form}}(t)$), while "L" indicates a single-side model relying only on post-missing segment information and external signals (i.e., solely $\widehat{\boldsymbol{Y}}_{\boldsymbol{latt}}(t)$). Evidently, the proposed Bi-ESN-R model achieves higher accuracy (MSE values of 0.0384 and 0.0888 for the two scenarios) compared to both the standard ESN and single-side Bi-ESN variants. The largest errors occur in the ESN-F model results (MSE values of 0.1058 and 0.1116), which rely exclusively on $\widehat{\boldsymbol{Y}}_{\boldsymbol{form}}(t)$ encoded based on single-side information for reconstruction. After incorporating the reverse-encoded state sequence $\overleftarrow{\boldsymbol{R}}(t)$ (i.e., Bi-ESN-F), the MSE decreases by over 20%. A similar accuracy improvement is observed between ESN-L and Bi-ESN-L, with the MSE for reconstructing S-1-2 dropping from 0.1376 to 0.1099, and for E-1-2-P from 0.0619 to 0.0412. These results confirm that integrating future information enhances the reconstruction accuracy of ESN-based models. Given the inherent autoregressive properties in ESN-based models, though error accumulation for multi-step predictions remains unavoidable (albeit less pronounced than in ARIMAX or KF methods), Figs. 4.9 and 4.10 reveal that ESN-F and Bi-ESN-F provide more accurate reconstructions at the beginning of the missing segment, while ESN-L and Bi-ESN-L perform better toward the end. Thus, optimizing the model by weighting the reliability of Bi-ESN-F and Bi-ESN-L outputs based on their proximity to the reconstruction start point delivers the best performance. Consequently, the Bi-ESN-R model reduces errors by over 10% compared to unidirectional Bi-ESN variants.

For the $\boldsymbol{ST}$ reconstruction task, the accuracy of ESN-F and ESN-L is nearly identical, and the performance of Bi-ESN-F and Bi-ESN-L is also comparable, indicating that pre- and post-missing-segment data contribute almost equally to reconstruction. However, for the $\boldsymbol{EQ}$ task, ESN-L and Bi-ESN-L significantly outperform ESN-F and Bi-ESN-F,

suggesting that incorporating $\widehat{\boldsymbol{Y}}_{\boldsymbol{latt}}(t)$ yields more accurate results. This discrepancy arises because the ***ST*** dataset lacks significant external excitation, maintaining consistent temporal correlations between the missing segment and its adjacent regions. In contrast, the ***EQ*** dataset involves seismic excitation, which complicates inference from pre-missing data, while the post-missing signals primarily reflect structural recovery, which is a phase with stronger correlations to the missing segment. Additionally, regardless of the ESN variant, ***EQ*** reconstruction accuracy surpasses that of ***ST***, particularly in recovering high-frequency components. This is attributed to residual stochastic noise in the filtered signals, which weakens inter-signal correlations, even after denoising.

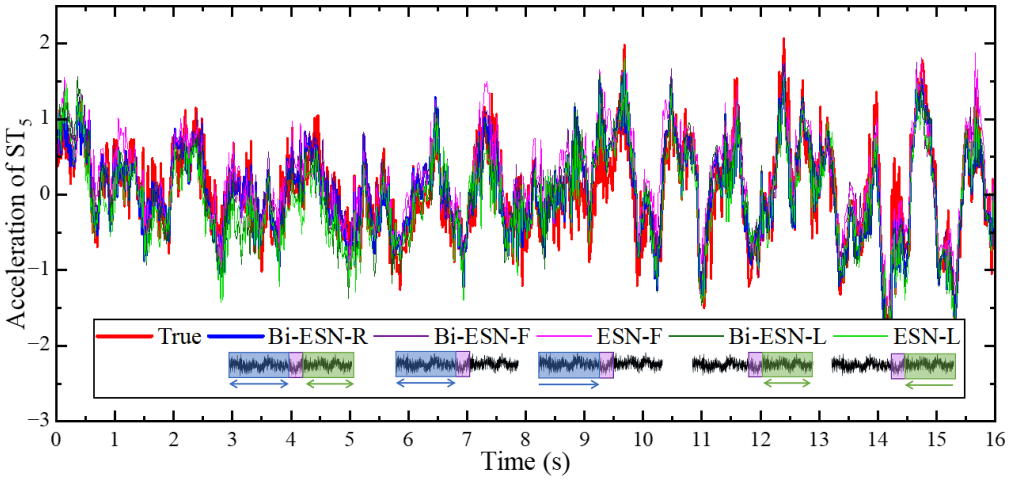


Fig. 4.9 Impact of dual-side and bidirectional structure on stationary data reconstruction

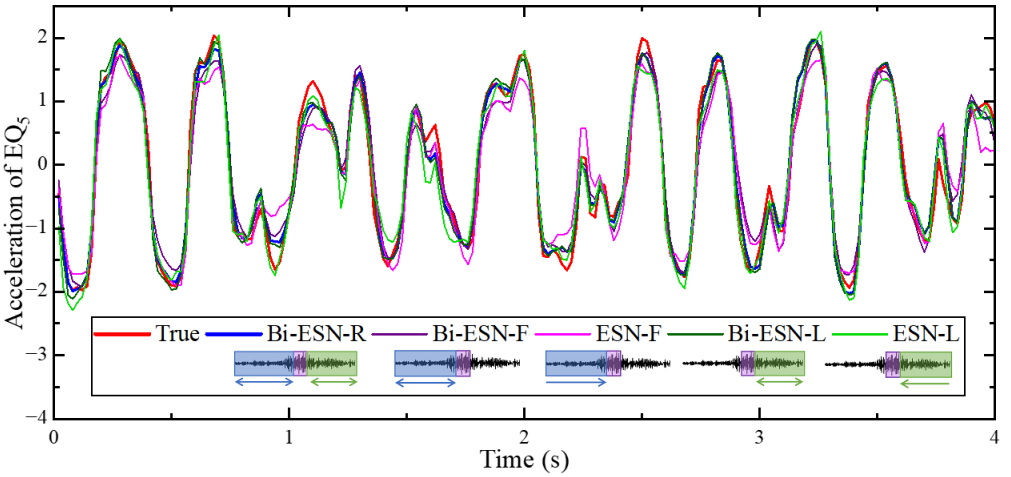


Fig. 4.10 Impact of dual-side and bidirectional structure on non-stationary data reconstruction

The presence of uncorrelated temporal noise in raw signals can mislead the network during training, necessitating an evaluation of the impact of IKF-based denoising on reconstruction performance. All parameters in the IKF method used in this study are inferred from known information or determined by sensor specifications, without relying on arbitrary assumptions, making it more rational than conventional low-pass filters. Figs. 4.11 and 4.12 illustrate the time- and frequency-domain representations of raw and filtered signals for S-1-2 and E-1-2-P, with reconstructed signals derived from these datasets shown below. Signal frequencies are obtained via FFT.

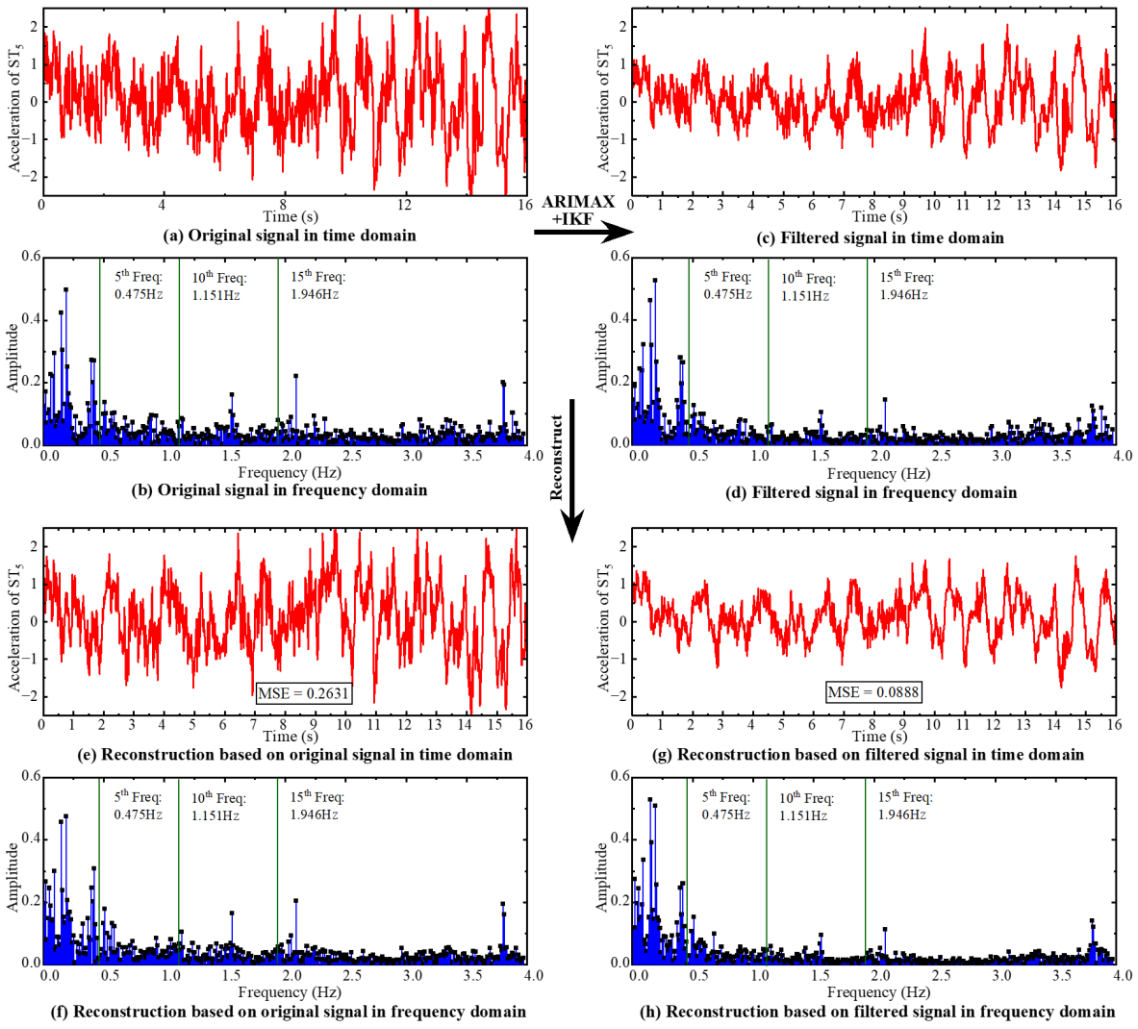


Fig. 4.11 IKF denoising performance for stationary data

The 5th, 10th, and 15th natural frequencies of the Canton Tower are 0.475 Hz, 1.151 Hz, and 1.946 Hz, respectively (Ni et al., 2009). During free vibration, acceleration signals are dominated by the first five modal frequencies. Under seismic excitation, the 5th to

10th modes are also activated. Both $\boldsymbol{ST}$ and $\boldsymbol{EQ}$ datasets contain non-negligible frequency components exceeding the 10th or even 15th modes, complicating the selection of an appropriate cutoff frequency for simple low-pass filters. In contrast, the IKF algorithm preserves all high-amplitude components regardless of their frequency while effectively attenuating insignificant high-frequency noise. Specifically, it removes components definitively identified as noise while retaining potentially valid ones. Furthermore, the algorithm automatically adjusts denoising intensity based on the SNR, calculated from $\boldsymbol{Q_\omega}$ and $\boldsymbol{Q_s}$: for $\boldsymbol{EQ}$ data with higher SNR, it applies minimal smoothing, whereas for $\boldsymbol{ST}$ data with lower SNR, it filters more aggressively to enhance structural vibration features.

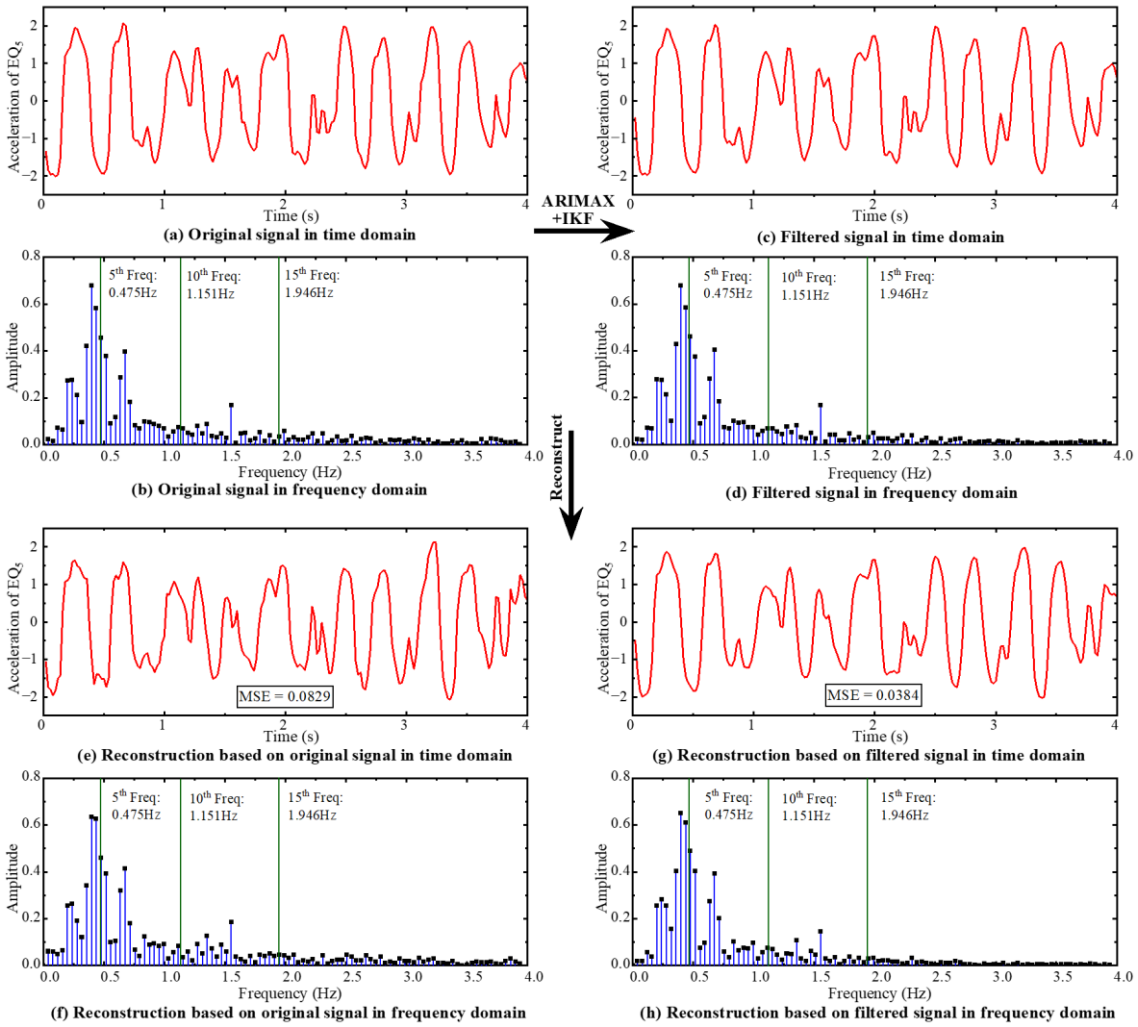


Fig. 4.12 IKF denoising performance for non-stationary data

From the perspective of denoising aiding reconstruction, the model's learning capacity is

inherently limited, leading to insufficient recovery of minor high-frequency components. The ultimate manifestation is a reduced proportion of these challenging-to-learn components in the reconstructed signals. However, this is not detrimental, as most of these components are considered noise, and this limitation effectively acts as an additional denoising step. Although the Bi-ESN-R model retains the ability to reconstruct high-frequency signals, its reconstruction error for filtered signals remains lower than that for raw signals. Specifically, denoising reduces the reconstruction error for ***EQ*** by approximately 50% and for ***ST*** by around 70%. Since the IKF algorithm cannot entirely eliminate specific frequency components, its smoothing performance is inferior to that of low-pass filters. Therefore, in scenarios lacking evidence for the validity of high-frequency components, low-pass filters may yield better results, as illustrated in Fig. 4.13.

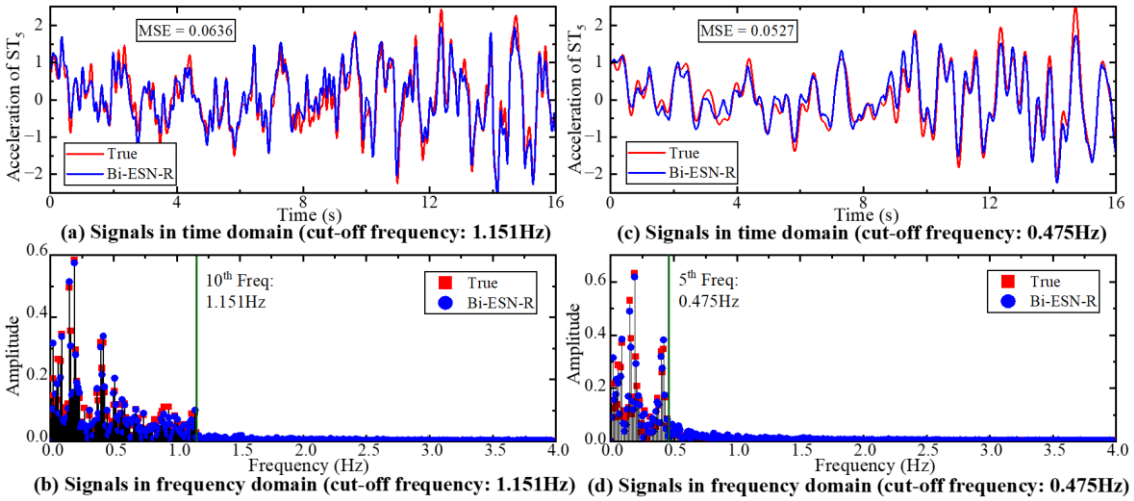


Fig. 4.13 Comparison between IKF and low-pass filter

4.3.4.2 Robustness of Bi-ESN-R

The stochastic nature of lost data introduces complexity to real-world reconstruction tasks, which often deviate from the idealized assumptions discussed in the previous sections. This section examines the model's robustness under extreme missing-data scenarios.

A critical challenge arises with prolonged data missing: as the missing segment lengthens, the central portion becomes increasingly distant from the available data at both ends, inevitably degrading reconstruction accuracy. This issue is exacerbated in non-stationary

data scenarios, where the statistical relationships between the missing segment and existing data in the same channel are likely to shift, further impairing reconstruction quality. However, since external information from other channels remains intact (in single-channel missing-data tasks), the model's accuracy does not degrade indefinitely. To evaluate this, test cases E-1-1-P to E-1-6-P are analyzed, representing missing data proportions ranging from 4% to 50%, with corresponding reconstruction results illustrated in Figs. 4.14 and 4.15.

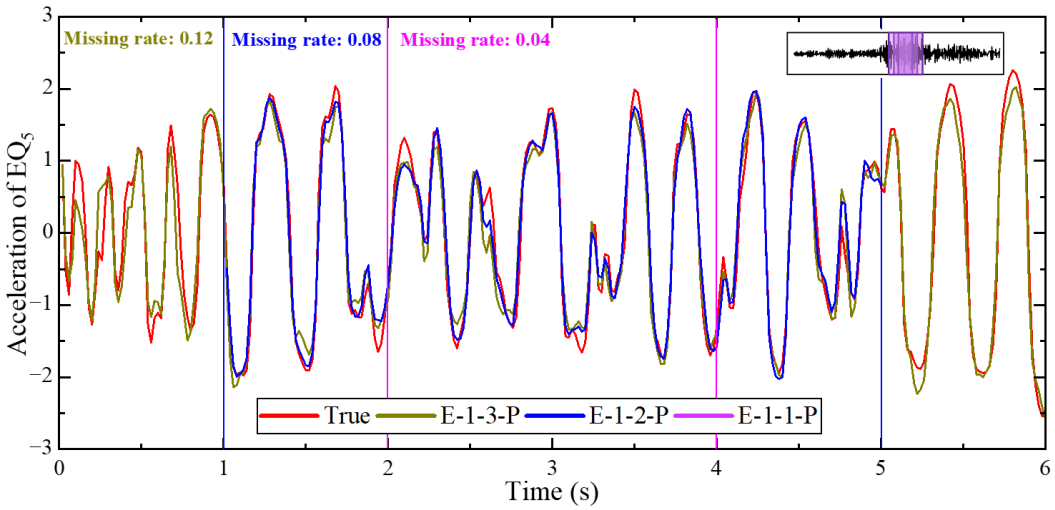
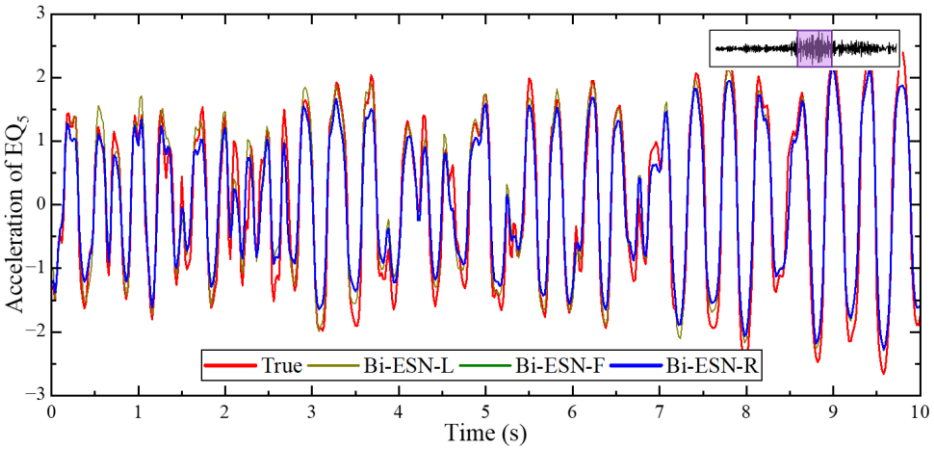


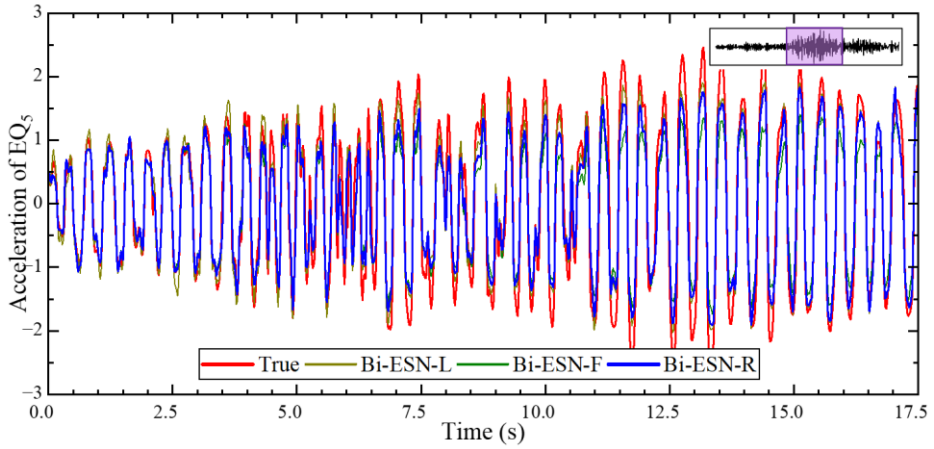
Fig. 4.14 Reconstruction of short-term lost data

For ***EQ*** data, as the missing rate increases from 0.04 to 0.12, the reconstruction accuracy declines slightly (MSE rises from 0.0739 to 0.0951), yet the MSE remains below 0.1. This indicates that when sufficient data segments resembling the missing portion exist before and after the lost data, the Bi-ESN-R model can rapidly reconstruct missing data accounting for approximately 10% of the total dataset with relatively high accuracy. As the missing rate further increases, the alignment between reconstructed and original signals becomes notably better at the start and end segments compared to the middle portion. This occurs because error accumulation is minimal in the boundary regions, whereas the central segment relies almost entirely on synchronized information from other channels, with autoregressive features largely introducing interference. When the missing rate exceeds 0.35, the gap spans the entire vibration peak phase induced by the seismic event, rendering the current channel's data nearly ineffective due to the absence

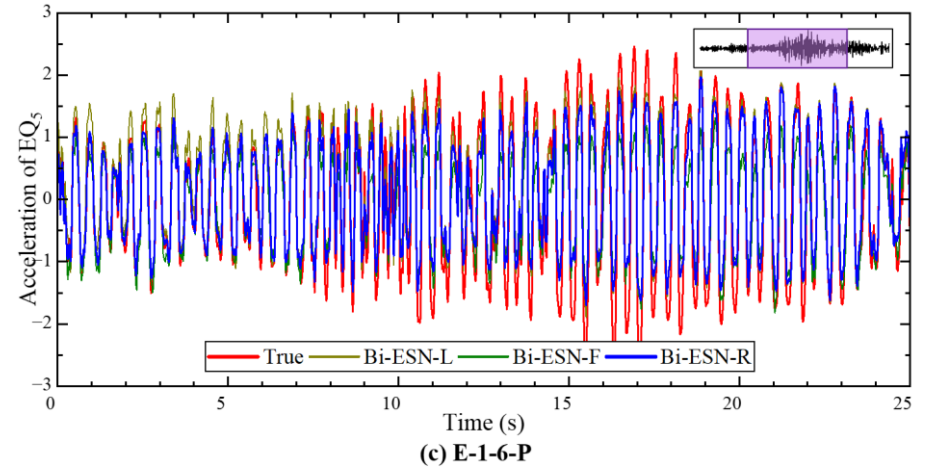
of earthquake-related information. Consequently, the MSE gradually increases to 0.1469. For signals normalized to an amplitude range of $[-2, 2]$, this error remains acceptable, demonstrating the Bi-ESN-R model's capability to handle moderately long-term data gaps.



(a) E-1-4-P



(b) E-1-5-P



(c) E-1-6-P

Fig. 4.15 Reconstruction of long-term lost data

The second consideration involves special gap locations, specifically for the $\boldsymbol{EQ}$ dataset. Beyond the scenario shown in Fig. 4.14 where gaps coincide with seismic peak phases, this section focuses on cases where missing segments occur either before earthquakes (influenced solely by early P-waves) or after earthquakes (free vibration with amplitude decay), labeled as E-1-1/2/3-I and E-1-1/2/3-E. Results for these cases are presented in Figs. 4.16 and 4.17.

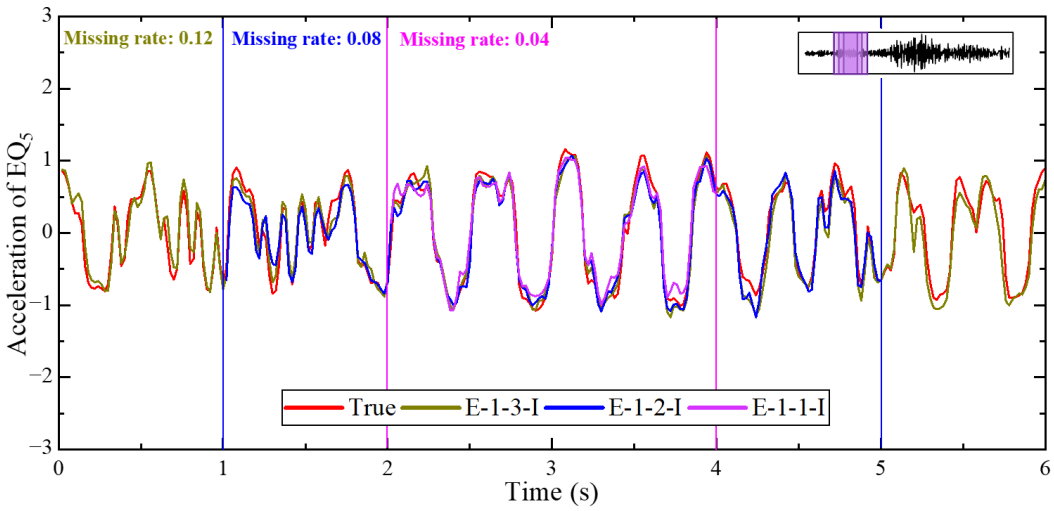


Fig. 4.16 Reconstruction of the starting phase of $\boldsymbol{EQ}$ dataset

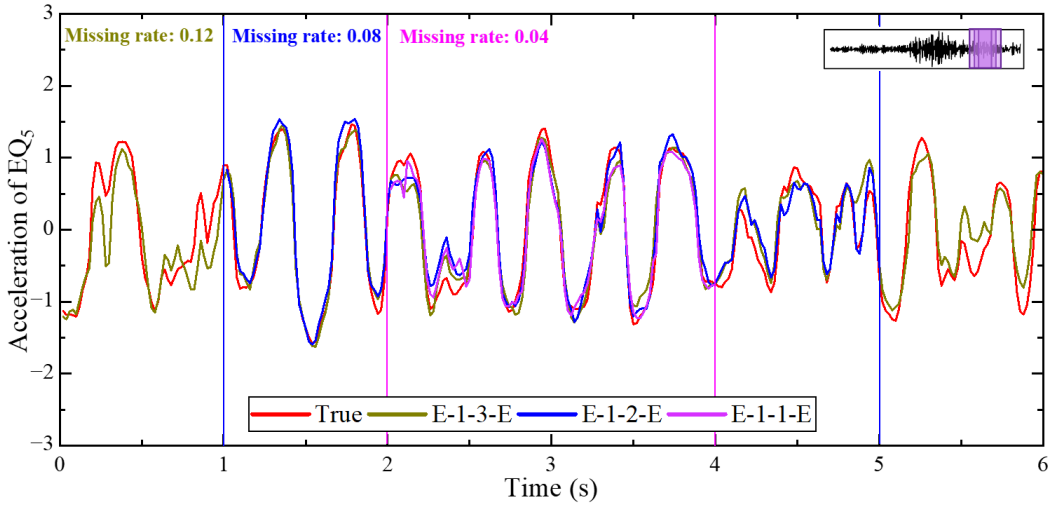


Fig. 4.17 Reconstruction of the ending phase of $\boldsymbol{EQ}$ dataset

When reconstructing gaps at different locations, the Bi-ESN-R model consistently maintains low MSE values (0.02 to 0.07) due to the availability of external synchronized signals and autoregressive modeling using adjacent in-channel data. Comparatively,

reconstruction errors are highest for "P" locations, followed by "E" and then "I". This hierarchy stems from increasing temporal correlation across the three segments. The "P" phase involves continuous high-level environmental excitation, causing signal dynamics to deviate from historical patterns. The "I" phase achieves exceptional reconstruction stability across missing rates, benefiting from low and stable external excitation amplitudes.

It becomes evident that using the entire available dataset for reconstruction is unnecessary, as distant signals contribute minimally to gap recovery. In practical SHM systems, where data segments flanking gaps are virtually infinite, this approach would prolong training time and negate the ESN's rapid-training advantage. To address this, we further analyze how reconstruction performance for benchmark cases (S-1-2 and E-1-2-P) varies with the length of training data on either side of gaps. This relationship, governed by the ESN's learning capacity and the similarity between training data and missing segments, is quantified using the ratio of gap duration to training data duration (R_{rt}). The correlations between R_{rt}, training duration, and MSE are illustrated in Fig. 4.18.

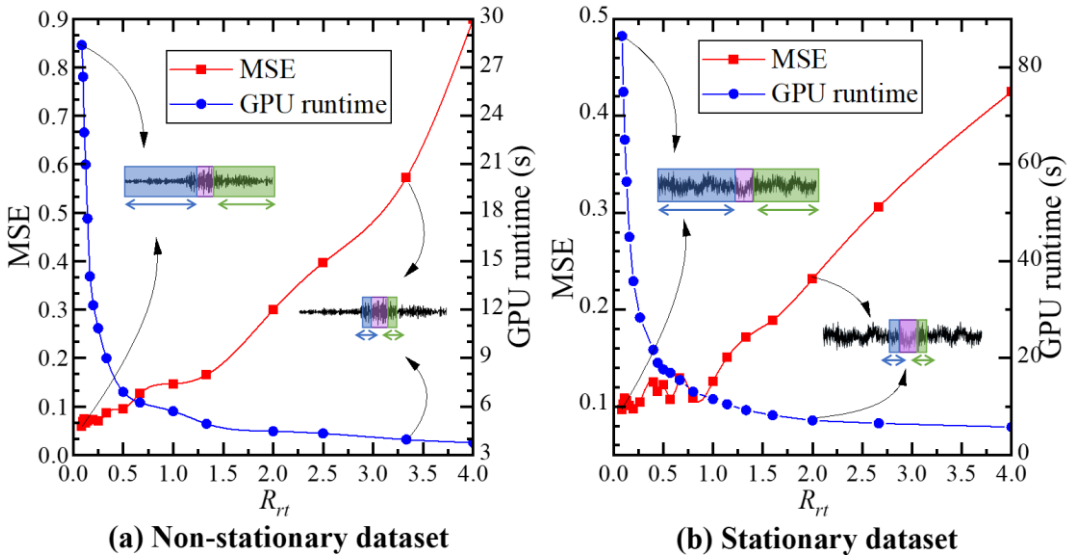


Fig. 4.18 Variation of model performance and training difficulty with available data amount

As R_{rt} decreases from its maximum value, the MSE initially enters a plateau phase. During this period, the accuracy of the Bi-ESN-R model remains stable despite reduced training data, with MSE fluctuating between 0.05 and 0.15. This indicates that the

"effective information" from data segments adjacent to the missing gap is sufficient to optimize and determine the network's weight parameters. However, this plateau phase ends for both $\boldsymbol{ST}$ and $\boldsymbol{EQ}$ datasets when R_{rt} approaches approximately 1.0. Beyond this point, the reconstruction capability of the Bi-ESN-R model deteriorates significantly, suggesting that the training data volume becomes insufficient for the model to learn meaningful patterns. Concurrently, as the input signal length decreases, the training time required for the Bi-ESN-R model also reduces, eventually approaching a minimum threshold dictated by data transfer latency between CPU and GPU. These observations collectively identify the computationally optimal training data length, balancing reconstruction accuracy with rapid deployment efficiency.

Another extreme scenario involves multi-channel data loss. Taking reconstruction with a missing rate of 0.08 as an example, cases S-5-2 and E-5-2-P represent scenarios where half of the channels are simultaneously impaired, while S-9-2 and E-9-2-P involve the loss of nine channels except for adjacent ones. The reconstruction MSEs for all missing channels in these four cases are shown in Fig. 4.19.

Fig. 4.19 reveals that spatial distribution between intact and missing channels influences reconstruction accuracy, even when the number of missing channels is identical. For instance, signals from $\boldsymbol{ST_5}$ and $\boldsymbol{ST_4}$, two accelerometers at the same height, exhibit high similarity, enabling accurate reconstruction of $\boldsymbol{ST_5}$ using intact data from $\boldsymbol{ST_4}$ despite other missing channels. In contrast, channels with weaker spatial correlations to $\boldsymbol{ST_4}$ show lower accuracy. Similarly, in cases E-5-2-P and S-5-2, the relatively low MSEs for $\boldsymbol{EQ_9}$ and $\boldsymbol{ST_9}$ stem from the availability of highly similar and fully intact signals from $\boldsymbol{EQ_{10}}$ and $\boldsymbol{ST_{10}}$. However, when $\boldsymbol{EQ_{10}}$ and $\boldsymbol{ST_{10}}$ are also missing (as in E-9-2-P and S-9-2), reconstruction errors surge. Notably, $\boldsymbol{EQ_1}$ consistently exhibits poor reconstruction accuracy due to its limited adjacency (only one neighboring channel) and unique structural dynamics: its accelerometer, positioned near the ground, captures vibration patterns closer to ground motion rather than structural responses, differing significantly from its adjacent channel $\boldsymbol{EQ_2}$.

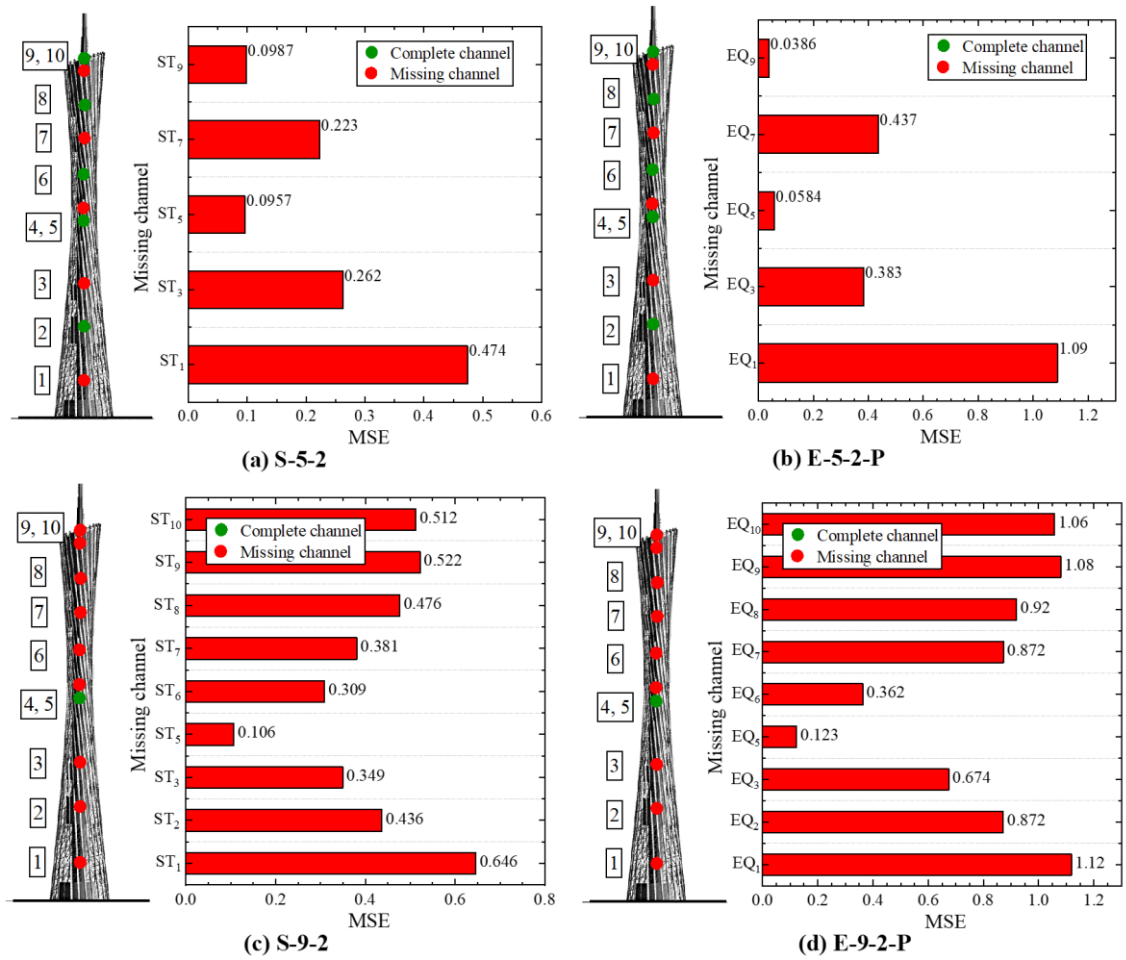


Fig. 4.19 Bi-ESN-R reconstruction results under multi-channel missing conditions

4.3.4.3 Hyperparameter study of Bi-ESN-R

Hyperparameter selection critically influences NN's performance. In the ESN framework, elements within the mapping matrices $\boldsymbol{W_{in}}$, $\boldsymbol{W_{res}}$, and $\boldsymbol{W_{back}}$ are predefined without training, effectively constituting hyperparameters. However, optimizing individual elements of these matrices is neither feasible nor necessary. As these matrices extract features through extensive nonlinear transformations, their collective behavior—rather than individual element values, which is governed by macroscopic hyperparameters. For $\boldsymbol{W_{res}}$, key macroscopic metrics include its spectral radius $\rho(\boldsymbol{W_{res}})$, sparsity $spar(\boldsymbol{W_{res}})$, and reservoir size N_{res}, where N_{res} is measured in units of T, and $spar(\boldsymbol{W_{res}})$ is expressed relative to N_{res}.

Three reservoir sizes were evaluated: $N_{res} = 2.0T$, $N_{res} = 3.0T$, and $N_{res} = 4.0T$. For

each configuration, grid searches were performed over $\rho(\boldsymbol{W_{res}}) \in [0.1, 12]$ and $spar(\boldsymbol{W_{res}}) \in [1.0N_{res}, 3.0N_{res}]$, with step sizes of 0.1 and $0.5N_{res}$, respectively. As exemplified by case E-1-2-P in Fig. 4.20, a distinct MSE "valley" emerges across all reservoir sizes during parameter variation, a phenomenon previously documented (Jiang & Lai 2019). Contrary to expectations, enlarging the reservoir narrows this trough rather than broadening it, suggesting that for fixed input signal lengths, the learnable internal features are inherently limited. Excessive reservoir scaling may therefore introduce irrelevant features, degrading accuracy. Furthermore, while sparsity minimally affects accuracy near the trough, smaller sparsity values help maintain lower MSE in regions with large spectral radii, particularly outside the optimal parameter range.

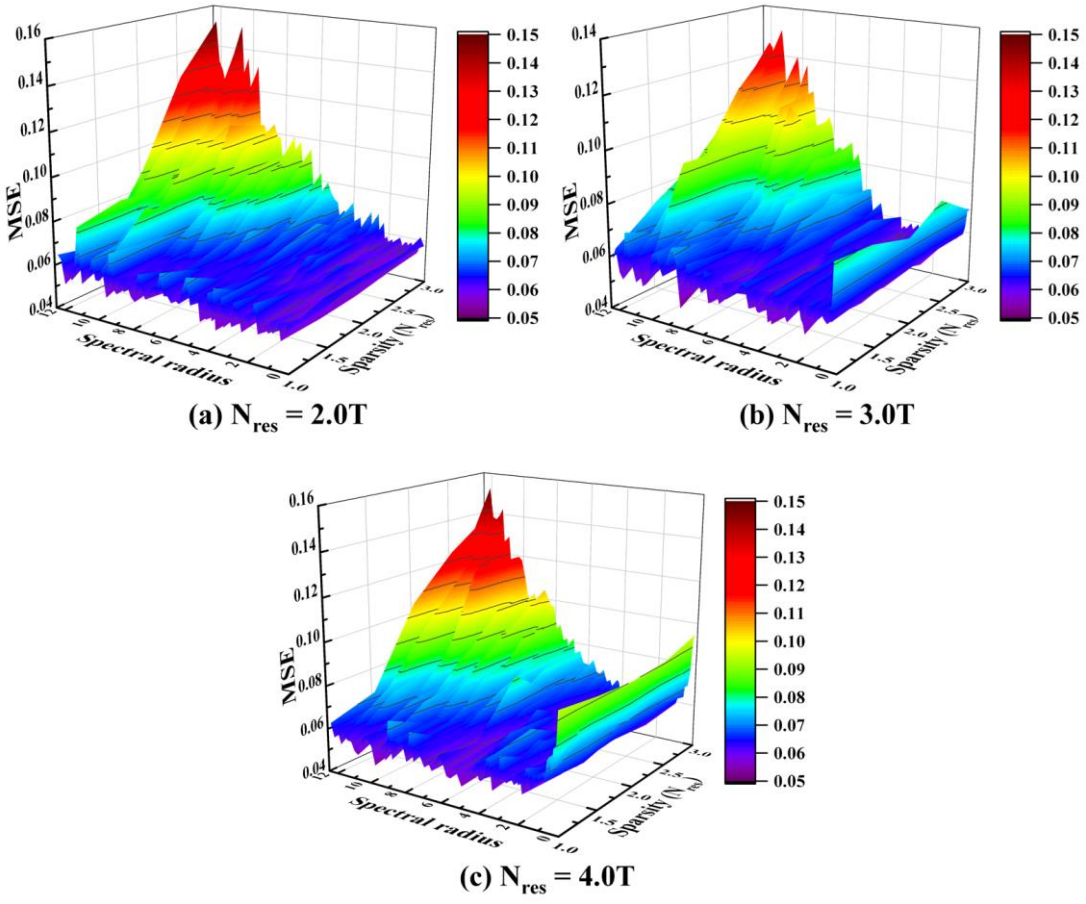


Fig. 4.20 Reservoir-related hyperparameter searching results

4.4 Outlier detection and recovery

Signals exhibiting significant deviations from true values due to external factors are

termed outliers. Unlike noise, outliers involve larger, potentially biased deviations. These anomalies often follow identifiable patterns (e.g., periodicity, linear trends) due to recurring external excitations. Segments with clustered outliers display distinct "abnormality" resembling structural damage or state changes targeted by monitoring algorithms, making outlier interference a critical challenge in structural state assessment. Automated outlier detection and recovery thus constitute an essential preprocessing step in SHM systems. However, traditional methods struggle to address the diversity of outliers, which vary in quantity, severity, and distribution patterns.

This section introduces a unified detection-recovery method applicable to arbitrary outliers, leveraging signal generation principles. The approach exploits the weak logical relationship between outlier-corrupted segments and normal data, using generation errors to identify anomalies. Since outliers arise from unpredictable external factors, they invariably exhibit high local generation errors, detectable LOF analysis. To enhance generation accuracy, a frequency decomposer is integrated before conventional time series-targeted NNs, forming a frequency-decomposed echo state network (FD-ESN). The proposed single-stage framework directly replaces detected outliers with generated values, enabling simultaneous identification and recovery.

4.4.1 FD-ESN and its frequency decomposer

The computational workflow of FD-ESN is illustrated in Fig. 4.21. The complete dataset after filtering and missing-data reconstruction, denoted as the raw data $\boldsymbol{Z}(t)$ to be processed in this section, may contain outliers across all channels. FD-ESN iteratively processes each channel. In a single iteration, the signal of the target channel is defined as the network output $\boldsymbol{Y} \in \mathbb{R}^{1\times T}$ (i.e., $N_{out} = 1$), while signals from the remaining channels serve as network inputs $\boldsymbol{X} \in \mathbb{R}^{N_{in}\times T}$, where $N_{in} = N_z - 1$. The FD-ESN generates an estimate $\widehat{\boldsymbol{Y}}$ of $\boldsymbol{Y}$ based on $\boldsymbol{X}$. This generative capability arises from the intrinsic mapping relationships between vibration signals measured at different structural locations, governed by the structure's vibrational modes. Since $\widehat{\boldsymbol{Y}}$ excludes contributions from outliers, the generation error sequence $\boldsymbol{e_y} = \widehat{\boldsymbol{Y}} - \boldsymbol{Y}$ exhibits low values near zero at

normal signal points but exceptionally high values at outlier-contaminated points. These sparse high-magnitude errors are amplified via the LOF, introduced in Section 4.4.2, enabling clear separation through simple thresholding. This procedure is repeated N_z times to cover all channels.

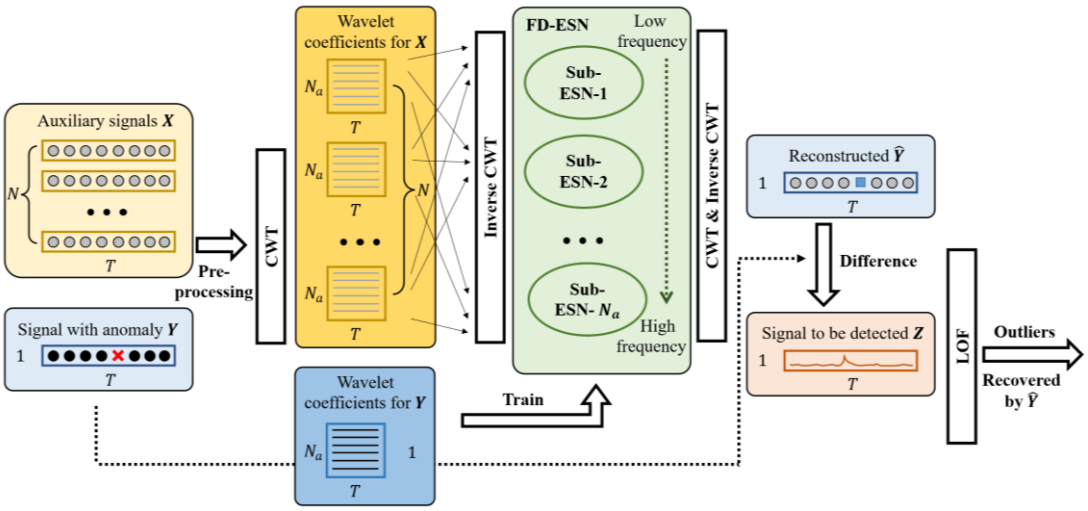


Fig. 4.21 FD-ESN

Learning direct mapping relationships between $\boldsymbol{X}$ and $\boldsymbol{Y}$ remains challenging for RNN architectures like ESN due to the complex components embedded within $\boldsymbol{X}$ and $\boldsymbol{Y}$. Training a single network to approximate such intricate mappings inherently requires trade-offs and balancing of learned patterns, inevitably introducing errors. Conversely, when input and output signals exhibit purer frequency-specific components, the required learning task simplifies, allowing the network to focus on fitting a limited set of mapping rules, thereby improving accuracy.

To address this, FD-ESN first decomposes signals (both $\boldsymbol{X}$ and $\boldsymbol{Y}$) into a set of frequency-specific components via frequency decomposition. Each component represents a narrow frequency band's amplitude. A series of parallel sub-networks are then introduced, with each subnetwork dedicated to learning the mapping relationship between a single pair of $\boldsymbol{X}$- and $\boldsymbol{Y}$-components. Since time-domain processing remains necessary post-decomposition, the continuous wavelet transform (CWT), which is a time-frequency analysis method, is employed for this decomposition.

CWT is a powerful tool that decomposes a signal into sub-signals with spectral

characteristics closely matching a set of designed wavelet bases through inner-product operations. Each wavelet basis has a central frequency f_c, defining its primary spectral coverage, and its function values decay rapidly away from the central point, preserving time-localized features without loss. Taking a single-channel $\boldsymbol{Y}$ as an example, its wavelet coefficients can be calculated as:

$$W_y(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} Y(t)\,\overline{\psi_{a,b}(t)}dt \tag{4.30}$$

where, a and b represent the scale factor and translation factor, respectively, which determine the time-frequency characteristics of the wavelet basis $\psi_{a,b}(t)$. Typically, the resolution and range of b align with the original signal, $b = 1,2,\dots,T$. The number of scale factors N_a, partitioned across the frequency range of interest, defines the granularity of decomposition and the number of resultant components. In this study, a decaying geometric progression is selected for the values of a, ensuring that the central frequencies of the converted wavelet bases fall within the primary frequency range of structural vibration signals. Letting r denote the common ratio, the sequence of aa can be expressed as ($i = 2,3,\dots,N_a$):

$$\left[\frac{f_c f_s}{f_e} \quad \dots \quad \frac{f_c f_s}{f_e} r^{(i-1)} \quad \dots \quad \frac{f_c f_s}{f_e} r^{(N_a-1)}\right] \tag{4.31}$$

where, f_e represents the upper limit of the vibration signal's frequency range, typically set to the structure's 10th order natural frequency or similar values. As recommended by the previous research (Kankanamge, Hu & Shao 2020), the Morlet wavelet (with a central frequency $f_c = 0.8125\,Hz$) is selected as the mother wavelet function for signals in SHM systems, defined as:

$$\psi(t) = e^{-t^2/2} e^{\mathrm{j}(2\pi f_c)t} \tag{4.32}$$

Thus, the wavelet coefficients of $\boldsymbol{Y}$ can be obtained as $\boldsymbol{W_y} \in \mathbb{R}^{N_a \times T}$, and those of $\boldsymbol{X}$ as $\boldsymbol{W_x} \in \mathbb{R}^{N_{in} \times N_a \times T}$. Since the wavelet coefficients retain time-domain characteristics, each row can be viewed as a sub-signal of the original signal. These coefficients are then grouped into N_a sets based on the scale factor a, expressed

as $[\boldsymbol{W}_{x,1}(a); \ldots; \boldsymbol{W}_{x,N_{in}}(a); \boldsymbol{W}_{y}(a)]$, where each group contains input-output signals with similar spectral characteristics. The first N_{in} elements within each group are collectively denoted as $\boldsymbol{W}_{x}(a) \in \mathbb{R}^{N_{in} \times T}$. Conversely, for wavelet coefficients corresponding to a specific a, their inverse continuous wavelet transform (ICWT) can be computed using Eq. (4.33).

$$Y_a(t) = \frac{1}{C_\psi} \int_0^T \frac{1}{a^2} W_y(a,b)\, \psi_{a,b}(t) db \tag{4.33}$$

where, C_Ψ is a constant determined by the wavelet function. By performing the inverse transform on each group of wavelet coefficients sharing the same a, N_a groups of $\boldsymbol{Y}_{a} \in \mathbb{R}^{1 \times T}$ and $\boldsymbol{X}_{a} \in \mathbb{R}^{N_{in} \times T}$ can be obtained. The FD-ESN framework is formed by establishing individual sub-ESN models for each of these N_a groups to learn the $N_{in} \times T \rightarrow 1 \times T$ mappings. The core principles of ESNs, as introduced in Section 4.3.1, are retained for each sub-network. However, since the entire sequence $\boldsymbol{Y}_{a}$ is generated directly, the autoregressive structure is deemed unnecessary and thus removed. To maintain the model's memory retention capability while leveraging dependencies on $\boldsymbol{X}_{a}$, residual connections, inspired by networks like residual network (ResNet) (He et al., 2016), are incorporated into the output layer, along with a backward time window for state retention. Consequently, the encoding and output processes of each sub-model in FD-ESN follow Eqs. (4.34) and (4.35).

$$\boldsymbol{R}_{a}(t) = (1-\alpha)\boldsymbol{R}_{a}(t-1) + \alpha f\left(\boldsymbol{W}_{in,a}\boldsymbol{X}_{a}(t) + \boldsymbol{W}_{res,a}\boldsymbol{R}_{a}(t-1) + \boldsymbol{b}_{r,a}\right) \tag{4.34}$$

$$\widehat{\boldsymbol{Y}}_{a}(t) = \boldsymbol{W}_{out,a}[\boldsymbol{X}_{a}(t-t_x); \ldots; \boldsymbol{X}_{a}(t); \boldsymbol{R}_{a}(t)] + \boldsymbol{b}_{y,a} \tag{4.35}$$

Here, t_x denotes the number of backward steps for residual connections in the output layer. Since each sub-network learns distinct features, their encoding matrices $\boldsymbol{W}_{in,a}$ and $\boldsymbol{W}_{res,a}$ should be generated separately, and their hyperparameters must be individually optimized following the methodology described in Section 4.3.4.3. The loss function for each sub-network remains the MSE, which will not be reiterated here. The parallel architecture of FD-ESN is illustrated in Fig. 4.22.

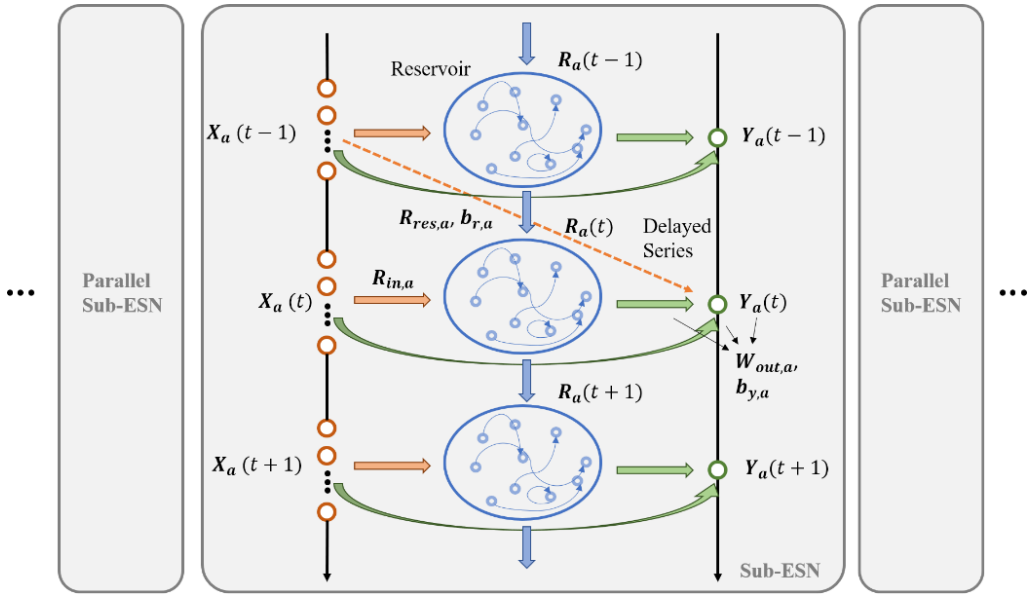


Fig. 4.22 Parallel sub-ESNs

4.4.2 LOF algorithm

The LOF method does not directly label data points as outliers; it fundamentally serves as an anomaly amplification technique. More objectively, there is no strict boundary between outliers and non-outliers, so nearly all outlier detection algorithms focus on quantifying and ranking anomaly scores, followed by applying criteria or simple thresholds to identify outliers. These criteria may be manually defined or learned through ML methods, but they are only relatively appropriate for the given dataset and cannot generalize to other problems.

Taking structural vibration signals as an example, these signals inherently fluctuate within a certain range and are significantly influenced by external loads. Outliers in such signals may not exhibit clear amplitude deviations compared to other temporal points. Thus, the FD-ESN model introduced in Section 4.4.1 essentially performs standardized normalization, transforming the easily confounded task of "detecting anomalous signal values" into one of "detecting anomalies in signal generation errors", which is predominantly influenced by non-structural information in the signals. Before applying LOF amplification, the generation error $\boldsymbol{e_y}$ already exhibits a pattern where most values fluctuate near zero, while a few shows larger magnitudes. When LOF is applied to $\boldsymbol{e_y}$,

the inherent distinction between outliers and non-outliers is further amplified, often to a degree where threshold-based separation becomes feasible, thereby avoiding contentious outlier criteria.

LOF is a method for detecting outliers by calculating the relative density of all samples (Alghushairy et al., 2020). When applied to time-series data (e.g., $\boldsymbol{e_y}$), the $\hbar$-distance of a data point $e_y(t)$ can be defined as the distance to its $\hbar$th nearest neighbor, denoted as $D_\hbar(t)$. The reachability distance from $e_y(t)$ to another data point $e_y(t')$ is then computed as:

$$D_{\hbar,r}(t,t') = max\big(D_\hbar(t), D(t,t')\big), \tag{4.36}$$

where, $(t,t') = \big((t-t')^2 + (e_y(t) - e_y(t'))^2\big)^{1/2}$ represents the Euclidean distance between two data points. For stationary signals, the temporal dimension can be ignored or assigned lower weight to exclude additional distances caused purely by temporal shifts. A simplified form $D(t,t') = \big|e_y(t) - e_y(t')\big|$ may be adopted. Clearly, $D(t,t')$ serves as the lower bound for $D_{\hbar,r}(t,t')$, and the fewer data points near $e_y(t)$, the larger $D_{\hbar,r}(t,t')$ becomes. Consequently, for the data point $e_y(t)$ at time t, the set of all timestamps t' satisfying $D_{\hbar,r}(t,t') \le D_\hbar(t)$ is defined as $\Omega_\hbar(t)$. The local reachability density (LRD) for the data point at time t can then be calculated as:

$$LRD_\hbar(t) = \frac{1}{E_{t'\in\Omega_\hbar(t)}\left(D_{\hbar,r}(t,t')\right)} \tag{4.37}$$

The $LRD_\hbar(t)$ is an absolute measure of reachability density. However, considering that different data categories (different types of anomalies) may exhibit varying densities, the final LOF is obtained by normalizing $LRD_\hbar(t)$ to the average level of its neighboring points, expressed as:

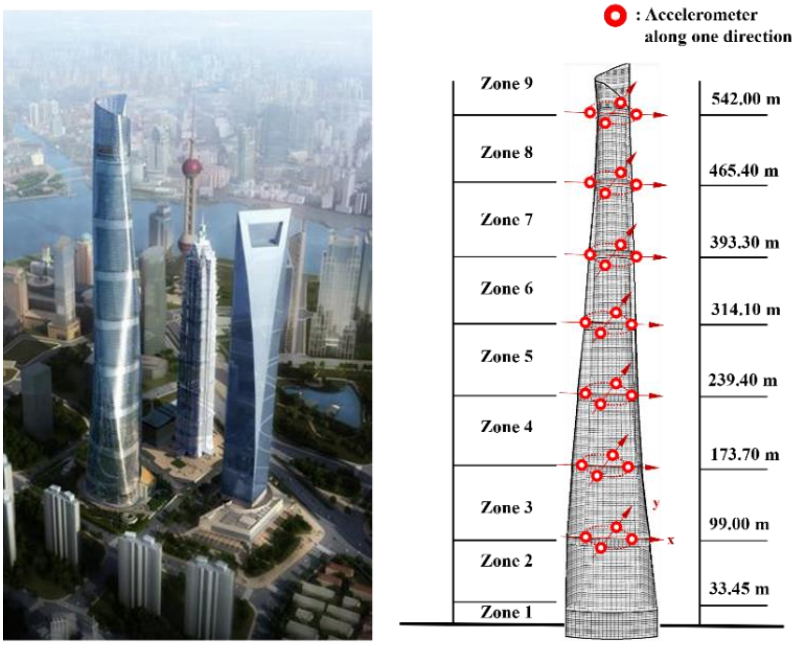
$$LOF_\hbar(t) = \frac{E_{t'\in\Omega_\hbar(t)}\big(LRD_\hbar(t')\big)}{LRD_\hbar(t)} \tag{4.38}$$

In summary, $LOF_\hbar(t)$ represents the relative density of neighboring data points around

the sample at timestep t. When this value is significantly higher than the average level, $(t, e_y(t))$ is identified as an anomaly. The calculation of $LOF_k(t)$ relies entirely on the sequence itself, requiring no manually labeled data for training, thereby enabling automatic detection without prior knowledge.

4.4.3 Application: Outlier detection of Shanghai Tower SHM data

The Shanghai Tower, located in the Lujiazui financial district of Shanghai, China, is the tallest skyscraper in China, with a structural height of 580 m and an architectural height of 632 m. Its structural system adopts a mega-frame-core tube-outrigger configuration. The square high-strength concrete core tube has a side length of 30 m, while the external mega-frame consists of eight mega-columns, four corner columns, and two-story-tall box-shaped ring spatial trusses. The outer frame and glass curtain wall form a triangular-like shape in plan and are anchored to the central concrete shear wall core tube via six cantilevered steel trusses. The outer frame tapers and twists approximately 120° vertically, creating a streamlined facade. The tower features eight strengthening layers internally, dividing the structure into nine zones.



(a) Shanghai Tower (b) Accelerometer layout

Fig. 4.23 Accelerometer layout of Shanghai Tower

Over 400 sensors are installed on the Shanghai Tower to measure variables such as acceleration, displacement, dip angle, strain, wind pressure, wind direction, and temperature. A hybrid wired and wireless data acquisition and transmission system relays electrical signals to storage terminals. Additionally, a signal preprocessing system integrated into the SHM system performs preliminary noise removal (Su et al., 2013) This section continues to use acceleration data for algorithm validation. For accelerometer placement, each of the four sides of the concrete core tube at every strengthening layer is equipped with an accelerometer (with a sampling frequency f_s of 100 Hz), as shown in Fig. 4.23.

Based on observations of massive data from the Shanghai Tower and other monitoring systems, five common types of outliers can be broadly categorized, as shown in Fig. 4.24. The most frequent type is spikes, characterized by extreme signal values (either maximum or minimum) at a single timestep. These spikes may occur individually, periodically, or entirely randomly. When outliers cluster, they form sustained elevations (plateaus) or complete signal absences. While reconstruction methods for missing signal segments were introduced in Section 4.3, this section focuses on automatically detecting such missing segments. Additionally, if abnormal states persist irreversibly, baseline shifting or linear and nonlinear trends may emerge. Finally, loose installations or external disturbances can introduce transient vibrational components, termed invaded components. To validate the effectiveness of FD-ESN, data from the Shanghai Tower on August 17, 2018, prior to the landfall of Typhoon Rumbia in Shanghai, is used. The dataset inherently contains some outliers from the measurement process. To comprehensively evaluate the method's capability in identifying all outlier types, artificial outliers and outlier segments are deliberately added. The outlier detection algorithm processes each channel individually. For simplicity, this section presents results for the west-side accelerometer in Direction-1 of the 4th strengthening layer as the target variable $\boldsymbol{Y}$, while other accelerometers in the same direction are treated as inputs $\boldsymbol{X}$ (i.e., $N_{in} = 13$). A 20,000-timestep segment (200 s) is selected: the first 10,000 timesteps (100 s) serve as the

training set to establish the mapping between $\boldsymbol{X}$ and $\boldsymbol{Y}$, while the remaining 10,000 timesteps (100 s) are augmented with various outlier types for network testing.

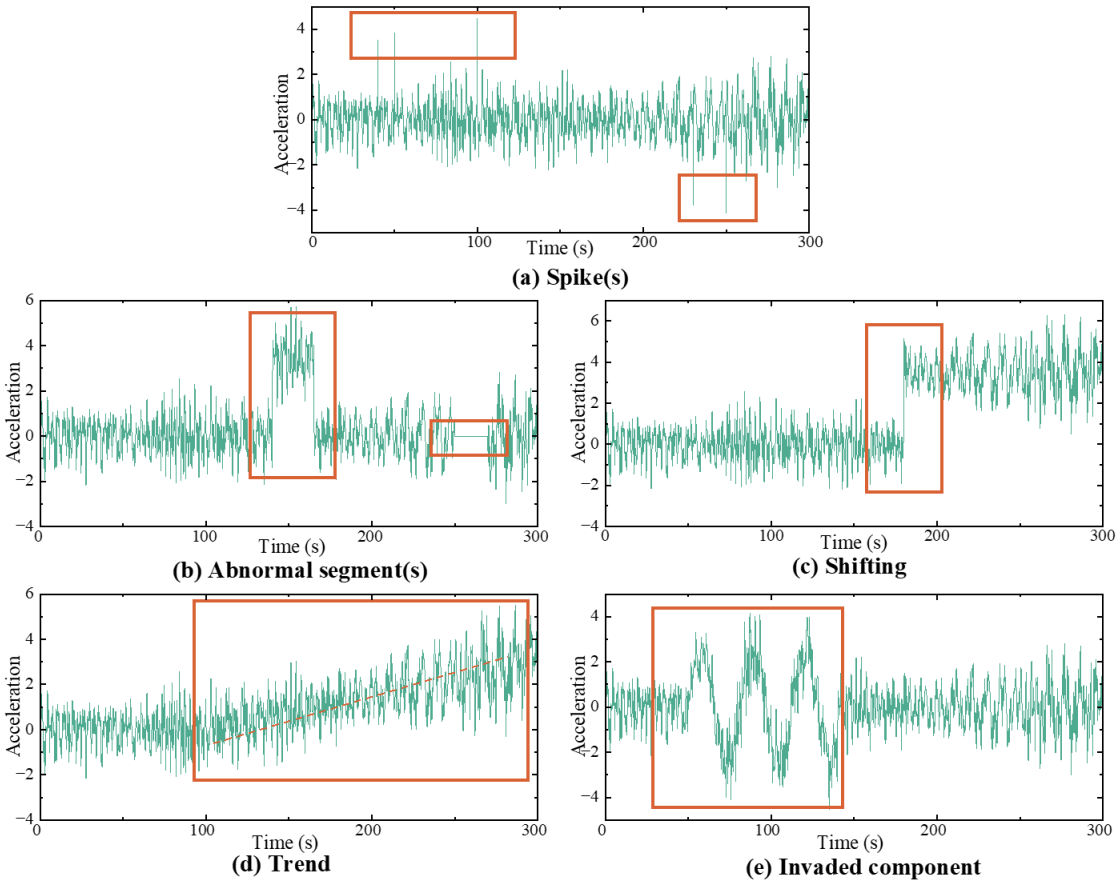


Fig. 4.24 Different types of outliers

4.4.4 Outlier detection results and discussion

This section discusses the detection capability of the FD-ESN method for outliers in practical SHM datasets. First, the generation accuracy of the generative network with a frequency decomposer is demonstrated and compared. Subsequently, the proposed model's ability to detect the five aforementioned outlier types and its influencing factors are analyzed in detail.

The primary task remains filtering as described in Section 4.2. Based on field measurements and FEM, the 1st and 10th natural frequencies of the Shanghai Tower are 0.107 Hz and 1.52 Hz, respectively. Therefore, when decomposing the tower's acceleration data using CWT, the cutoff frequency f_e is set to 2.0 Hz, the frequency

resolution to 0.02 Hz, and the r to 1.668. After balancing training costs and frequency-domain resolution, N_a is determined as 10, establishing 10 parallel sub-networks to generate target signal components with distinct frequencies. The hyperparameters of each sub-ESN are optimized using the method outlined in Section 4.3.4.3, with results listed in Table 4.3. All sub-ESNs adopt identical training strategies: a learning rate $lr = 0.0005$, $\eta = 0.001$, and training cycles lasting 300 epochs. Generation accuracy is evaluated using MSE.

Table 4.3 Hyperparameters of sub-models in FD-ESN

a	$spar(W_{res,a})$	$\rho(W_{res,a})$	α	a	$spar(W_{res,a})$	$\rho(W_{res,a})$	α
1	2.0	1.2	0.9	6	2.0	0.4	0.1
2	2.0	1.6	0.9	7	2.0	0.3	0.6
3	2.0	1.1	0.9	8	2.0	0.3	0.5
4	2.0	0.8	0.3	9	1.0	0.5	0.5
5	2.0	1.2	0.5	10	1.0	0.2	0.1

4.4.4.1 Generation accuracy of FD-ESN

To validate the effectiveness of the added frequency decomposer, a standard ESN model was established using the same hyperparameter selection method. This model directly accepts undecomposed data for training, and its generation results are compared with those of FD-ESN. During the training phase, FD-ESN already demonstrated advantages. The training loss-epoch curves of some representative sub-models are shown in Fig. 4.25. Due to the complexity of the mapping relationships to be learned, the training loss reduction rate of the standard ESN model ranked below average among all models, with its final training loss nearly three times higher than those of the sub-models. Each ESN module required under two minutes of training time, while RNN-based models demanded

at least half an hour, primarily because ESN's recurrent structure is shifted to a non-trainable reservoir. Furthermore, FD-ESN's inference time for a test data segment accounted for only 4% of its duration (100 s), enabling real-time generation.

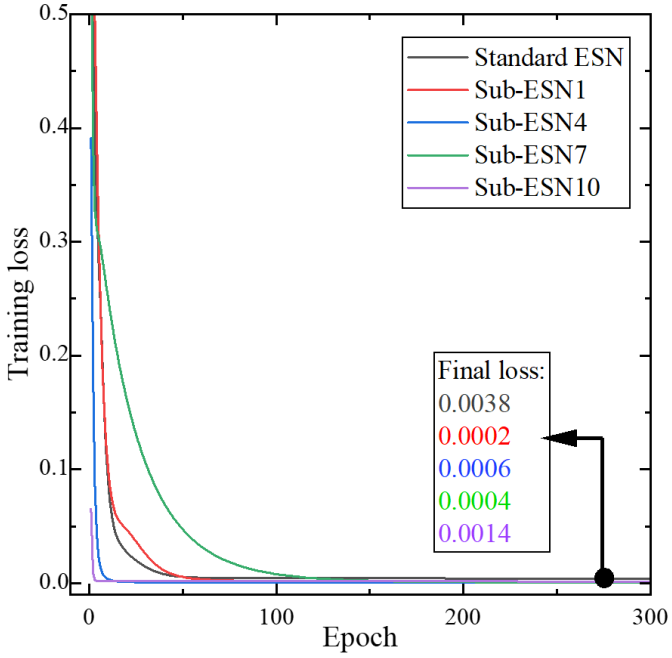


Fig. 4.25 Training loss-epoch curve of FD-ESN

Fig. 4.26(a) illustrates the characteristics of the target signal in both time and frequency domains, with its primary frequency components concentrated between 0.1~0.6 Hz and 1.5~1.7 Hz. As shown in Fig. 4.26(b), both the standard ESN and FD-ESN demonstrate superior performance in signal generation accuracy, with their outputs closely resembling the ground truth in both time and frequency domains. This observation suggests that the reservoir's extensive collection of simple state sequences can reproduce all common frequency components of signals collected by SHM systems. However, the proposed FD-ESN model exhibits notably higher precision than the standard ESN. Specifically, while the standard ESN-generated signals accurately capture the overall trend of the original signal, they show relatively larger errors at extremum points or abrupt transitions. For the same target signal segment, the FD-ESN achieves an MSE of 0.0015, approximately one-fifteenth of the standard ESN's MSE (0.0228). In the frequency domain, the FFT of FD-ESN's output aligns more closely with the original signal compared to the standard ESN's

results. Furthermore, FD-ESN achieves rational division of labor in learning frequency-domain information, avoiding the generation of spurious frequency components absent in the original signal. In contrast, when standard ESN outputs undergo low-pass filtering, frequency components above f_e exhibit residual artifact amplitudes rather than complete suppression to zero.

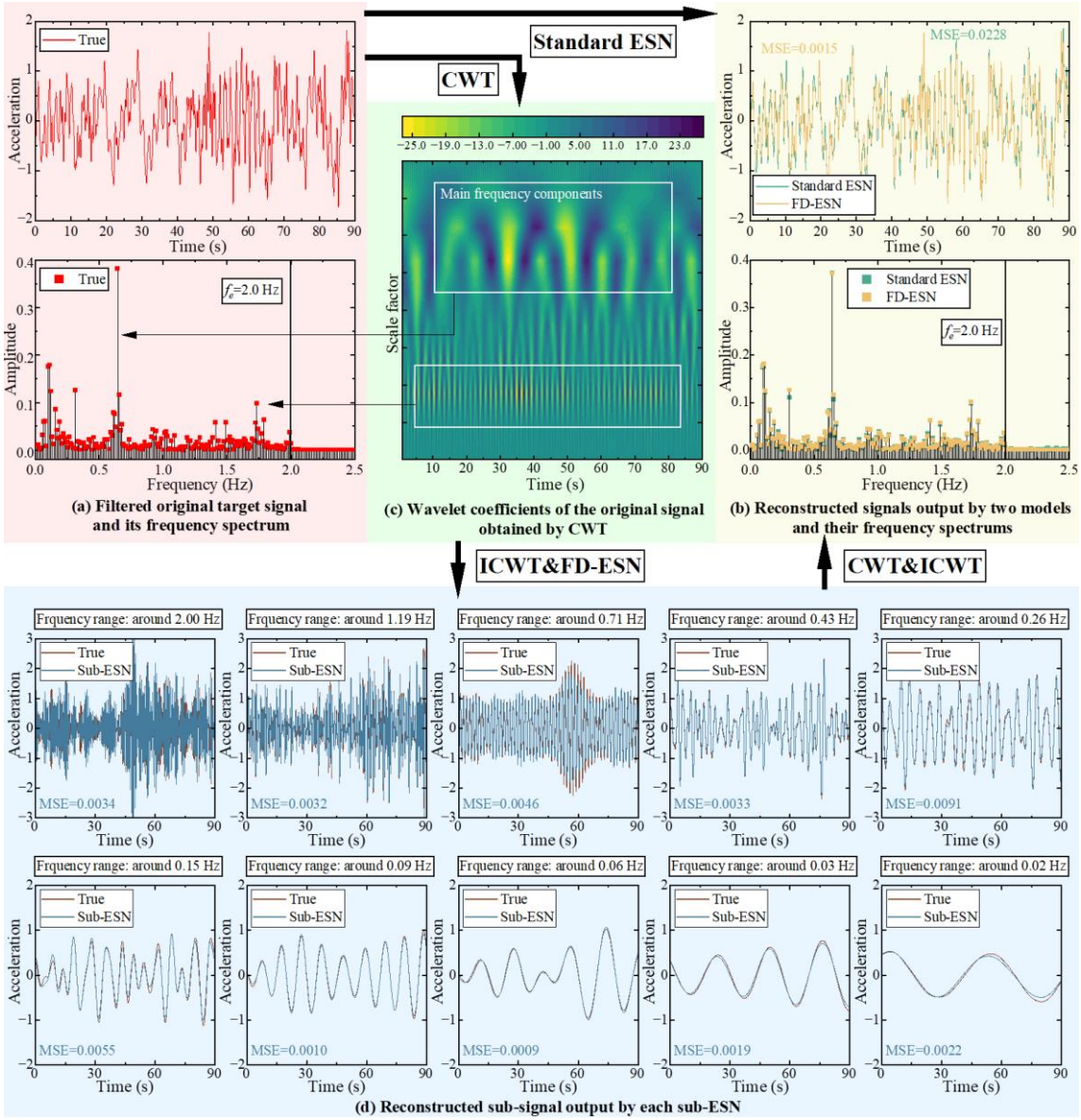


Fig. 4.26 Generation results of FD-ESN

Fig. 4.26(c) displays the CWT coefficients of the original signal. Two regions circled in the coefficient colormap represent the locations of the primary frequency components, which align well with the spectrum shown in Fig. 4.26(a). By splitting the coefficient

matrix according to a and performing ICWT separately, 10 sub-signals are reconstructed, as illustrated in the figure. From the original signal's waveform, it can be inferred that low-frequency components dominate the first 40 s and final 20 s, while high-frequency components prevail between the 40th and 70th seconds. This characteristic is clearly visible in the two subplots corresponding to the 1.19~2.00 Hz frequency range, where high-frequency components are concentrated in the middle and sparse at both ends. Additionally, the vibration amplitude gradually increases over time, a pattern observable in nearly all sub-signals. The reconstructed signals from the 10 sub-ESNs are also shown in the figure. Their MSE values range from 0.0009 to 0.0091, all significantly lower than the generation error of the standard ESN model. This suggests that ESN models excel at generating signals with simple frequency components, likely because the mapping relationships between such components are easier to learn, enabling direct correspondence in value variations. The analyzed signal segment was recorded before a typhoon event, exhibiting non-stationary characteristics. The results demonstrate that the FD-ESN model, through frequency decomposition and parallel processing, not only learns features of stationary time series but also effectively handles non-stationary time series.

4.4.4.2 Outlier detection results based on LOF

Fig. 4.27(a) illustrates the detection results of spike-like outliers. With the variance of the target signal normalized to 1.0 and the embedded spikes set to a base amplitude of 1.0, the parameter k in the LOF method is chosen as 10. The figure shows the detection process on a signal without outliers, including signal generation and $\boldsymbol{e_y}$ calculation (left), the $\boldsymbol{e_y}$ distance matrix (middle), and the derived LOF values (right). As analyzed earlier, $\boldsymbol{e_y}$ fluctuates around zero with no significant anomalies. Between the 40th and 70th seconds, larger $\boldsymbol{e_y}$ fluctuations emerge, reflected in both the distance matrix and $\boldsymbol{D_k}$ sequences. However, such error variations caused by normal signal fluctuations become negligible in $\boldsymbol{LRD_k}$ and $\boldsymbol{LOF_k}$ after scaling and reciprocal operations. This indicates that the LOF method does not misinterpret normal signal fluctuations as outliers.

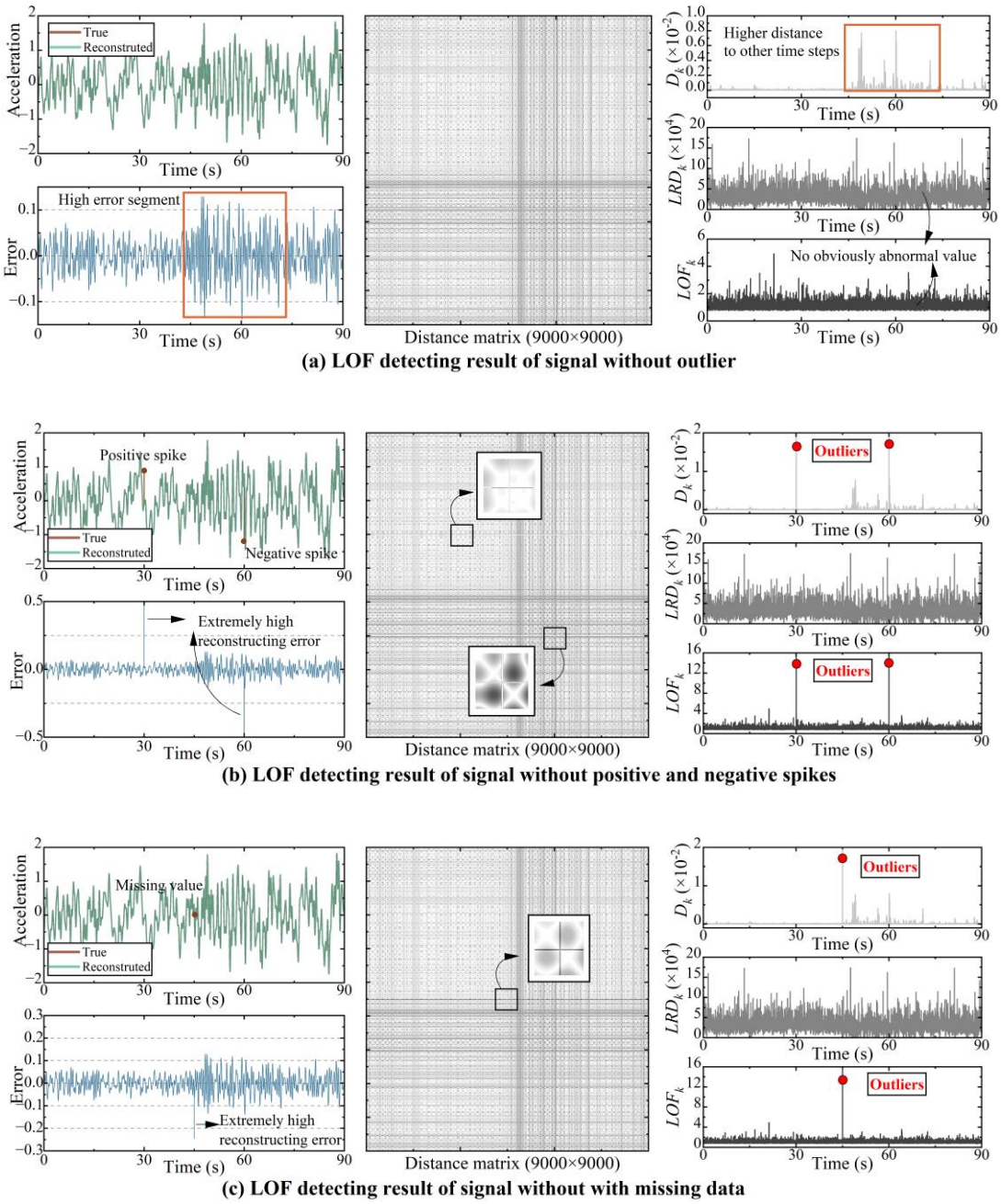


Fig. 4.27 Single spike detection results

Fig. 4.27(b) displays detection results after adding a positive spike at the 3000th time step and a negative spike at the 6000th time step. Despite the subtle spike magnitudes, the high-precision generation of FD-ESN still produces sharp deviations in $\boldsymbol{e_y}$ at these positions. Both spikes exhibit $LRD_k(t)$ and $LOF_k(t)$ values at least two orders of magnitude higher than normal time steps (anomalous $LOF_k(t)$ values: 513.84 and 401.45 vs. the maximum normal $LOF_k(t)$ of 4.93). This significant disparity enables

straightforward outlier identification by setting an appropriate threshold on the $\boldsymbol{LOF_k}$ sequence. Additionally, signal loss (zero value) can also be treated as a spike. Fig. 4.27(c) shows detection results after setting the 4500th time step value to zero. The corresponding $LOF_k(t)$ still shows a notable value (38.17). Since the original value at the 4500th time step was 0.24, reducing it to zero introduces a smaller error compared to spikes with magnitudes of 1.0, resulting in $LOF_k(t)$ being only one order of magnitude higher than normal values.

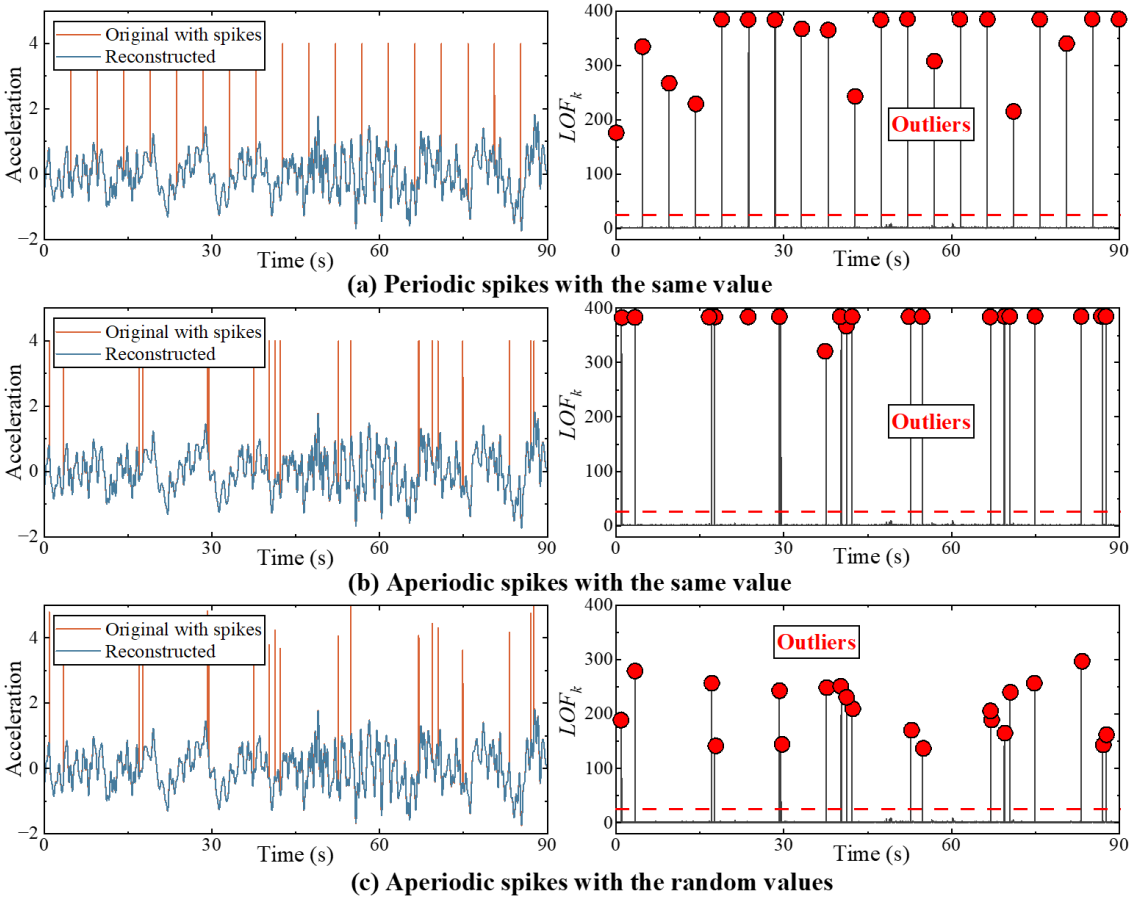
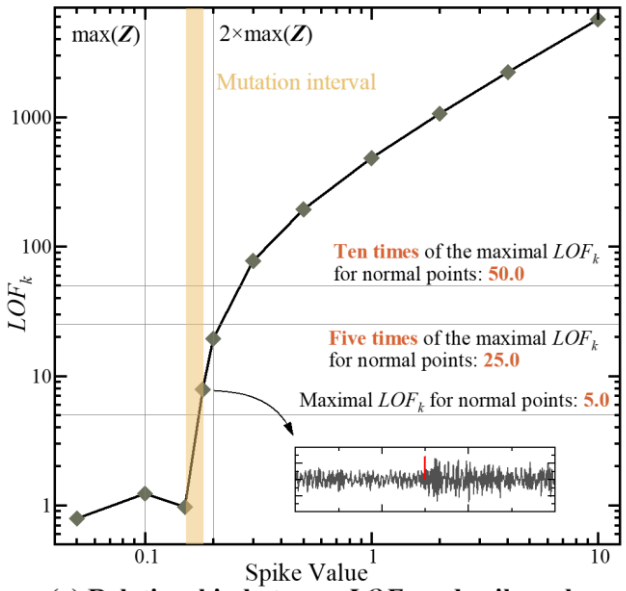


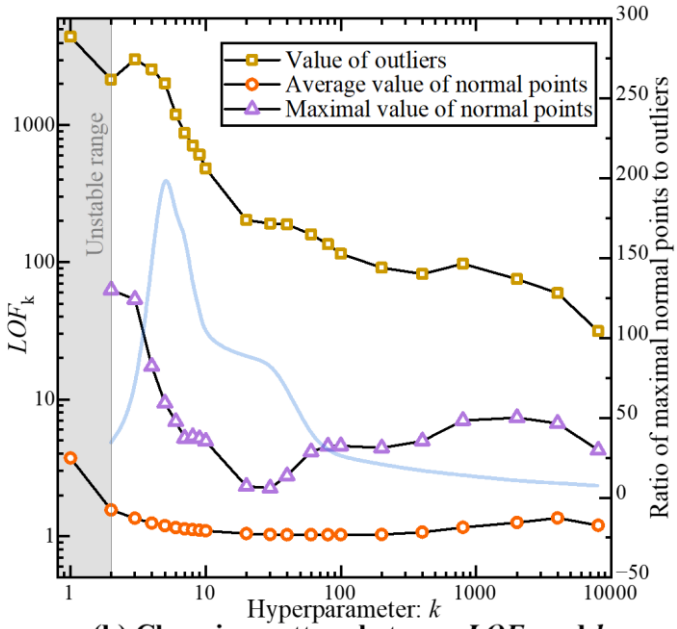
Fig. 4.28 Grouped spike detection results

Three additional spike detection tasks were designed: periodic equal-magnitude spike sequences (Fig. 4.28(a)), aperiodic equal-magnitude spike sequences (Fig. 4.28(b)), and aperiodic varying-magnitude spike sequences (Fig. 4.28(c)), each containing 20 spikes. For all three scenarios, the $LOF_k(t)$ values at outlier positions exceeded normal time steps by tens to hundreds of times, reaching up to 1600. A threshold of 20, approximately five times the maximum $LOF_k(t)$ of normal steps, effectively distinguished outliers. The

results in Fig. 4.28 confirm the method's capability to detect dense spikes with varying magnitudes.



(a) Relationship between LOF_k and spike value



(b) Changing pattern between LOF_k and k

Fig. 4.29 Parameter study of LOF-guided FD-ESN detecting spikes

The detection performance fundamentally depends on the disparity between outlier and normal $LOF_k(t)$ values. When spike magnitudes fall below a critical level, their induced deviations become comparable to generation errors, rendering $LOF_k(t)$ indistinguishable from normal fluctuations. For the single spike at the 4500th time step

(Fig. 4.29(a)), $LOF_{\hbar}(t)$ exhibits a transition zone where outlier detection becomes viable only when spike magnitudes exceed an upper threshold of 0.18 for this specific case. For spikes at other time steps, this critical upper bound fluctuates between the non-outlier series' $max(\boldsymbol{e_y})$ and $2max(\boldsymbol{e_y})$ under extreme scenarios. Consequently, a universal detectable deviation lower limit of 0.2 was established for normalized signals, determined by engineering requirements and FD-ESN performance. This threshold, already stricter than many existing methods, ensures reliable detection of all spikes exceeding this baseline magnitude. Actually, it has to be set manually. Thus, to avoid misleading, our method proposes to amplify the distance between the normal points and outliers (by FD-ESN calculating error sequence $\boldsymbol{e_y}$). When the distance is large enough, other simplex thresholds (e.g., 3σ) can also facilitate the method.

Fig. 4.29(b) illustrates the impact of the method's only hyperparameter $\hbar$, with the outlier fixed at the 4500th time step and a spike magnitude of 1.0 (exceeding the aforementioned transition zone). When $\hbar = 1$, although the outlier's $LOF_{\hbar}(t)$ reaches the maximum value in the sequence, the calculation of $LOF_{\hbar}(t)$ for all points becomes unstable. Thus, only values of $\hbar$ greater than 2 should be considered ideal. By definition, as $\hbar$ increases, $LOF_{\hbar}(t)$ decreases because a larger $\Omega_{\hbar}(t)$ diminishes the anomaly's distinctiveness. The optimal $\hbar$ should maximize the ratio between the outlier's $LOF_{\hbar}(t)$ and the maximum $LOF_{\hbar}(t)$ of normal points to facilitate clear distinction. Consequently, the recommended range for $\hbar$ in spike detection is determined as $[2,10]$. Its value is strongly affected by the scale of the data amount (length of one testing slice). The detection results are not sensitive to this hyperparameter. Within the reasonable range (around $0.01T$), the LOF will always work.

The second common type of outliers involves anomalous segments, such as sustained shifts or constant values. Despite differing data distributions between anomalous and normal segments, values within an anomalous segment are internally similar, potentially misclassified as homogeneous. From the LOF algorithm's perspective, the detection accuracy for continuous anomalous segments heavily depends on the segment's length

and the choice of k. If k is smaller than the anomaly length, $\Omega_k(t)$ will fail to include normal points, resulting in $LOF_k(t)$ values for anomalies not significantly exceeding normal points. For longer anomalous segments, smaller k values fail to bridge the anomaly's influence. To validate this, a 50-step elevated segment (shifted by 1.0) and a 50-step constant-zero segment were analyzed with k values of 20, 50, 100, and 200. The previously proposed threshold remains effective, and k must align with the above reasoning, as shown in Fig. 4.30.

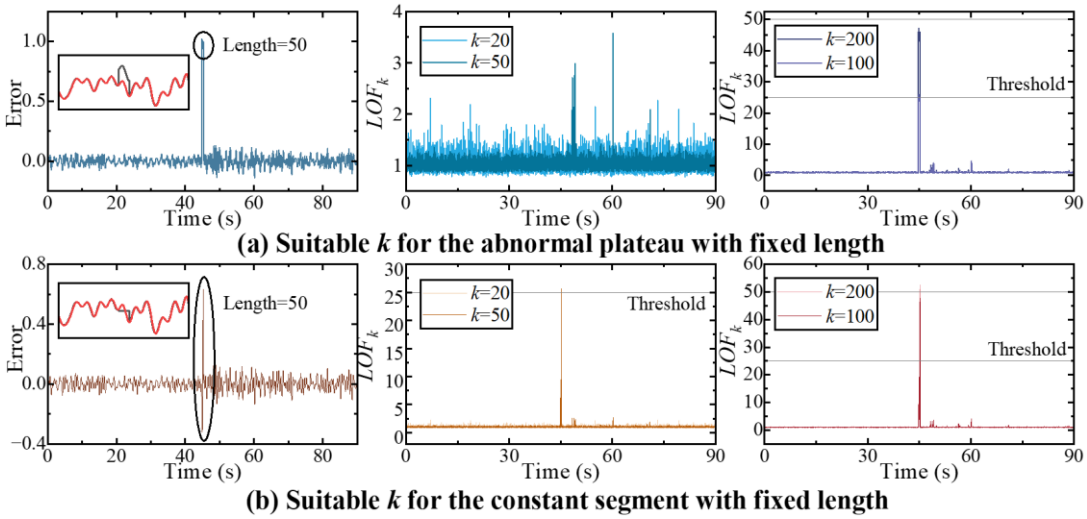
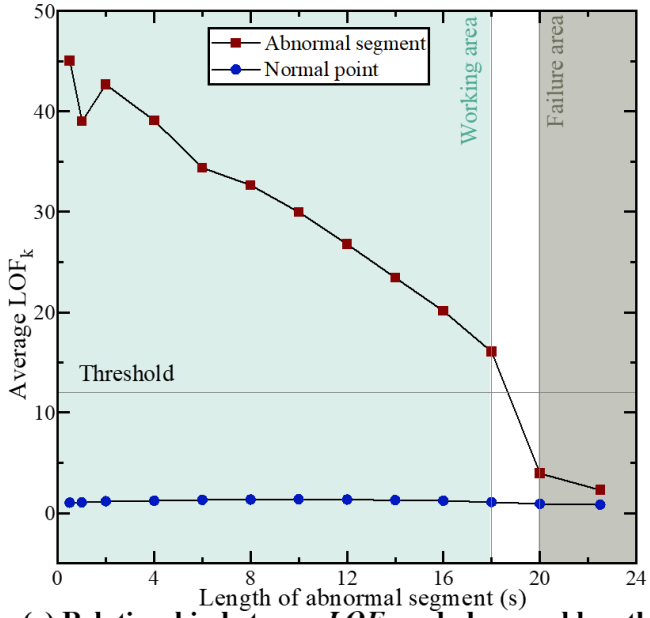


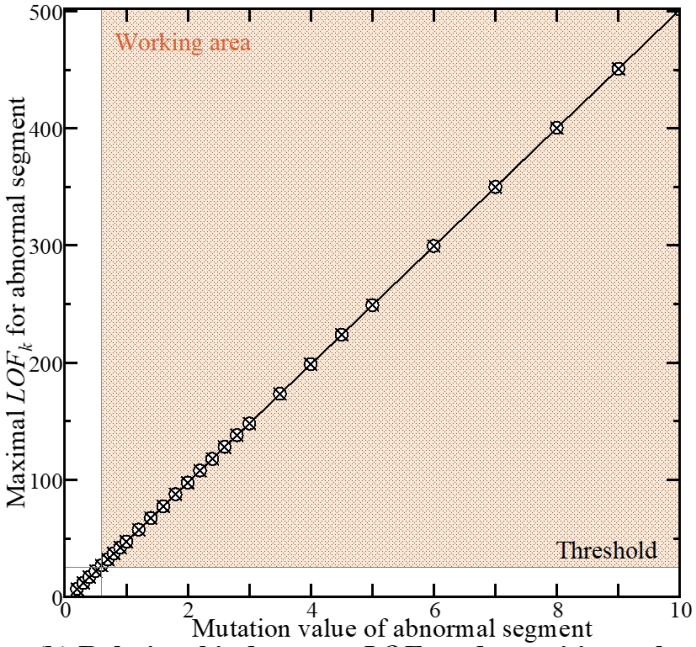
Fig. 4.30 Anomalous segment detection results

This implies a finite detectable maximum length for anomalous segments. When anomaly lengths exceed this limit, LOF fails due to comparable quantities of normal and anomalous samples. This limit is illustrated in Fig. 4.31(a), derived by increasing segment lengths with appropriate k. For anomalies shorter than 30% of the total signal duration, normal points dominate, maintaining $LOF_k(t)$ between 1.0~1.5 for normal points and above 10.0 for anomalies. However, as anomaly length increases, average $LOF_k(t)(t)$ for anomalies declines linearly, experiencing a sharp drop when anomalies exceed 70% of the signal. This decline arises because anomalous $LOF_k(t)$ is diluted by small intra-anomaly distances. Only by increasing k (expanding $\Omega_k(t)$ to include more normal points) can a clear distinction between normal and anomalous $LOF_k(t)$ be restored. The anomaly's severity determines the contribution of normal points within $\Omega_k(t)$, as shown in Fig. 4.31(b). Worsening anomalies linearly amplify $LOF_k(t)$ values across the

segment, but due to dilution from internal homogeneity, the minimum detectable deviation increases from 0.2 (spike detection) to approximately 0.5. This demonstrates that the selected k and thresholds enable nearly equivalent detection capabilities for diverse anomaly types.



(a) Relationship between LOF_k and abnormal length



(b) Relationship between LOF_k and mutaition value

Fig. 4.31 Parameter study of LOF-guided FD-ESN detecting anomalous segments

Compared to spikes and anomalous segments, trends, baseline drift, and intrusive signal components are more detrimental, as these outliers rarely self-resolve, leading to broader

impacts. Automatic outlier detection algorithms inherently assume sufficient normal samples to highlight anomalies. In continuous SHM systems, normal data volume is typically assured. Thus, a 10,000-step signal was used as the dataset, with the last 2000 steps assumed anomalous due to external influences (in real projects, normal-to-anomaly ratios are often higher).

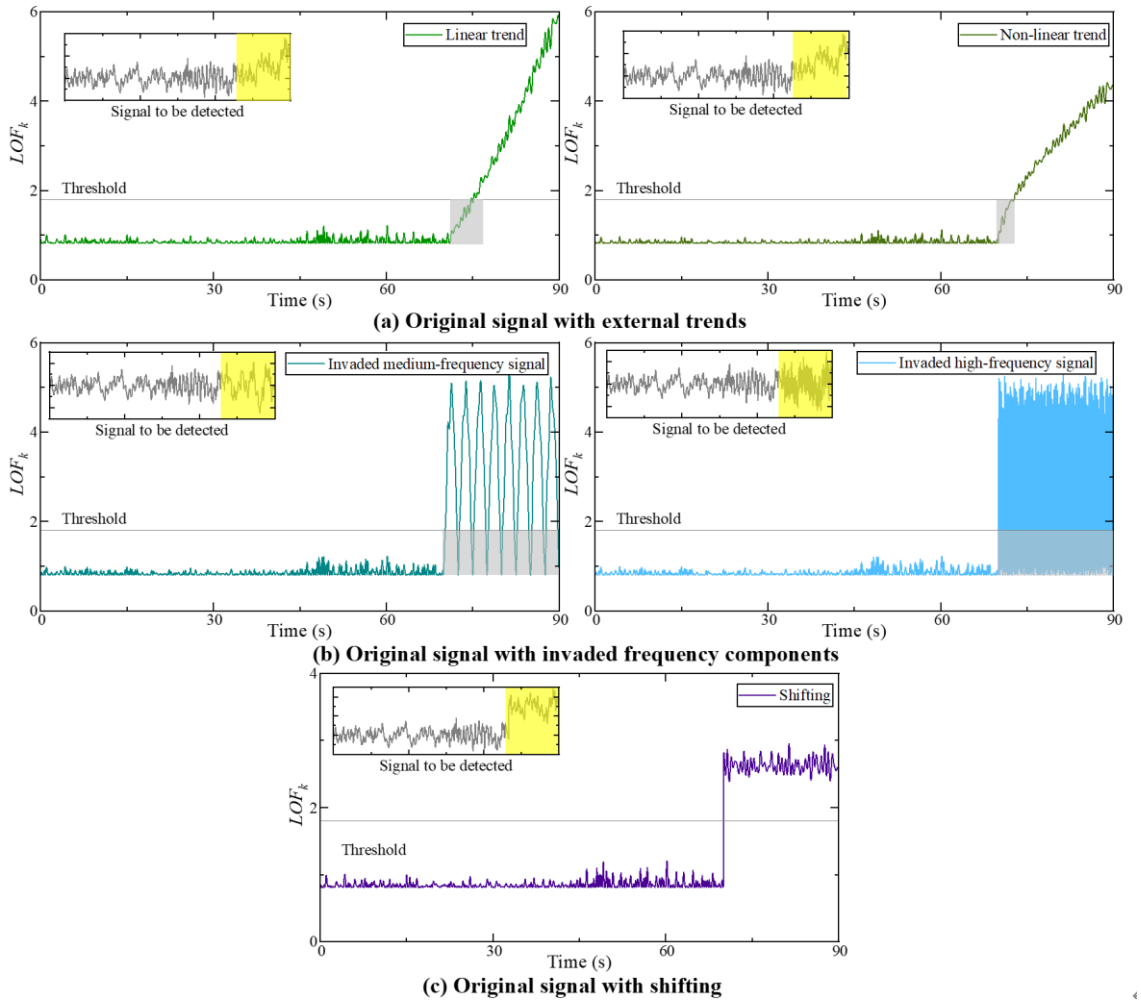


Fig. 4.32 Other types of outlier detection results

When linear or nonlinear trends are added to the signal (Fig. 4.32(a)), $LOF_k(t)$ increases following a pattern similar to $e_y(t)$, gradually amplifying the disparity from normal values. A well-calibrated threshold can separate most anomalies, though early-stage trend anomalies may remain undetected due to insufficient cumulative deviation. Intrusive signal components, viewed as a specific type of trend, pose challenges for filter-based methods when frequencies overlap with the original signal. However, the proposed

method detects intrusive components regardless of frequency. As shown in Fig. 4.32(b), mid-frequency (matching the signal's base frequency) and high-frequency (exceeding the low-pass filter's cutoff frequency) intrusions are identifiable, while low-frequency intrusions (overlapping with nonlinear trends) are omitted. Notably, intrusions with amplitudes near zero induce negligible deviations, evading detection. Such low-severity anomalies, if retained, cause minimal harm. For thorough outlier removal, iterative algorithm execution with progressively lowered thresholds, after replacing initially detected anomalies with reconstructed values, is recommended. Baseline drift, akin to platform segments, is detectable with appropriate k values (Fig. 4.32(c)).

4.5 Conclusions

This chapter proposes solutions for three common issues in raw SHM system data. For general noise, an IKF denoising method using ARIMAX as the state transition matrix is introduced; for data missing, a reconstruction model Bi-ESN-R based on a bidirectional and bilateral NN architecture is developed; and for outliers, an FD-ESN network embedding LOF for signal generation is proposed. The effectiveness, superiority, and robustness of these methods are validated using field data from the Guangzhou Tower and Shanghai Tower SHM systems, with key conclusions as follows:

(1) ARIMAX and standard time-series NNs can serve as state transition matrices for the IKF method. The optimal combination of model outputs $\widehat{\boldsymbol{Z}}^{+}(t)$ and $\widehat{\boldsymbol{Z}}^{-}(t)$ enables noise reduction while preserving valuable low- and high-frequency components.

(2) The proposed Bi-ESN-R model learns patterns from bidirectional historical and future input signals, outperforming traditional methods (ARIMAX and KF) and standard ESN in reconstructing signals of varying shapes, including free vibration, forced vibration, and amplitude-decaying recovery signals. Bi-ESN-R also handles long-term missing data if the ratio of known-to-unknown segment lengths is larger than 1.0. Sensor proximity between complete and reconstructed signals significantly affects accuracy, especially during synchronous data loss.

(3) By fitting and recombining signals within individual frequency bands, FD-ESN generates signals closer to ground truth (MSE reduced from 0.0228 to 0.0015) compared to standard ESN. Analyzing the generation error $\boldsymbol{e_y}$ isolates outliers while mitigating signal variation patterns, and LOF further amplifies anomaly severity. With appropriate k and thresholds, spikes of varying magnitudes, number, and distributions are detectable. However, increasing continuous similar outliers linearly dilutes $LOF_k(t)$, imposing a detection upper limit where anomalies must occupy less than 25% of the total signal length. Similar requirements apply to diagnosing signals with linear or nonlinear trends, intrusive frequency components, and baseline drift, except for early-stage trend accumulation or intrusive oscillations near zero, where method accuracy remains acceptable.

For application, the three components of the denoising strategy can be used as a pipeline model. The raw data is firstly filtered through ARIMAX+IKF to remove the common noise, while also reducing the difficulty of downstream tasks by purifying. Afterwards, FD-ESN+LOF is applied to identify and remove the outliers. During this process, the missing segments are also discovered, which can be further reconstructed by Bi-ESN-R.

DATA INFERENCE BASED ON TIME SERIES-TARGETED NEURAL NETWORKS

5.1 Introduction

Data inference, an implicitly defined module in the post-processing chain of SHM systems, addresses tasks where even properly denoised, reconstructed, and corrected monitoring data cannot be directly utilized for final decision-making. In practical structural operations, critical decisions often require deriving actionable insights from existing data before final judgments. For example, risk warnings typically rely on predicting key physical quantities or indicators, a form of data inference that estimates future values based on historical data. Additionally, some critical parameters may lack direct sensor measurements due to safety or feasibility constraints, necessitating indirect inference from spatially or functionally related data. These two inference types inherently involve unavoidable estimation due to data inaccessibility. A third common inference task in SHM post-processing is “cross-verification” of already measured data, where redundant inference aims to detect damage by comparing inferred results with measurements. Such tasks essentially solve physical systems, such as dynamic forward or inverse problems (e.g., estimating responses from measured excitations and contrasting them with actual responses).

This chapter proposes time series-targeted NN-based solutions for the three inference tasks, including a prediction algorithm using memory-augmented LSTM networks for relatively long-term temporal forecasting; a Transformer-based method with modified positional embedding for unmeasured and complex variable mapping; and a time series-targeted NN architecture modified by a general generalization performance enhancing module as a solver for complex nonlinear dynamic systems. All methods are validated against conventional approaches using real-world structural or dynamic system datasets, while their robustness is also determined.

5.2 Future data prediction

Time series prediction has been extensively researched, with signal forecasting methods playing a critical role in structural intelligent operations such as risk early warning. The fundamental logic of all prediction algorithms involves modeling signal fluctuations using current and historical data, estimating future values (single-step or multi-step), and iteratively advancing the prediction horizon. While NN-based approaches generally outperform traditional methods, their efficacy remains constrained by memory capacity. For mainstream RNN architectures, long-term dependency problems restrict memory capacity, often limiting inferences to recent data points. Information from distant time steps becomes noise due to gradient vanishing during training.

This section proposes a novel prediction model based on enhancements to LSTM networks. Although LSTM itself improves memory retention over vanilla RNNs through gating mechanisms, this study further augments the forget gate with an attention mechanism, creating a memory-enhanced long short-term memory (eLSTM) architecture. Recognizing that prediction tasks often involve limited input channels (even single-channel signals), this chapter introduces a dynamic echo state network (dESN) as a training-free adaptive feature encoder preceding the eLSTM. This hybrid architecture, termed dESN-eLSTM, aims to enhance feature learnability from low-dimensional inputs. The remainder of this section evaluates its predictive performance against state-of-the-art benchmarks.

5.2.1 Long-term memory and eLSTM

To illustrate the memory limitations inherent in RNN-based networks, this section revisits the basic RNN model introduced in Section 3.2.1, simplified to a single-layer architecture. Here, the hidden state $\boldsymbol{H}(t)$ is encoded through a fully connected linear layer with parameters $\boldsymbol{W}$, processing both the current input $\boldsymbol{X}(t)$ and the previous state $\boldsymbol{H}(t-1)$. This minimalist RNN unit encapsulates the quintessential recursive structure of such networks. During the training of $\boldsymbol{W}$, the core task involves computing the partial

derivatives of the network loss $L\left(\boldsymbol{Y}(t), \widehat{\boldsymbol{Y}}(t)\right)$ (abbreviated as $L(t)$ with respect to each parameter:

$$\frac{\partial L(t)}{\partial \boldsymbol{W}} = \sum_{\tau=1}^{t} \frac{\partial^{+} \boldsymbol{H}(\tau)}{\partial \boldsymbol{W}} \frac{\partial L(t)}{\partial \boldsymbol{H}(\tau)} = \sum_{k=1}^{t} \frac{\partial^{+} \boldsymbol{H}(\tau)}{\partial \boldsymbol{W}} \delta_{t,\tau} \cong \sum_{k=1}^{t} \frac{\partial^{+} \boldsymbol{H}(\tau)}{\partial \boldsymbol{W}} \varsigma^{t-\tau} \delta_{t,t} \tag{5.1}$$

where, ∂^{+} denotes the direct partial derivative without considering the chain rule, and $\delta_{t,\tau}$ represents the derivative of the loss at time t with respect to the hidden layer's net input at time τ, which can be approximated as a constant ς. Eq. (5.1) reveals that the partial derivative of the loss for any time step must be computed recursively back to the initial step due to the network's recurrent structure. Consequently, the final partial derivative becomes a summation of t terms, where the τth term corresponds to the derivative at step τ. Crucially, terms farther from t inherit a high-order multiplicative factor $\varsigma^{t-\tau}$. If $\varsigma > 1$, derivatives risk gradient explosion, causing numerical instability and training failure; If $\varsigma < 1$ (can be realized by setting proper activation function), gradient vanishing occurs, where distant terms decay to near-zero despite meaningful direct derivatives. Thus, in trained RNNs, parameters $\boldsymbol{W}$ effectively disregard long-term historical inputs, rendering distant temporal contributions to $\boldsymbol{H}(t)$ negligible, which is the fundamental cause of limited memory capacity.

To enhance memory retention, researchers developed the LSTM network. Each LSTM unit introduces a cell state $\boldsymbol{C} \in \mathbb{R}^{N_h \times T}$, propagated alongside $\boldsymbol{H}$ across time. LSTMs employ three gating mechanisms (forget gate, input gate, and output gate) to regulate information flow. These gates are N_h-dimensional vectors with elements in $[0,1]$, selectively filtering information via element-wise multiplication with states. The forget gate $\boldsymbol{O}(t)$ controls how much of the current cell state $\boldsymbol{C}(t)$ is retained in the hidden state $\boldsymbol{H}(t)$:

$$\boldsymbol{H}(t) = \boldsymbol{O}(t) \odot f\left(\boldsymbol{C}(t)\right) \tag{5.2}$$

While standard RNNs directly compute the current state $\boldsymbol{H}(t)$ from the input $\boldsymbol{X}(t)$ and previous state $\boldsymbol{H}(t-1)$, LSTMs first generate a candidate cell state $\widetilde{\boldsymbol{C}}(t)$ through Eq.

(5.3), which serves as an intermediate memory representation before gating operations.

$$\widetilde{\boldsymbol{C}}(t) = f(\boldsymbol{W_c}[\boldsymbol{H}(t-1); \boldsymbol{X}(t)] + \boldsymbol{b_c}) \tag{5.3}$$

where, $\boldsymbol{W_c} \in \mathbb{R}^{N_h \times (N_h + N_{in})}$ is the weight matrix, and $\boldsymbol{b_c}$ is the bias term. The final cell state $\boldsymbol{C}(t)$ integrates two information components through gated selection: one is the filtered historical information (the previous cell state $\boldsymbol{C}(t-1)$ modulated by the forget gate $\boldsymbol{F}(t)$; and another is the selected candidate memory (the cell state $\widetilde{\boldsymbol{C}}(t)$ regulated by the input gate $\boldsymbol{I}(t)$). This integration is formally expressed as:

$$\boldsymbol{C}(t) = \boldsymbol{F}(t) \odot \boldsymbol{C}(t-1) + \boldsymbol{I}(t) \odot \widetilde{\boldsymbol{C}}(t) \tag{5.4}$$

Thus, it becomes evident that the introduction of the cell state $\boldsymbol{C}(t)$ provides the model with an additional dimension and mechanism for information propagation. Functionally, $\boldsymbol{C}(t)$ acts as the model's memory, preserving comprehensive historical information, while the hidden state $\boldsymbol{H}(t)$ resembles the model's "response" derived from both memory and current inputs.

The three gating mechanisms responsible for generating $\boldsymbol{H}(t)$ and $\boldsymbol{C}(t)$ must remain dynamic to adapt to evolving temporal patterns. Conventionally, these gates are computed via trainable linear transformations of the state and input, as formalized in Eqs. (5.5), (5.6), and (5.7):

$$\boldsymbol{O}(t) = f(\boldsymbol{W_o}[\boldsymbol{H}(t-1); \boldsymbol{X}(t)] + \boldsymbol{b_o}) \tag{5.5}$$

$$\boldsymbol{F}(t) = f\big(\boldsymbol{W_f}[\boldsymbol{H}(t-1); \boldsymbol{X}(t)] + \boldsymbol{b_f}\big) \tag{5.6}$$

$$\boldsymbol{I}(t) = f(\boldsymbol{W_i}[\boldsymbol{H}(t-1); \boldsymbol{X}(t)] + \boldsymbol{b_i}) \tag{5.7}$$

where, $\boldsymbol{W_o}$, $\boldsymbol{W_f}$, and $\boldsymbol{W_i}$ are weight matrices with dimensions $\mathbb{R}^{N_h \times N_{in}}$, $\boldsymbol{b_o}$, $\boldsymbol{b_f}$, and $\boldsymbol{b_i}$ are bias terms. While the LSTM architecture incorporates a larger number of strategically distributed linear layers (as shown in Fig. 5.1), increasing model complexity and fitting capacity, its core recursive encoding mechanism remains unchanged. Consequently, long-term dependency challenges persist, with LSTMs retaining only short-term memory capabilities.

Recent studies (Zhao et al., 2020) reveal that the forget gate $\boldsymbol{F}(t)$ governs the retention rate of historical information, thereby dictating LSTM's memory capacity. Building on this insight, this study suggests modifying the forget gate by integrating an attention-like mechanism. Instead of solely relying on the immediate prior cell state $\boldsymbol{C}(t-1)$, our enhanced forget gate incorporates a sequence of Γ historical cell states. The retained historical information is now a weighted sum of these states, where the combination coefficients are determined by the similarity between $\widetilde{\boldsymbol{C}}(t)$ and each historical state. This transforms the first term on the right-hand side of Eq. (5.4) into:

$$\sum_{\tau=1}^{\Gamma} \boldsymbol{F}_{\tau}(t) \odot \boldsymbol{C}(t-\tau) \tag{5.8}$$

where, , $\boldsymbol{F}_{\tau}(t)$ represents a term in the combinatorial coefficients of the Γ groups, which was previously considered as a set of non-trainable parameters derived from Γ in prior studies. In this work, the author incorporates it into the set of parameters to be determined, where its value is to be obtained via a linear mapping. Consequently, the network is upgraded to eLSTM, and the corresponding Eq. (5.6) is revised to the following form:

$$\boldsymbol{F}_{\tau}(t) = \prod_{t=0}^{\tau-1} \frac{t - \boldsymbol{W}_{\boldsymbol{f}}[\widetilde{\boldsymbol{C}}(t); \boldsymbol{C}(t-\tau); \boldsymbol{X}(t); \boldsymbol{X}(t-\tau)]}{t+1} \tag{5.9}$$

Specifically, the value of $\boldsymbol{F}_{\tau}(t)$ decays continuously as the retrospective time increases, achieved primarily by multiplying a fraction (less than 1) $(\tau-1)$ times. The negative term in the numerator of this fraction governs the decay rate, which is derived from a linear layer parameterized by $\boldsymbol{W}_{\mathbf{f}} \in \mathbb{R}^{N_h \times 2(N_h+N_{in})}$. The inputs to this linear layer are the hidden states and inputs at both the current time step and the τth retrospective step. The output layer of the eLSTM network aligns with that of a standard RNN, and its loss function is constructed using the MSE, both of which have been detailed in preceding chapters and are not reiterated here. The architecture of the eLSTM network is illustrated in Fig. 5.2.

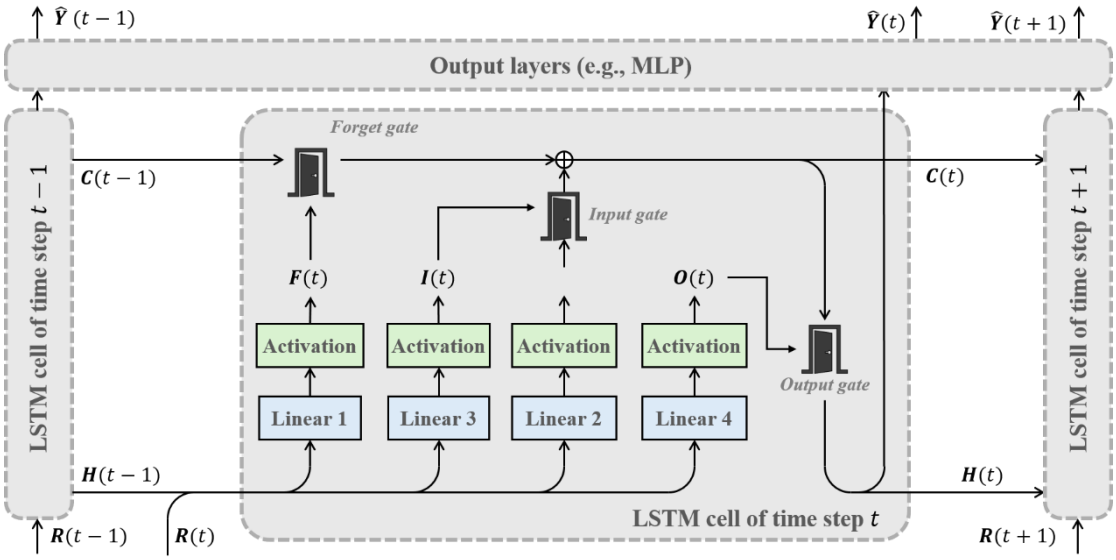


Fig. 5.1 Standard LSTM

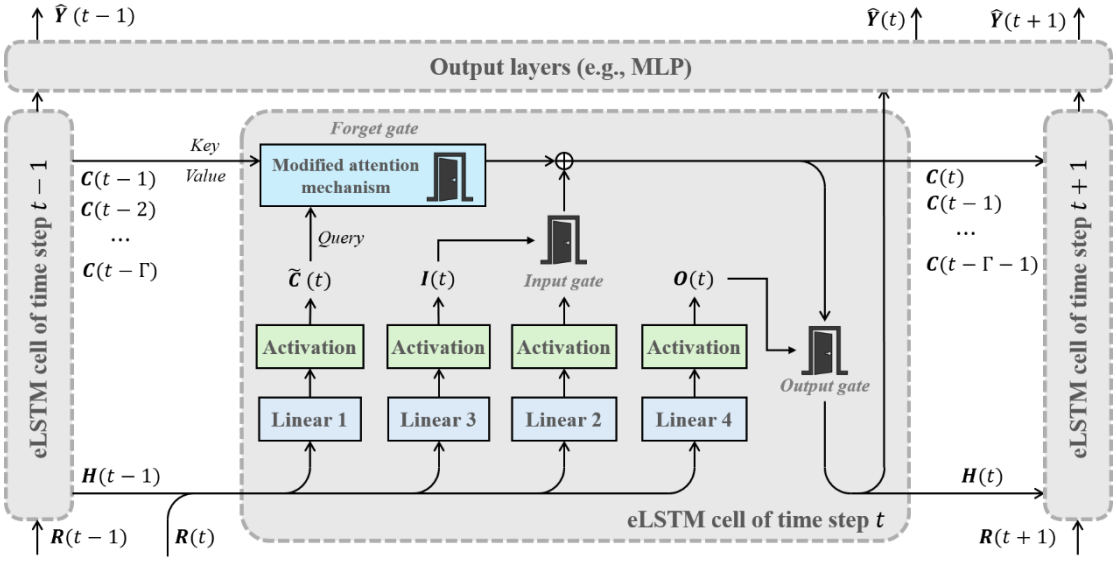


Fig. 5.2 eLSTM

5.2.2 Dynamic encoding by dESN

In prediction tasks, the intrinsic channel information often holds higher value. However, when historical data from this channel is input into RNN-type architectures such as eLSTM for encoding, it is mapped into an N_h-dimensional state space, forming a "one-to-many" mapping. To mitigate issues such as low feature extraction efficiency and high homogenization, a pre-processing feature extractor is required. Considering the inherent training difficulty of RNN-type networks, the added feature extractor should minimally increase the training workload. Hence, the ESN is revisited. The ESN (as described in

Section 4.3.1) employs a large, sparsely connected, and non-trainable reservoir for encoding, which can also be viewed as a $N_{in} \to N_{res}$ mapping. The encoded results $\boldsymbol{R}(t)$ contain state sequences of simple shapes, reflecting features of the input observed from different perspectives, making it a qualified base model for feature extraction.

Since training is required, setting an excessively high number of neurons in the eLSTM hidden layer is impractical. This results in a mismatch with the channel number of $\boldsymbol{R}(t)$ generated by ESN encoding. To reduce N_{res} in the ESN to conventional network levels, the efficiency of the remaining encoding components must be further enhanced, necessitating meticulous design of $\boldsymbol{W_{res}}$, $\boldsymbol{W_{in}}$, and $\boldsymbol{W_{back}}$. Typically, elements of $\boldsymbol{W_{in}}$ and $\boldsymbol{W_{back}}$ are sampled element-wise from a uniform distribution $\mathcal{U}[0,1]$, where each row represents a distinct encoding pattern for $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t-1)$. To ensure stable auto-encoding and equal contributions from inputs and outputs under each encoding mode, both matrices are normalized row-wise as follows:

$$W_{in/back,ij} = \frac{W_{in/back,ij}}{\sum_{k=1}^{N_{in}} W_{in/back,ik}} \tag{5.10}$$

where, i and j denote the row and column indices, respectively. More importantly, the fully randomized nature of $\boldsymbol{W_{res}}$ significantly impacts the encoding performance. Therefore, the values of $\boldsymbol{W_{res}}$ should be dynamically adjusted stepwise based on the fluctuation patterns of the input signal. Previous studies on neuronal activation patterns in biological systems revealed that neuronal activity dynamically changes under the influence of activation states of neighboring neurons. This variability in synaptic connection strength, termed synaptic plasticity (Wang, Jin & Hao 2019), has inspired its application to ESNs. Specifically, after each encoding step, the elements of $\boldsymbol{W_{res}}$ are updated according to the input from the previous step to enhance encoding stability in subsequent steps. The dynamic adjustment increment for any element in this matrix can be expressed as:

$$\Delta W_{res,ij} = -\alpha_{sp} X_i(t-1)\left(X_j(t-1) - X_i(t-1)W_{res,ij}(t-1)\right) \tag{5.11}$$

Here, α_{sp} is a controllable hyperparameter that modulates the strength of the synaptic plasticity mechanism. Additionally, neuronal activity is influenced by historical self-activation values, a mechanism known as intrinsic plasticity (Schrauwen et al., 2008). Therefore, after the ESN model completes the preliminary encoding with the dynamic $\boldsymbol{W_{res}}$ to generate $\boldsymbol{R}(t)$, it undergoes a subsequent correction determined by $\boldsymbol{R}(t-1)$ and $\boldsymbol{X}(t-1)$, as follows:

$$\boldsymbol{R}(t) = \boldsymbol{W_{in}}\boldsymbol{R}(t) + \boldsymbol{W_{res}}(t)\boldsymbol{R}(t-1) + \boldsymbol{W_{back}}\boldsymbol{Y}(t-1) \tag{5.12}$$

$$\boldsymbol{R}(t) \leftarrow f\left(\boldsymbol{k_r}(t) \odot \boldsymbol{R}(t) + \boldsymbol{b_r}(t)\right) \tag{5.13}$$

The preliminary encoding largely follows the standard ESN framework but excludes nonlinear activation functions and bias terms, which are primarily incorporated during the correction process. Here, $\boldsymbol{k}_{\mathbf{r}}(t) \in \mathbb{R}^{N_{res}}$ and $\boldsymbol{b}_{\mathbf{r}}(t) \in \mathbb{R}^{N_{res}}$ represent the correction mapping's slope and bias, respectively, determined by the historical activation values of individual neurons. These parameters are calculated via Eqs. (5.14) and (5.15), respectively.

$$\Delta\boldsymbol{k_r}(t) = \frac{\alpha_{ip}}{\boldsymbol{k_r}(t-1)} + \boldsymbol{X}(t-1) \odot \Delta\boldsymbol{b_r}(t-1) \tag{5.14}$$

$$\Delta\boldsymbol{b_r}(t) = \alpha_{ip}(1 - 2\boldsymbol{X}(t-1) + \boldsymbol{X}(t-1)^2) \tag{5.15}$$

Since the calculation of $\boldsymbol{k}_{\mathbf{r}}(t)$ and $\boldsymbol{b}_{\mathbf{r}}(t)$ involves only element-wise multiplication, the correction values for each channel depend solely on their own historical inputs, reflecting the intrinsic sustained response characteristics of the neurons. This encoding procedure, termed dESN, is more complex than the standard ESN but aligns better with signal fluctuation dynamics while remaining trainable-free for deployment, as illustrated in Fig. 5.3. In the proposed dESN-eLSTM prediction architecture, the input $\boldsymbol{X}(t)$ first undergoes dynamic encoding to generate $\boldsymbol{R}(t)$, which encapsulates multi-dimensional features extracted from the input. Subsequently, $\boldsymbol{R}(t)$ is treated as the model input and fed into the eLSTM network to complete the prediction (i.e., replacing $\boldsymbol{X}(t)$ in Section

5.2.1 with $\boldsymbol{R}(t)$, for which the matrix dimensions must be adjusted accordingly).

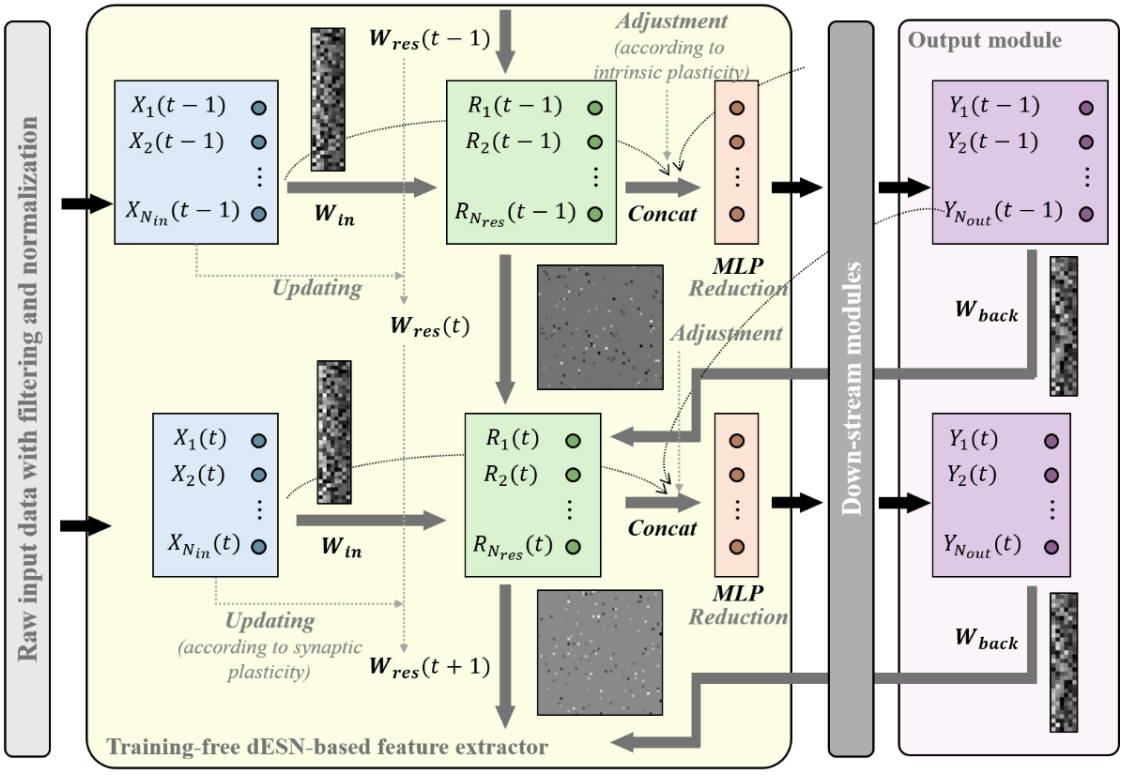


Fig. 5.3 dESN

5.2.3 Application: Wake field prediction of Lanzhou-Xinjiang HSR

5.2.3.1 Project overview

As shown in Fig. 5.4(a), the Baili Wind Zone, located in northwestern China, experiences frequent strong winds of magnitude 8 or higher year-round, with a maximum ten-minute average wind speed of 40 m/s and extreme gusts exceeding 60 m/s. The wind direction is nearly fixed, flowing from north to south. The Lanzhou-Xinjiang HSR traversing this region from east to west, is subjected to intense crosswinds. To ensure the safety of high-speed trains (HST), two types of windproof structures (WSs) have been constructed along the railway: solid windbreak walls on embankments and bilateral porous wind barriers (with a porosity of 30%) on bridges. While these facilities are designed and deployed according to specifications, their wake flow fields, particularly under extreme wind conditions, remain poorly characterized. Furthermore, the transition segment between the two types of WSs involves complex flow boundary conditions, making it uncertain

whether the actual wind fields on the railway meet safety requirements for HST operation.

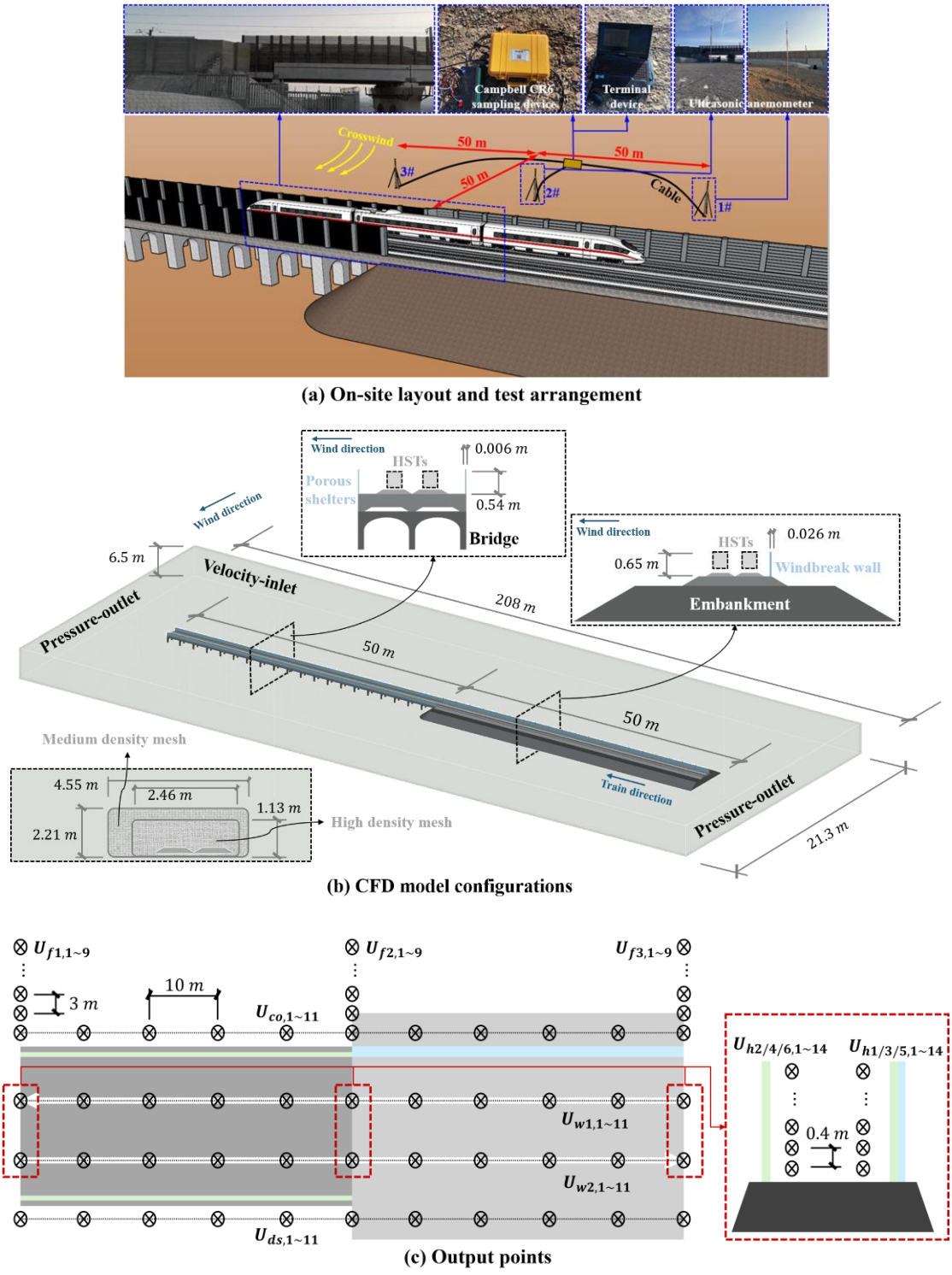


Fig. 5.4 WS transition segment of Lanzhou-Xinjiang HSR in Baili Wind Zone

Direct real-time monitoring of wind fields on the track using anemometers is the most straightforward approach. However, railway authorities typically prohibit installing

sensors near tracks due to operational safety concerns. As an alternative, external wind fields are monitored outside the WSs to infer internal wind conditions. The internal and external wind fields can be viewed as sequential states of fluid flow passing through the exterior and interior regions of the WSs, establishing a predictive mapping relationship governed by the underlying PDEs of fluid dynamics. While computational fluid dynamics (CFD) models can theoretically capture this relationship, their computational complexity, required to ensure accuracy, renders them impractical for real-time prediction. Even simulating a one-minute flow field on conventional workstations may take days or weeks. To address this, a prediction method for internal wind fields behind HSR WSs is proposed, based on the dESN-eLSTM model, as shown in Fig. 5.5. This approach leverages short-term CFD-generated data as training samples to extract spatiotemporal correlations between external and internal flow states at sequential time points. Once trained, the model enables real-time prediction of internal wind signals from new external flow field measurements.

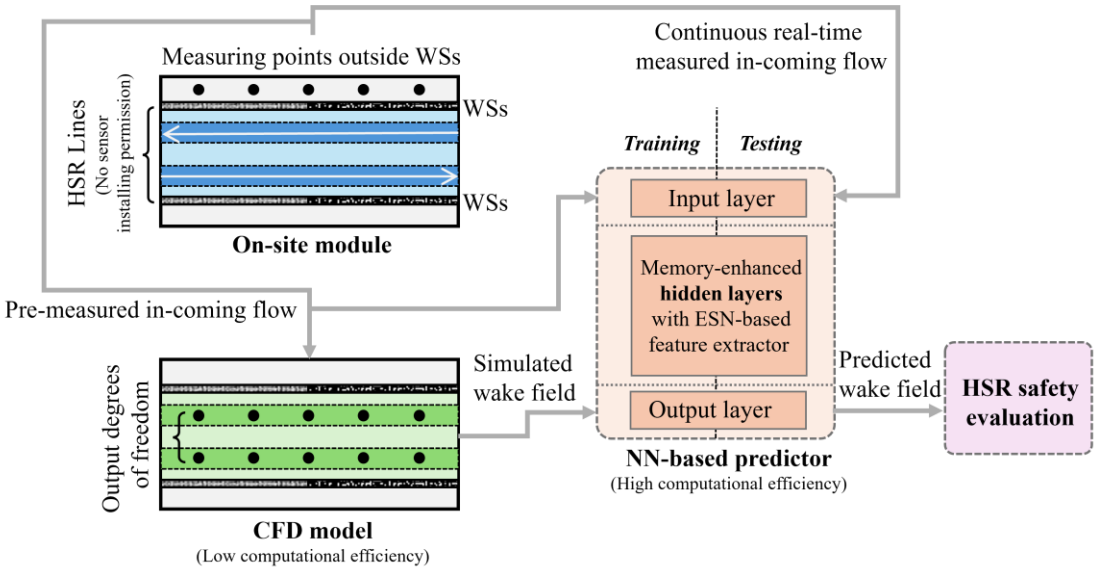


Fig. 5.5 HSR wake field prediction scheme

5.2.3.2 CFD model and the dataset acquisition

A full-scale three-dimensional CFD model of the transition segment between WSs, as shown in Fig. 5.4(b), was established in Fluent 2020 R2. The measured or synthesized

incoming flow signal $\boldsymbol{U_{co}}$ was applied as the input to compute the internal wind field $\boldsymbol{U_w}$ and the downstream WS-external signal $\boldsymbol{U_{ds}}$. The purpose of this CFD model is to generate high-fidelity wake flow data by solving the governing PDEs of fluid motion, thereby providing sufficient input-output pairs (external: $\boldsymbol{U_{co}}$ and $\boldsymbol{U_{ds}}$; internal: $\boldsymbol{U_{w1}}$ and $\boldsymbol{U_{w2}}$) for training the NN. Detailed geometric dimensions of the model are illustrated in Fig. 5.4(b). In the CFD setup, the income boundary was defined as a "Velocity-Inlet," while the two adjacent surfaces to the income side and the wake-side surface were configured as "Pressure-Outlet" boundaries to accurately simulate crosswind conditions. The top surface was assigned a "Symmetry" boundary. Solid surfaces, including the ground, bridge structures, and WSs, were modeled as "No-Slip Wall" boundaries to account for air-surface interactions.

A hierarchical meshing strategy was employed to discretize the wake region of the WS transition segment under strong crosswinds. The area above the HSR tracks utilized dynamically refined grids controlled by user-defined functions (UDFs). The final mesh, generated using mosaic grid techniques, comprised approximately 56 million cells. Specifically, grid sizes for railway support structures ranged from 3 mm to 8 mm, while those for WSs varied between 50 mm and 100 mm. To enhance the resolution of leeward vortices, two refinement zones with progressively increasing grid sizes were added near the no-slip boundaries. The first and second refinement zones had maximum grid sizes of 2 mm and 10 mm, respectively, with the outermost boundary grid size set to 40 mm. The delayed detached eddy simulation (DDES) approach, coupled with the shear-stress transport (SST) model, was adopted to resolve near-wall flows. Eleven boundary layers were configured in this region, with the first-layer thickness varying from 0.2 mm to 1 mm to maintain y^+ values below 10.0. Additional mesh details can be found in prior published work (Deng et al., 2023b).

The CFD model was solved using a high-resolution, total variation diminishing (TVD) temporal discretization scheme. A time step size of 0.0001 s and 50 iterations per time step were implemented. A moving average process reduced the output sequence sampling

rate to 100 Hz, which aligns with typical real-time sensor frequencies for engineering applications. In this case, the CFD input was a on-site measured wind speed signal (150 s) $\boldsymbol{U_{co}}$ outside the WSs. Output points were arranged as shown in Fig. 5.4(c): four groups of anemometers were positioned along the railway's operational direction. Two groups ($\boldsymbol{U_{w1}}$ and $\boldsymbol{U_{w2}}$) were located inside the WSs facilities, while the other two ($\boldsymbol{U_{co}}$ and $\boldsymbol{U_{ds}}$) were placed 5 m outside the track. Within each group, sensors were evenly spaced at 10 m intervals. Three additional cross-directional measurement groups ($\boldsymbol{U_{f1}}$, $\boldsymbol{U_{f2}}$, and $\boldsymbol{U_{f3}}$) were defined with 3 m spacing. All seven groups were positioned at 1.38 m above ground or bridge level, corresponding to the centroid height of HSTs. Six vertical measurement groups ($\boldsymbol{U_{h1}}$~$\boldsymbol{U_{h6}}$) were established at the starting point of the solid windbreak wall ($\boldsymbol{U_{h1}}$ and $\boldsymbol{U_{h2}}$), the endpoint of the porous wind barrier ($\boldsymbol{U_{h3}}$ and $\boldsymbol{U_{h4}}$), and the transition segment ($\boldsymbol{U_{h5}}$ and $\boldsymbol{U_{h6}}$), with 0.4 m height intervals. The model was validated against wind tunnel experiments, and all CFD computations were performed on the Hong Kong Polytechnic University's supercomputing cluster over approximately 30 days.

The computational results are presented in Fig. 5.6, which shows flow field cross-sections at five longitudinal positions (10 m, 30 m, 50 m, 70 m, and 90 m east of the model's eastern edge). Left-side contour maps depict wind speed, while right-side maps illustrate turbulence intensity $\boldsymbol{I}$. Key observations can be stated as follows. The wake fields in embankment sections with solid windbreak walls (Profiles A-A, B-B, and C-C) exhibit greater complexity than those in bridge sections with porous barriers (Profiles D-D and E-E). The embankment regions feature multiple vortical structures and higher turbulence intensities, exemplified by small-scale vortices in Profile C-C and three dominant vortices in Profile B-B, corresponding to dark regions in turbulence length contours. The wake fields of the bridge section demonstrate more uniform flow with smaller vortices localized near the ground and barrier tops, accompanied by lower turbulence intensities fluctuating between $[0,1]$. Wind speeds in the train operational zone are significantly reduced compared to the incoming flow, validating the efficacy of the WSs. After CFD

computations, wind speed signals from the measurement points in Fig. 5.4(c) were used to train (first 100 s, including hyperparameter optimization) and test (last 50 s) the dESN-eLSTM predictor.

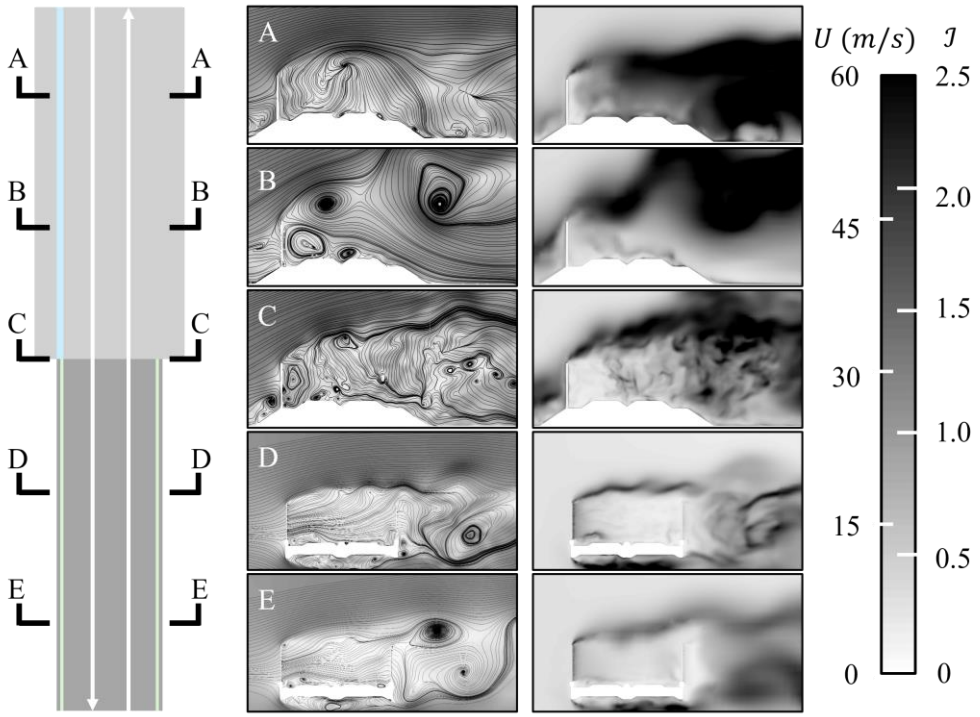


Fig. 5.6 CFD results

5.2.4 Prediction results and discussion

The predictive performance of the dESN-eLSTM model for WS wake fields is the focus of this section. In addition to evaluating its fundamental prediction capability for in-plane flow fields, the study investigates the influence of the number, distribution, and distance of external measurement points on network performance. Crucially, to validate the model's reliability under extreme conditions, its generalization ability is tested against flow velocities significantly exceeding those in the training data. Furthermore, the model's predictive capacity for out-of-plane wind fields is also examined.

Hyperparameters with significant impacts on the dESN-eLSTM results were selected using the grid search method detailed in Section 4.3.4.3, while others were empirically determined. All hyperparameter values are listed in Table 5.1. A two-layer MLP was implemented as the output module, with the first layer containing 20 neurons and

applying dropout with a probability of 0.2. The SGD remained the optimization method for model training, with a learning rate $lr = 0.0005$, and $\eta = 0.001$, and a training duration of 300 epochs. To account for the complexity of wind speed signals, evaluation metrics included not only MSE but also mean absolute wind speed $\overline{\boldsymbol{U}}$ and turbulence intensity $\boldsymbol{\mathcal{I}}$.

Table 5.1 Hyperparameters of dESN-eLSTM

N_{res}	N_h	α_{sp}	α_{ip}	$\rho(W_{res})$	$spar(W_{res})$	Γ
2000	200	0.000001	0.00001	0.85	$1.5N_{res}$	50

5.2.4.1 In-plane wake field prediction results

The model's predictive performance is first evaluated under dense external measurement conditions, where all data points from $\boldsymbol{U_{co}}$ and $\boldsymbol{U_{ds}}$ serve as network inputs. Representative channels ($\boldsymbol{U_{w1,6}}$ and $\boldsymbol{U_{w2,6}}$) near the transition segment of the WSs are selected for comparative analysis. Their prediction results are illustrated in Figs. 5.7(a) and 5.7(b), while their power spectral densities (PSDs) are compared in Figs. 5.8(a) and 5.8(b). These channels, located closest to the transitional segment, exhibit the most complex wake dynamics and thus represent the most challenging prediction targets. The dESN-eLSTM model achieves MSEs of 5.2436 and 4.1071 for $\boldsymbol{U_{w1,6}}$ and $\boldsymbol{U_{w2,6}}$, corresponding to relative errors of approximately 11.44% and 12.65%, respectively. In the frequency domain, both CFD simulations and network outputs for these channels demonstrate similar spectral characteristics, with energy predominantly concentrated below 1.0 Hz. Notably, the model also captures high-frequency components with reasonable consistency.

For mean absolute wind speed ($\overline{\boldsymbol{U}}$), the model predicts 3.63 m/s and 2.90 m/s for $\boldsymbol{U_{w1,6}}$ and $\boldsymbol{U_{w2,6}}$, compared to reference values of 3.67 m/s and 2.66 m/s, achieving prediction accuracies of 96% and 91%, respectively. This indicates an unbiased temporal prediction, as evidenced by the high accuracy in mean value estimation. Turbulence intensity ($\boldsymbol{\mathcal{I}}$)

values derived from CFD simulations are 1.13 and 1.15 for the two channels, while the model predicts 1.11 and 1.22, demonstrating estimation precision near 95% for wind speed fluctuations. Instantaneous maximum wind speeds, critical for train safety assessments, are also evaluated. For $\boldsymbol{U_{w1,6}}$, the simulated maximum northward and southward wind speeds during the computational period are 12.51 m/s and 7.41 m/s, respectively. The corresponding predicted maxima are 12.77 m/s and 8.31 m/s, confirming the model's capability to accurately bound wind speed variations.

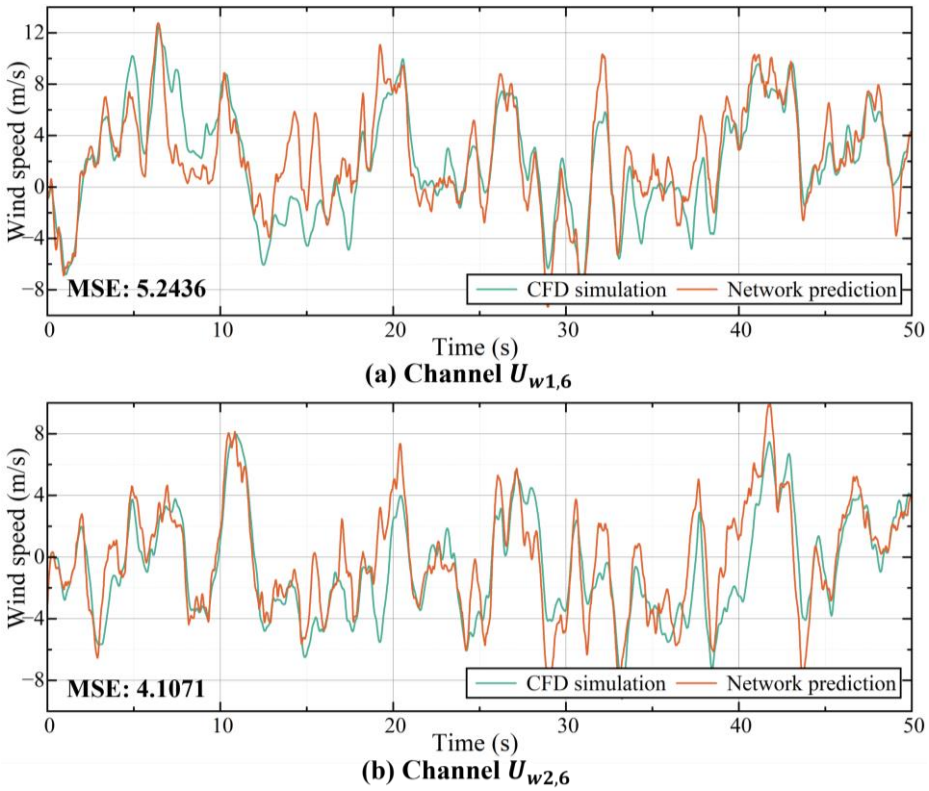


Fig. 5.7 In-plane wake field prediction in time domain

The prediction results for other channels are summarized in Table 5.2. Considering the varying fluctuation ranges across channels, the lowest MSEs for the $\boldsymbol{U_{w1}}$ and $\boldsymbol{U_{w2}}$ measurement groups occur at $\boldsymbol{U_{w1,9}}$ and $\boldsymbol{U_{w2,9}}$, with mean relative errors accounting for approximately 15% and 7% of their absolute mean wind speeds, respectively. In contrast, the highest MSEs are observed at $\boldsymbol{U_{w1,3}}$ and $\boldsymbol{U_{w2,6}}$. Comparative analysis reveals that regions with poorer prediction performance primarily coincide with the initial segment of the solid windbreak wall and the transition segment of the WSs. This is attributed to the

superior wind-blocking efficiency of the solid wall, which forces airflow to circumvent the structure entirely before entering the railway zone, resulting in more chaotic wake dynamics that challenge network learning.

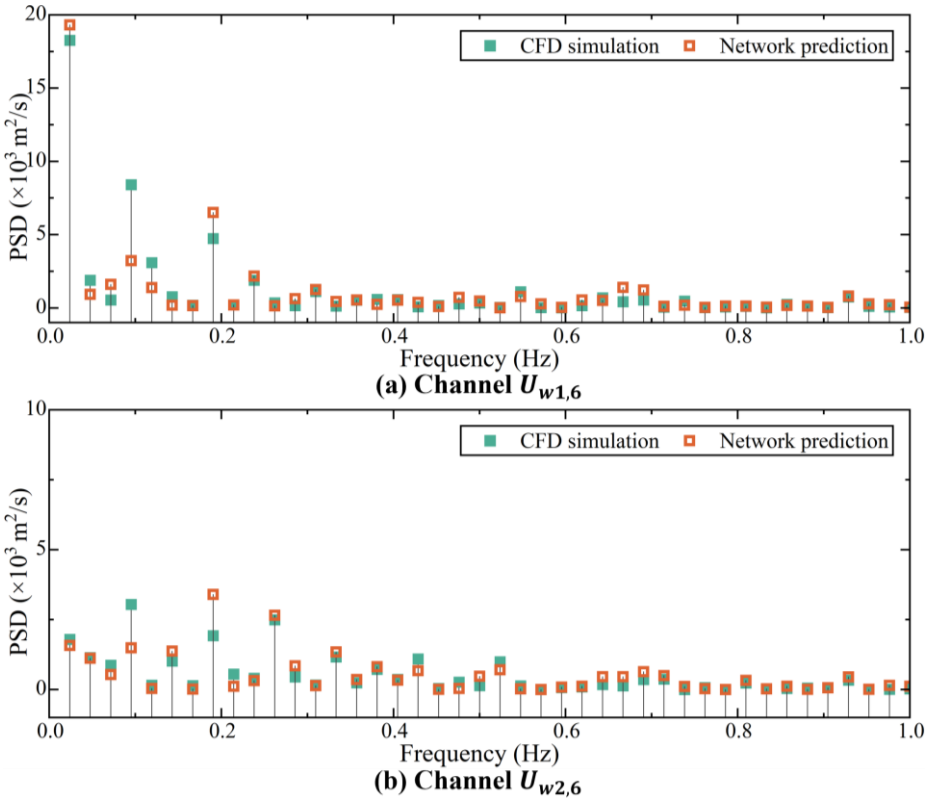


Fig. 5.8 In-plane wake field prediction in frequency domain

Additionally, at the initial segment of the solid wall, input channels (measurement points) are predominantly located on its western side. Asymmetric source information and limited eastern-side inputs contribute to degraded model performance in this region. Conversely, in areas protected by porous wind barriers ($\boldsymbol{U_{w1,7\sim11}}$ and $\boldsymbol{U_{w2,7\sim11}}$), the flow fields exhibit greater regularity, leading to improved model performance. For instance, the five measurement points in $\boldsymbol{U_{w1,7\sim11}}$ achieve an average MSE of 2.5237. Furthermore, as channels become more distant from the northern wind barrier, the flow field simplifies and wind speeds increase, reducing the average MSE for $\boldsymbol{U_{w2,7\sim11}}$ to 2.0956. Despite variations in prediction accuracy across channels and temporal instances, the model achieves approximately 90% accuracy in estimating both absolute mean wind speed and turbulence intensity for in-plane wake fields, meeting engineering precision requirements.

Notably, the dESN-eLSTM model completes inference for a 5,000-time step wake signal (22 wind speed channels) in 1.17 s, corresponding to 50 s of real-world data. This demonstrates the model's capability for real-time wake field prediction, in stark contrast to the computationally intensive CFD simulations.

Table 5.2 In-plane wake field prediction results for all channels

Channel	MSE	$\overline{U}/\overline{\hat{U}}$	$\overline{J}/\overline{\hat{J}}$	Channel	MSE	$\overline{U}/\overline{\hat{U}}$	$\overline{J}/\overline{\hat{J}}$
$U_{w1,1}$	9.9321	8.36/8.50	0.49/0.54	$U_{w2,1}$	9.8901	5.49/5.99	0.98/0.95
$U_{w1,2}$	11.0052	8.06/8.96	0.52/0.54	$U_{w2,2}$	7.4308	5.03/5.38	0.69/0.73
$U_{w1,3}$	14.7891	9.02/9.28	0.52/0.57	$U_{w2,3}$	9.8438	4.61/5.66	1.11/0.98
$U_{w1,4}$	9.6232	8.17/8.62	0.48/0.49	$U_{w2,4}$	8.8077	6.81/7.93	0.46/0.51
$U_{w1,5}$	7.8094	6.63/7.06	0.64/0.68	$U_{w2,5}$	4.2529	5.76/6.14	0.62/0.59
$U_{w1,6}$	5.2436	3.67/3.53	1.13/1.11	$U_{w2,6}$	4.1071	2.90/2.66	1.15/1.22
$U_{w1,7}$	2.9738	6.83/6.41	0.26/0.31	$U_{w2,7}$	5.6497	12.13/10.85	0.14/0.17
$U_{w1,8}$	3.0161	7.76/7.34	0.23/0.28	$U_{w2,8}$	2.6740	10.65/9.58	0.15/0.19
$U_{w1,9}$	1.6447	8.31/8.10	0.16/0.20	$U_{w2,9}$	0.6209	10.83/10.35	0.06/0.10
$U_{w1,10}$	2.0973	6.87/6.69	0.21/0.28	$U_{w2,10}$	1.0346	10.22/9.84	0.08/0.11
$U_{w1,11}$	3.1726	7.55/7.41	0.15/0.18	$U_{w2,11}$	0.4991	9.64/9.76	0.11/0.13

To validate the superiority of the dESN-eLSTM model and the necessity of its architectural modifications, comparative performance analyses were conducted against multiple baseline models: standard ESN, LSTM, dESN, and eLSTM. All reference models were trained and tested on the aforementioned dataset, with results summarized in Table 5.3. The layer configurations and neuron number for the ESN, dESN, LSTM,

and eLSTM models align with those listed in Table 5.1, where ESN-based architectures incorporate a 2000-neuron reservoir, and LSTM-based models employ a 200-neuron hidden layer. The number of trainable parameters for each model is also provided in Table 5.3. In this application, the temporal dependencies inherent to fluid dynamics impose heightened demands on model memory capacity. The eLSTM model addresses this challenge through its enhanced forget gate and attention-like mechanism, which enable balanced consideration of multiple prior time steps during encoding, thereby establishing extended memory retention. Consequently, memory-enhanced models exhibit significantly superior performance in this task compared to networks lacking long-term memory capabilities.

Table 5.3 Model comparison on prediction performance

Models	Average MSE of U_{w1}	Average MSE of U_{w2}	Parameter number	Training time
ESN	21.3252	18.6865	44022	6.61s/epoch
LSTM	25.0165	20.0025	182022	25.02s/epoch
dESN	17.1258	13.2898	44022	6.61s/epoch
eLSTM	10.3825	7.9896	182022	27.87s/epoch
Transformer	8.2434	4.7529	646320	19.26s/epoch
dESN-eLSTM	6.4824	4.9828	626022	35.53s/epoch

Additionally, considering the challenges of sensor deployment and maintenance along HSR lines, this section investigates the impact of measurement point density on model performance, as illustrated in Fig. 5.9. By progressively reducing the number of input measurement points from 11 to 3 per group outside the WSs, the prediction MSEs for all channels in $\boldsymbol{U_{w1}}$ and $\boldsymbol{U_{w2}}$ are observed to increase, reaching up to three times their

original values. Despite this error escalation, the simplified configuration requiring only three measurement points per side remains practically acceptable. The results in Fig. 5.9 further reveal two critical patterns: channels protected by porous wind barriers exhibit more predictable wake fields with higher signal similarity across measurement points, resulting in smaller error increments when reducing sensor density; prediction accuracy improves substantially when measurement points exist at positions directly opposite to target channels, regardless of their spatial locations.

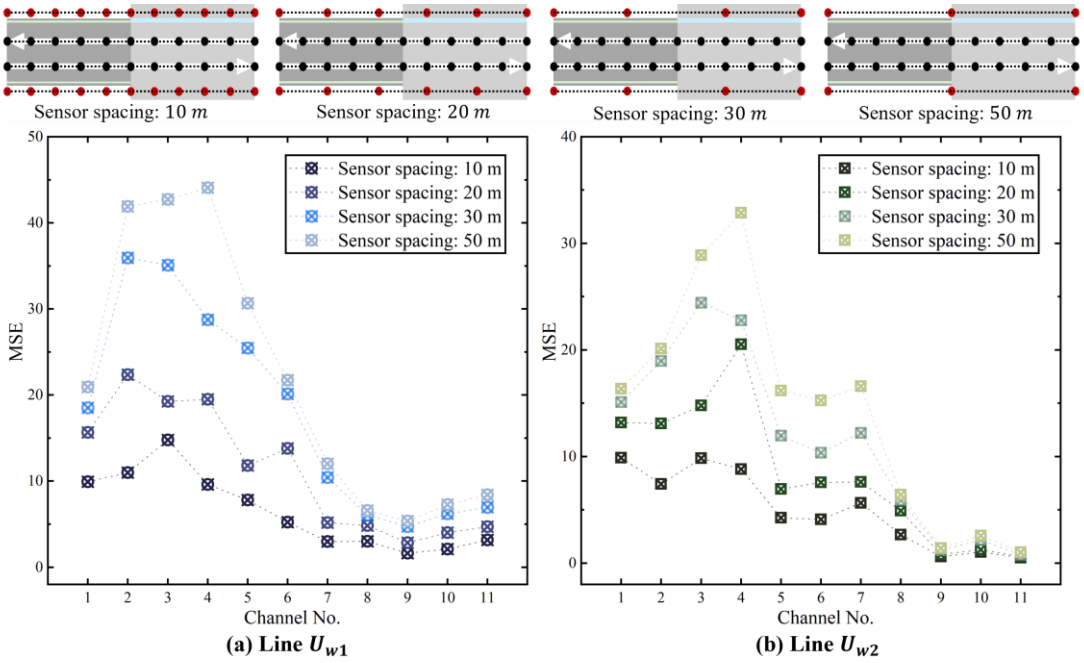


Fig. 5.9 Influence of measurement intensity on prediction accuracy

The study further evaluates the contributions of income- and downstream-side measurements to wake field determination by comparing three sensor configurations: bilateral, income-side only, and downstream-side only (Fig. 5.10). Income-side measurements demonstrate significantly greater value for wake characterization than downstream-side data. For the critical transitional segments ($\boldsymbol{U_{w1,6}}$ and $\boldsymbol{U_{w2,6}}$), removing downstream-side measurements increases mean relative errors by 5.2% and 50.1%, respectively, while eliminating income-side sensors causes error increments of 74.5% and 102.2%. Similar trends are observed across most channels. This discrepancy likely stems from airflow disturbances occurring at the opposite-side wind barriers, where

complex interactions degrade correlations with the railway flow field. These findings confirm the practical feasibility of unilateral measurement configurations while emphasizing the superior information content of income-side data for operational implementations.

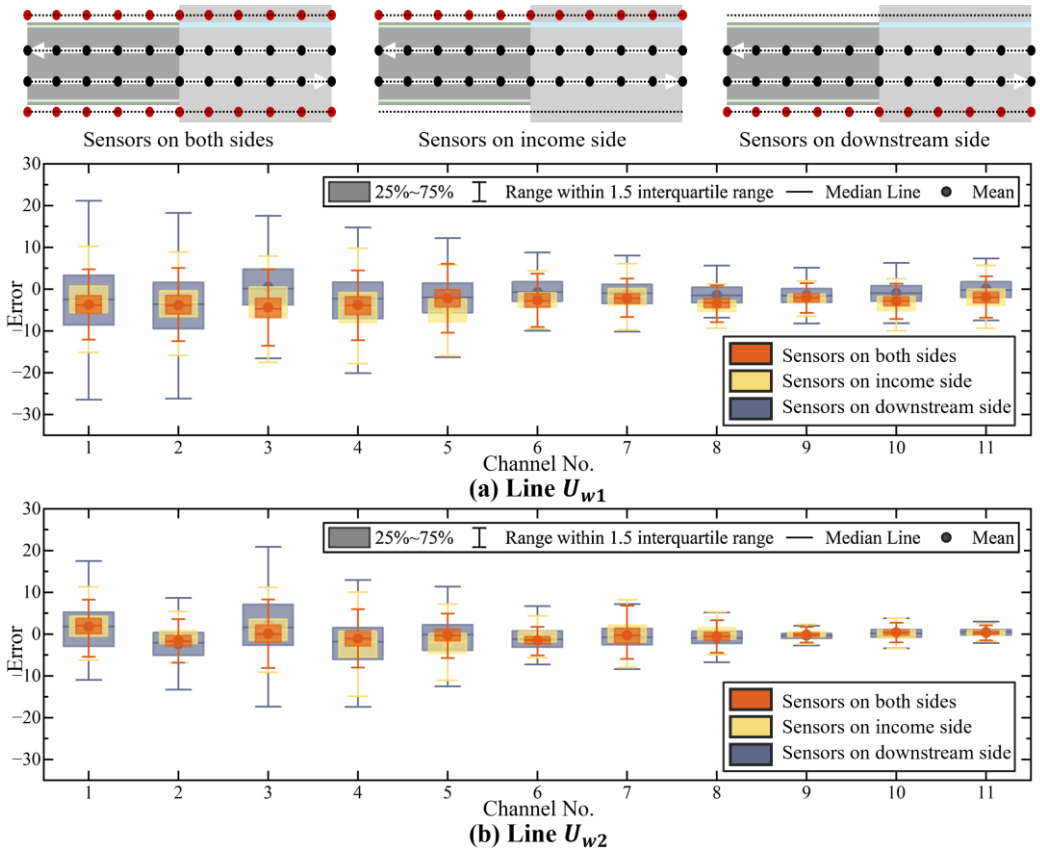


Fig. 5.10 Influence of measurement location on prediction accuracy

Finally, the influence of measurement point distance from WSs on network performance is investigated in Fig. 5.11. Increased measurement distance implies prolonged fluid transit time from external sensors to the railway interior, corresponding to predictions targeting more distant future states with elevated difficulty. Building upon previous feasibility analyses, the optimal unilateral three-sensor configuration is adopted while progressively increasing sensor distances from 3 m to 30 m. Results demonstrate that distant fluid paths exhibit minimal cross-sectional variations along the flow direction due to unobstructed motion. In contrast, near-field flow undergoes drastic divergent alterations when circumventing WSs. CFD simulations confirm higher similarity between

wake fields and near-field flows, with intense fluid discontinuities and broader influence ranges observed near solid windbreak walls due to their superior blocking efficiency. Consequently, abrupt MSE increases occur as sensors initially move away from windbreaks, showing amplification factors of approximately 2.0 and 1.4 for embankment and bridge segments, respectively. Beyond 10 m from WSs, MSE values stabilize within limited fluctuation ranges without further escalation.

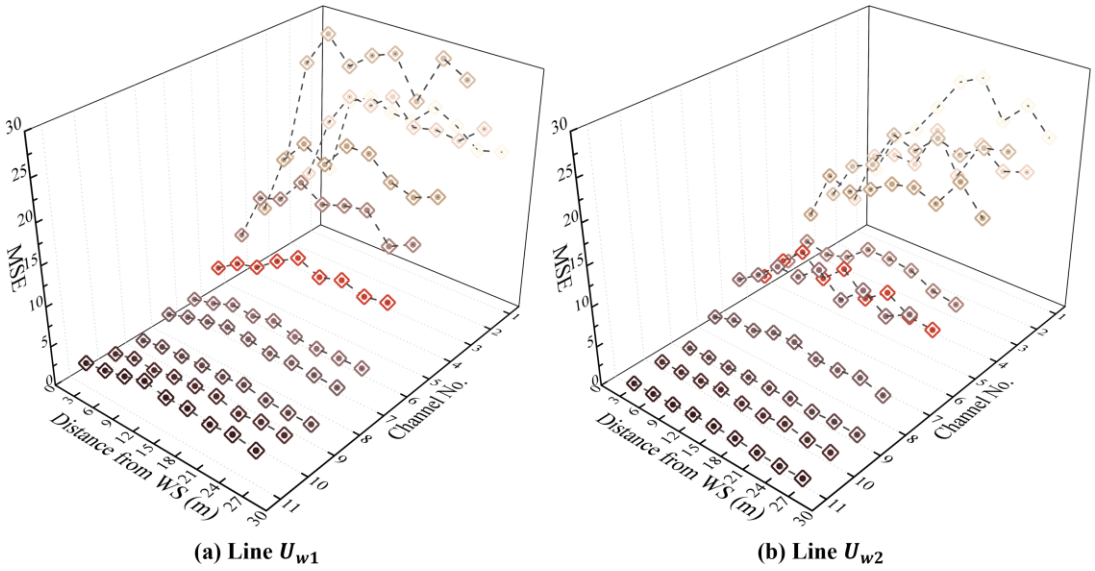


Fig. 5.11 Influence of measurement distance on prediction accuracy

5.2.4.2 Out-of-plane wake field prediction results

Beyond in-plane wind fields, the vertical wind profile from rail level to train roof exhibits significant height-dependent variations due to complex turbulence effects, making its prediction crucial for operational safety assessments. A dedicated dESN-eLSTM model was developed for wind profile predictions ($\boldsymbol{U_{h1}}$~$\boldsymbol{U_{h6}}$), maintaining the configuration specified in Table 5.1 while expanding input channels to incorporate in-plane prediction results ($\boldsymbol{U_{w1}}$ and $\boldsymbol{U_{w2}}$) derived from Section 5.2.4.2. Fig. 5.12 illustrates the transitional segment wind profile ($\boldsymbol{U_{h4}}$) with detailed temporal prediction comparisons across channels, while Fig. 5.13 presents equivalent results for the remaining five profiles. The $\boldsymbol{U_{h4}}$ predictions demonstrate the model's capability to reconstruct full wind profiles using inputs from the train's centroid height. Within windbreak-protected zones (below 5 m), wind speeds stabilize around 15 m/s with mean values below 6 m/s, where prediction

accuracy remains high due to minimal inter-channel speed variations. However, near or above WS's heights, shielding effectiveness diminishes markedly, resulting in speed increments exceeding 15 m/s and enriched high-frequency components. Exclusive reliance on reference plane measurements (1.38 m height) reduces prediction accuracy to approximately 85%. As shown in Fig. 5.13, performance consistency across all six profiles is maintained, with maximum prediction errors (2.3 m/s) occurring in $\boldsymbol{U_{h4}}$ uppermost channel. All vertical profile predictions achieve over 85% accuracy, confirming the method's efficacy in resolving vertical wind variations across solid windbreaks, porous barriers, and their transitional segments.

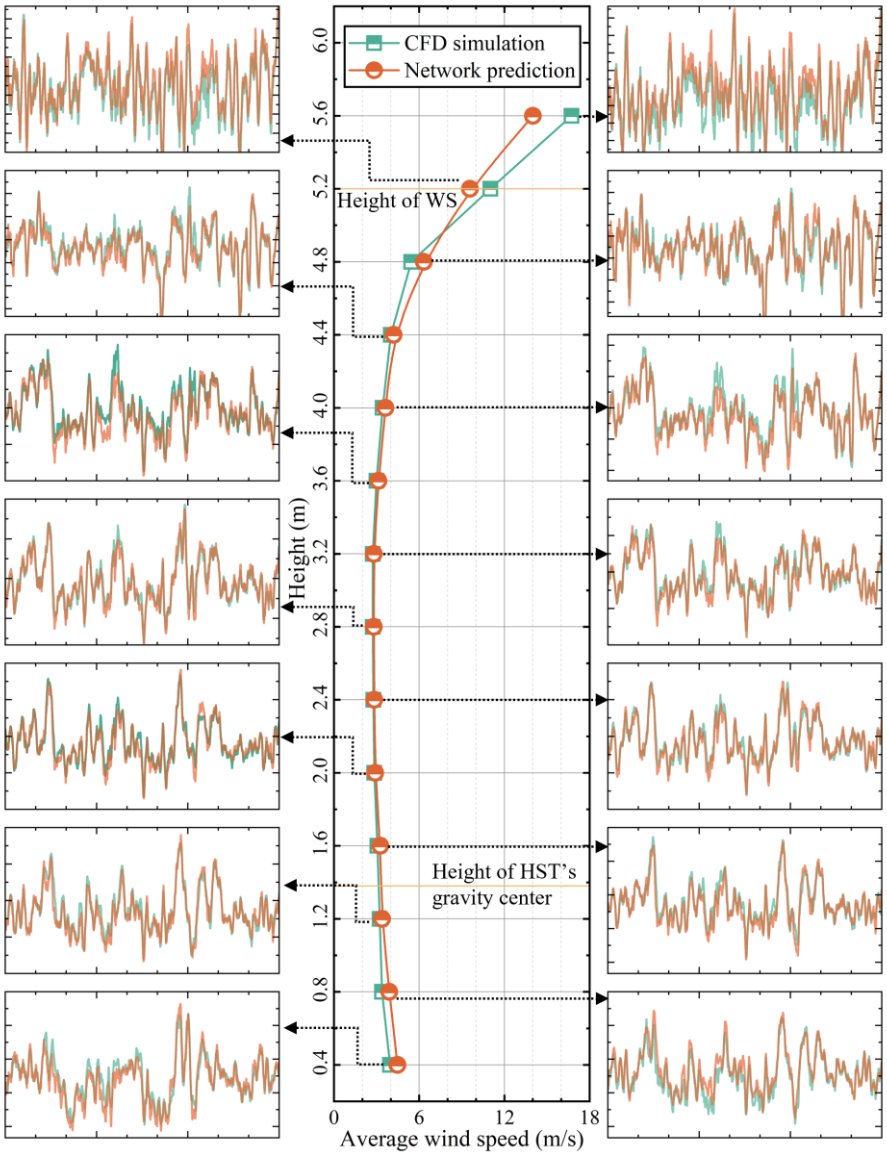


Fig. 5.12 Wake field prediction of wind profile $\boldsymbol{U_{h4}}$

5.2.4.3 Generalization performance under extreme income flow

The primary objective of the dESN-eLSTM predictor is to rapidly derive wake fields around WSs (including transitional segments) along HSR lines, particularly when sensors detect incoming wind speeds exceeding safety thresholds or exhibiting hazardous features, thereby enabling timely warnings for approaching trains. A critical requirement lies in its extrapolation capacity under extreme scenarios where incoming wind speeds surpass the maximum values in the training dataset, as CFD simulations cannot deliver real-time solutions. To evaluate this capability, scaled incoming wind speed signals are generated by applying scaling factors ($\gamma = 1.1$, $\gamma = 1.2$, and $\gamma = 1.5$) to amplify $\boldsymbol{U_{w1}}$, $\boldsymbol{U_{w2}}$, $\boldsymbol{U_{co}}$, and $\boldsymbol{U_{ds}}$. The original dESN-eLSTM predictor trained on unmodified datasets is then employed to determine wake field characteristics under these amplified inputs. For analytical clarity, only the first 2000-time steps from the testing dataset are output and analyzed, with results comprehensively illustrated in Fig. 5.14.

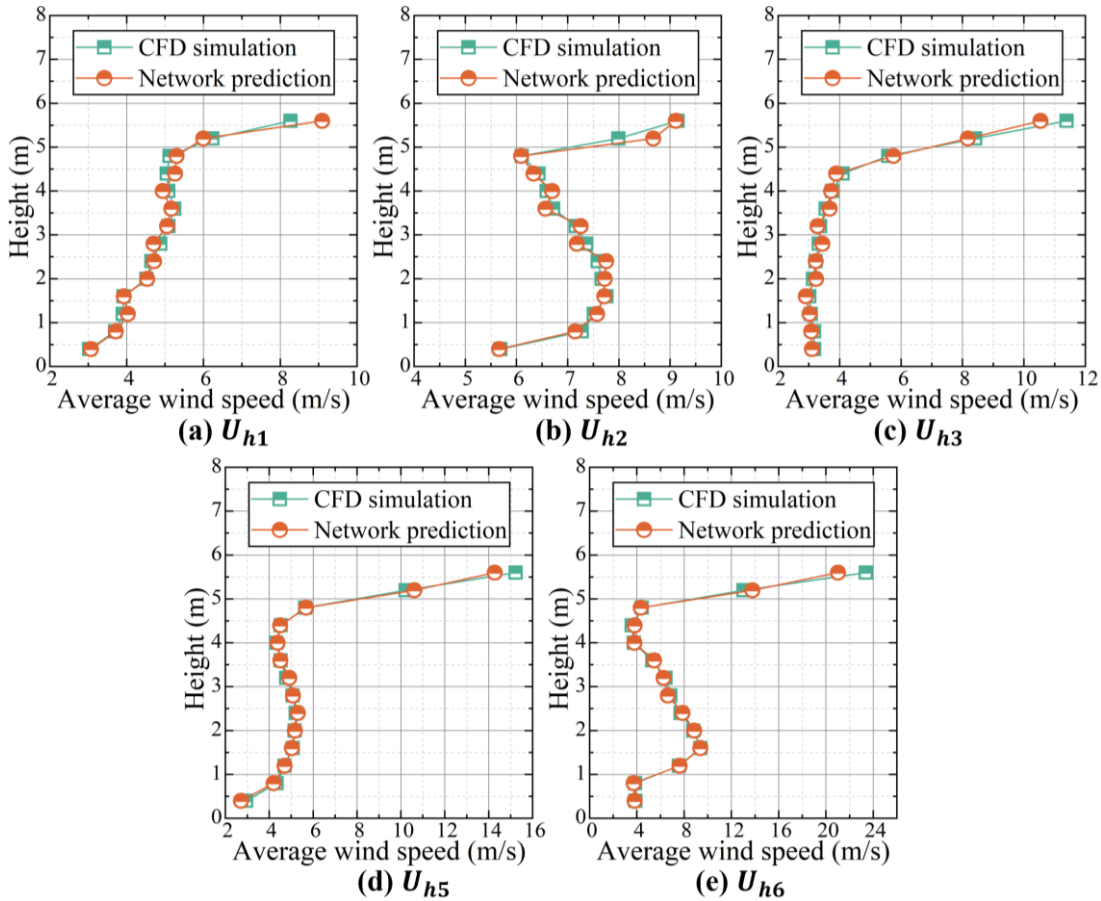
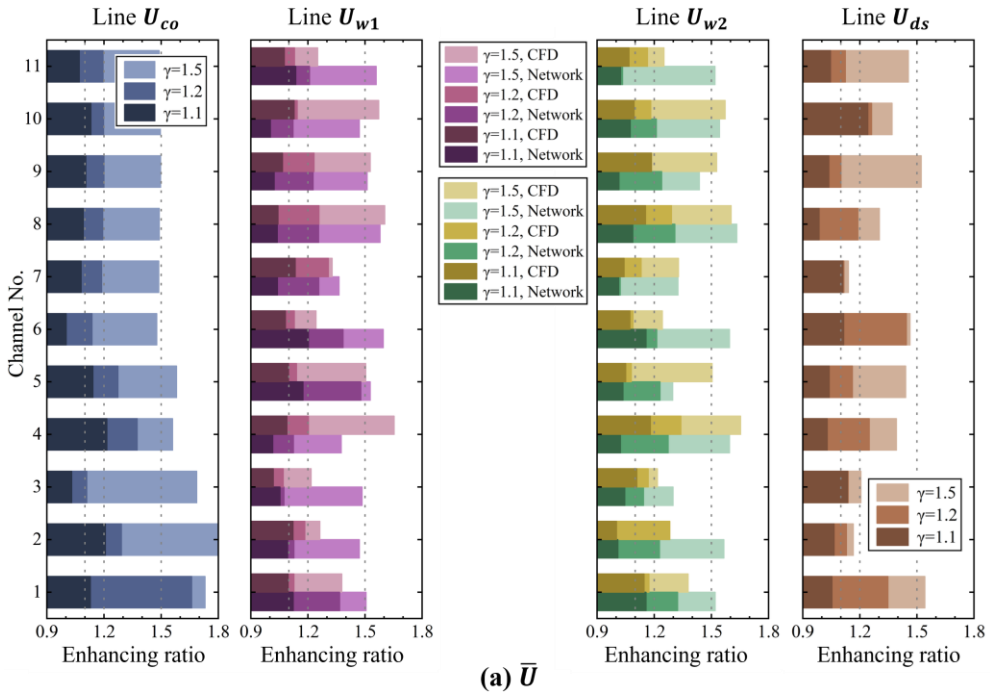


Fig. 5.13 Wake field prediction results for other wind profiles

The presence of WSs introduces deviations in the statistical parameters of wind speed measured at $\boldsymbol{U_{co}}$, which fail to scale proportionally with boundary conditions. Instead, these parameters exhibit fluctuating and nonlinearly growing characteristics. Among $\boldsymbol{U_{co}}$ measurement points, regions shielded by porous wind barriers demonstrate the least nonlinearity, while other measurement groups display more pronounced nonlinear behavior. When the dESN-eLSTM predictor is employed for wake field reconstruction, higher accuracy is observed in regions where flow responses to wind speed variations remain stable ($\boldsymbol{U_{w1,7\sim11}}$ and $\boldsymbol{U_{w2,7\sim11}}$). This stability primarily arises from the near-proportional scaling of model inputs $\boldsymbol{U_{co}}$ at these locations.

In contrast, flow fields inside and outside WSs within transitional segments exhibit heightened turbulence, leading to increasingly complex mapping relationships. For instance, at $\gamma = 1.5$, the $\overline{\boldsymbol{U}}$ in the transitional segment increases by over 60%, while turbulence intensity $\boldsymbol{\mathcal{I}}$ rises nearly 1.8-fold, resulting in degraded model performance. Nevertheless, the predictor retains approximately 70% accuracy even when incoming wind speeds exceed the upper limit of the training dataset by 50%. This observation underscores the model's robust generalization performance, demonstrating its resilience under adverse weather conditions.



(a) $\overline{\boldsymbol{U}}$

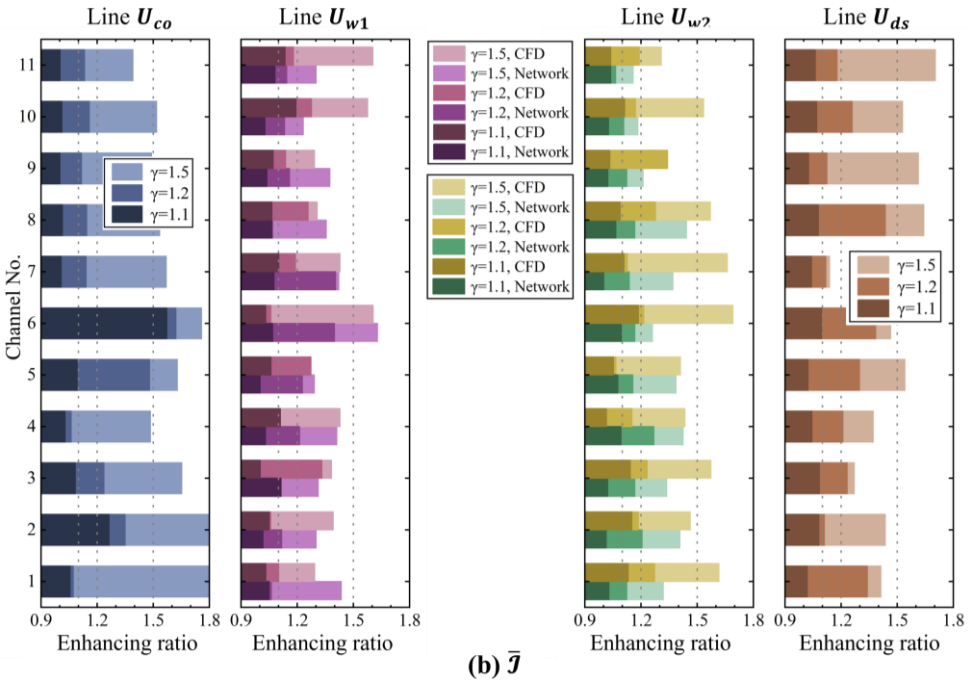


Fig. 5.14 extrapolation performance of dESN-eLSTM

5.3 Synchronized data mapping

Compared to prediction tasks, solving synchronous sequence-to-sequence (Seq2Seq) mapping relationships between coupled signals is relatively straightforward. In prediction problems, only historical data is available as input for estimating future signal values, whereas synchronized mapping leverages concurrent known information to infer outputs at each timestep. However, the precision expectations for these two tasks differ significantly in engineering applications. Absolute accuracy in prediction is inherently unattainable due to the unknowable nature of future states; even statistically accurate parameter predictions often suffice for complex systems. Conversely, synchronized mapping tasks in SHM systems typically involve inferring strongly correlated variables for applications such as indirect measurement or damage detection benchmarks, necessitating higher precision requirements.

This section proposes a solution for synchronous Seq2Seq problems involving variables with intricate mapping relationships. A modified Transformer architecture, a prominent framework for temporal data modeling, serves as the foundation. To accommodate the many-to-many mapping prevalent in engineering scenarios, the original positional

embedding is replaced with a random reservoir mechanism, enhancing adaptability to arbitrary input configurations. Furthermore, recognizing the frequency-domain characteristics inherent to physical quantities in engineering systems, a frequency-aware attention mechanism (Qin et al., 2021) is integrated to complement the standard attention framework. This enhanced Seq2Seq architecture, termed Echoformer, is evaluated for its mapping inference capabilities and compared with existing methods through practical case studies.

5.3.1 Multi-head attention mechanism and Transformer

To comprehensively address the long-range dependency issues discussed in Section 5.2.1 and enhance network memory capacity, researchers in computational science introduced the self-attention mechanism. This approach mimics human attentional capabilities to selectively extract valuable information from large datasets. For time series, self-attention evaluates the "importance" of signals at each timestep and encodes them via weighted summation. These weights are derived by measuring the similarity between the current timestep's signal and those at other timesteps, enabling Transformers to uniformly attend to the entire sequence during encoding, thereby improving memory capacity. However, this encoding method, which focuses solely on similarity and computes weights independently for each timestep, fails to preserve the sequential order of the time series. For instance, shuffling the sequence would yield identical encoded results under the attention mechanism. To retain the strict and valuable temporal relationships between signals, positional encoding must be applied prior to encoding. This involves embedding the temporal position t of each input $\boldsymbol{X}(t)$ into the signal itself, as formalized in Equation (4.16).

$$\boldsymbol{X}(t) \leftarrow \boldsymbol{X}(t) + \boldsymbol{PE}(t) \tag{5.16}$$

where, $\boldsymbol{PE}(t) \in \mathbb{R}^{N_{pe}}$ is a position-encoded vector with arbitrary input length, typically determined by sine-cosine functions (Vaswani et al., 2017). To ensure distinct positional information, N_{pe} is often set to tens or higher. However, this encoding method is

primarily suitable for single-channel inputs, as vector addition becomes infeasible for multi-channel scenarios.

Following positional encoding, attention computation begins, where $\boldsymbol{X}(t) \in \mathbb{R}^{N_{pe}}$. Taking the initially proposed temporal attention $\boldsymbol{A_t}(t) \in \mathbb{R}^{N_{att}}$ as an example, the final attention values derive from a weighted aggregation of the entire input sequence $\boldsymbol{X}(1) \sim \boldsymbol{X}(T)$. This operation is not performed directly on $\boldsymbol{X}$ but involves projecting it into an N_{att}-dimensional space to generate three critical components for temporal attention: the query ($\boldsymbol{Q_t} \in \mathbb{R}^{N_{att} \times T}$), key ($\boldsymbol{K_t} \in \mathbb{R}^{N_{att} \times T}$), and value ($\boldsymbol{V_t} \in \mathbb{R}^{N_{att} \times T}$). The projection matrices are often implemented as simple single-layer linear transformations, typically excluding bias terms and nonlinear activation functions, i.e.:

$$\boldsymbol{Q_t}(t) = \boldsymbol{W_{qt}}\boldsymbol{X}(t) \tag{5.17}$$

$$\boldsymbol{K_t}(t) = \boldsymbol{W_{kt}}\boldsymbol{X}(t) \tag{5.18}$$

$$\boldsymbol{V_t}(t) = \boldsymbol{W_{vt}}\boldsymbol{X}(t) \tag{5.19}$$

where, $\boldsymbol{W_{qt}}$, $\boldsymbol{W_{kt}}$, and $\boldsymbol{W_{vt}}$ are weight matrices, each with dimensions $\mathbb{R}^{N_{att} \times N_{pe}}$. To compute the temporal attention $\boldsymbol{A_t}(t)$ at a specific timestep, the query value $\boldsymbol{Q_t}(t)$ for that timestep is first dot-producted with the key values $\boldsymbol{K_t}$ across all timesteps to measure similarity. The resulting scores are normalized to form weights, which are then combined with the value data $\boldsymbol{V_t}$ from all timesteps. This process is formulated in matrix notation as follows:

$$\boldsymbol{A_t} = softmax\left(\frac{\boldsymbol{Q_t}\boldsymbol{K_t}^T}{\sqrt{N_{att}}}\right)\boldsymbol{V_t} \tag{5.20}$$

where, $softmax(\cdot)$ represents the normalization operation, with scaling applied to regulate the magnitude of the dot product. Studies have shown (Vaswani et al., 2017) that a single attention mechanism exhibits limited focus perspectives. To address this, multiple sets of query-key-value pairs, mapped through distinct transformations, are employed to form a multi-head self-attention mechanism. Within this framework, the network encodes multiple temporal attention values $\boldsymbol{A_t^{(1)}}, \boldsymbol{A_t^{(2)}}, \ldots, \boldsymbol{A_t^{(N_{head})}}$ following the same

methodology as Eqs. (4.17), (4.18), (4.19), and (4.20). These values correspond to diverse similarity computation patterns. The final multi-head temporal attention $\boldsymbol{MA_t}(t)$ is subsequently derived through a linear projection, expressed as:

$$\boldsymbol{MA_t}(t) = \boldsymbol{W_{mat}}\left[\boldsymbol{A_t^{(1)}}(t); \boldsymbol{A_t^{(2)}}(t); \dots; \boldsymbol{A_t^{(N_{head})}}(t)\right] \tag{5.21}$$

where, $\boldsymbol{W_{mat}} \in \mathbb{R}^{N_{pe} \times (N_{head} N_{att})}$ is the projection matrix, with N_{att} typically set as N_{pe}/N_{head}. The length of $\boldsymbol{MA_t}(t)$ is constrained to N_{pe} to enable residual connections between $\boldsymbol{MA_t}$ and the input $\boldsymbol{X}(t)$, which stabilizes deep network training. Additionally, the output from the attention module is further processed by a feed-forward neural network (FFNN) to produce the final encoder output. The FFNN, comprising two fully connected layers, first maps the input to a higher-dimensional space and then reduces it back, primarily introducing nonlinearity and increasing network depth. The final encoder output can thus be expressed as:

$$norm\big(\boldsymbol{W_{fn2}} f\big(\boldsymbol{W_{fn1}} norm\big(\boldsymbol{MA_t}(t) + \boldsymbol{X}(t)\big) + \boldsymbol{b_{fn1}}\big) + \boldsymbol{b_{fn2}}\big) \tag{5.22}$$

Here, $norm(\cdot)$ denotes the layer normalization operation; $\boldsymbol{W_{fn1}} \in \mathbb{R}^{N_{fn} \times N_{pe}}$ and $\boldsymbol{W_{fn2}} \in \mathbb{R}^{N_{pe} \times N_{fn}}$ represent the two projection matrices, with $\boldsymbol{b_{fn1}}$ and $\boldsymbol{b_{fn2}}$ being the bias terms.

The above describes the computational flow of a single encoder in the Transformer. In practice, multiple encoders can be stacked, which necessitates the repeated application of normalization in Eq. (5.22). The original Transformer architecture also includes a decoder, primarily designed for prediction tasks through autoregressive structures and masked attention mechanisms. However, in synchronized mapping problems, the output $\boldsymbol{Y}(t)$ can be directly obtained from the value at the tth step of the final encoder. This implies a structural equivalence between the encoder and decoder in such scenarios, which is not elaborated here. The output layer of the model employs a fully connected network (MLP), and the loss function is constructed using the MSE, both of which have been introduced in preceding chapters and are thus not reiterated. The architecture of the standard Transformer is illustrated in Fig. 5.15.

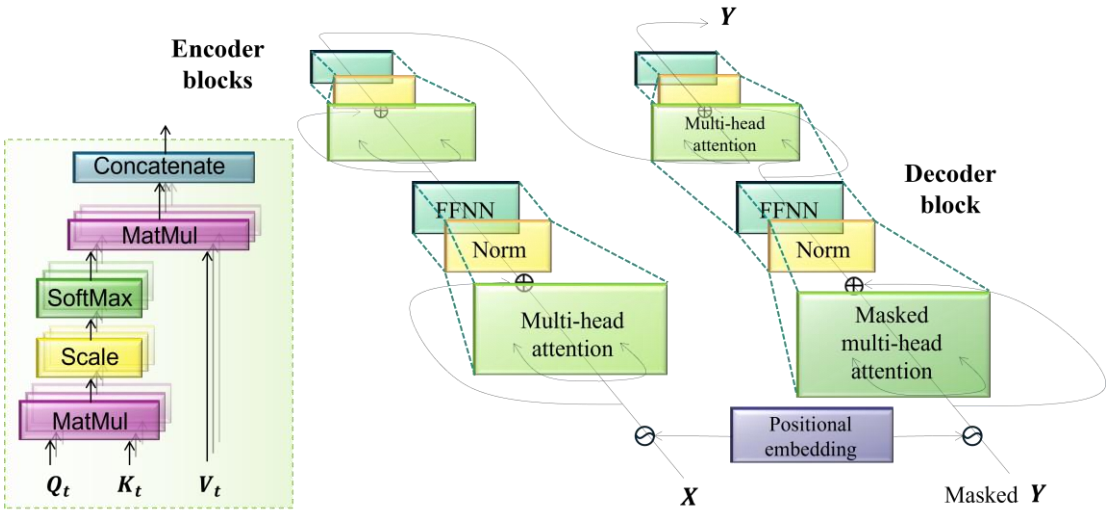


Fig. 5.15 Transformer

5.3.2 Echoformer

The Echoformer proposed in this study introduces two key improvements to the standard Transformer architecture. First, frequency-domain information (Qin et al., 2021) is incorporated into the attention mechanism by converting the signal into the frequency domain via FFT before computing attention. The frequency-domain attention mechanism focuses on the similarity of amplitude values across different frequency points, which inherently models numerical relationships between these amplitudes. By capturing and fitting these relationships, the model can more accurately identify fundamental patterns in signal fluctuations.

To the learn frequency-domain information, an additional set of query-key-value pairs ($\boldsymbol{Q_f}$, $\boldsymbol{K_f}$, and $\boldsymbol{V_f}$) with the same dimensionality is introduced. These are derived through linear transformations of the spectral sequence using trainable weight matrices, that is, $\boldsymbol{W_{qf}}$, $\boldsymbol{W_{kf}}$, and $\boldsymbol{W_{vf}}$, which share the same dimensions as their time-domain counterparts. Specifically:

$$\boldsymbol{Q_f}(t) = \boldsymbol{W_{qf}}\mathcal{F}\big(\boldsymbol{X}(t)\big) \tag{5.23}$$

$$\boldsymbol{K_f}(t) = \boldsymbol{W_{kf}}\mathcal{F}\big(\boldsymbol{X}(t)\big) \tag{5.24}$$

$$\boldsymbol{V_f}(t) = \boldsymbol{W_{vf}}\mathcal{F}\left(\boldsymbol{X}(t)\right) \tag{5.25}$$

where $\mathcal{F}(\cdot)$ stands for the FFT operation. Afterwards, the frequency-domain attention $\boldsymbol{A_f}$ can be derived by Eq. (5.26).

$$\boldsymbol{A_f} = \mathcal{F}^{-1}\left(softmax\left(\frac{\boldsymbol{Q_f}{\boldsymbol{K_f}}^T}{\sqrt{N_{att}}}\right)\boldsymbol{V_f}\right) \tag{5.26}$$

where, $\mathcal{F}^{-1}(\cdot)$ denotes the inverse FFT operation. The above equation describes the single-head attention mechanism. The Echoformer retains the multi-head concept in the frequency domain, and thus the frequency-domain multi-head self-attention can be formulated as:

$$\boldsymbol{MA_f}(t) = \boldsymbol{W_{maf}}\left[\boldsymbol{A_f^{(1)}}(t); \boldsymbol{A_f^{(2)}}(t); \dots; \boldsymbol{A_f^{(N_{head})}}(t)\right] \tag{5.27}$$

Here, $\boldsymbol{W_{maf}} \in \mathbb{R}^{N_{pe}\times(N_{head}N_{att})}$ is the projection matrix. To prevent interference between time- and frequency-domain information, N_{pe} is doubled, i.e., $N_{pe} \leftarrow 2N_{pe}$, such that $N_{att} = N_{pe}/2N_{head}$, ensuring numerical consistency with the configuration prior to the introduction of frequency-domain attention. Finally, the final output of each encoder should be modified from Eq. (5.22) to the following form:

$$norm\left(\boldsymbol{W_{fn2}}f\left(\boldsymbol{W_{fn1}}norm\left(\left[\boldsymbol{MA_t}(t); \boldsymbol{MA_f}(t)\right] + \boldsymbol{X}(t)\right) + \boldsymbol{b_{fn1}}\right) + \boldsymbol{b_{fn2}}\right) \tag{5.28}$$

Another improvement in Echoformer addresses the aforementioned positional encoding limitations. The original model employs a fixed set of sine and cosine functions for encoding, where each position is represented by a vector of length N_{pe}, comprising $N_{pe}/2$ pairs of sine and cosine values at timestep t with varying frequencies (Vaswani et al., 2017). This encoding, $\boldsymbol{PE}(t)$, offers advantages such as bounded values and uniqueness for specific timesteps. However, these benefits rely on a sufficiently large N_{pe}, while the input dimension N_{in} of $\boldsymbol{X}(t)$ is often determined by the problem, leading to dimensionality mismatch. To resolve this, dimensional expansion is typically required, but arbitrarily increasing N_{pe} significantly impacts model complexity and parameter count, which is impractical.

A viable solution involves determining N_{pe} based on signal duration and hardware constraints, followed by encoding and expanding $\boldsymbol{X}(t)$ to match N_{pe}. Two critical requirements for this expansion are: the expanded signal must retain all information from $\boldsymbol{X}(t)$, preferably with enhanced feature separation and extraction; and the expansion process must be computationally decoupled from training to avoid increasing network complexity. Traditional approaches, such as RNN-based recursive structures for dimensional expansion, risk undermining Transformer advantages if their parameters are trainable.

To address this, Echoformer employs a random reservoir for dimensionality expansion. As detailed in Section 4.3.1, random reservoir encoding enables feature extraction without requiring training. By setting $N_{res} = N_{pe}$, $\boldsymbol{X}(t)$ is mapped to a higher-dimensional representation R(t)$R(t)$, where each channel in $\boldsymbol{R}(t)$ captures distinct interpretable features of $\boldsymbol{X}(t)$ from varying perspectives. When the reservoir mapping matrix is well-designed, $\boldsymbol{R}(t)$ itself exhibits uniqueness at each timestep, potentially serving directly as positional encoding. Simplified by removing state-to-output connections (as in Section 4.3.1), the positional encoding process in Echoformer is formalized in Eqs. (5.29) and (5.30).

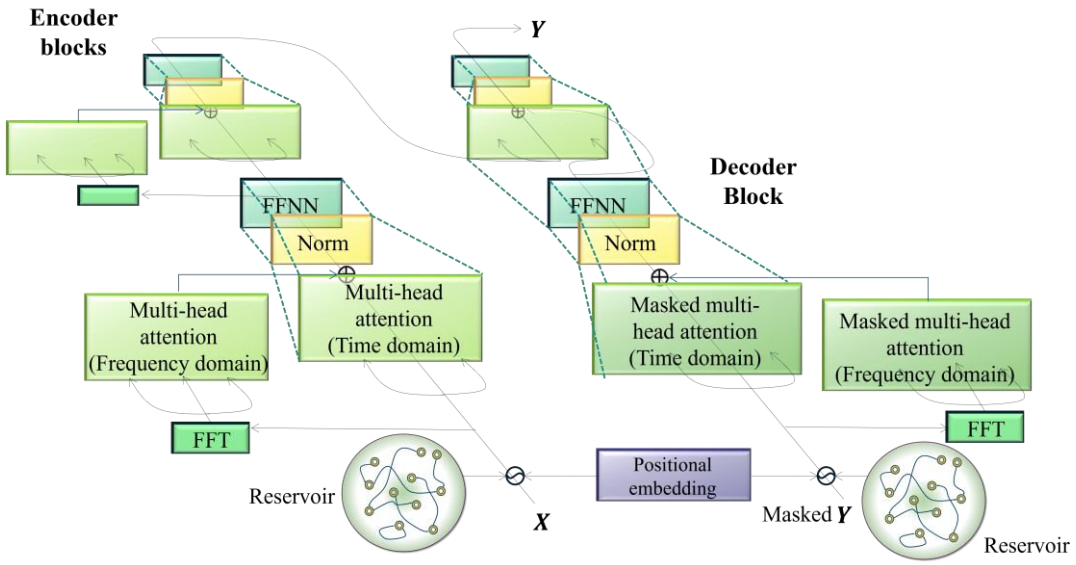


Fig. 5.16 Echoformer

$$\boldsymbol{R}(t) = \boldsymbol{W_{in}}\boldsymbol{R}(t) + \boldsymbol{W_{res}}(t)\boldsymbol{R}(t-1) \tag{5.29}$$

$$\boldsymbol{X}(t) \leftarrow \theta\boldsymbol{R}(t) + (1-\theta)\boldsymbol{PE}(t) \tag{5.30}$$

where, θ is a hyperparameter that controls the combination of the two encoding schemes. Given that $\boldsymbol{R}(t)$ and $\boldsymbol{PE}(t)$ exhibit similar magnitude distributions, θ is typically set to approximately 0.5. The architecture of Echoformer is illustrated in Fig. 5.16.

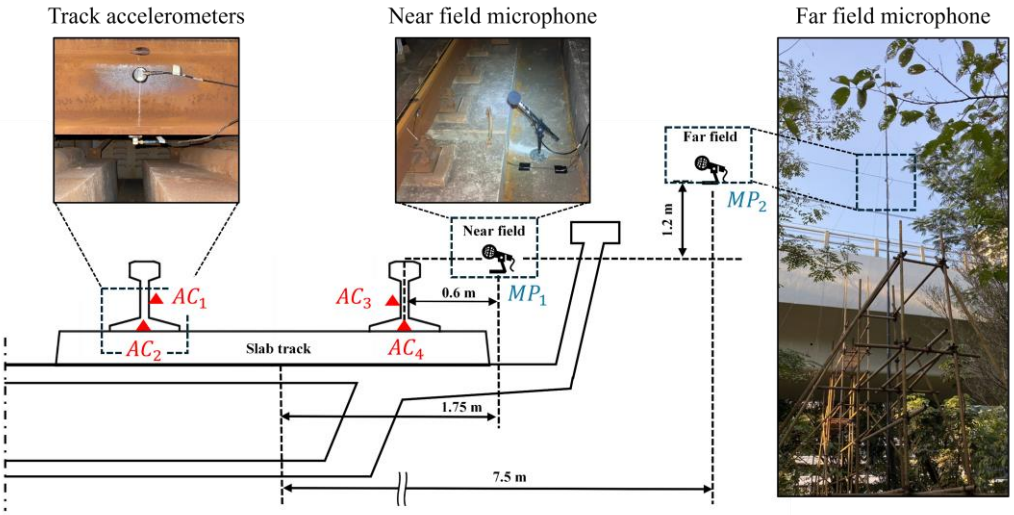
5.3.3 Application: Noise test of Shenzhen Metro Line 5

Noise has become a primary source of complaints against metro operators. As illustrated in Fig. 5.17(a), the section between Tanglang Station and Daxuecheng Station on Shenzhen Metro Line 5 exemplifies this issue. This segment, adjacent to a residential area, is classified as a noise-sensitive zone. The trains operating on this line consist of six-carriage A-type metro units (as defined by the code TB 10624-2020), traversing the monitored section at approximately 70 km/h. Given the straight alignment of this metro segment, the dominant noise source is rolling noise generated by wheel-rail contact-induced track vibrations, which propagate radially. Noise affecting nearby residents primarily constitutes far-field noise, attenuated by sound barriers along the tracks and medium dispersion over distance. To validate compliance of far-field noise with regulatory standards, characterize noise propagation mechanisms, and inform scientifically optimized mitigation strategies, simultaneous monitoring of track vibration, near-field noise, and far-field noise was conducted in this section.

Four triaxial accelerometers (Dytran, 3023A4) were deployed to measure vibration responses, with their signals denoted as $\boldsymbol{AC_1}$~$\boldsymbol{AC_4}$. These sensors were installed on the rail webs and bases of both tracks. The deployment configuration is illustrated in Fig. 5.17(b). Although the accelerometers recorded responses in three directions (vertical, lateral, and longitudinal), only vibrations in the vertical (ground-normal, Direction-1) and lateral (Direction-2) directions were analyzed. For instance, the two internal channels of the first accelerometer were labeled $\boldsymbol{AC_{1,1}}$ and $\boldsymbol{AC_{1,2}}$. Additionally, two microphones (B&K, Type 4189) were utilized to monitor rolling noise near the tracks. The first

microphone ($\boldsymbol{MP_1}$) was positioned at railhead height, 0.6 m from the track, to capture near-field noise. The second microphone ($\boldsymbol{MP_2}$) was installed 1.2 m above the railhead and 7.5 m from the track centerline to measure far-field noise. To facilitate the installation of $\boldsymbol{MP_2}$, a scaffold was erected near the viaduct to support a lightweight pole. During the field test, all sensors were sampled at 10,000 Hz, as the dominant frequency components of the rolling noise were below 2,000 Hz.

(a) Railway segment



(b) Sensor arrangement

Fig. 5.17 Noise test programme

The noise data measured during a train pass-by event is shown in Fig. 5.18, with the dominant noise occurring at approximately 630 Hz. The peak sound pressure levels (SPLs) reached 111.0 dB(A) in the near field and 93.8 dB(A) in the far field, while the overall SPLs were 95.8 dB(A) and 78.4 dB(A) for the near and far fields, respectively. Concurrently recorded acceleration data, presented in Fig. 5.19, revealed that rail vibration energy predominantly concentrated in Direction-1 based on time-frequency

spectral analysis. Although minor differences existed in vibrations between the two rails, acceleration signals at distinct positions (rail web and base) on the same rail were nearly identical, attributed to the rigid cross-section of the rail. Distinct pressure spikes coincided with wheel passages through the monitored section. Data segmentation was performed according to wheel presence or absence to extract train-induced signals. The training set comprised 3.65 s (27,500 data points) of data spanning the passage of the first four wheels, including brief intervals without wheel contact. The subsequent 2.75 s (27,500 data points) of signals, covering from the fifth wheel passage until the final wheel exited the monitoring zone, formed the testing set.

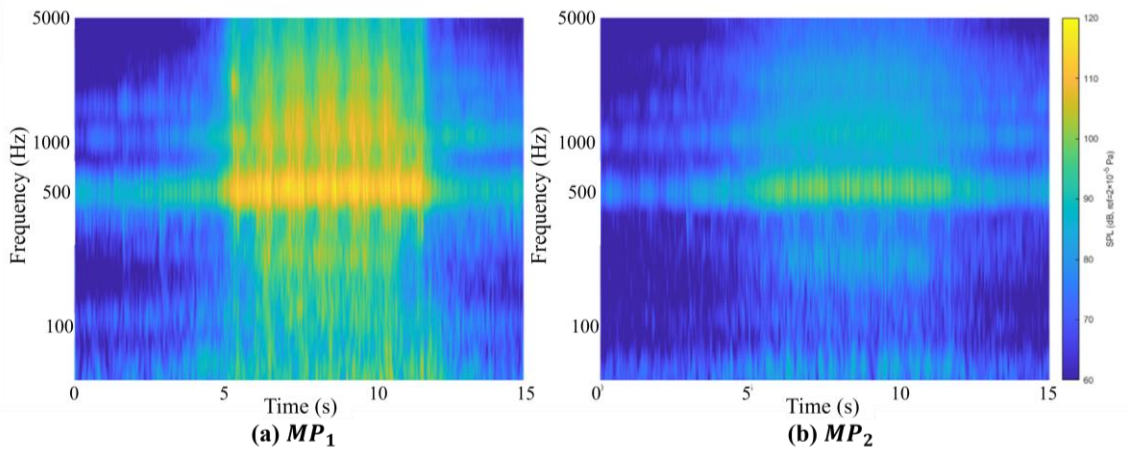


Fig. 5.18 Time-frequency diagram of noise measurement

The installation and long-term operation of microphones are economically impractical, particularly given that scaffolding for far-field noise measurements can only be temporarily erected. This necessitates a more stable noise monitoring solution. Notably, the dominant frequency ranges of track vibrations align with those of wayside noise, indicating a strong correlation between track vibration and rolling noise. This observation underscores the potential for predicting wayside noise based on track vibration. After time-shifting the far-field noise data according to the sound propagation velocity, the task of inferring near- and far-field wayside noise from track vibration accelerations becomes a synchronous data mapping problem. Furthermore, the successful development of such a mapping algorithm could enable quantitative assessment of the proportion of regular

track vibration contributions to wayside noise during operational phases. This capability may further support applications such as wheel defect diagnosis and rail surface damage evaluation.

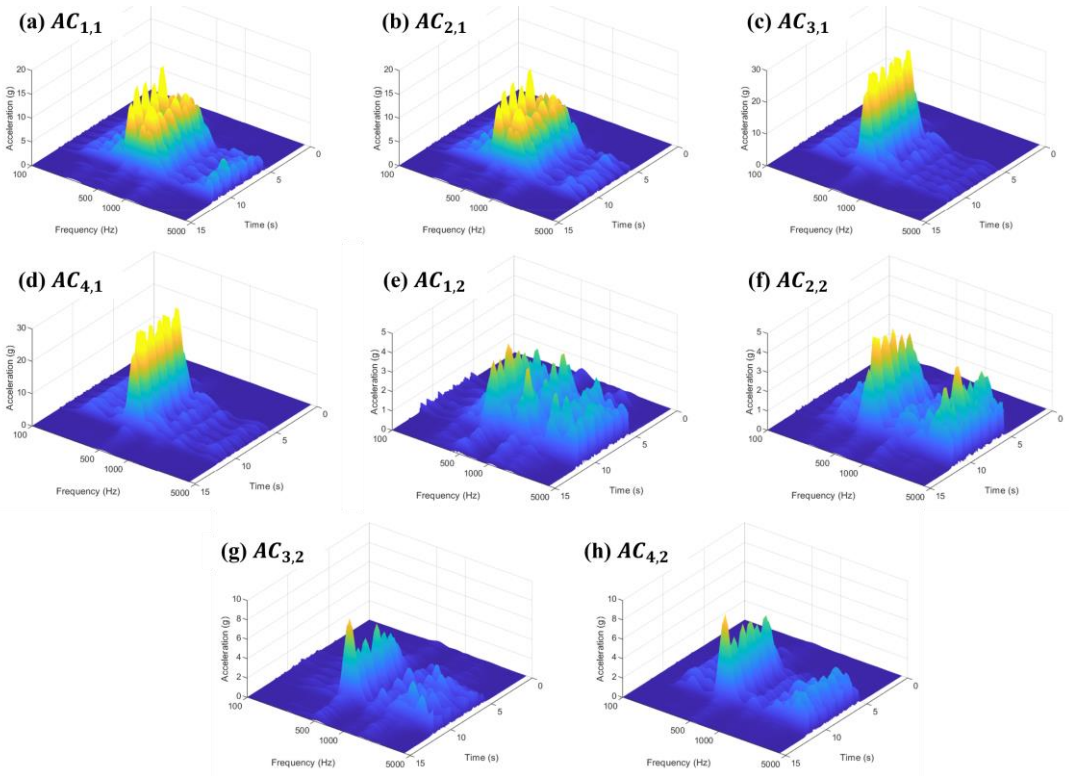


Fig. 5.19 Time-frequency diagram of acceleration measurement

5.3.4 Synchronized mapping results and discussion

This section focuses on evaluating the performance of Echoformer in synchronously mapping near- and far-field wayside noise based on track vibrations and train speed, with comparisons against baseline models including the original Transformer. Additionally, the robustness of the proposed method is investigated for scenarios involving data contaminated by environmental noise or measurement errors.

Critical hyperparameters in Echoformer were determined through a grid search approach as described in Section 4.3.4.3, with optimization results detailed in Table 5.4. The model architecture was configured with two encoder and two decoder layers, while the number of attention heads (N_{head}) was set to 10. Consequently, the temporal and spectral attention modules each contained $N_{att} = 50$ parameters. Training employed the SGD optimizer

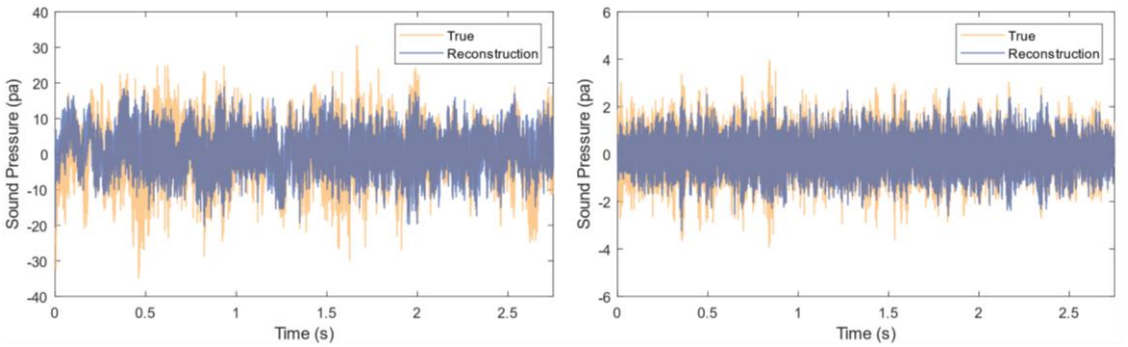
with a learning rate $lr = 0.0001$ and $\eta = 0.001$, and the training spanned 50 epochs.

Table 5.4 Hyperparameters of Echoformer

N_{res}	α	$\rho(W_{res})$	$spar(W_{res})$
1000	0.9	0.8	$2.0N_{res}$

5.3.4.1 Synchronized noise mapping results

Fig. 5.20 illustrates the inference results of Echoformer across the entire test dataset. At the experimental site, the near-field acoustic environment exhibited greater complexity than the far field due to incomplete attenuation of multiple acoustic components. Consequently, the sound pressure amplitudes in the far field remained relatively stable, which facilitated simpler noise mapping, as shown in Fig. 5.11(b). Taking the near-field noise as an example, zoomed-in segments (0.1 s duration) of the inferred results are presented in Figs. 5.21 and 4.22, corresponding to scenarios with and without wheel passages, respectively. These figures also demonstrate the spectral agreement between the test signals and inferred signals.



(a) MP_1 (b) MP_2

Fig. 5.20 Full-time mapping results

The acoustic energy was observed to concentrate and slightly increase during wheel passages. However, the time window of wheel passage was too brief to reveal significant differences in the operational conditions depicted in Figs. 5.20 and 5.21. For both scenarios, Echoformer successfully captured the temporal variations of the signals while

maintaining strong spectral alignment. Based on the mapped noise signals, SPLs in 1/3-octave bands were calculated, along with equivalent continuous sound levels (L_{eq}) and peak sound levels (L_{10}) (Namba, Kuwano & Kato 1978). Given the measured sound pressure p_s and reference pressure $p_0 = 2 \times 10^{-5}\, Pa$, the SPL can be computed Eq. (5.31).

$$SPL = 20\, log_{10}\left(\frac{p_s}{p_0}\right) \tag{5.31}$$

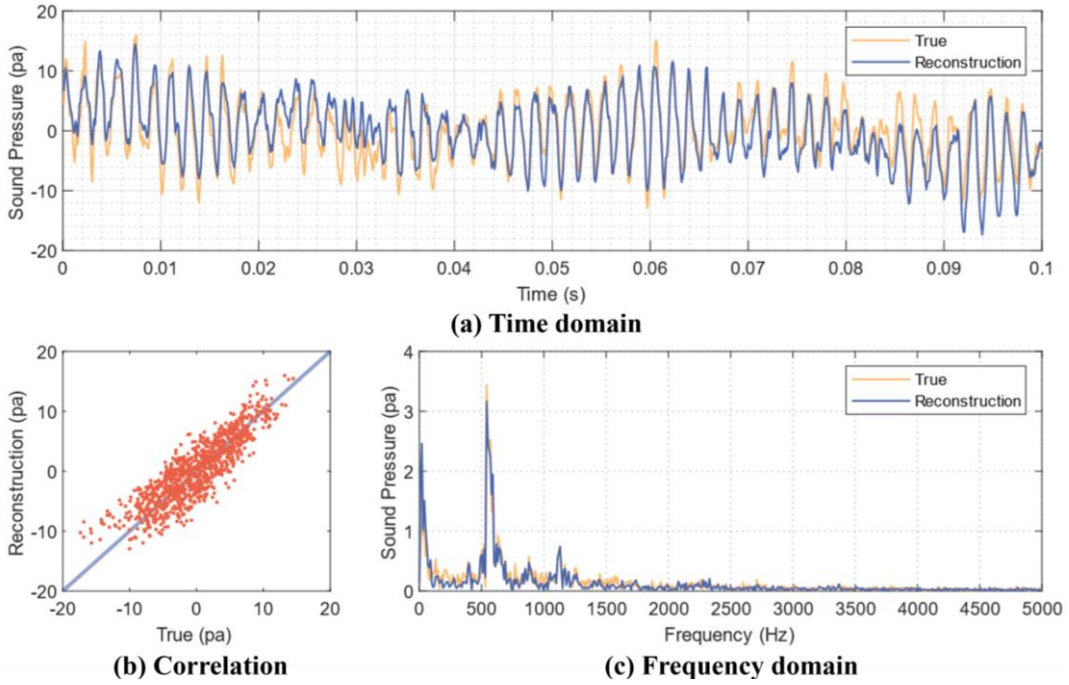


Fig. 5.21 MP_1 mapping results when wheel passing

The SPL analysis results are presented in Figs. 5.23 and 5.24. Transformer, ESN, and LSTM were selected as baseline models to validate the superiority of Echoformer. Additionally, since Echoformer can be interpreted as an ESN appended with a Transformer as the output layer, a variant of ESN followed by a standard MLP, termed echo state deep network (ESDN), was included in the comparative analysis to highlight the advantages of attention-based architectures.

Due to the vanishing gradient problem, the standard LSTM failed to retain long-term historical dependencies, which hindered its ability to learn long-term temporal features. Consequently, LSTM exhibited inferior performance in both near- and far-field noise

inference, with its predicted L_{eq} and L_{10} values being largely unacceptable. In contrast, the other two RNN-based methods, ESN and ESDN, demonstrated improved performance. However, the noise magnitudes inferred by ESN and ESDN in the one-third-octave bands were generally underestimated compared to the ground truth, and their mapping errors increased with frequency. Nevertheless, both methods successfully captured the overall trends of the target signals across the one-third-octave bands. Notably, the performance of ESDN showed no significant improvement over ESN, which employs only a simple linear regression output layer, as evidenced by the results in the figures. This observation suggests that replacing the linear output layer with an MLP is unnecessary, further confirming the effectiveness of the reservoir computing paradigm in this task.

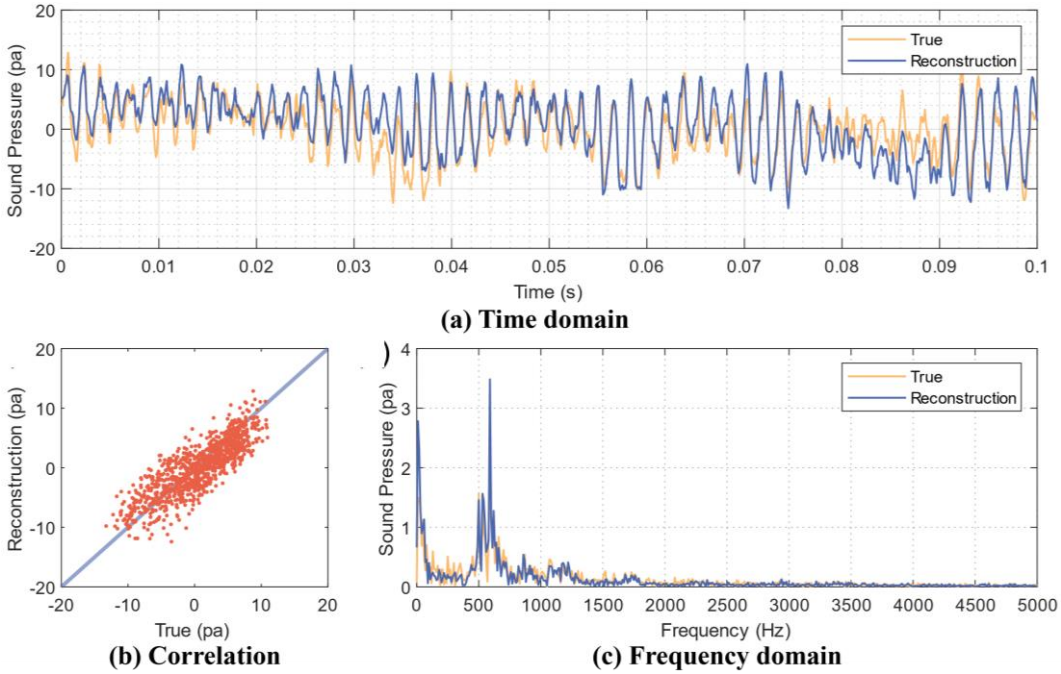


Fig. 5.22 MP_2 mapping results when wheel passing

The Transformer-based architecture achieved relatively accurate predictions of L_{eq} and L_{10}, indicating its potential for coarse analysis of railway noise. However, closer inspection of the one-third-octave band results reveals limitations in the Transformer's ability to capture the frequency characteristics of near-field noise. Additionally, in far-field noise mapping, the Transformer overestimated amplitudes in frequency ranges

above 2,000 Hz. The proposed Echoformer integrates the strengths of both approaches: the reservoir's frequency feature extraction capability and the Transformer's deterministic noise-level estimation. Experimental results demonstrate that Echoformer delivers the highest-quality noise mapping for both near- and far-field scenarios in terms of frequency and amplitude accuracy. For near-field noise with wheel passages, the inferred L_{eq} and L_{10} values were 108.37 and 113.1, respectively, closely matching the ground truth values of 111.6 and 116.0. Similarly, for far-field noise, the inferred and true values were 91.64 vs. 93.32 (L_{eq}) and 95.92 vs. 97.72 (L_{10}). Comparable performance was observed in scenarios without wheel passages.

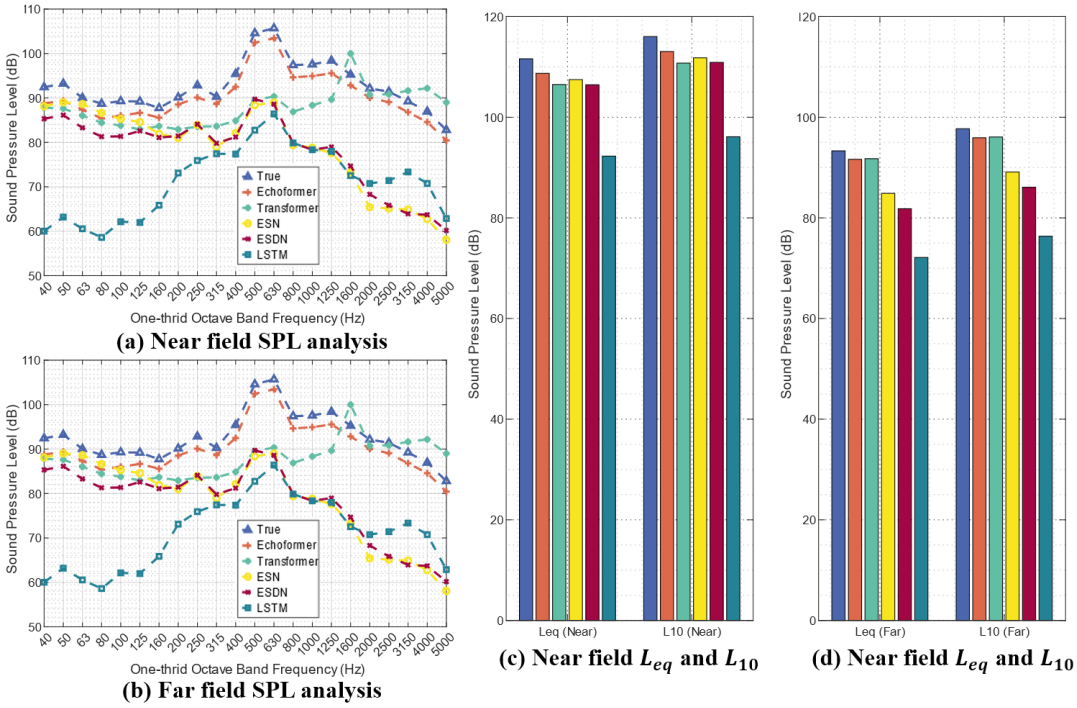


Fig. 5.23 Noise mapping accuracy comparison when wheel passing

As Echoformer is primarily designed as a cost-effective monitoring alternative to mitigate operational expenses and address potential sensor failures in wayside noise monitoring systems, the necessity of eight input channels from four sensors was further investigated. This analysis also evaluates Echoformer's performance under reduced sensor configurations. As shown in Table 5.5, the model's inference capability was tested by progressively removing measurement points. Tested configurations included the original

setup (T-F-2), configurations with only rail web ("W") or base ("B") sensors (T-W-2 and T-B-2), and setups with single-side rail web or base sensors (T-W-1 and T-B-1), where the suffix digit indicates the number of measurement points per location. Configurations combining single-side web and base sensors were excluded due to their high signal similarity, which would yield results nearly identical to T-W-1 and T-B-1.

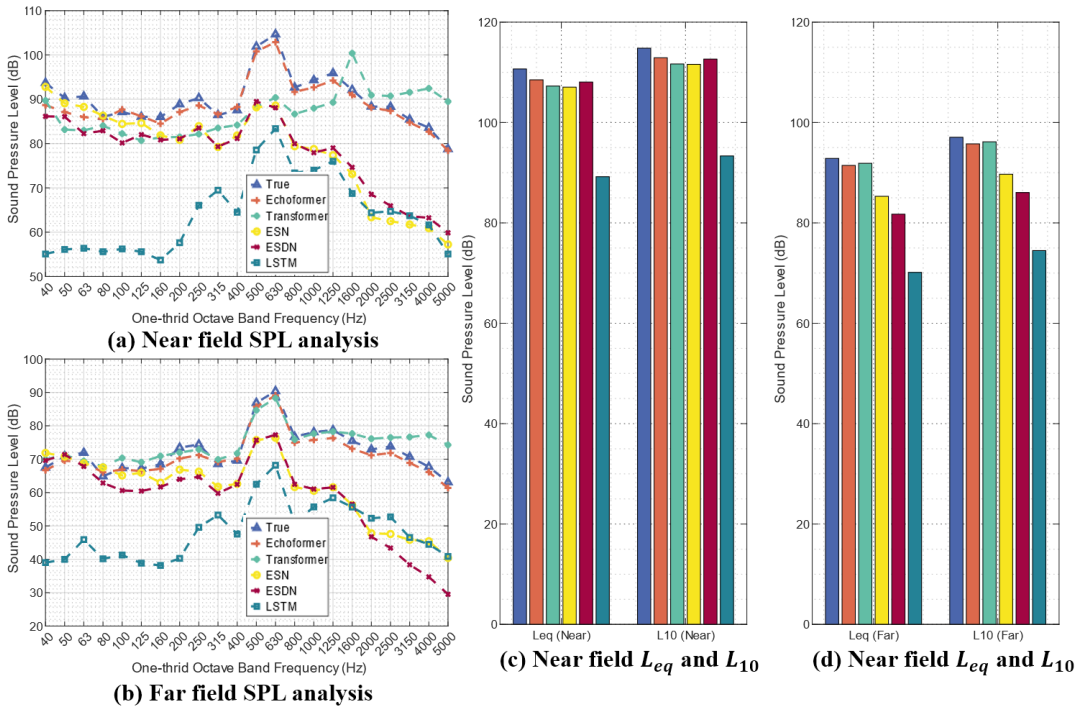


Fig. 5.24 Noise mapping accuracy comparison when no wheel passing

For visualization purposes, the total 2.75 s test dataset was segmented into 55 intervals, each spanning 0.05 s (500 data points). The segment-wise MSE calculated for the five models is illustrated in Fig. 5.25. Evidently, Echoformer consistently outperformed all other methods across the five tested configurations with varying data richness. However, for the more challenging near-field noise inference task, Echoformer exhibited greater error variability compared to the Transformer. In this task, ESN and ESDN also demonstrated substantial error standard deviations, indicating that reservoir-based methods occasionally produced segments with anomalously high (or low) MSE values. This observation suggests that the stochastic nature of the reservoir may compromise temporal stability in inferred signals. Notably, LSTM remained the worst-performing

model practically. Across different sensor configurations, removing sensor information did not lead to significant performance variations. The proposed method maintained robust inference capabilities even when only a single accelerometer was utilized as the input signal. This resilience likely stems from the strong physical correlation between track vibrations and railway rolling noise, enabling sufficient information extraction from a single accelerometer for synchronized noise mapping.

Table 5.5 Noise mapping conditions

Dataset	Known channel	Unknown channel	Input channel number
T-F-2	AC_1 AC_2, AC_3, AC_4	/	8
T-W-2	AC_1, AC_3	AC_2, AC_4	4
T-B-2	AC_2, AC_4	AC_1, AC_3	4
T-W-1	AC_2	AC_1, AC_3, AC_4	2
T-B-1	AC_4	AC_1, AC_2, AC_3	2

5.3.4.2 Influence of measurement error

In SHM systems, noise contamination in measurement data is inevitable. In railway noise data, the reduced differentiability between extraneous noise and valid information further diminishes the effectiveness of conventional denoising methods. In this section, the robustness of Echoformer is investigated under conditions where input signals are contaminated with random white noise. The intensity of white noise added to the training dataset is incrementally increased to evaluate Echoformer's performance under varying contamination levels. The model's capability and stability in inferring rail noise are examined by comparing signal clarity against noise interference levels. The white noise intensity was determined based on the energy of each input channel. Four noise levels

were considered, with white noise components accounting for 10%, 30%, 50%, and 70% of the input signal energy.

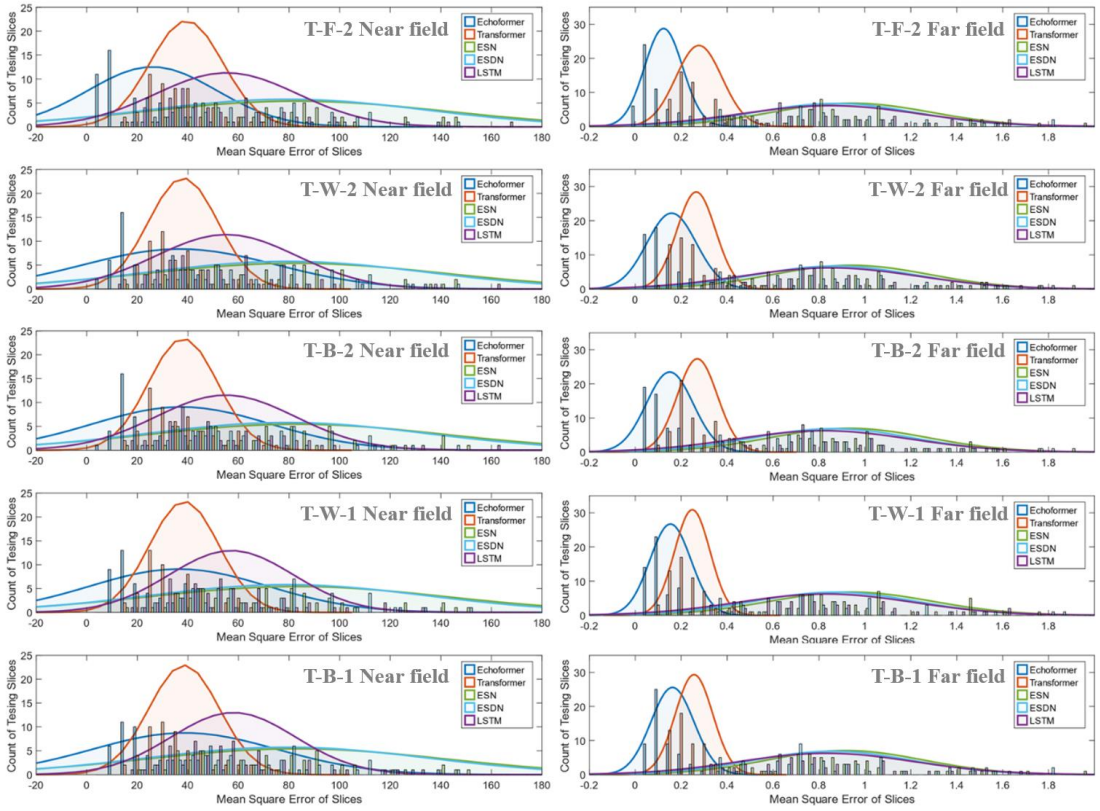


Fig. 5.25 Noise mapping results of different data integrity

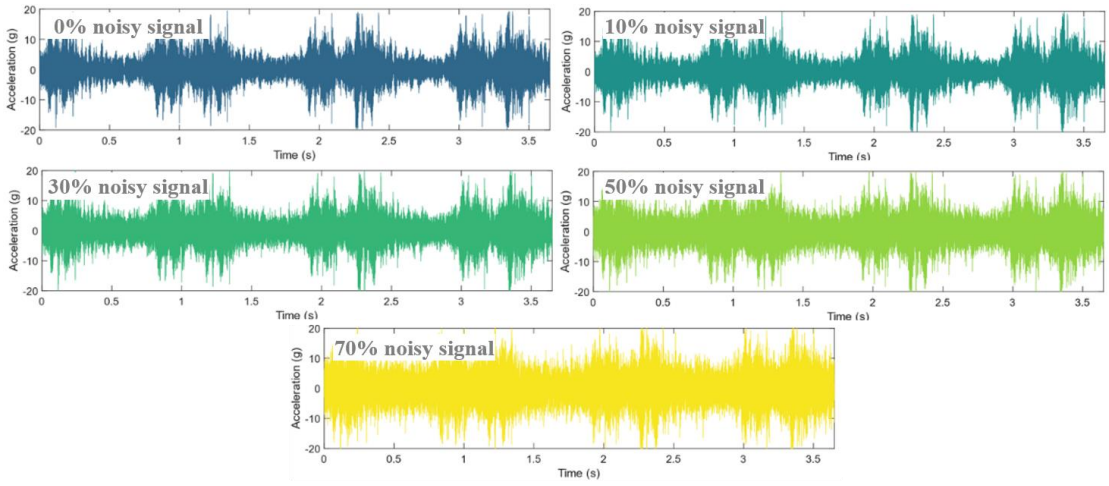
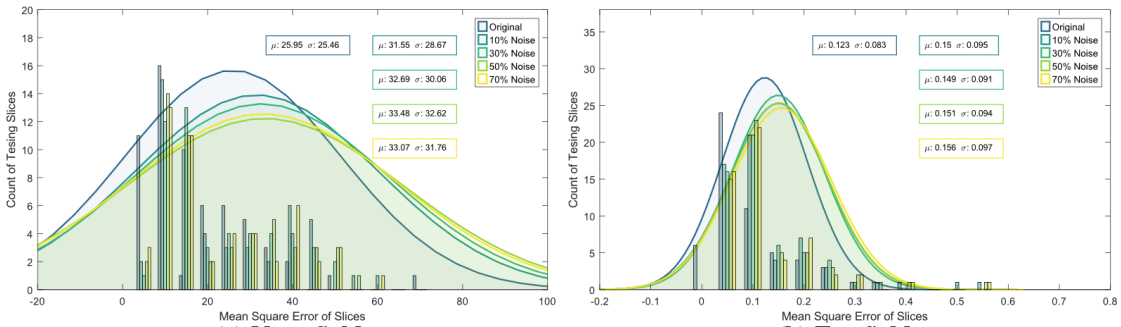


Fig. 5.26 Input series after pollution

Taking $\boldsymbol{AC_{1,1}}$ as an example, Fig. 5.26 illustrates the acceleration signals of this channel under different white noise contamination levels. Following the methodology described

earlier, the test set was segmented into 55 intervals for evaluation. Statistical results of Echoformer's MSE for near- and far-field noise inference across these segments are presented in Fig. 5.27. Notably, the impact of white noise on prediction accuracy was uniformly distributed across the test set, with mapping errors generally increasing with higher noise levels. Overall, the degradation in model performance was insignificant for both near- and far-field inference tasks. Two factors likely contributed to this robustness: First, the uniform spectral energy distribution of white noise minimally disrupts the stochastic reservoir's frequency-specific feature extraction, as localized frequency components remain relatively unaffected. Second, the self-attention mechanism in Echoformer inherently focuses on intra-signal correlations, which are preserved in the original data and remain largely unaltered by uncorrelated noise. This architecture exhibits natural insensitivity to noise interference, as the algorithm prioritizes intrinsic signal relationships over stochastic perturbations. Consequently, Echoformer demonstrates strong noise resistance and maintains reliable performance under contaminated input conditions. The extremely high noise resisting ability of the method mainly comes from two aspects. Firstly, the acquisition noise is only added into the input of the training data, which means the model trained by this strategy inherently has ability to identify valuable information from polluted data. Secondly, the added noises are white noise with uniformly distributed frequency components. But the railway noise dominantly contains components of a narrow frequency band, which weakens the effect introduced by acquisition noise.



(a) Near field **(b) Far field**

Fig. 5.27 Noise mapping results after polluted

5.4 Dynamic system solving

In general, any inference task (prediction, synchronized mapping, etc.) between physically correlated signals can be regarded as solving the governing equations of the corresponding physical system. For instance, in the wind and acoustic field propagation cases discussed in Sections 5.2 and 5.3, the target outputs are uniquely determined by fluid motion equations and wave propagation equations given inputs and boundary conditions. However, due to challenges in defining boundary conditions, computational complexity, lack of analytical solutions, and incomplete understanding of equation formulations, the adopted solution employs NNs to approximate transfer functions using measurement data. In such scenarios, where inferences involve homogeneous physical quantities (e.g., historical-to-future wind velocities or near-field-to-far-field vibrations), the mapping relationships exhibit relative simplicity and stability, enabling effective implementation. Physical systems, however, often involve heterogeneous physical quantities, as exemplified by common post-processing tasks in SHM systems. For instance, mappings between excitation and response signals (e.g., wind pressure to acceleration, seismic motion to strain) are governed by system-specific control equations. These relationships dynamically evolve with system modifications, such as reduced stiffness due to structural damage or resonance emergence at specific frequencies, rendering conventional networks ineffective for such inferences.

When learning input-output mappings mediated by physical systems, model development must account not only for signal correlations but also for temporal variations in mapping relationships caused by system evolution. This necessitates prioritized consideration of model generalization performances. This section investigates approaches for enhancing NN's generalization performance in solving typical dynamic systems through time-series modeling. A generalized architecture termed generalization performance expansion structure (GPES) is proposed and implemented in a temporal NN framework, resulting in an enhanced example network (expanded echo state network (Ex-ESN)). The modified network is subsequently applied to practical dynamic system modeling to evaluate its

accuracy improvements.

5.4.1 Physical system varying with key variables and GPES

The fundamental approach of NN-based models in solving physical systems lies in learning the mapping relationship (transfer function) between input and output variables, primarily encoded in trained weight parameters. However, these parameters remain fixed post-training, rendering them unsuitable for practical physical systems, which often exhibit dynamic variations requiring adaptive linear or nonlinear mapping relationships. Typically, physical systems evolve according to key variables with discernible patterns, such as structural stiffness degradation due to damage or temperature variations, or damper behavior modulated by excitation frequency. This study focuses exclusively on systems governed by such variable-dependent patterns, excluding completely chaotic or stochastic systems. For such systems, fixed-weight models inherently demonstrate limited generalization performances.

To address this limitation, we propose GPES, a universal plug-in for expanding model generalization performance, applicable to nearly all NN-based architectures. In this framework, the model is conceptualized as a continuous generalized function of key variables, where each weight parameter becomes a real-valued function of these variables. Consider a simplified scenario with a single key variable (denoted as $\Psi \in \mathbb{R}$) and a generic time-series NN comprising N_l layers, where the weight parameters of the l th layer are denoted as $\boldsymbol{W}^{(l)}$ $(l = 1,2,\dots,N_l)$. The output of the model under this framework is expressed by Eq. (5.32).

$$\boldsymbol{Y} = f\left(\boldsymbol{W}^{(N_l)} f\left(\cdots f\left(\boldsymbol{W}^{(1)}\boldsymbol{X}\right)\right)\right) \tag{5.32}$$

In conventional model training approaches, all sampled data pairs $(\boldsymbol{X}, \boldsymbol{Y})$ are aggregated for training regardless of the key variable Ψ values during data acquisition. However, in practice, variations in Ψ correspond to distinct underlying mapping relationships within the samples. To address this, GPES initiates the training process with dataset partitioning based on Ψ values. Suppose the dataset contains observations collected from systems

under N_ψ distinct Ψ values, denoted as $\Psi_1, \Psi_2, \ldots, \Psi_{N_\psi}$. The dataset is then divided into N_ψ subgroups, where the ψ-th subgroup containing dataset $(\boldsymbol{X_\psi}, \boldsymbol{Y_\psi})$ corresponds to observations acquired under the key variable value Ψ_ψ. These subgroups are subsequently used to independently and in parallel train N_ψ sub-models. Each sub-model can be formally expressed as:

$$\boldsymbol{Y_\psi} = f\left(\boldsymbol{W_\psi^{(N_l)}} f\left(\cdots f\left(\boldsymbol{W_\psi^{(1)}} \boldsymbol{X_\psi}\right)\right)\right) \tag{5.33}$$

Consequently, N_ψ groups of weight matrices with identical dimensions are obtained, each representing the input-output mapping relationships of the physical system under specific values (and their neighborhoods) of the key variable Ψ. Since this approach involves only dataset partitioning without altering the sub-model architecture, the computational cost remains equivalent between training a single aggregated model and training partitioned sub-models. All sub-models employ identical training protocols, and the training process can be parallelized across multiple devices. The core rationale of GPES lies in the assumption that sufficiently fine-grained partitioning of Ψ ensures negligible behavioral discrepancies between adjacent Ψ values. Under this condition, each sub-model learns stable mapping relationships without requiring trade-offs between conflicting mapping patterns in the training data, thereby achieving higher convergence efficiency and accuracy.

After training all N_ψ sub-models, they are interpolated to construct a continuous NN. During this fitting process, the varying mapping relationships are treated as functions of Ψ, with each sub-model representing a discrete point on the function curve, as shown in Fig. 5.28. The weight parameters of this continuous NN, denoted as $\boldsymbol{W^{(1)}}(\Psi) \sim \boldsymbol{W^{(N_l)}}(\Psi)$, are derived by interpolating the weight matrices of the sub-models. These functions take the parameter values of the corresponding pre-trained sub-model at each discrete point Ψ_ψ, such that the element in the ith row and jth column can be calculated as:

$$W_{ij}^{(l)}(\Psi_\psi) = W_{\psi,ij}^{(l)} \tag{5.34}$$

For the parameter at lth layer, ith row, and jth column, the fitting process involves interpolating N_ψ discrete points into a curve in the Ψ-W space. Given that the weight parameter matrices at each layer l have dimensions $h_l \times b_l$, the total number of parameter fittings required is $N_W = \sum_{l=1}^{N_l} h_l \times b_l$. Consequently, the GPES-adapted network no longer contains N_W scalar values but instead N_W continuous functions. Thus, for any Ψ value within the predefined range, the GPES-adapted network can input Ψ as an independent variable into these N_W functions to generate N_W weight parameters specifically tailored to that Ψ value.

The fitting of univariate real-valued functions has been extensively studied. Typically, this involves selecting M basis functions $g_1(\Psi) \sim g_M(\Psi)$ based on the characteristics of the target function, followed by determining the combination coefficients of these basis functions using sampled data points, as given by Eq. (5.35).

$$W_{ij}^{(l)}(\Psi) = \sum_{m=1}^{M} c_{m,ij}^{(l)} \, g_m(\Psi) \tag{5.35}$$

where, $c_{m,ij}^{(l)}$ denotes the undetermined combination coefficient for the mth basis function during the fitting process of the parameter at lth layer, ith row, and jth column. By applying the least squares method, the fitting problem can be reformulated as solving a system of linear equations. To avoid singular coefficient matrices during the solving process, $M = N_\psi$ is typically adopted.

Furthermore, when combining discrete sub-models into a continuous model, their reliability and significance vary due to differences in performance and accuracy. Specifically, even under identical training protocols, the final training errors of individual sub-models may differ. Sub-models with higher performance are deemed more reliable, and their parameters should contribute more significantly to the fitting process. Consequently, the final GPES fitting procedure incorporates uncertainties at discrete points by treating them as noisy samples, where the noise magnitudes are normalized using the sub-models' training losses L_ψ. The optimization objective of the GPES fitting

process is thus defined as:

$$\underset{c_{m,ij}^{(l)}}{argmax} \prod_{\psi=1}^{N_\psi} P\left(W_{\psi,ij}^{(l)} \left| \left(g_1(\Psi_\psi) \sim g_M(\Psi_\psi), c_{m,ij}^{(l)}, L_\psi\right)\right.\right) \tag{5.36}$$

If the noise caused by training loss is Gaussian noise, then any item of the above joint distributions can be written as:

$$P_\psi = \frac{1}{\sqrt{2\pi {L_\psi}^2}} e^{\left(-\frac{1}{2{L_\psi}^2}\left(W_{\psi,ij}^{(l)} - \Sigma_{m=1}^{M} c_{m,ij}^{(l)} g_m(\Psi_\psi)\right)^2\right)} \tag{5.37}$$

This optimization task can be accomplished by computing the partial derivatives of the probability function containing undetermined coefficients and setting them to zero. After completing the fitting process, when a new sample is input into the model, all parameters in the weight matrices automatically adjust to probabilistically optimal levels based on the sample's corresponding Ψ value, regardless of whether this Ψ value was included in the training set.

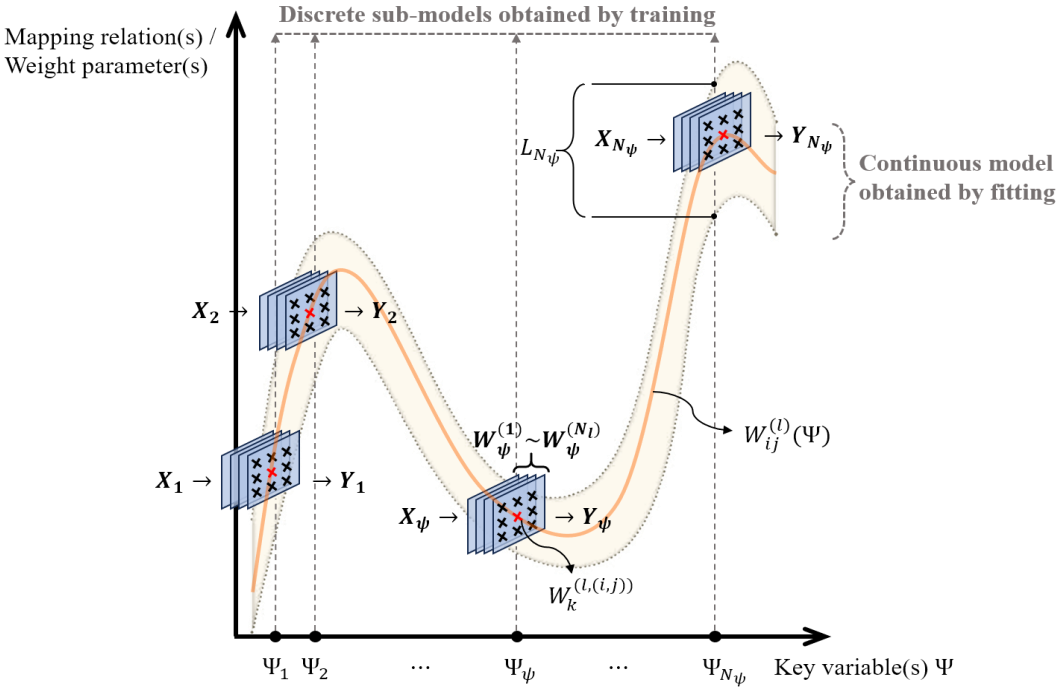


Fig. 5.28 GPES

5.4.2 Multi-layer ESN and Ex-ESN

GPES is applicable to nearly all NN models with well-defined architectures. However,

the ESN introduced in Section 4.3.1 is still selected as a case study for extending generalization performances. This preference stems from the structural compatibility between ESNs and GPES. Given that any fitting process introduces errors, fitting errors in weight parameters may accumulate progressively during model inference for networks with multiple layers. In this context, ESNs (with $N_l = 1$), where only a single linear output layer is trained, are particularly suitable for GPES-based generalization enhancements.

While structural extensions alone cannot guarantee generalization performance, the intrinsic fitting capability of a model to learning objectives remains critical. For well-architected models, increasing depth is the most common and effective approach to enhance performance. Expanding model parameters raises the upper bound of target function complexity that the model can fit, enabling the recognition of diverse patterns in the target mapping and thereby improving generalization. Notably, increasing the depth of ESNs, achieved by cascading multiple reservoirs, does not elevate training complexity. Leveraging the concept of deep ESNs proposed in prior work (Ma, Shen & Cottrell 2020), a two-layer ESN architecture is adopted as the foundational framework. Its encoding and output mechanisms are governed by Eqs. (5.38), (5.39), and (5.40).

$$\boldsymbol{R}^{(1)}(t) = (1-\alpha)\boldsymbol{R}^{(1)}(t-1) + \alpha f\left(\boldsymbol{W}_{\boldsymbol{in}}^{(1)}\boldsymbol{X}(t) + \boldsymbol{W}_{\boldsymbol{res}}^{(1)}\boldsymbol{R}^{(1)}(t-1) + \boldsymbol{b}_{\boldsymbol{r}}^{(1)}\right) \quad (5.38)$$

$$\boldsymbol{R}^{(2)}(t) = (1-\alpha)\boldsymbol{R}^{(2)}(t-1) + \alpha f\left(\boldsymbol{W}_{\boldsymbol{in}}^{(2)}\boldsymbol{R}^{(1)}(t) + \boldsymbol{W}_{\boldsymbol{res}}^{(2)}\boldsymbol{R}^{(2)}(t-1) + \boldsymbol{b}_{\boldsymbol{r}}^{(2)}\right) \quad (5.39)$$

$$\widehat{\boldsymbol{Y}}(t) = \boldsymbol{W}_{\boldsymbol{out}}\left[\boldsymbol{X}(t); \boldsymbol{R}^{(1)}(t); \boldsymbol{R}^{(2)}(t)\right] + \boldsymbol{b}_{\boldsymbol{y}} \quad (5.40)$$

where, $\boldsymbol{R}^{(1)}(t)$ and $\boldsymbol{R}^{(2)}(t)$ represent the states of two reservoirs of equal size, $\boldsymbol{W}_{\boldsymbol{in}}^{(1)}$, $\boldsymbol{W}_{\boldsymbol{res}}^{(1)}$, $\boldsymbol{W}_{\boldsymbol{in}}^{(2)}$, and $\boldsymbol{W}_{\boldsymbol{res}}^{(2)}$ denote the encoding matrices for the two reservoirs, respectively. Except for $\boldsymbol{W}_{\boldsymbol{in}}^{(2)} \in \mathbb{R}^{N_{res} \times N_{res}}$, the dimensions of the remaining matrices align with those of a single-layer ESN. The terms $\boldsymbol{b}_{\boldsymbol{r}}^{(1)}$ and $\boldsymbol{b}_{\boldsymbol{r}}^{(2)}$ are bias vectors, and $\boldsymbol{W}_{\boldsymbol{out}} \in \mathbb{R}^{N_{out} \times (2N_{res} + N_{in})}$ represents the trainable parameters of the two-layer ESN. The model employs the MSE as its loss function, which has been discussed previously

and will not be reiterated here.

In this base architecture, only $\boldsymbol{W_{out}}$ requires training. When adapting it using GPES, N_ψ parallel instances of $\boldsymbol{W_{out,\psi}}$ are trained. After fitting, these yield $\boldsymbol{W_{out}}(\Psi)$, which internally comprises $N_{out} \times (2N_{res} + N_{in})$ real-valued functional curves parameterized by Ψ. The resulting model, represented by $\boldsymbol{W_{out}}(\Psi)$ and endowed with high generalization performance across the Ψ-dimension, is termed the Ex-ESN, as illustrated in Fig. 5.29.

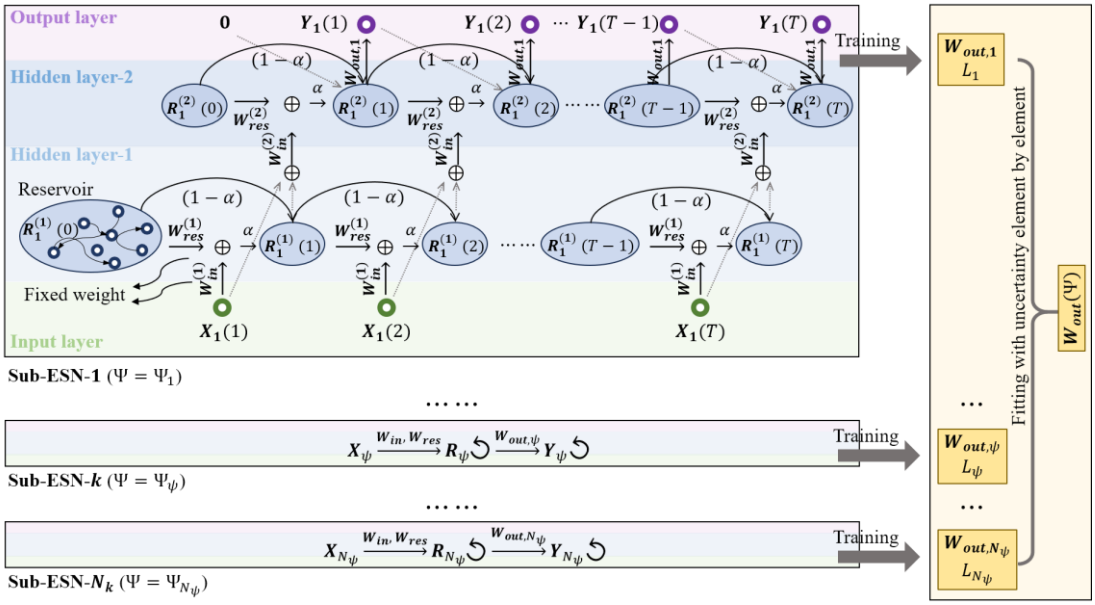


Fig. 5.29 Ex-ESN

5.4.3 Application: Reaction force solving of particle damper system

Particle dampers (PDs) are effective energy-dissipation devices and are widely applied in vibration control (Chen et al., 2001). When a PD vibrates with the host structure, energy is dissipated through collisions and friction between the filler particles and between the particles and the damper walls. The performance of different PDs can be evaluated by measuring the relationship between the input excitation and the output reaction force of the system, which also serves as a key criterion for determining the layout of PDs in practical applications. Previous researchers proposed a two-phase flow equivalent viscosity model (Wu et al., 2004) to simulate the damper system. However, this model is insufficient to fully characterize the comprehensive performance of PDs and can only serve as auxiliary information for PD experiments and analysis. In recent PD studies (Ye

et al., 2022), researchers found that the excitation intensity exhibits only a linear influence on the damper response within a certain range, while the excitation frequency significantly and nonlinearly affects PD performance. Particularly in certain sensitive frequency ranges, the mapping between the system reaction force and excitation becomes entirely unstable. For safety considerations, the performance of PDs intended for real-world applications must be validated across the entire range of potential excitation frequencies, which drastically increases the workload of high-precision hysteresis experiments. Nevertheless, laboratory testing is still limited to discrete excitation frequencies and cannot exhaustively cover all possibilities. For PDs with specific parameters, there is an urgent need to develop a high-generalization Seq2Seq model to estimate PD performance under untested excitation frequencies within the target range, based on discrete experimental results, thereby reducing the calibration effort for vibration control effectiveness across frequencies prior to deployment.

In this study, an aluminum alloy cylindrical cavity with a height of 30 mm and a diameter of 50 mm was fabricated to construct the PD and conduct experiments. Tungsten powder with a particle diameter of approximately 0.2 mm was selected as the filler, occupying 30% of the cavity's total volume. The apparent density of the filler was $8.9\ g/cm^3$. The fabricated PD was bolted to an acceleration-controlled shaking table to achieve synchronous vibration. Therefore, the actual input excitation (acceleration signals) applied to the PD could be directly measured using a vertical impedance head fixed between the PD and the shaking table. Simultaneously, the output response (reaction force signals) was recorded via corresponding sensors. The experiments involved applying a series of vertical sinusoidal acceleration excitations with varying amplitudes and frequencies. Specifically, the acceleration amplitude was incrementally increased from 5 g to 25 g in 5 g steps. For each amplitude, 20 discrete frequencies were evenly selected within the 100 Hz to 2,000 Hz range, resulting in a total of $5 \times 20 = 100$ excitation sequences, each lasting 0.1 s, with corresponding reaction force outputs recorded. The sampling rate was 10 kHz, yielding 1,000 data points per excitation. The inertial force of

the PD casing was subtracted from the recorded force values. The PD and experimental setup are illustrated in Fig. 5.30.

The acceleration signal and the velocity signal derived from its integration are treated as the input $\boldsymbol{X}$. Based on these two signals and a pre-existing model proposed in prior studies, an estimated response can also be computed. This estimated response, which can be explicitly pre-computed (albeit with lower accuracy but incorporating material and geometric information of the PD) prior to testing, is incorporated into $\boldsymbol{X}$, resulting in Nin=3*N*in=3. The target PD performance, represented by the experimentally measured reaction force, is defined as $\boldsymbol{Y}$, corresponding to $N_{out} = 1$. The key variable Ψ is defined as the excitation frequency. The hysteresis loops obtained from testing are shown in Fig. 5.31, which clearly indicate that the hysteretic behavior of the PD exhibits negligible dependence on excitation amplitude but significant variation with frequency Ψ. When employing NNs or other models to characterize the dynamic time-series mapping relationship, challenges arise due to limited generalization performance.

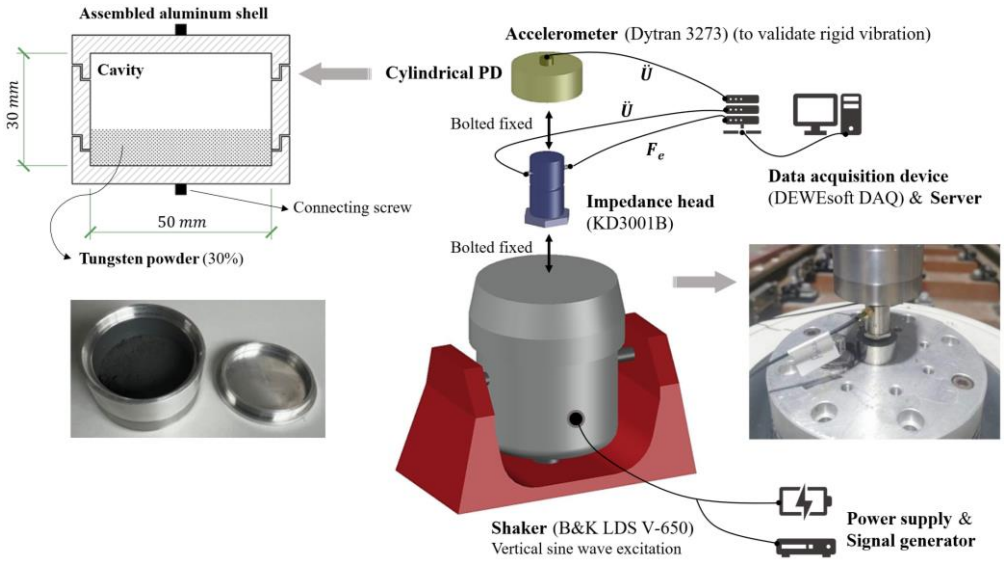


Fig. 5.30 PD and its performance experiment

5.4.4 Dynamic system solving results and discussion

This section focuses on demonstrating and discussing the accuracy and generalization performance limitations of a single model in solving the PD system, the enhanced generalization performances achieved by modified Ex-ESN, and ultimately revealing the

role of GPES in addressing complex nonlinear systems. The process involves first establishing 20 Seq2Seq sub-models to learn the system behavior under specific excitation frequencies. The Ex-ESN model is then constructed by fitting the parameters of a subset of these sub-models using GPES, and its performance is validated by applying it to untrained excitation frequencies corresponding to the excluded sub-models.

Following the hyperparameter screening method outlined in Section 4.3.4.3, all sub-models share identical architectures to ensure compatibility for GPES fitting. The finalized hyperparameter selections are summarized in Table 5.6. The MSE is adopted as the model accuracy metric, and SGD with a learning rate $lr = 0.00005$ and $\eta = 0.0001$ remains the optimization method for training. The training process iterates for 1,000 epochs.

Table 5.6 Hyperparameters of Ex-ESN

$\boldsymbol{\rho(W_{res})}$	$\boldsymbol{spar(W_{res})}$	$\boldsymbol{N_{res}}$	$\boldsymbol{\alpha}$
1.1	$\boldsymbol{4.0N_{res}}$	8000	0.9

5.4.4.1 Sub-model solving accuracy and its generalization performance

The 20 sub-ESN models were each trained using the first 80% of the dataset corresponding to their respective excitation frequencies. The training sets included five distinct acceleration amplitudes, and their performance in predicting the reaction force of the PD under the corresponding frequency excitation was tested using the remaining 20% of the sequences. Within each training epoch, each sub-model received training data encompassing all five acceleration amplitudes, leading to fluctuations in the training loss (calculated via MSE) across samples associated with different amplitudes. Consequently, the average training loss was adopted to plot the loss-epoch curves for the 20 sub-ESNs, as shown in Fig. 5.32. Nearly all training losses decreased during training (except for sub-ESN-7) and stabilized after 600 epochs, with final model loss values ranging from 0.04 to 4.40. Due to variations in SNRs and the weakly nonlinear influence of excitation

amplitudes on energy dissipation effects, the gradients of training loss for certain samples occasionally opposed those of others. This resulted in an increase in training loss for specific samples within individual sub-models as training.

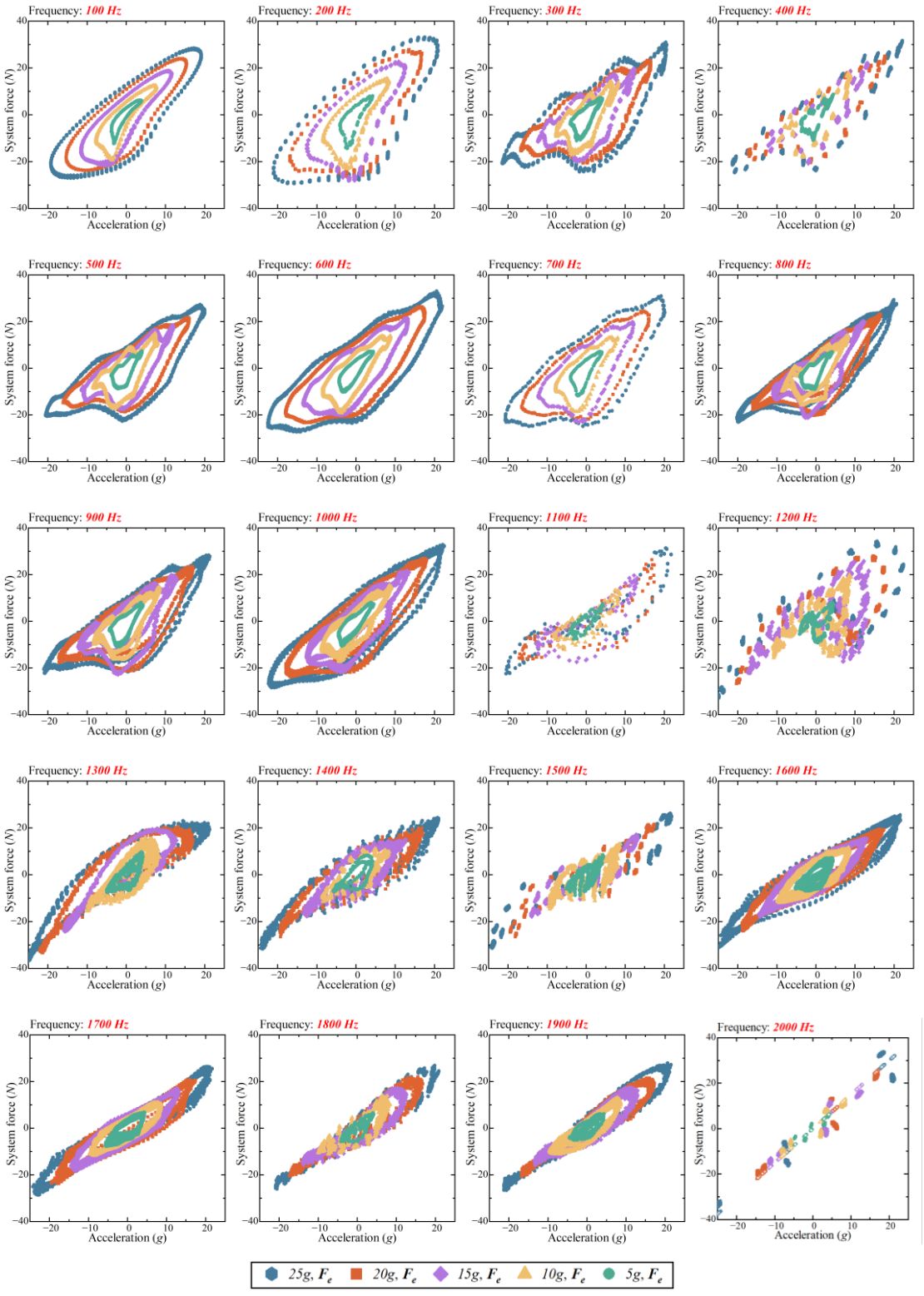


Fig. 5.31 Hysteresis curve of PD under different excitations

The reaction force results of the 20 sub-models under their respective excitation frequencies are illustrated in Fig. 5.33, where samples corresponding to different acceleration amplitudes are plotted separately to avoid interference. Additionally, the detailed results of two representative test samples are shown in Fig. 5.34: the prediction of sub-ESN-1 under a 5 g acceleration and that of sub-ESN-10 under a 25 g acceleration. Overall, for these selected cases, the sub-models generate sequences that align closely with the ground truth in both time and frequency domains, indicating that the hysteresis curves of the PD forces predicted by the sub-ESNs approximate the experimental results. Even under relatively lower time-domain accuracy, the ESNs at least ensure consistency in the fluctuation range and trend of PD reaction forces with respect to acceleration. Specifically, the MSE of the predicted forces across all independent sub-ESNs ranges between 1.0 and 10.0, generally increasing with higher excitation amplitudes. However, this trend is not statistically significant, and instances of reduced prediction errors are observed at elevated acceleration amplitudes (particularly under 20 g and 25 g external excitations). Since the outputs are not normalized, the absence of a quadratic increase in MSE with excitation amplitude implies a reduction in relative error.

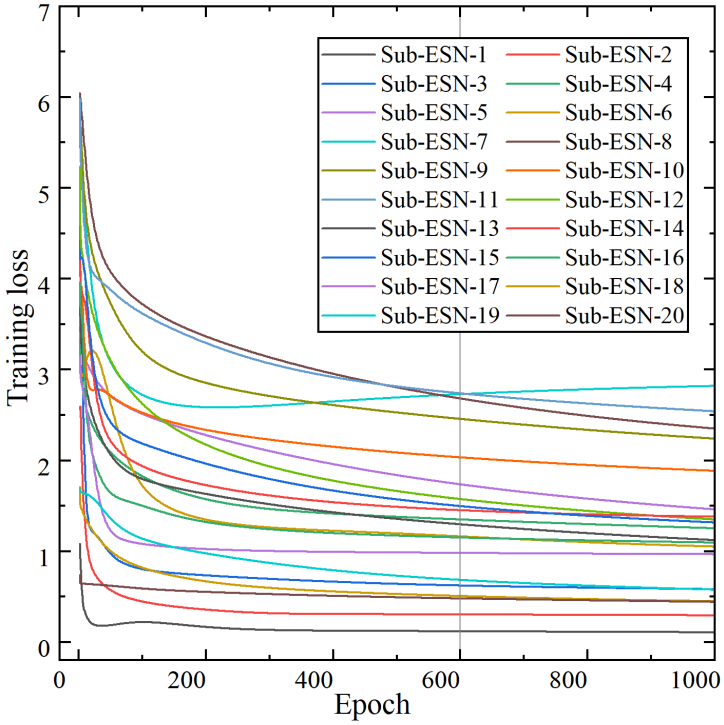
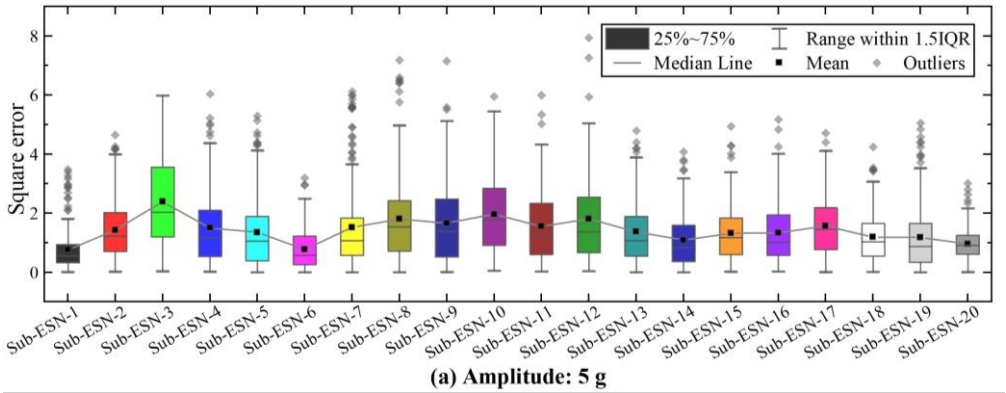


Fig. 5.32 Training loss-epoch curves of sub-ESN models

In other words, compared to low-amplitude excitations, where PD testing is prone to vibrational instability and high measurement noise, the reaction forces measured under larger accelerations more purely reflect the energy dissipation mechanisms of the PD and exhibit a more learnable mapping relationship. Therefore, training each sub-model with five amplitude-varied sample groups not only reduces the required number of sub-models but also leverages complementary information between samples to more effectively determine network weights within reasonable bounds. A notable observation is that, in certain low-amplitude samples, despite comparable prediction accuracy on the test set, the computed reaction forces lack high-frequency components present in the ground truth (see Fig. 5.34(a), where the sample exhibits the lowest MSE but not the lowest relative error). Based on theoretical analysis of the PD, these high-frequency components are more likely attributable to measurement noise. Consequently, the sub-ESN models inherently act as filters, producing output signals with purer frequency spectra.

On the other hand, under identical acceleration amplitude excitations, the MSE of all sub-ESN models (frequency range: from 100 Hz to 2,000 Hz) exhibits minor fluctuations. Comparatively, sub-ESN-9~12 (frequency range: 900 Hz~1,200 Hz) demonstrate reduced capability in determining the dynamic reaction forces of the PD. Furthermore, the error distributions of these four sub-models are more dispersed and contain a higher number of outliers. As illustrated in the time-domain curves (Fig. 5.34(b)), when the PD is subjected to sinusoidal waves within this frequency range, its reaction force sequences display more complex frequency components, regardless of variations in excitation amplitude.



(a) Amplitude: 5 g

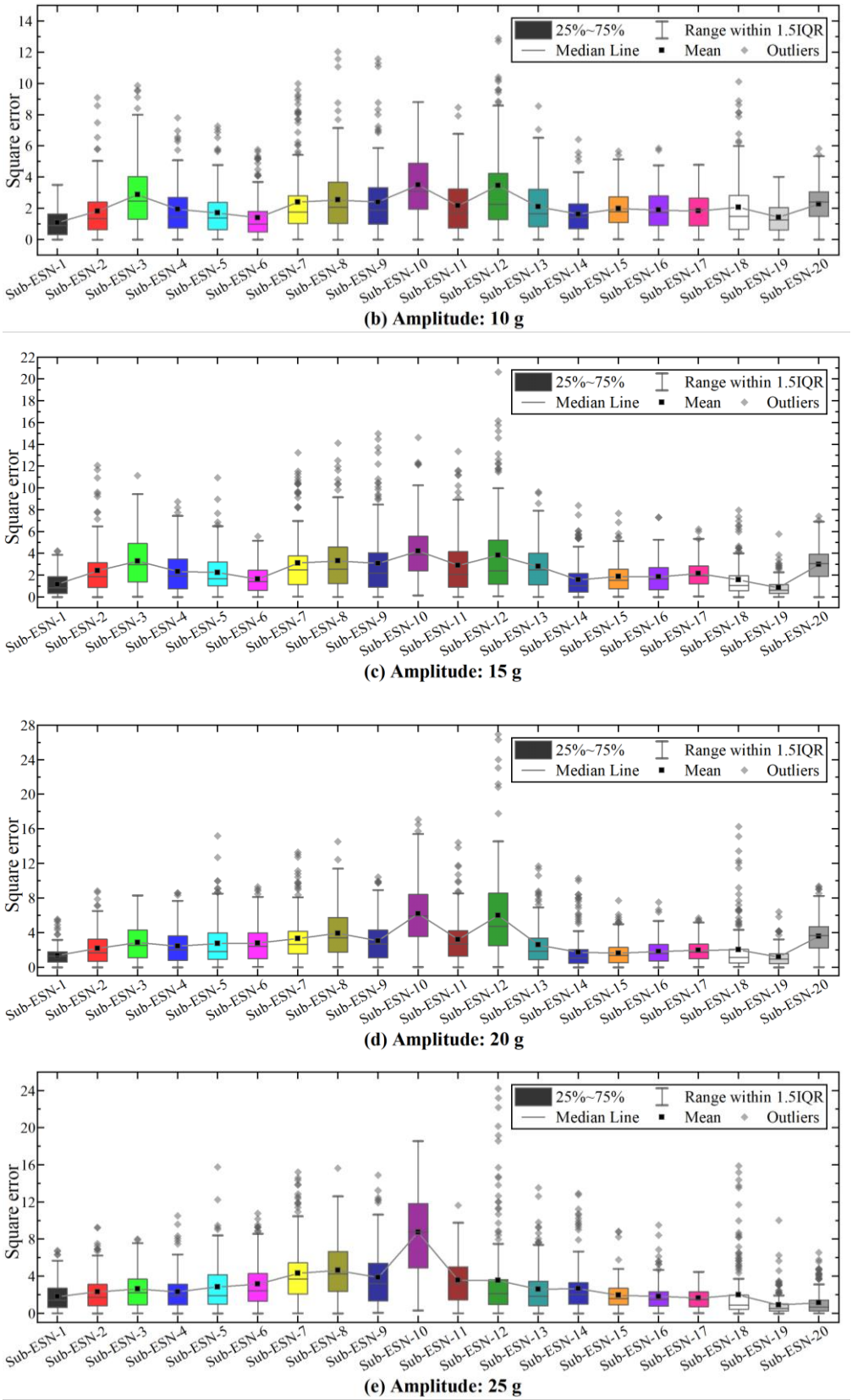


Fig. 5.33 Solving results of sub-ESNs

Additionally, for nearly all test samples, the error sequences between the model outputs and ground truth remain periodic, with reduced accuracy observed at the peaks of the

sinusoidal signals, leading to sharp spikes in the error sequences at these positions. Sharper peak segments correlate with poorer approximation performance by the sub-ESNs. This may stem from the fact that the temporal evolution of the RNN-based model's output states depends on delayed feedback loops in the hidden layer, which inherently operate with lagged responses. The aforementioned frequency range is defined as the sensitive frequency band of the PD. In subsequent fitting procedures, the credibility of the weight parameters for the corresponding sub-models will be reduced accordingly.

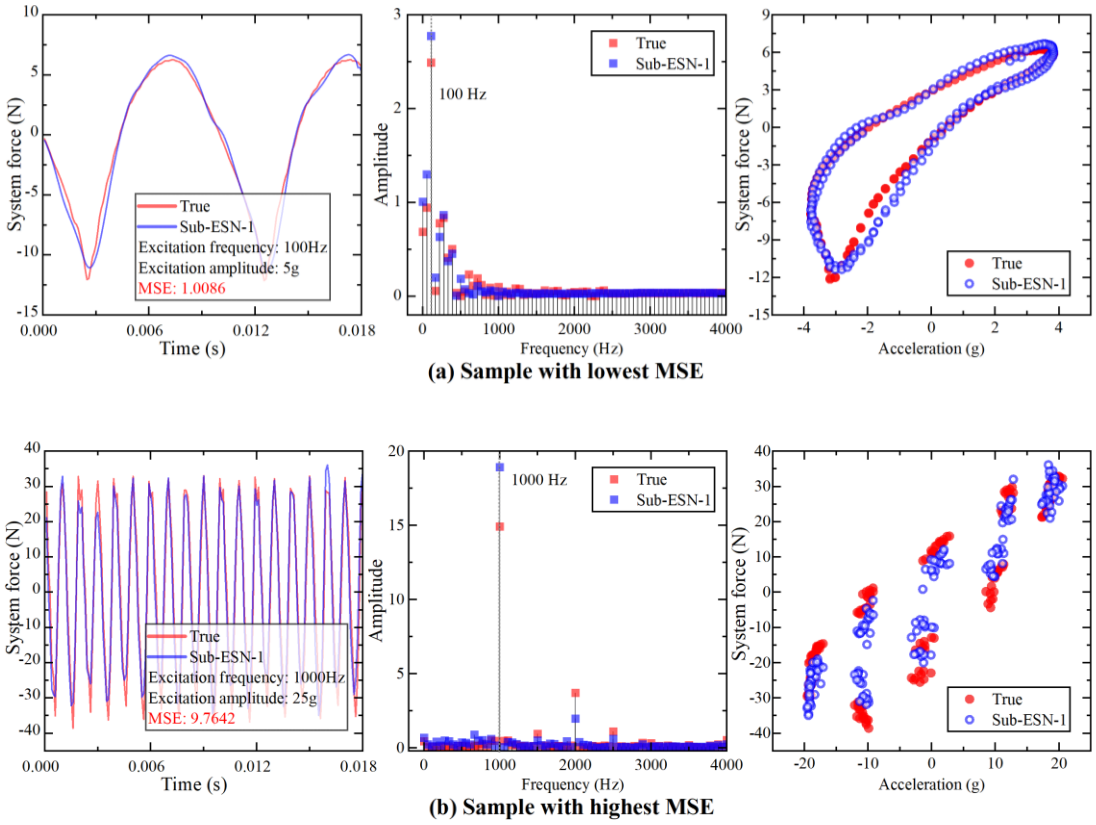
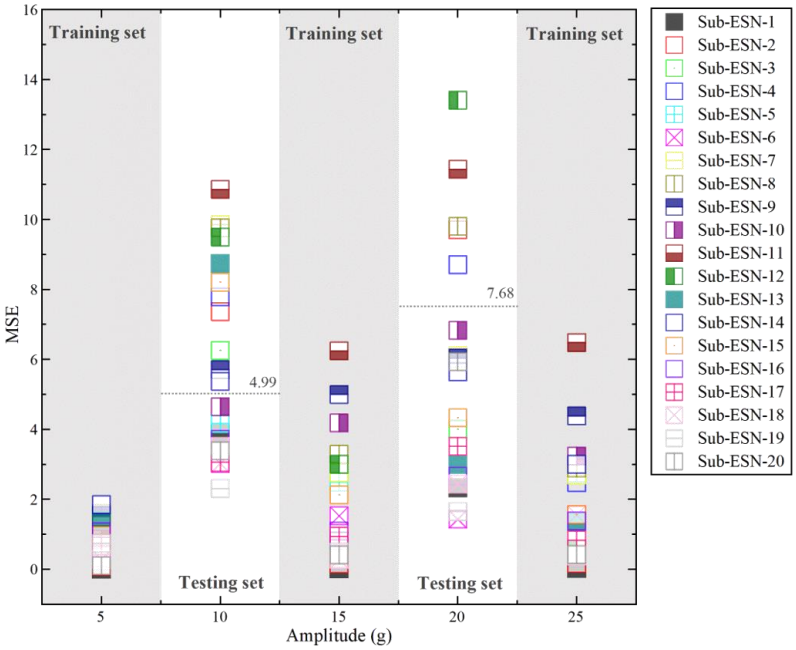


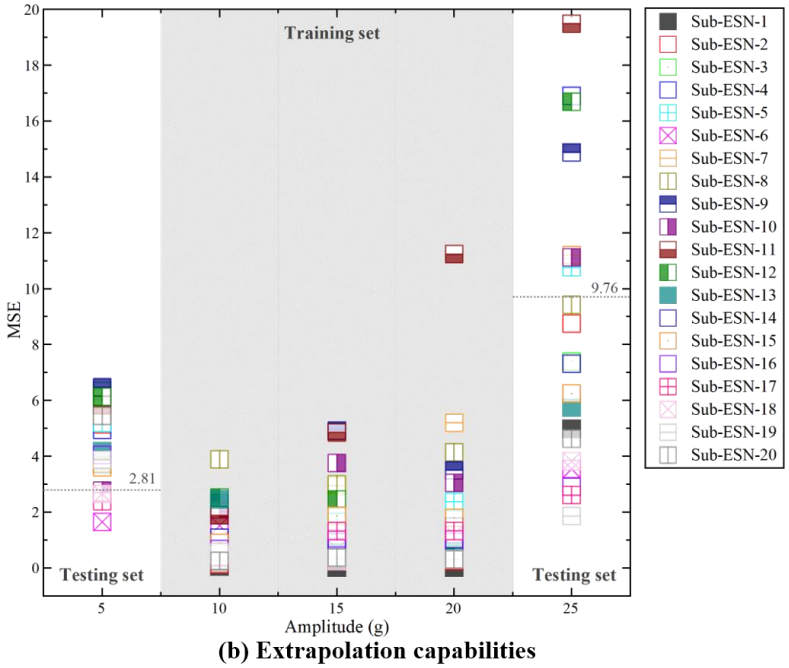
Fig. 5.34 Representative sub-ESN model solving details

Although no explicit regularization was applied, the sub-models inherently possess basic generalization performances to handle test samples that are not drastically dissimilar from the training data. For instance, it has been demonstrated that the PD's response varies smoothly with excitation amplitudes. A competent Seq2Seq model should thus exhibit strong interpolation ability, accurately predicting reaction forces for any amplitude within the trained range, as well as moderate extrapolation capability to maintain reasonable accuracy for amplitudes slightly beyond the training scope. To evaluate this, a new set of

sub-models was retrained using samples with 5 g, 15 g, and 25 g excitation amplitudes and tested on 10 g and 20 g cases. The results, shown in Fig. 5.35(a), confirm that the prediction accuracy degrades when the excitation amplitude is absent from the training set. For the two "unseen" amplitudes in the test set, the maximum MSE of the corresponding sub-models increased from 4.99 and 7.68 to 9.85 and 13.50. However, their relative accuracy remained above 85%, indicating robust interpolation performance.



(a) Interpolation capabilities



(b) Extrapolation capabilities

Fig. 5.35 Basic generalization performance of standard ESN

During the PD's service life, installations in real-world structures inevitably experience external loads exceeding experimental ranges, particularly during accidents or disasters. This necessitates evaluating the models' extrapolation ability. Consequently, another set of sub-models was trained using only 10 g, 15 g, and 20 g amplitudes and tested on 5 g and 25 g cases, as shown in Fig. 5.35(b). Similar to the interpolation scenario, model accuracy declines when excitation amplitudes exceed the training range's maximum or fall below its minimum. Nevertheless, the results remain acceptable for engineering applications (relative accuracy larger than 87%). Notably, under 5 g excitation, sub-ESNs 17~18 trained with this scheme outperformed models trained with all amplitude samples, further suggesting that the latter are perturbed by experimental noise. Additionally, during the generalization tests for varying amplitudes, sub-models corresponding to the sensitive frequency band (as defined earlier) consistently underperformed compared to others.

5.4.4.2 Fitting process and generalization performance of Ex-ESN

Although the sub-models exhibit moderate generalization performances, the energy dissipation behavior of the PD varies too abruptly with frequency, leading to unacceptable accuracy degradation when a sub-model trained on data from one frequency is applied to samples of another frequency, even if the frequencies are adjacent, as illustrated in Fig. 5.38. To address this, the GPES framework is employed to extend the model by constructing a continuous Ex-ESN through a fitting process. Each sub-model contains 8,021 weight parameters (including bias terms), necessitating the fitting of 8,021 real-valued functions. Here, this study first demonstrate the baseline scenario where sub-ESNs trained at excitation frequencies of 200 Hz, 400 Hz, up to 2000 Hz (i.e., sub-models selected at intervals of 2 from the 20 sub-models, $N_\psi = 10$) serve as known models. The basis functions for the fitting process are determined by analyzing the distribution of sampled weight parameters. For instance, the weight parameter in the first row and first column of $\boldsymbol{W_{out}}(\Psi)$ exhibits distinct peaks and periodic oscillations along the frequency dimension (Fig. 5.36). Similar observations of discrete point distributions across all parameters motivate the selection of a cubic polynomial as the trend term, combined with

sinusoidal and peak terms to form the basis functions, as specified in Eq. (5.41). The credibility of each sub-model's weight parameters is then determined based on their final training losses using Eq. (5.42).

$$W_{out,ij}(\Psi) = c_{1,ij} + c_{2,ij}\Psi + c_{3,ij}\Psi^2 + c_{4,ij}\Psi^3 + c_{5,ij}sin\left(c_{6,ij}\Psi + c_{7,ij}\right) + c_{8,ij}e^{-\frac{\left(\Psi - c_{9,ij}\right)^2}{2c_{10,ij}^2}} \tag{5.41}$$

$$L_\psi \leftarrow \left(\frac{L_\psi}{max(L)}\right)^\phi \tag{5.42}$$

where, ϕ is a hyperparameter controlling the contribution of low-accuracy sub-model parameters. A fitting algorithm based on maximum likelihood estimation, without considering the varying credibility of sub-models, can yield a relatively reasonable approximation function. In this case, the fitted curve achieves 85% accuracy for weight parameters at 10 additional frequencies. However, both the weight values and fitting accuracy within the sensitive frequency band exhibit anomalies, and conventional algorithms often sacrifice the accuracy of normal points to achieve acceptable fitting results for these aberrant points. Since the weight parameters of sub-models in the sensitive frequency band are inherently less trustworthy than those of other sub-models, their uncertainties must be explicitly accounted for. After encoding the reliability of each sub-model's weight parameters into the fitting algorithm, the average fitting accuracy improves despite the degraded performance in the sensitive band. This approach not only avoids detrimental effects but can also be interpreted as a correction mechanism for poorly trained model weights.

To obtain a continuous matrix $\boldsymbol{W_{out}}$, which together with the shared reservoir constitutes the Ex-ESN model, all weight parameters in $\boldsymbol{W_{out}}$ are fitted. Consequently, when encountering excitation frequencies (e.g., 300 Hz) absent from the training set, the PD mapping relationship corresponding to that frequency (i.e., $\boldsymbol{W_{out}}(300)$) can be directly sampled from the continuous matrix to generate outputs. A comparison of the parameter distributions between $\boldsymbol{W_{out}}(\Psi_\psi)$ and the output matrix $\boldsymbol{W_{out}}$ of sub-ESN-ψ is

presented in Fig. 5.37. Due to the high fitting accuracy of individual weight parameters, the distributions of the 8,021 weight parameters across the ten fitted models closely resemble those of the sub-models obtained through independent training. Since elements of the mapping matrix in the reservoir are scaled to $[-1,1]$, the $\boldsymbol{W_{out}}$ parameters for any frequency are nearly symmetric around zero. Parameters in the sensitive frequency band exhibit more uniform distributions, while those in the normal frequency range are more concentrated near zero. Statistically similar distributions allow partial fitting errors to be offset during the final weighted summation process, thereby relaxing the stringent requirements on the fitting algorithm.

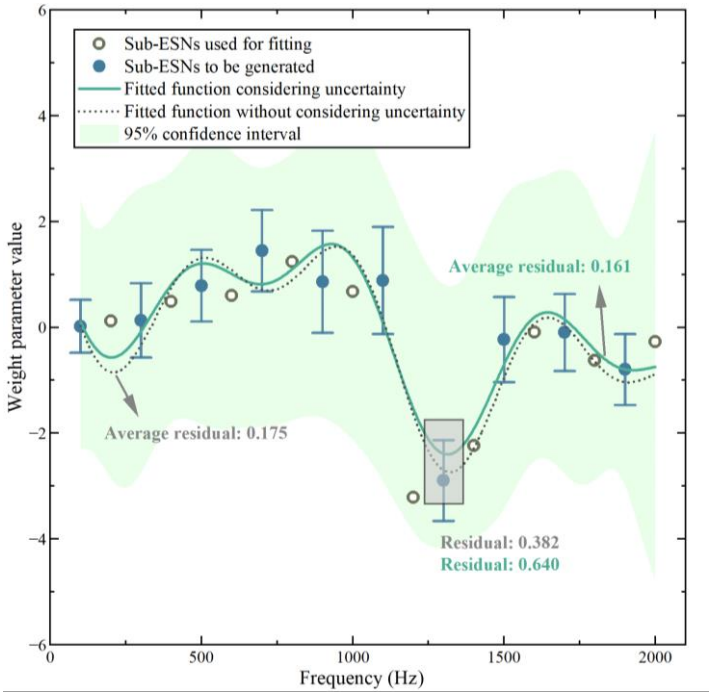
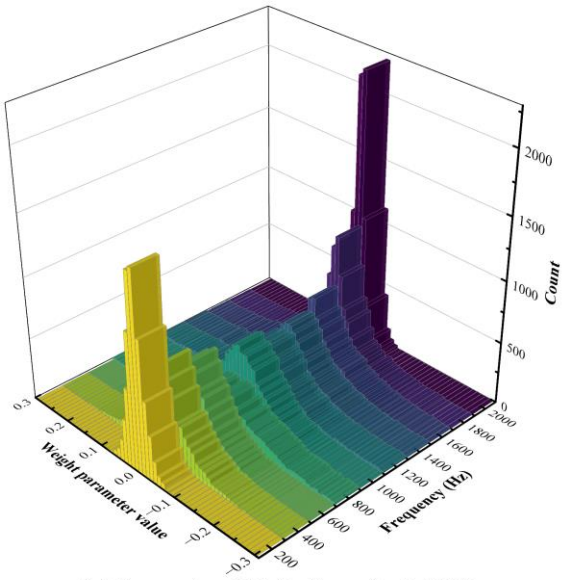


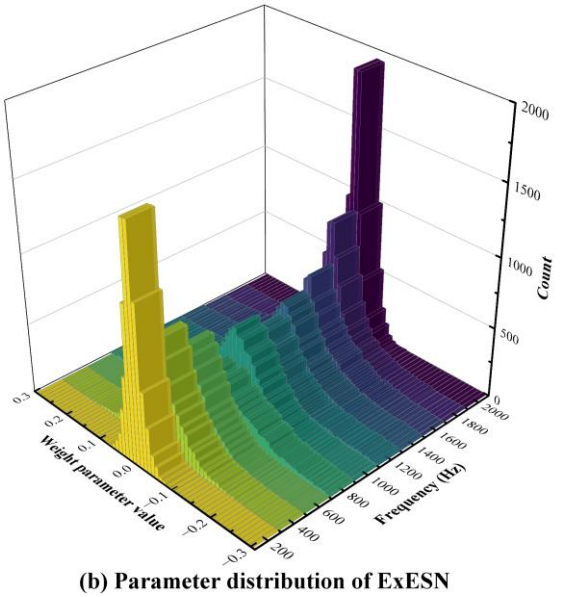
Fig. 5.36 Example of parameter fitting

As shown in Fig. 5.38, the fitted Ex-ESN model demonstrates acceptable capability in computing PD responses under excitation frequencies not included in the training set, compared to the poor accuracy of using non-corresponding frequency sub-models. This confirms that the generalization capability of the Seq2Seq model has been extended. In the Ex-ESN model, performance at the lower and upper frequency bounds (200 Hz and 2,000 Hz) is relatively weaker, in addition to the higher MSE observed in the sensitive frequency band. This phenomenon may arise from the lack of adjacent supplementary information at the start and end points of the fitting range. Although the accuracy of the

Ex-ESN model is generally lower than that of the corresponding sub-ESNs, its relative accuracy remains above 82% even in the worst-case scenario (MSE of 18.77), staying within the same order of magnitude as the corresponding sub-ESN models. Considering the time and economic costs of refined experiments, the GPES-extended model already satisfies most engineering requirements. More importantly, for dynamic systems with continuously evolving mapping relationships, it is impractical to exhaustively cover all scenarios encountered in real-world applications through discrete experiments. In such cases, a continuous Seq2Seq model capable of accurately estimating mapping relationships under any level of Ψ influence becomes indispensable.



(a) Parameter distribution of sub-ESNs



(b) Parameter distribution of ExESN

Fig. 5.37 Parameter distribution of W_{out}

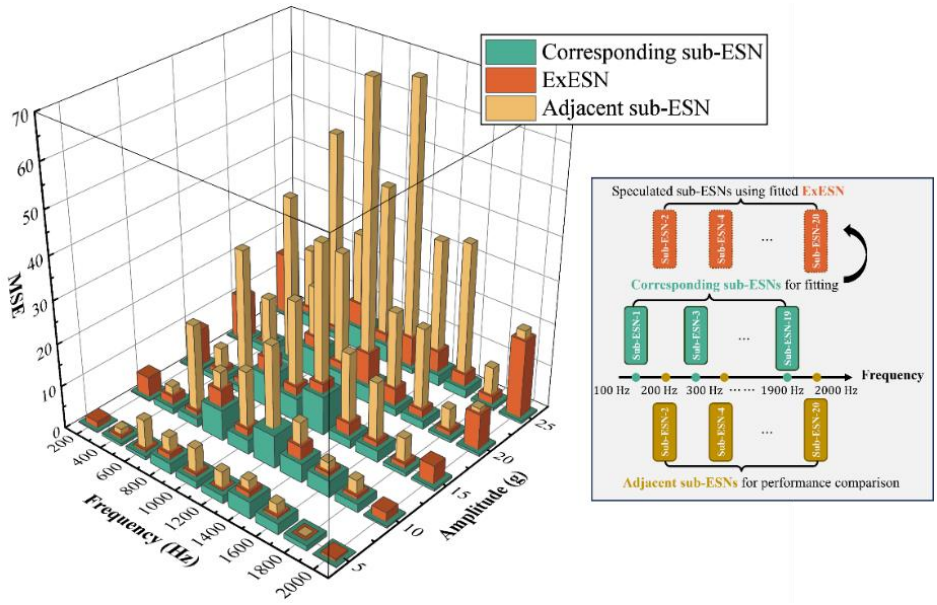
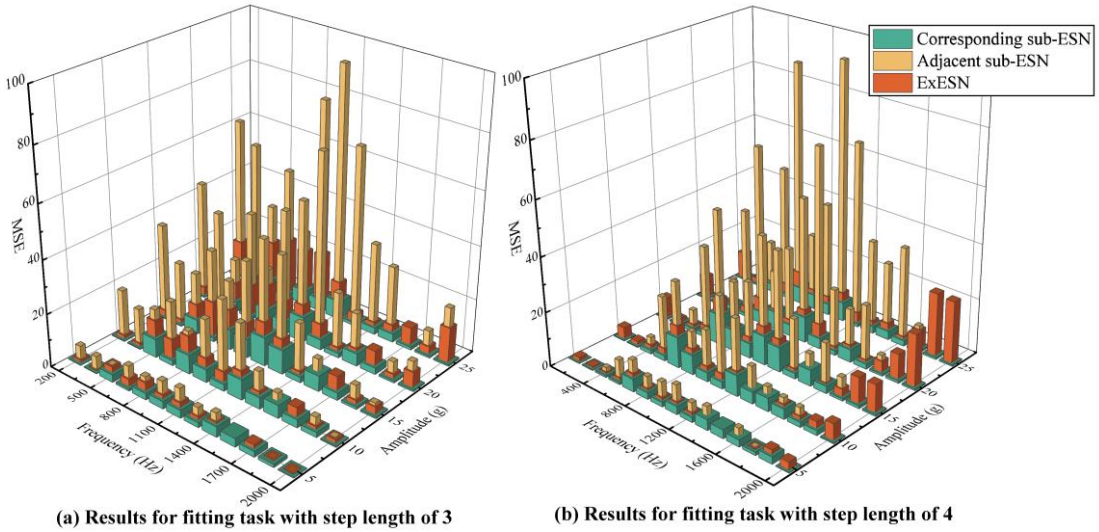


Fig. 5.38 Solving results of Ex-ESN



(a) Results for fitting task with step length of 3

(b) Results for fitting task with step length of 4

Fig. 5.39 Fitting process investigation of Ex-ESN

To further explore the performance limits of the Ex-ESN, numerical experiments with varying fitting difficulties were conducted. The number of known sub-models was gradually reduced, increasing the unfitted regions. The test results for Ex-ESNs fitted using sub-models selected at intervals of 3 and 4 (employing seven and five sub-models, respectively) are shown in Fig. 5.39. As data sources become sparser along the frequency dimension, reduced information availability and simplified basis functions lead to declines in both accuracy and generalization performance of the Ex-ESN models. Notably,

when the highest frequency of the known sub-ESNs is limited to 1,700 Hz, the accuracy for computing PD responses under 2,000 Hz excitation drops to 80%. These results demonstrate that the Ex-ESN, constructed via the proposed generalization extension framework, can infer mapping relationships of dynamic systems under critical variable variations from discrete "sample points". For the PD case studied here, the GPES framework reduces experimental efforts by over half while maintaining near-comparable accuracy.

5.5 Conclusions

This chapter addresses the critical data inference tasks required for final decision-making based on preprocessed SHM data. Data inference is defined as the process of estimating unmeasured but valuable data by leveraging spatiotemporal correlations within datasets. Common inference tasks include: prediction problems, for which this study proposes the dESN-eLSTM predictor; synchronized mapping problems, addressed by the Echoformer network tailored for multi-channel inputs and complex mapping relationships; and dynamic system solutions (e.g., physical systems), for which the GPES method is introduced to extend the generalization capability of model learning for nonlinear dynamic behaviors. The effectiveness, superiority, and robustness of these methods are validated using real-world datasets, including wind fields from the Lanzhou-Xinjiang HSR, noise measurements from Shenzhen Metro, and hysteresis curves of a PD system. Key findings are summarized as follows:

(1) For prediction tasks, the dESN-eLSTM network was developed by integrating an attention-like mechanism to modify the forget gate of the LSTM network and incorporating a dynamic encoding feature extractor. This network was applied to predict wake fields in a sudden transition section of WSs within the hundred-mile wind zone of the Lanzhou-Xinjiang HSR, ensuring train operational safety. After training, the dESN-eLSTM predictor achieved MSE values of 5.2436 and 4.1071 for rapid in-plane wake predictions. It also demonstrated over 85% accuracy in determining vertical wind profiles for average wind speed and

turbulence intensity. The model maintained basic functionality under extreme conditions, such as inflow speeds exceeding 20% of the maximum training set values, missing downstream measurement points, and sparse or distant sensor configurations.

(2) The Echoformer model enhances synchronous sequence inference by modifying the positional encoding process of the Transformer network with a randomized reservoir, reducing structural dependencies on input data. Applied to infer track-side rolling noise from rail vibrations, Echoformer outperformed conventional models (e.g., standard Transformer, ESN, ESDN, and bidirectional LSTM) with MSE values of 25.95 for near-field and 0.123 for far-field noise. It exhibited robust reconstruction under incomplete or noise-contaminated inputs: reducing acceleration channels to a single sensor increased near-field and far-field MSE by only 12.79 and 0.039, respectively. Even with 70% noise contamination, MSE increments remained limited to 7.12 (near-field) and 0.033 (far-field).

(3) The GPES method extends the generalization of Seq2Seq models by incorporating uncertainty-aware fitting for pretrained sub-models. Leveraging compatibility with ESN architectures, the ExESN framework was developed to solve PD system reaction forces. When excitation frequencies were fixed at 20 discrete levels, individually trained sub-models achieved at least 91% accuracy and successfully learned linear mapping variations induced by excitation amplitudes for interpolation and extrapolation tasks. However, due to the PD system's frequency-sensitive behavior, neither standalone nor unified sub-models could estimate reactions under unseen excitation frequencies. The GPES-fitted ExESN model, comprising 8,021 real-valued functions, achieved over 80% accuracy in inferring input-output mappings for unfamiliar frequencies. This approach provides reliable references for characterizing damper performance and significantly reduces experimental calibration efforts prior to deployment.

DAMAGE DETECTION AND STATE EVALUATION BASED ON TIME SERIES-TARGETED NEURAL NETWORKS

6.1 Introduction

The fundamental objective of establishing an SHM system lies in structural damage diagnosis and condition assessment, which can be achieved through filtered measurement and essential inferred data. Early-stage damage and degradation in structures originate from localized plastic deformation, microcracks, material creep, interface detachment, fastener loosening, environmental corrosion, and other factors. These phenomena are inevitable during structural service life and irreversible without human intervention. Even incipient damage leads to detrimental structural states including deteriorated load-transfer mechanisms, stress concentration at damaged regions, and increased contact area with corrosive environments, thereby accelerating further damage accumulation. This progression may ultimately develop into severe functional impairment or structural failure. To prevent safety incidents, it is imperative to monitor damage evolution proactively, issue timely warnings, and implement structural reinforcement or evacuation measures when necessary. Such critical decisions rely on comprehensive information from both SHM systems and regular inspections. Early damage detection algorithms were predominantly based on the assumption that structural damage universally reduces stiffness, thereby altering macroscopic dynamic characteristics such as natural frequencies and modal parameters - essentially transforming damage diagnosis into parameter identification tasks. However, practical structures requiring SHM installations, particularly critical infrastructure, often exhibit substantial safety margins. Combined with environmental influences like temperature variations, this renders macroscopic dynamic characteristics insensitive to early localized damage. Critical information about incipient damage and degradation may instead reside in more microscopic features of

time-domain signals.

This chapter proposes time series-targeted NN-based solutions to address these challenges. The first approach involves a signal classifier embedded with multi-dimensional and multi-scale convolution (MDSC) layers, enabling supervised classification while accomplishing damage localization and high-precision severity quantification. The second solution establishes a DT architecture through a multi-head full-rank reservoir computing network, facilitating rapid real-time condition assessment based on data distribution variations to guide structural reinforcement. The proposed methodologies are validated through numerical simulations of various damage scenarios and experimental shake table datasets, with comparative analyses conducted against conventional approaches.

6.2 Supervised damage detection

Damage diagnosis methodologies must encompass at least two essential components: damage localization and damage severity quantification. The latter primarily relies on time-domain signal features captured near potential damage locations, while the former necessitates comprehensive analysis of spatial correlations among multiple monitoring points. Consequently, damage diagnosis approaches must address the spatiotemporal interdependencies inherent in SHM data. Inspired by the principles of TCN, this section introduces convolutional layers as robust tools for feature extraction and refinement. To accommodate variations in information sources across dimensions and scales, the conventional convolutional layers are extended into MDSC layers, which are subsequently integrated into an ESN architecture. Given that the reservoir in ESN remains untrained, two additional MDSC layers are incorporated: one with fixed weights and another with trainable weights. The resultant network, termed MDSC-ESN, operates as a supervised signal classifier trained on labeled damage data to simultaneously identify damage locations and quantify severity during signal classification. This section focuses on evaluating its diagnostic accuracy through comparative analyses with mainstream methodologies.

6.2.1 Temporal convolution and MDSC layer

Signals acquired by SHM systems, when arranged by channels (different sampling points), form a matrix containing both temporal and spatial dimensions of measurement data. Conventional RNNs, which focus primarily on sequential dependencies, exhibit limited capability in handling spatial or multi-dimensional combined information. In contrast, convolutional operations demonstrate superior effectiveness for feature extraction in multi-dimensional data processing. Convolutional layers, a fundamental technique in DL, extract features through weighted summation of input matrices, where the weights correspond to learnable convolutional kernels. By scanning the input matrix with one or multiple kernel sets, these layers learn composite features within specific receptive fields. For instance, consider an input matrix $\boldsymbol{X}$ containing spatiotemporal information, where adjacent rows represent spatially proximate sensors. Convolutional operations inherently preserve spatial relationships while compressing temporal dependencies. Additionally, convolutional layers serve as efficient dimensionality reduction tools. In ESNs, for example, the reservoir-generated state sequence $\boldsymbol{R}$ often becomes excessively high-dimensional for direct classification or regression tasks (due to impractical parameter counts in fully connected layers). By leveraging weight-sharing during kernel scanning, convolutional layers significantly reduce trainable parameters while compressing features transmitted from the reservoir to the output layer. Architectures employing temporal convolution for feature extraction, such as TCN, explicitly abandon recursive structures to circumvent long-range dependency issues through truncated receptive fields. However, achieving sufficient temporal memory necessitates network deepening. Taking a stride of 1 as an example, the convolution operation for the kth kernel $\boldsymbol{W_{c1,k}} \in \mathbb{R}^{U_k \times V_k}$ applied to $\boldsymbol{X}$ can be formulated as follows, with analogous operations applicable to $\boldsymbol{R}$ or other matrices:

$$X_{co,k,ij} = \sum_{u=1}^{U_k} \sum_{v=1}^{V_k} W_{c1,k,uv} \times X_{(j-u+1)(k-v+1)} \tag{6.1}$$

where, $\boldsymbol{X_{co,k}}$ represents the convolution output extracted by the kth kernel, where U_k

and V_k denote the temporal and spatial dimensions of the kernel, respectively. These dimensions determine the number of input elements involved in weighted summation during convolution, defining the receptive field, the spatiotemporal scope of local information captured by the kernel. While the receptive field expands with deeper network layers, the kernel's intrinsic size critically influences its functionality. For TCNs, if $U_k = 1$, the kernel performs 1D temporal convolution to extract intra-channel time-domain features; if $V_k = 1$, it executes 1D spatial convolution to capture inter-channel transient correlations; and if both $U_k \neq 1$ and $V_k \neq 1$, it conducts 2D spatiotemporal convolution to model complex hybrid patterns. In practical applications, all three convolution types contribute valuable information. A convolutional layer incorporating kernels of varying dimensions is termed a multi-dimensional convolutional layer, and when integrating kernels of different sizes, it becomes a MDSC Layer. For instance, temporal 1D kernels with smaller sizes (e.g., 1×3) emphasize high-frequency components, while larger ones (e.g., 1×9) capture low-frequency trends. Similarly, spatial 1D kernels with compact dimensions (e.g., 2×1) focus on local inter-channel relationships, whereas extended kernels (e.g., 4×1) aggregate global correlations. To ensure uniform output widths across differently sized kernels, asymmetric padding is applied.

While stacking MDSC layers alone resembles TCN architectures with limited capability to capture signal trends and periodicities, this study integrates MDSC modules with recurrent models. MDSC modules are typically deployed in two configurations. One is positioned before recurrent layers to perform preliminary feature extraction on inputs $\boldsymbol{X}$, referred to as MDSC-1; while the other one is placed after recurrent layers to summarize and reduce the dimensionality of hidden states $\boldsymbol{R}$, termed MDSC-2. For ESNs to be employed in Section 6.2.2, where reservoir weights remain fixed, the preceding MDSC-1 layer also adopts frozen parameters. This design aligns with feature hierarchy principles: shallow layers with smaller receptive fields encode low-level spatiotemporal patterns, which can be standardized based on established signal processing frameworks.

In the MDSC-1 implementation, temporal 1D kernels include four sizes (1×3, 1×5, 1×7, and 1×9) to extract multi-scale frequency and gradient features. Spatial 1D kernels use three sizes (2×1, 3×1, and 4×1) to compute channel-wise means and gradients (1st to 3rd order). Additionally, 2D kernels inspired by classical edge detectors (Roberts (Albdour & Zanoon 2020), Prewitt (Balochian & Baloochian 2022), Sobel (AS & Gopalan 2022), Laplacian D4 and D8 (Lorencin et al., 2020)) and Gaussian operators (Assirati, et al., 2014) are employed, with dimensions 2×2, 3×3, and 3×5. Kernel weights (illustrated in Fig. 6.1) are predefined to ensure robust feature extraction across varying signal amplitudes and fluctuation patterns, reflecting foundational signal processing principles. The MDSC-1 layer remains task-agnostic, representing a fixed low-level feature extraction process. Letting $\otimes$ denote the convolution operation, the MDSC-1 process is formalized as:

$$\boldsymbol{X_c} \leftarrow [\boldsymbol{X}; \boldsymbol{X_c}] = [\boldsymbol{X}; \boldsymbol{W_{c1}} \otimes \boldsymbol{X}] \tag{6.2}$$

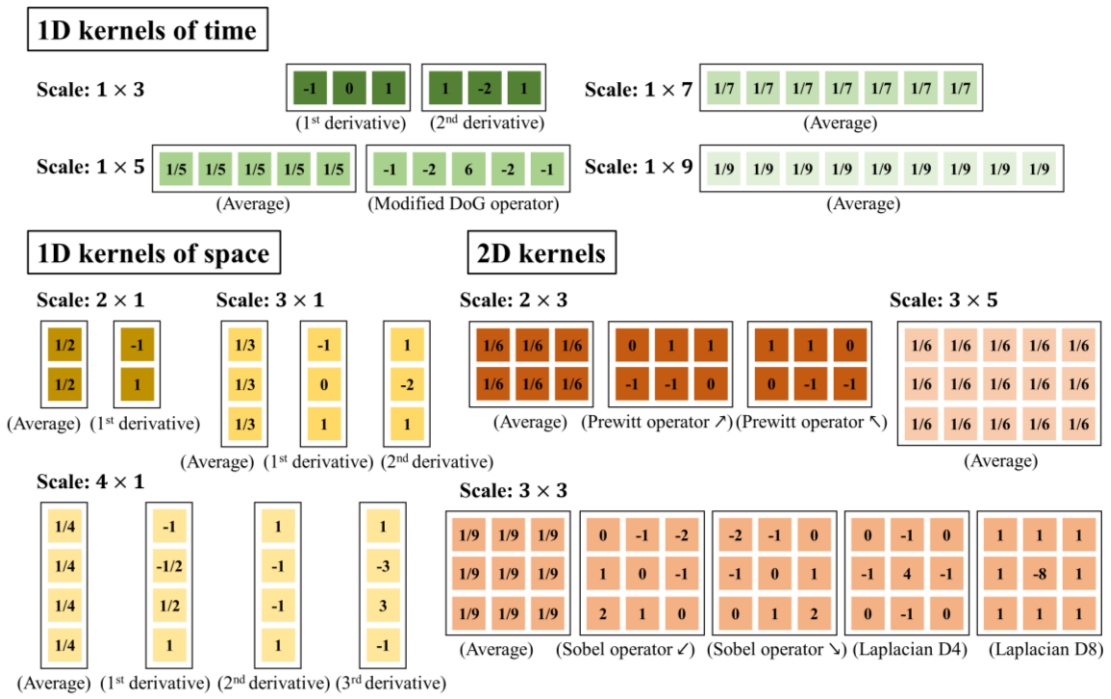


Fig. 6.1 Convolutional kernels of MDSC-1

The architecture of MDSC-2 is primarily determined by downstream tasks. In the damage diagnosis scenario discussed here, where the model's output must be a one-hot encoded vector representing damage categories, MDSC-2 focuses on reducing the dimensionality

of the reservoir state sequence $\boldsymbol{R}$ (with dimensions $N_{res} \times T$) to match this target format. Given that N_{res} and T share similar magnitudes, most convolutional architectures originally designed for image classification can serve as the backbone for MDSC-2, augmented with kernels of varying dimensions and scales. However, when MDSC-2 follows an ESN, the spatial correlations within $\boldsymbol{R}$ are inherently weaker due to the fixed, randomized reservoir. To address this, temporal 1D kernels should dominate the convolutional operations to emphasize time-dependent feature extraction. Furthermore, a stride of at least 2 is recommended for convolutional kernels to expedite dimensionality reduction while preserving critical temporal patterns.

6.2.2 MDSC-ESN and its training data selection module

The MDSC-ESN network is constructed by integrating the standard ESN introduced in Section 4.3.1 as the recurrent module, flanked by MDSC-1 and MDSC-2 layers at its input and output stages, respectively. For the damage identification classifier, the output layer employs a SoftMax activation function. The network architecture is illustrated in Fig. 6.2. To enrich the basis for damage identification, the 1st order frequency of the raw vibration signals is incorporated as supplementary features and concatenated with the flattened vector derived from MDSC-2 before being fed to the output layer.

$$\widehat{\boldsymbol{Y}} = \boldsymbol{W_{out}}[flat(\boldsymbol{W_{c2}} \otimes \boldsymbol{R}); fft_1(\boldsymbol{X})] \tag{6.3}$$

where, $flat(\cdot)$ denotes the matrix flattening operation; $fft_1(\cdot)$ represents the extraction of 1st order frequencies via the FFT algorithm, where multi-input arguments to this function yield a 1D vector composed of the first-order frequencies of each sequence; $\boldsymbol{W_{out}}$ and $\boldsymbol{W_{c2}}$ correspond to the trainable weight parameters of the output layer and MDSC-2, respectively. Unlike conventional classification tasks, damage detection requires not only localization but also severity quantification. Thus, the output dimension N_{out} equals the number of potential damage locations, and the values in the predicted output $\widehat{\boldsymbol{Y}}$ (as well as the ground truth $\boldsymbol{Y}$) represent the damage severity at the corresponding structural positions. The loss function remains the MSE, which is not

elaborated here.

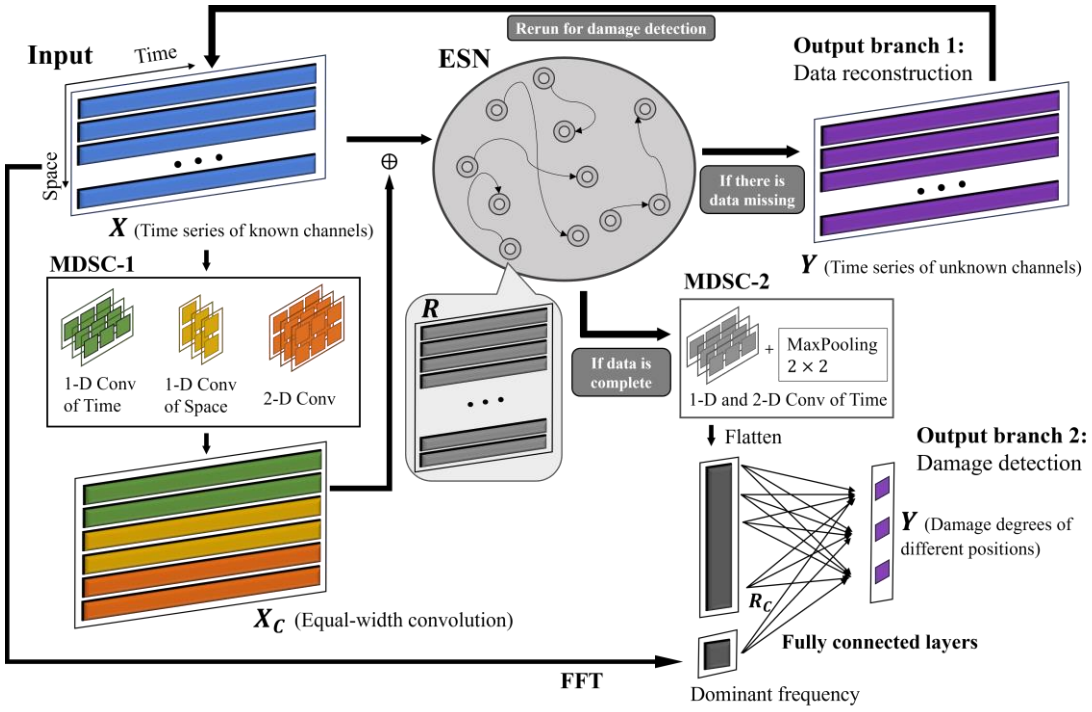


Fig. 6.2 MDSC-ESN

As a supervised model, MDSC-ESN requires labeled training data with both damage locations and severity annotations. While its accuracy surpasses unsupervised alternatives, a significant drawback lies in the scarcity of real-world labeled damage data for actual structures. To mitigate this, simulated damage datasets generated via FEM models are typically employed. Damage can be simulated by reducing elastic moduli, introducing cracks or notches, and applying operational excitations (e.g., wind pressure on buildings, vehicle loads on bridges) to structural models. The resulting responses at various locations under different damage scenarios are then recorded. To ensure data diversity, operational excitations, characterized by stochastic temporal fluctuations, are sampled extensively from their statistical distributions, either as numerous discrete data points or prolonged sequences. However, this approach often results in excessively large training datasets with high redundancy, which not only compromises training efficiency but may also misguide the model due to repetitive patterns. This challenge is not unique to MDSC-ESN but affects most supervised models in SHM data postprocessing. An optimal training

dataset should align class volumes with their probabilistic occurrences in real-world scenarios, where common operational conditions dominate the data distribution and rare events are proportionally represented. Consequently, meticulous curation of the virtually infinite FEM-generated data becomes imperative to balance representativeness and computational feasibility.

This study proposes a signal similarity-based data screening method, whose core idea is to split the dataset into segments, compute the similarity between each segment and all others, and then select representative and unique samples. Specifically, segments with higher aggregate similarity scores indicate representativeness, while those ranked lower in similarity reflect uniqueness. By selecting a portion of top-ranked samples (representative) and a small number of bottom-ranked samples (unique) based on a predefined ratio, efficient dataset curation is achieved. For a simulated dataset under any damage pattern, denoted as $\in \mathbb{R}^{N_{in} \times T}$, the segment length T_s is determined by the fundamental structural period (typically 2~10 cycles). The segmented dataset becomes $\boldsymbol{X} \in \mathbb{R}^{N_t \times N_{in} \times T_s}$, where $N_t = T/T_s$. Dimensionality reduction via principal component analysis (PCA) is applied to each segment, yielding its principal components as:

$$\boldsymbol{\Sigma_i} = cov(\boldsymbol{X_i}, \boldsymbol{X_i}) \tag{6.4}$$

$$\boldsymbol{\Sigma_i} = \boldsymbol{\Phi_i} \boldsymbol{\Lambda_i} \boldsymbol{\Phi_i}^T \tag{6.5}$$

where, $\boldsymbol{\Sigma_i} \in \mathbb{R}^{N_{in} \times N_{in}}$ denotes the covariance matrix of $\boldsymbol{X_i}$, $\boldsymbol{\Phi_i} \in \mathbb{R}^{N_{in} \times N_{in}}$ is the matrix formed by arranging the eigenvectors of $\boldsymbol{X_i}$, and $\boldsymbol{\Lambda_i} \in \mathbb{R}^{N_{in} \times N_{in}}$ represents the diagonal matrix with the eigenvalues of $\boldsymbol{X_i}$ as its diagonal elements. The column vectors of $\boldsymbol{\Phi_i}$, denoted as $\boldsymbol{\phi_{i,1}} \sim \boldsymbol{\phi_{i,5}}$, correspond to the principal components of $\boldsymbol{X_i}$. Based on these, the weighted Eros similarity between any two segments $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ can be calculated as shown in Eq. (6.6).

$$eros(\boldsymbol{X_i}, \boldsymbol{X_j}, \boldsymbol{\varpi}) = \sum_{k=1}^{N_{in}} \varpi_k \left| \boldsymbol{\phi_{i,k}}^T \boldsymbol{\phi_{j,k}} \right| \tag{6.6}$$

where, $\boldsymbol{\varpi}$ is a group of weights which meets $\sum_{k=1}^{N_{in}} \varpi_k = 1$. The specific values can be

obtained through:

$$\omega_k = \frac{\sum_{i=1}^{N_t} \Lambda_{i,kk}}{\sum_{k=1}^{N_{in}} \sum_{i=1}^{N_t} \Lambda_{i,kk}} \tag{6.7}$$

For slice $\boldsymbol{X_i}$, its representative score S_i can be determined as the average similarity between this slice and others ($eros(\boldsymbol{X_i}, \boldsymbol{X_j}, \boldsymbol{\varpi})$), as expressed in Eq. (6.8).

$$S_i = \frac{1}{N_t} \sum_{j=1}^{N_t} eros(\boldsymbol{X_i}, \boldsymbol{X_j}, \boldsymbol{\varpi}) \tag{6.8}$$

Finally, based on the hardware computational capacity, the maximum number of training data points T_{tr} (i.e., the maximum subset of data points selectable from the virtually unlimited generated T data points for training) is used to define two hyperparameters, N_{top} and N_{bot}, representing the number of high-scoring and low-scoring segments, respectively. These hyperparameters must satisfy the constraint: $(N_{top} + N_{bot})T_s = T_{tr}$.

6.2.3 Application: Damage detection of a three-span model bridge

The model validation in this section utilizes a damage dataset of a three-span bridge provided by the International Conference on Structural Health Monitoring (ICSHM 2022). The benchmark bridge features two platforms, with side spans of 6 m and a central span of 10 m. Five accelerometers were installed on the physical model to calibrate the FEM. As illustrated in Fig. 6.3, the FEM simplifies the two platforms as fixed and roller hinge supports, with a mesh size of 0.5 m, resulting in 44 elements. Boundary conditions, material parameters, and other properties of the FEM were refined using experimental data, ensuring its reliability. Mechanical analysis identifies the mid-span regions of the three spans (specifically Elements 7, 22, and 38) as vulnerable zones, implying potential damage localization near these areas (i.e., $N_{out} = 3$). Damage is simulated by reducing the Young's modulus of these elements, where the reduction magnitude corresponds to the damage severity. Specific damage pattern configurations are summarized in the Table 6.1.

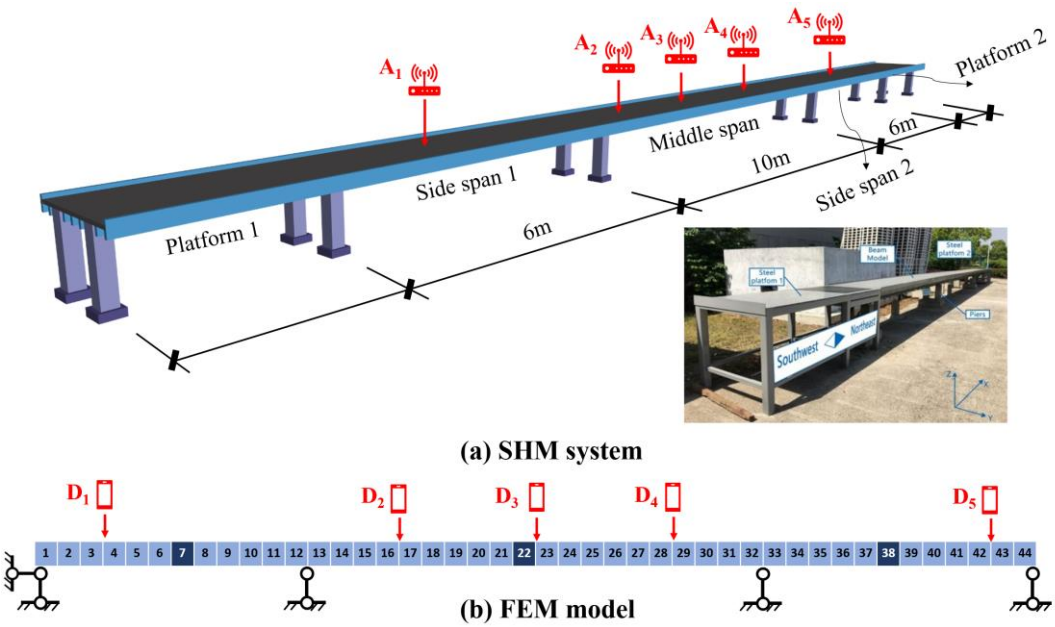


Fig. 6.3 Three-span model bridge and its FEM

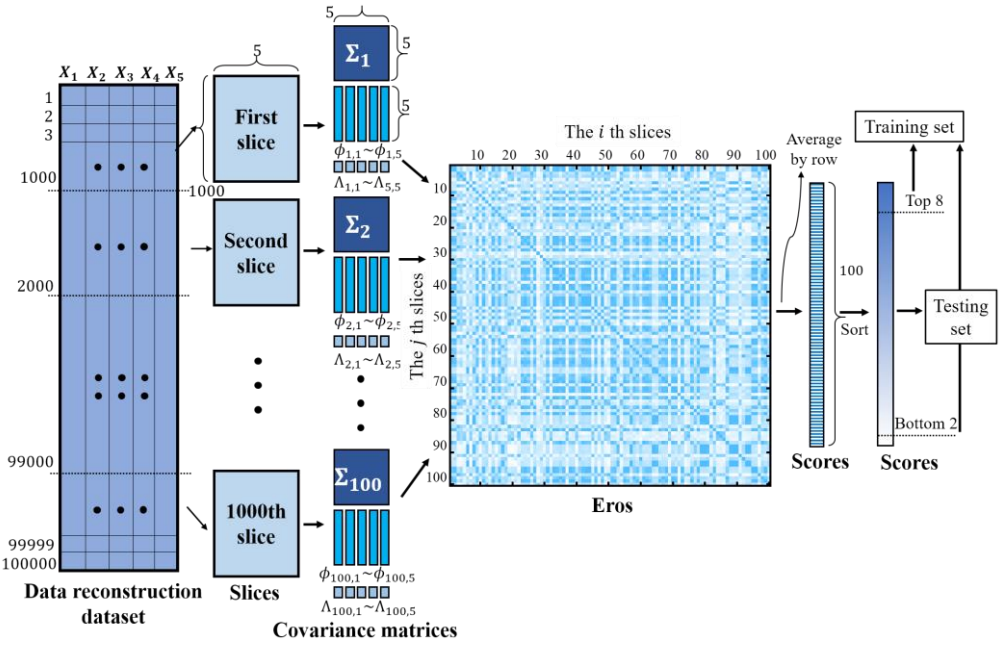


Fig. 6.4 Dataset selection

For each damage pattern, random vehicle loads were independently generated from the same distribution to excite the bridge. The excitation signals were sampled at 100 Hz over a duration of 1000 s (100,000 data points). Acceleration signals from the five sensor locations were recorded as simulated damage-sensitive SHM data ($N_{in} = 5$). The first natural frequency of the bridge, derived from the acceleration responses for each damage

pattern, is listed in Table 6.1. Based on preliminary analysis, the required training sample size was set to $T_{tr} = 10{,}000$. A test set was first randomly partitioned from each damage pattern dataset at a $10{:}1$ ratio. Following the methodology described in Section 6.2.2, training data segments of 10,000 steps were selected for each damage pattern. With $T_s = 1{,}000$, this yielded $N_t = 99$ segments. Hyperparameters N_{top} and N_{bot} were set to 8 and 2, respectively. The data selection process is visualized in Fig. 6.4.

Table 6.1 Three-span model bridge damage scenarios

Damage scenario No.	Damage on Element 7	Damage on Element 22	Damage on Element 38	1st order frequency (Hz)	Dataset
D0	0%	0%	0%	9.4367	$\boldsymbol{D_{D1}^{(0)}(t)}$~$\boldsymbol{D_{D5}^{(0)}(t)}$
D1	0%	20%	0%	9.3613	$\boldsymbol{D_{D1}^{(1)}(t)}$~$\boldsymbol{D_{D5}^{(1)}(t)}$
D2	0%	30%	0%	9.3093	$\boldsymbol{D_{D1}^{(2)}(t)}$~$\boldsymbol{D_{D5}^{(2)}(t)}$
D3	0%	40%	0%	9.2421	$\boldsymbol{D_{D1}^{(3)}(t)}$~$\boldsymbol{D_{D5}^{(3)}(t)}$
D4 (D11)	10% (0%)	0% (0%)	0% (10%)	9.4251	$\boldsymbol{D_{D1}^{(4)}(t)}$~$\boldsymbol{D_{D5}^{(4)}(t)}$
D5 (D12)	30% (0%)	0% (0%)	0% (30%)	9.3924	$\boldsymbol{D_{D1}^{(5)}(t)}$~$\boldsymbol{D_{D5}^{(5)}(t)}$
D6 (D13)	50% (0%)	0% (0%)	0% (50%)	9.3345	$\boldsymbol{D_{D1}^{(6)}(t)}$~$\boldsymbol{D_{D5}^{(6)}(t)}$
D7 (D14)	20% (0%)	20% (20%)	0% (20%)	9.3356	$\boldsymbol{D_{D1}^{(7)}(t)}$~$\boldsymbol{D_{D5}^{(7)}(t)}$
D8 (D15)	20% (0%)	40% (40%)	0% (20%)	9.2167	$\boldsymbol{D_{D1}^{(8)}(t)}$~$\boldsymbol{D_{D5}^{(8)}(t)}$
D9 (D16)	40% (0%)	20% (20%)	0% (40%)	9.2933	$\boldsymbol{D_{D1}^{(9)}(t)}$~$\boldsymbol{D_{D5}^{(9)}(t)}$
D10 (D17)	40% (0%)	40% (40%)	0% (40%)	9.1751	$\boldsymbol{D_{D1}^{(10)}(t)}$~$\boldsymbol{D_{D5}^{(10)}(t)}$

6.2.4 Damage detection results and discussion

This subsection first evaluates the fundamental performance of the MDSC-ESN damage detector, including its accuracy in damage localization and severity quantification. The MDSC-ESN is then benchmarked against damage detection models based on alternative signal classifiers. Additionally, the generalization capability of the MDSC-ESN, specifically its diagnostic accuracy for damage patterns not included in the training set, is analyzed. Finally, the robustness of the model is tested under scenarios involving noise-contaminated data.

Table 6.2 Hyperparameters of MDSC-ESN

N_{res}	α	$\rho(W_{res})$	$spar(W_{res})$
4000	0.9	0.9	$1.0N_{res}$

Table 6.3 Architecture of MDSC-2

Layer No.	Input size	Output size	Kernel size	Stride	Padding
Conv1	$1 \times 1800 \times 1714$	$128 \times 901 \times 857$	$1 \times 5,\ 3 \times 3$	2	1×0
	Activation: $tanh(\cdot)$, Operation: $MaxPool(2 \times 2)$				
Conv2	$128 \times 450 \times 429$	$256 \times 150 \times 143$	5×5	3	1
	Operation: $MaxPool(2 \times 2)$				
Conv3	$256 \times 150 \times 143$	$512 \times 75 \times 72$	3×2	2	0
Conv4	$512 \times 75 \times 72$	$1024 \times 37 \times 36$	3×3	2	0
	Operation: $Dropout(0.2)$, $Flatten(1, 3)$				
Linear	$1024 \times 37 \times 36$	313344×3	None	None	None

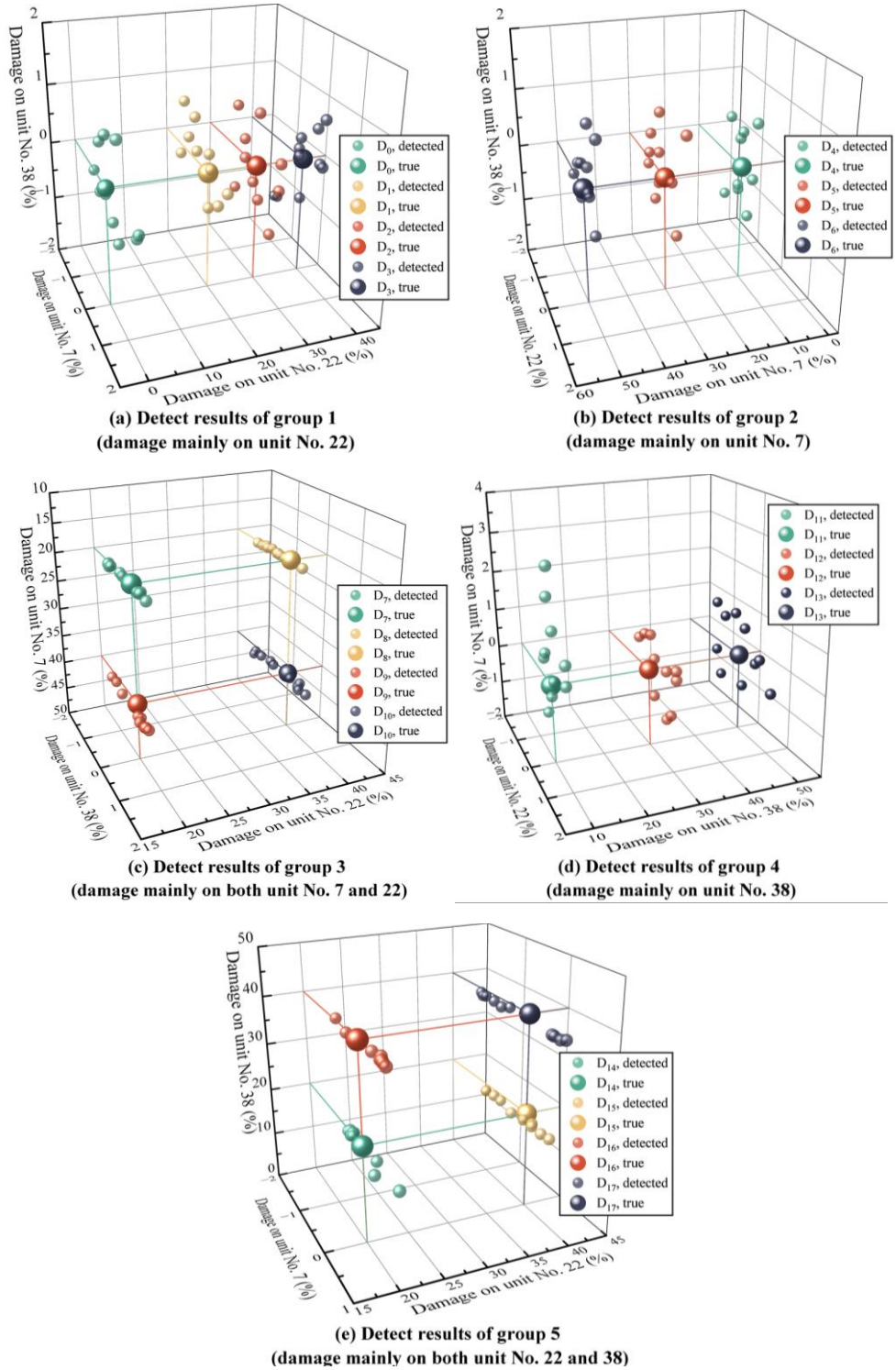
In the MDSC-ESN framework, MDSC-1 employs generic fixed-weight convolutional kernels, as depicted in Fig. 6.1. Hyperparameters for the ESN module are selected via the grid search methodology outlined in Section 4.3.4.3, with detailed results provided in Table 6.2. The architecture of MDSC-2, summarized in Table 6.3, represents a lightweight MDSC configuration with fewer parameters and higher inference speed. However, it is emphasized that the specific structure of MDSC-2 is flexible and can adopt any established convolutional architecture (e.g., ResNet (He et al., 2016), visual geometry group network (VGG) (Szegedy et al., 2015), Inception (Szegedy et al., 2017), ShuffleNet (Zhang et al., 2018), you only look once (YOLO) model (Redmon et al., 2016)) or a custom design tailored to specific tasks. The presented architecture utilizes four convolutional layers, where only the first layer employs a nonlinear hyperbolic tangent activation function to standardize feature distributions. Integration of dropout and pooling layers effectively mitigates overfitting risks. Model training employs SGD with a learning rate $lr = 0.00005$, and $\eta = 0.001$ over 200 epochs.

6.2.4.1 Damage detection results

The proposed MDSC-ESN model was trained to detect potential damage in Elements 7, 22, and 38. For visualization purposes, the 18 damage patterns listed in Table 6.1 were categorized into five groups, as shown in Fig. 6.5. Across nearly all test set diagnoses, the model's predicted damage severities closely aligned with their true values (points identified as 0% damage severity indicate undamaged states, enabling dual outputs for damage localization and quantification). While data acquisition remains challenging in supervised learning frameworks, the model achieves an average accuracy of 1% in estimating equivalent Young's modulus reductions across multiple locations, a level of precision difficult to attain with unsupervised methods.

Specifically, Group 1 exhibited the highest detection errors, with MSEs) of 1.17 and 1.53 for damage severity monitoring in patterns D2 and D3, respectively. This discrepancy may stem from the relatively narrow true-value ranges within Group 1, complicating the distinction between adjacent damage scenarios. The remaining four groups demonstrated

MSEs of 0.64, 0.24, 0.35, and 0.93. Damage pattern D12 showcased the model's highest detection performance, achieving an MSE of 0.018. Overall, the MDSC-ESN model attained an average relative detection error of 0.9% when evaluating Young's modulus reductions across the three mid-span regions for all 18 damage patterns in this study.



(a) Detect results of group 1 (damage mainly on unit No. 22)

(b) Detect results of group 2 (damage mainly on unit No. 7)

(c) Detect results of group 3 (damage mainly on both unit No. 7 and 22)

(d) Detect results of group 4 (damage mainly on unit No. 38)

(e) Detect results of group 5 (damage mainly on both unit No. 22 and 38)

Fig. 6.5 MDSC-ESN damage detection results

It is noteworthy that during training data preparation, leveraging the structural symmetry of the bridge, the training data for damage patterns associated with Element 38 were generated by channel-flipping datasets from damage patterns of the same severity occurring at Element 7. This approach serves two purposes: first, to evaluate the model's ability to learn and interpret physical symmetries; and second, to test its dependency on comprehensive datasets, as practical scenarios may require substituting incomplete damage data (e.g., mirrored data in this case) when location-specific measurements are unavailable. Consequently, the results reveal that the MSE for damage detection at Element 22 (mid-span) in Group 2 (0.06) is consistently lower than that in Group 4 (0.64). This disparity suggests that the model's weaker performance in detecting Element 38 damage compared to Element 7 likely stems from the additional uncertainties introduced by channel-flipped surrogate data. For damage patterns involving simultaneous damage at two locations, the MDSC-ESN model exhibits reduced accuracy in detecting Element 22 damage, with a maximum MSE of 2.25. The error distributions in these cases align with those observed in Group 1, indicating relatively lower precision in identifying mid-span damage. While the model's accuracy fluctuates when diagnosing multi-location damage combinations, its overall results remain valuable for guiding targeted bridge maintenance and reinforcement strategies.

To validate the effectiveness of MDSC-ESN, a series of networks based on recurrent and convolutional architectures were selected for performance comparison. The recurrent architectures include a standard two-layer RNN and LSTM network, both with a state sequence dimension of 50. Additionally, a standard ESN network was included as a reference, with hyperparameters selected according to Table 6.2. All these networks employ a two-layer MLP as their output layer. For convolutional architectures, the same MDSC-1 and recurrent layers used in MDSC-ESN are adopted for feature encoding, though these components remain frozen during training. The comparison primarily focuses on MDSC-2 and other CNN frameworks, including VGG19, ResNet-50, and Inception V2, whose architectures are detailed in their respective references. The

comparative results of these models are summarized in Table 6.4.

Table 6.4 Model comparison on damage detection performance

Group	Standard ESN	Standard RNN	LSTM	MDSC-ESN architectures			
				VGG19	ResNet-50	Inception	MDSC-2
1	29.23	36.69	24.55	20.51	20.23	19.47	19.96
2	24.68	31.26	18.11	16.79	15.12	14.85	14.74
3	27.69	34.18	23.49	20.63	18.33	18.31	18.70
4	20.83	26.80	14.93	10.98	10.85	11.08	11.14
5	25.91	33.76	21.32	19.53	18.55	17.10	16.88

The average MSE values in Table 6.4 represent the mean of MSEs across all damage patterns within each group for damage severity detection at three locations. Notably, the MDSC-ESN-based model demonstrates superior performance in all damage scenarios, achieving at least a 15% improvement in average MSE compared to other architectures. This underscores the capability of deep CNN modules to effectively extract valuable features from structured data (e.g., matrices). In contrast, purely RNN-based models exhibit inferior accuracy in damage detection tasks requiring spatial-dimensional awareness. Specifically, while the ESN outperforms LSTM in pure temporal-domain tasks (e.g., signal reconstruction and prediction), its reservoir-generated features prove overly sparse and disordered in spatial dimensions for damage detection, leading to lower precision than LSTM and approximately 5% errors in Young's modulus reduction predictions. This validates the necessity of the MDSC framework, under which all four CNN modules achieve nearly identical accuracy. However, VGG19 [170] is evidently impractical due to its excessive parameterization and limited computational efficiency. The comparable performance of the remaining three models highlights the flexibility of

MDSC-2 for task-specific fine-tuning. Future integration of more powerful or lightweight CNN modules into this architecture remains feasible.

6.2.4.2 Robustness of MDSC-ESN

This section first discusses the dataset dependency of the MDSC-ESN model, a widely criticized limitation of supervised damage detection algorithms, which require existing labeled damage data for training. However, for most target structures in SHM systems, damage has not yet occurred, and even if it occurs, quantifying damage severity remains challenging. Consequently, damage datasets are typically generated via numerical simulations or scaled-down experiments. Common practices include establishing FEMs of the structure with stiffness reductions at specific locations or creating artificial defects (e.g., notches, thinning) on scaled-down models. These damaged models are then subjected to operational loads to generate responses as simulated damage datasets. Nevertheless, neither prolonged dynamic FEM computations nor limited defect types in scaled experiments can exhaustively cover all potential structural damage scenarios. Thus, supervised damage identification algorithms must exhibit robustness to incomplete datasets, aiming to achieve competent detection capabilities with limited training data. To evaluate this, the model's outputs for each test sample were categorized into corresponding damage scenarios based on distance metrics, converting the regression task into a classification problem during visualization. Fig. 6.6 presents the confusion matrices of the classification results under the following scenarios: sparsifying the damage severity levels within the same location (i.e., same group) in the training set, and including only single-location damage patterns (i.e., Groups 1, 2, and 4) in the training set.

As shown in Fig. 6.6(a), when all 17 damage patterns were included in training (as described in Section 6.2.4.1), the classification accuracy reached 90%, with the MDSC-ESN model demonstrating strong performance in distinguishing damage patterns except for Group 1. When the number of training damage scenarios was reduced to include only the most and least severe cases per suspected location (e.g., removing the 20% severity data in Group 1), the overall classification accuracy dropped to 78% (Fig. 6.6(b)),

particularly for excluded patterns like D2. When the number of damage patterns per group was reduced to one (retaining only the most severe case, e.g., D3 in Group 1), the accuracy further declined to 63% (Fig. 6.6(c)). For damage patterns involving two locations (e.g., D7~D10 and D14~D17), the accuracy further decreased to approximately 50%.

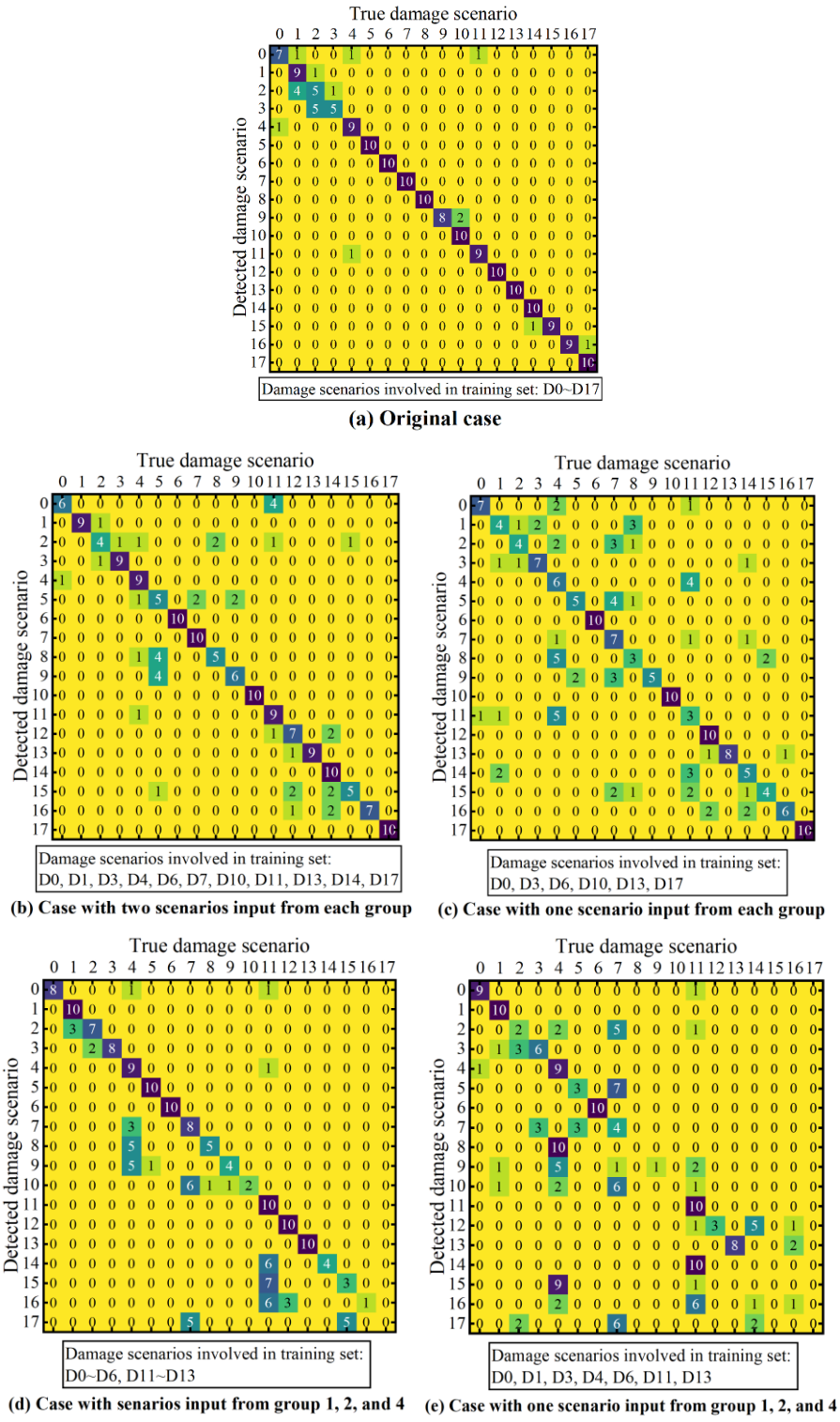


Fig. 6.6 Confusion matrices of damage detection based on datasets with different integrities

Furthermore, Fig. 6.6(d) presents diagnostic results when only single-damage patterns (D1~D6 and D11~D13) are included in the training data, with complete intra-group data for Groups 1, 2, and 4. In this scenario, while the accuracy of damage severity estimation by the MDSC-ESN model becomes unacceptable, damage localization remains largely accurate. Fig. 6.6(e) illustrates an even more extreme case where not only are only single-damage patterns included, but each group contains merely two damage patterns corresponding to the most and least severe cases. Under such conditions, the model struggles to reliably identify multi-damage scenarios. Only when damage severity is sufficiently high (e.g., D10, D17) can the model achieve localization, though the predicted severity values significantly underestimate ground-truth levels. Overall, the model demonstrates reasonable interpolation capabilities for damage severity estimation. However, when trained exclusively on single-location damage data, it primarily retains localization functionality for multi-location damage scenarios. This capability remains valuable, suggesting that in practical applications, collecting separate training data for critical weak points enables the trained model to localize combined multi-location damage. In other words, letting the model see more possible damage location combinations is more useful than just letting it see more damage degrees of a fixed damage location.

On the other hand, simulated damage datasets generated via FEM or scaled experiments are relatively noise-free, whereas real-world structural monitoring inevitably involves measurement noise. This section investigates the robustness of the MDSC-ESN model under varying SNRs, defined here as the ratio of squared signal amplitude to squared noise amplitude. The original dataset, generated through simulation with normalized amplitudes, inherently exhibits infinite SNR. Training data were progressively degraded by adding white noise of increasing variance to reduce SNR. Fig. 6.7 shows the classification accuracy for structural health and 17 damage patterns under different SNR levels. To mitigate randomness, the noise injection process was repeated five times per SNR level, with the results in Fig. 6.7 representing the averaged outcomes.

Clearly, as the noise proportion in the dataset increases, the model's damage detection capability gradually deteriorates. However, when the SNR exceeds 100 (i.e., the noise amplitude is less than 10% of the signal amplitude), the model still achieves average classification accuracy above 85%, demonstrating its robust noise-handling capacity. If the noise proportion surpasses 20% (equivalent to an SNR of 4), the accuracy declines significantly, dropping below 50%, and even to 20% in certain scenarios, indicating severe performance degradation that renders the model impractical without additional preprocessing. In such cases, introducing a pre-filter becomes necessary to mitigate noise interference. Among different damage groups, the model exhibits the highest noise tolerance for Group 1 damage patterns, where accuracy remains minimally affected even at an SNR as low as 25. Conversely, for Group 5 damage patterns, the model shows heightened sensitivity to low-level noise, resulting in a substantial decline in accuracy. Thus, to ensure detection accuracy, filtering is needed, as discussed in Section 4.2.

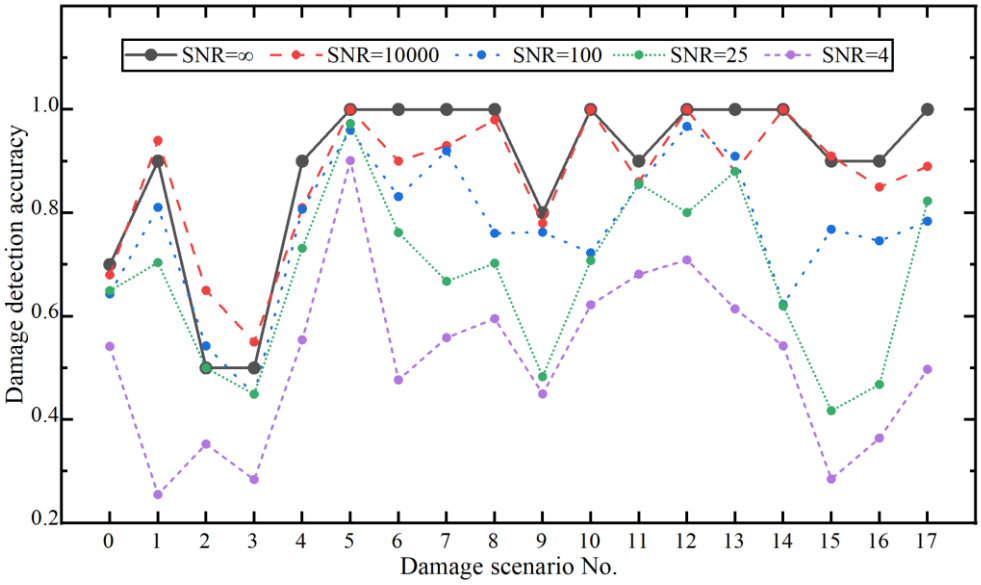


Fig. 6.7 Damage classification accuracy under different SNRs

6.3 DT and structural state assessment

Building upon the aforementioned damage detection framework, this section proposes a DT architecture designed to rapidly process massive and continuous monitoring data while providing real-time feedback on structural conditions. Compared to conventional

damage detection methods, this DT architecture offers two key advantages: bidirectional information interaction between physical and virtual entities, enabling mutual influence, and real-time capability to assess structural states at any operational moment. The core innovation lies in replacing traditional FEM in conventional DTs with a multi-head full-rank reservoir computing (MFRC) network. This substitution allows the virtual entity to simulate structural responses to environmental excitations with significantly faster inference speed than FEM-based dynamic iterative computations. Moreover, while FEM struggles to adapt to slow structural degradation and requires complex model updating, the NN-based virtual entity can be efficiently adjusted through transfer learning fine-tuning. Following damage occurrence or material degradation, the discrepancy between physical and virtual entities manifests as changes in the joint distribution of model-predicted and measured data. Structural state assessment is achieved by quantifying these distributional shifts through Wasserstein Distance (WD) (Panaretos & Zemel 2019) measurements.

As illustrated in Fig. 6.8, the proposed DT architecture adheres to the canonical three-component definition (Ye et al., 2014) for civil structures: the physical entity represented by SHM-acquired structural behavior, the virtual entity comprising the MFRC network that estimates real-time structural responses based on current states, and the connection module handling model updating and damage identification. The system primarily monitors dynamic responses (e.g., acceleration) under seismic excitations, though the framework remains applicable to other excitation types. The MFRC-based virtual model construction initiates during the design phase (Fig. 6.8(a)), where preliminary training utilizes excitation-response datasets from numerical simulations or scaled experiments. Post-construction SHM deployment enables model validation through field-testing data, serving as project acceptance criteria. Initial model calibration involves retraining the network's final layer to minimize measurement noise and modeling errors.

During operational phases (Fig. 6.8(b)), the SHM system continuously streams measured excitations and responses, while the MFRC network generates parallel predicted

responses. A probabilistic damage identification module evaluates structural conditions by analyzing real-time error distributions between measured and predicted responses. When detected damage severity exceeds predefined thresholds and undergoes field verification, the architecture triggers structural reinforcement protocols, establishing information flow from virtual to physical entities. For sub-threshold degradations, transfer learning-based parameter updating automatically adapts the MFRC network to current structural states, facilitating physical-to-virtual information transfer. Notably, this parameter adaptation mechanism remains effective for post-rehabilitation synchronization when structural retrofitting alters mechanical properties.

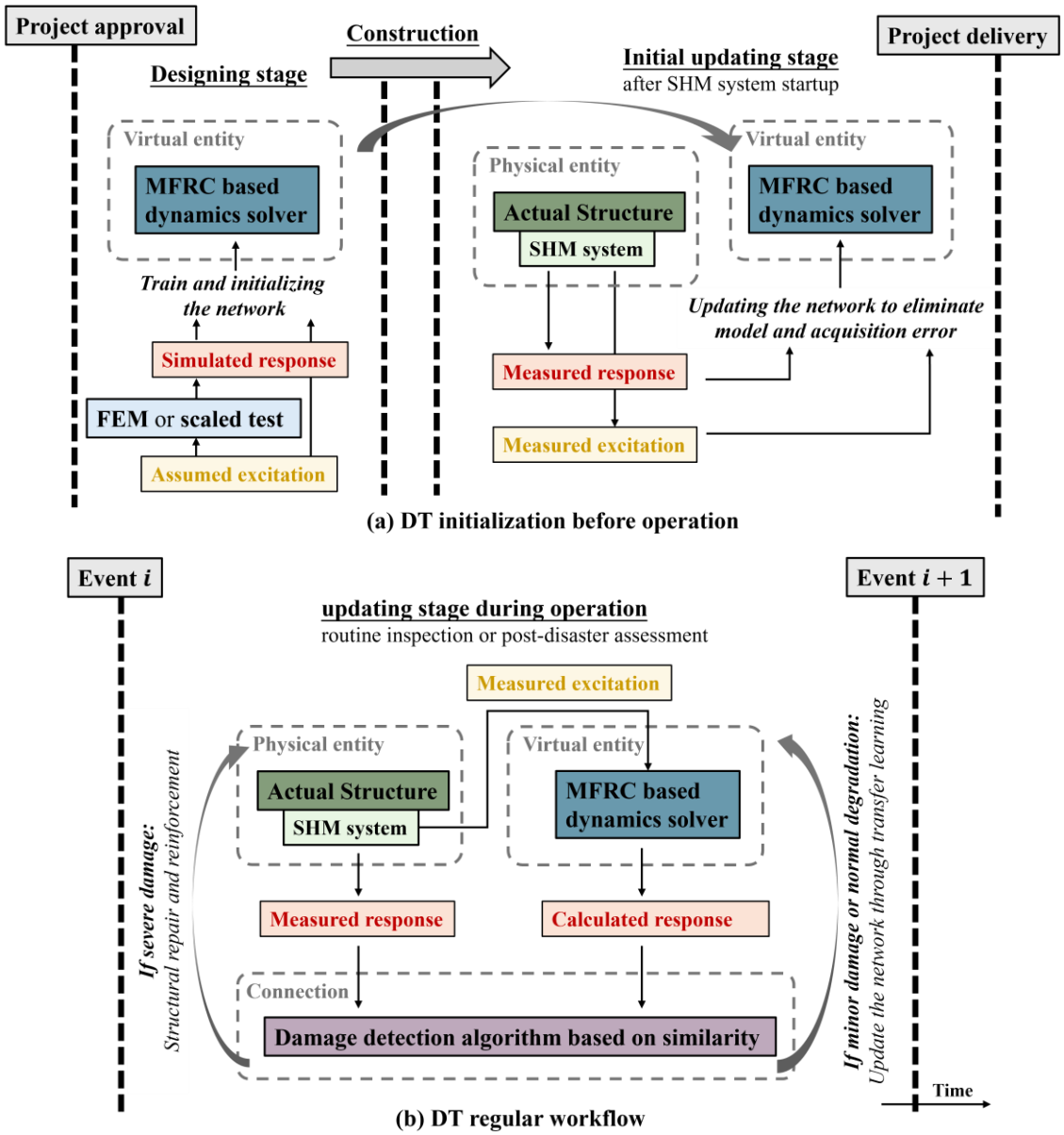


Fig. 6.8 Architecture of the proposed DT

6.3.1 MFRC and its updating

The excitation, response, and initial structural states satisfy dynamic equations, traditionally solved through step-by-step integration methods such as the Newmark-β (Pourzeynali et al., 2021) and Wilson-θ (Ozkul 2023) algorithms. These methods explicitly integrate historical excitations and states, calculating current responses based on previous states and current excitations. To ensure numerical stability, extremely small iteration steps are typically required, leading to high computational costs. Additionally, solution accuracy heavily depends on the quality of dynamic equation modeling (e.g., mesh resolution, material properties, and damping assumptions in FEM). With the rise of ML, it becomes evident that the architecture of RNNs resembles iterative solvers, as they encode input sequences into state sequences through delayed feedback loops to address Seq2Seq problems. Unlike traditional methods, RNN architectures can be flexibly defined, with their mapping matrices determined by training on existing input-output datasets. In contrast, conventionally computed or measured excitation-response datasets provide no direct assistance for solving responses to new excitations, though they inherently encapsulate sufficient input-output relationships for network training and prediction.

This work adopts the previously validated ESN as the foundational architecture for the virtual entity, where the input $\boldsymbol{X}$ represents structural excitation sequences and the output $\boldsymbol{Y}$ corresponds to structural response sequences. The encoding process aligns with Eq. (4.34). As discussed in Section 4.3.1, the number of reservoir channels $\boldsymbol{R}$ must substantially exceed the input channels $\boldsymbol{X}$ to avoid meticulous design of individual elements in the encoding matrices ($\boldsymbol{W_{in}}$ and $\boldsymbol{W_{res}}$). Input sequences and historical state information are combined and propagated through distinct rows of these matrices. Given their inherent sparsity, the reservoir's topology remains computationally tractable, with each state sequence effectively capturing input features from specific "perspectives". In conventional ESNs, $\boldsymbol{W_{res}}$ and $\boldsymbol{W_{in}}$ are randomly generated based on hyperparameters (e.g., N_{res} for matrix dimensions, $\rho(\boldsymbol{W_{res}})$ for spectral radius control, $spar(\boldsymbol{W_{res}})$

for sparsity regulation). However, randomly generated matrices, particularly $\boldsymbol{W_{res}}$, may contain entirely zero rows or linearly dependent rows when sparsity is high, resulting in redundant or ineffective reservoir channels. To enhance ESN efficiency, this study proposes an algorithm for generating full-rank $\boldsymbol{W_{res}}$, as formalized in Eqs. (6.9), (6.10), (6.11), and (6.12).

$$\boldsymbol{W_{res}} = diag(N_{res}, \mathcal{U}[0,1]) \tag{6.9}$$

$$\boldsymbol{W_{res}} \leftarrow \pi(\boldsymbol{W_{res}}, \boldsymbol{\mathcal{O}}) \tag{6.10}$$

$$\boldsymbol{W_{res}} \leftarrow supp(\boldsymbol{W_{res}}, spar, \mathcal{U}[0,1] | rank(\boldsymbol{W_{res}}) = N_{res}) \tag{6.11}$$

$$\boldsymbol{W_{res}} \leftarrow \rho \frac{\boldsymbol{W_{res}}}{\rho(\boldsymbol{W_{res}})} \tag{6.12}$$

The generation of $\boldsymbol{W_{res}}$ proceeds through four stages. First, a diagonal matrix is constructed by sampling diagonal elements from a uniform distribution. Next, a permutation operation $\pi(\cdot)$ reorders the rows of this diagonal matrix according to an index vector $\boldsymbol{\mathcal{O}} \in \mathbb{R}^{N_{res}}$, forming a sparse matrix with full rank ($rank(\boldsymbol{W_{res}}) = N_{res}$) and an initial sparsity pattern. To achieve the target sparsity level $spar$, additional non-zero elements are incrementally introduced by replacing zero entries selected via the support function $supp(\cdot)$. Each replacement ensures the matrix retains full rank, with newly added elements sampled from the same uniform distribution. Finally, the matrix undergoes spectral radius scaling to meet the target value ρ.

While fixed $\boldsymbol{W_{res}}$ matrices represent single encoding modes, they may not be optimal. Inspired by multi-head attention mechanisms, this study further proposes a multi-head reservoir architecture comprising multiple mapping matrices. Distinct reservoirs with varied dimensions, hyperparameters, and sampling configurations perform N_{head} independent state sequence computations, which are concatenated via Eq. (6.13). Compared to using a monolithic reservoir with a channel count equal to the sum of all heads, this approach substantially reduces computational costs by leveraging block-diagonalized matrix operations. Simultaneously, it enables projection of input $\boldsymbol{X}$ into diverse subspaces, extracting richer and more heterogeneous state features.

$$\boldsymbol{R} = [\boldsymbol{R_{head1}}; \boldsymbol{R_{head2}}; \dots; \boldsymbol{R_{headN_h}}] \tag{6.13}$$

The output layer maps reservoir states to target sequences through a trainable mapping matrix $\boldsymbol{W_{out}}$, which linearly combines $\boldsymbol{R}$ to fit the target sequences, analogous to Eq. (4.35). To enhance the memory capacity of the MFRC, residual connections and autoregressive feedback can be incorporated into the output layer. Additionally, a multi-step lookback mechanism may be adopted, further demonstrating the superior flexibility of ML-based approaches over FEM-based methods. The estimated target sequence ($\widehat{\boldsymbol{Y}}$) is computed as follows:

$$\widehat{\boldsymbol{Y}}(t) = \boldsymbol{W^{out}}\big[\boldsymbol{X}(t), \dots, \boldsymbol{X}(t - t_x), \boldsymbol{R}(t), \boldsymbol{Y}(t-1), \dots, \boldsymbol{Y}\big(t - t_y\big)\big] + \boldsymbol{b_Y} \tag{6.14}$$

When employing MSE as the loss function, it is observed that its gradient exhibits a linear dependence on the estimation error. However, such errors may also originate from noise in the ground truth, rendering MSE overly sensitive to outliers in the target sequences. This compels the model parameters to overfit low-quality data segments at significant cost. Considering the heterogeneity, diversity, and noisy nature of lifecycle data encountered in DT applications, as well as potential instability in the initial phase caused by random initial states, this section modifies the smooth quadratic loss (SQL) (Wang et al., 2024) proposed in prior work and defines the MFRC loss function as follows:

$$L\big(\widehat{\boldsymbol{Y}}, \boldsymbol{Y}\big) = \sum_{k=1}^{N_{out}} \left(\begin{array}{c} \dfrac{\vartheta_1}{(T - T_a)} \displaystyle\sum_{t=T_a}^{T} \dfrac{\big(Y_k(t) - \widehat{Y}_k(t)\big)^2}{\big(Y_k(t) - \widehat{Y}_k(t)\big)^2 + \vartheta_2} + \\ \dfrac{1 - \vartheta_1}{(T - T_a)} \displaystyle\sum_{t=T_a}^{T} \big|Y_k(t) - \widehat{Y}_k(t)\big| \end{array} \right) + \eta \|\boldsymbol{W^{out}}\|^2 \tag{6.15}$$

where ϑ_1 and ϑ_2 are adjustable parameters.

Another key advantage of replacing FEM with MFRC lies in its adaptability for updates. In this study, transfer learning is employed to adapt virtual entities pre-trained on initial datasets to real-world measured data, with updates performed after each disaster event or structural reinforcement. Both the source task (pre-training) and target task (adaptation) in this transfer scenario involve labeled datasets within the same domain, aligning with

the fundamental requirements of fine-tuning in transfer learning (Weiss, Khoshgoftaar & Wang 2016). Assuming the virtual entity has effectively captured the mapping between $\boldsymbol{X}$ and $\boldsymbol{Y}$ (denoted as $\{\boldsymbol{X^{ini}}, \boldsymbol{Y^{ini}}\}$) during pre-training, minor discrepancies arising from structural damage or model errors across distinct service-phase dataset groups ($\{\boldsymbol{X^{ser}}, \boldsymbol{Y^{ser}}\}$, which may correspond to multiple structural states over time) can be corrected through parameter adjustments.

To achieve this, the output module of MFRC is replaced with an MLP that maintains similar neuron ratios between adjacent layers. After pre-training, parameters in the initial layers of the MLP are frozen, while those in the final two layers remain trainable and are re-calibrated through limited epochs of retraining on new datasets. A reduced learning rate is applied during updates to capture subtle shifts in mapping relationships. Additionally, a unified early stopping strategy and maximum epoch limit are implemented to accelerate convergence and prevent overfitting. This transfer learning strategy enables rapid parameter convergence and requires only minimal input data for updates, facilitating prompt post-disaster assessment and early warning in MFRC-based DT systems. The structural configuration of the virtual entity is illustrated in Fig. 6.9.

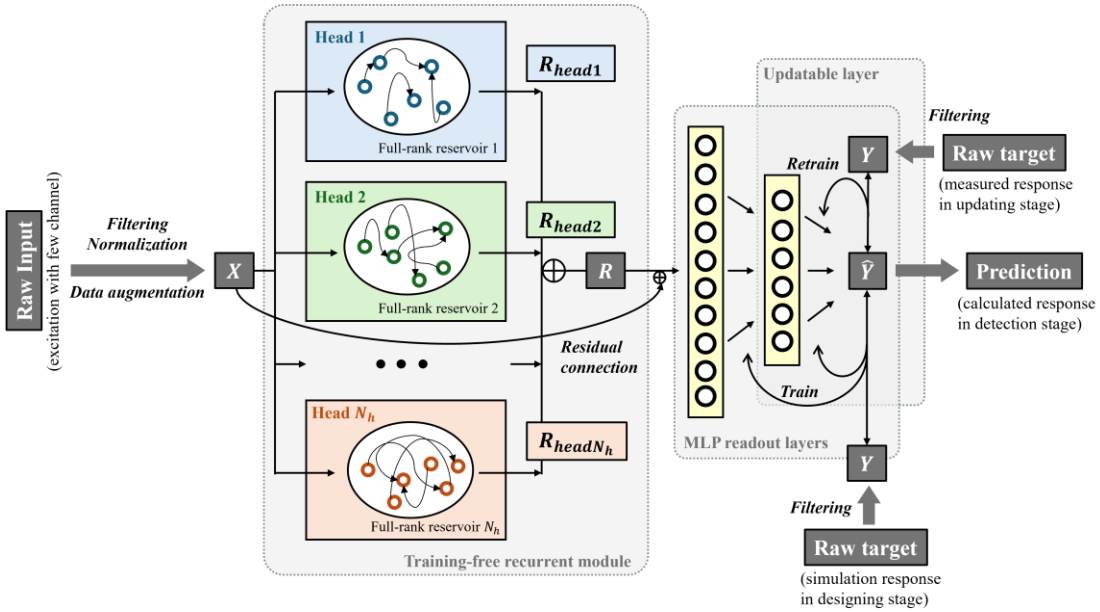


Fig. 6.9 MFRC

6.3.2 State assessment based on WD

When the structural state of the target remains unchanged, a well-trained virtual entity

maintains stable accuracy. Consequently, the joint distribution of the error between estimated and measured responses (denoted as $\boldsymbol{e_y} = \widehat{\boldsymbol{Y}} - \boldsymbol{Y}$) across its multiple channels, referred to as $\boldsymbol{\mathcal{D}_{ori}}$, remains nearly constant and unbiased, with variances significantly smaller than the vibration amplitudes. After structural state changes (induced by damage or reinforcement), the error $\boldsymbol{e_y}$ computed by the original virtual entity follows a distinct distribution ($\boldsymbol{\mathcal{D}_{ser}}$). Thus, the similarity (or distance) between $\boldsymbol{\mathcal{D}_{ori}}$ and $\boldsymbol{\mathcal{D}_{ser}}$ can serve as an indicator of damage, abnormality, or a signal of model obsolescence.

For the kth output channel, the distributions of $\boldsymbol{e_{y,k}}$ are denoted as $\mathcal{D}_{ori,k}$ and $\mathcal{D}_{ser,k}$. Based on the error sequence range, frequency distribution histograms of the two distributions can be plotted (Fig. 6.10) after defining a uniform number of bins (N_g). The central values of all statistical bins for $\mathcal{D}_{ori,k}$ and $\mathcal{D}_{ser,k}$ are $\mathcal{u}_{i,1}, \mathcal{u}_{i,2}, \ldots, \mathcal{u}_{i,N_g}$ and $\mathcal{v}_{i,1}, \mathcal{v}_{i,2}, \ldots, \mathcal{v}_{i,N_g}$, respectively. The distance between $\mathcal{D}_{ori,k}$ and $\mathcal{D}_{ser,k}$ is defined as the minimum effort required to transform the histogram of $\mathcal{D}_{ori,k}$ to that of $\mathcal{D}_{ser,k}$, quantified by the WD (Panaretos & Zemel 2019), as formulated in Eq. (6.16).

$$D_W\left(\mathcal{D}_{ori,k}, \mathcal{D}_{ser,k}\right) = \underset{\Pi_{ij}}{argmin}\left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(\Pi_{ij}\left\|\mathcal{u}_{k,i}, \mathcal{v}_{k,j}\right\|\right)\right) \tag{6.16}$$

where, Π_{ij} represents the effort (i.e., frequency) required to shift from the $\mathcal{u}_{k,i}$ bin to the $\mathcal{v}_{k,j}$ bin, where $\left\|\mathcal{u}_{k,i}, \mathcal{v}_{k,j}\right\|$ denotes the transport distance. Π can also be interpreted as a joint distribution with marginals $\mathcal{D}_{ori,k}$ and $\mathcal{D}_{ser,k}$, as illustrated in Fig. 6.10. The Wasserstein distance $D_W\left(\mathcal{D}_{ori,k}, \mathcal{D}_{ser,k}\right)$ is hereafter abbreviated as $D_{W,k}$. Consequently, the sequence $\boldsymbol{D_W}$ (comprising $D_{W,1}, D_{W,2}, \ldots, D_{W,N_{out}}$) is selected as the primary damage indicator in the MFRC-based DT framework.

Notably, the rationale for transitioning from direct time-domain accuracy comparisons to distribution similarity measurements lies in mitigating noise effects. While noise introduces fluctuations in solution accuracy, it imposes probabilistically equivalent impacts across all histogram bins, which are collectively neutralized when quantifying inter-distribution discrepancies. The sole requirement for this approach is a sufficiently

large sample size (i.e., signal segments for damage detection must be sufficiently long), a condition easily satisfied in long-term SHM systems.

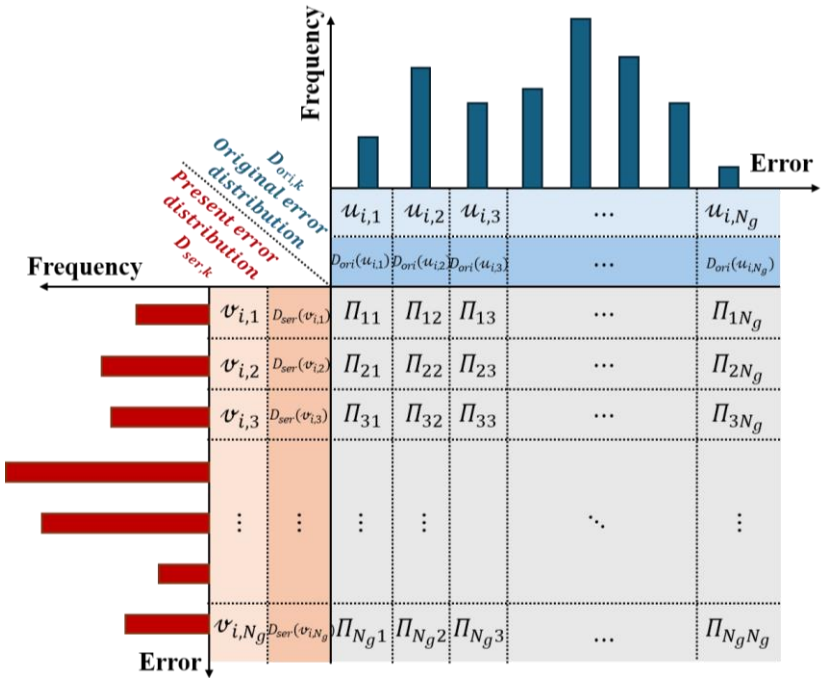


Fig. 6.10 WD calculation process

6.3.3 Application: High-rise building shaking table test

This section demonstrates the rapid state assessment capability of the MFRC-based DT using shaking table test data from a scaled-down high-rise building model (in practical applications, this would correspond to SHM-measured data from in-service structures). In the test, the model building was subjected to minor, moderate, major, and super major earthquakes, inducing varying degrees of damage. The establishment, training, updating, and diagnosis of the DT model were all conducted within this dataset.

The prototype of the scaled-down model is a reinforced concrete high-rise residential building featuring a core tube spanning 35 standard floors and a three-story equipment tower. The standard floors and upper sections are constructed above a four-story podium, connected via a thickened transfer slab. The concrete used has a strength of 40 MPa, with a reinforcement ratio of approximately 2% and steel bars exhibiting an ultimate strength exceeding 500 MPa, resulting in an in-plane lateral stiffness of approximately 2.2×10^6 kN/m per floor in both principal directions. Detailed structural specifications

are provided in reference (Li et al., 2006). This building represents one of the typical structures widely constructed under Hong Kong's BS8110 design code during the late 20th and early 21st centuries. Such structures, now aging and facing increasing risks of deterioration and damage, pose potential urban hazards, as illustrated in Fig. 6.11.

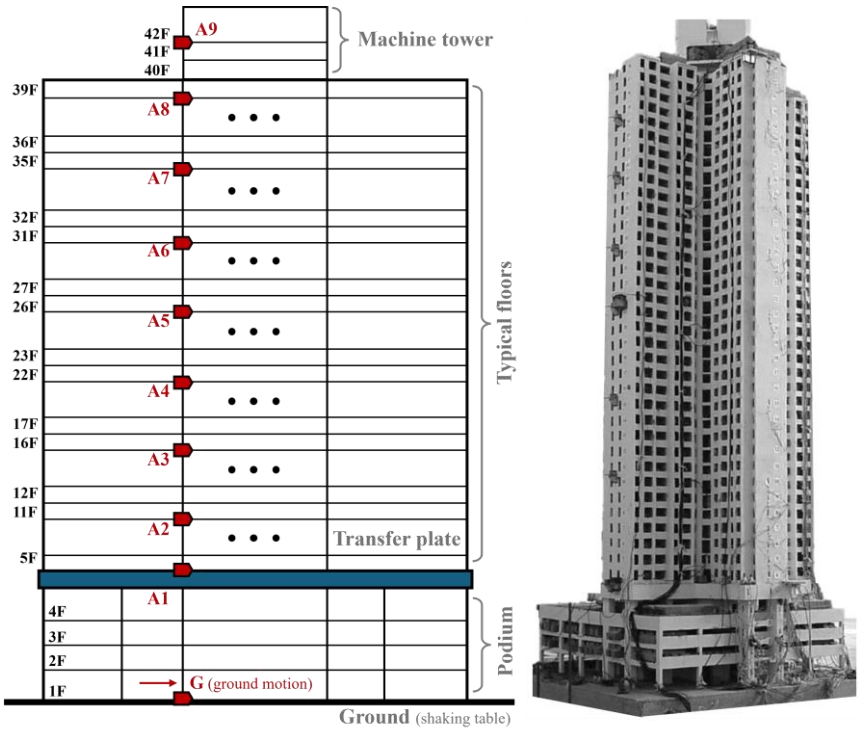


Fig. 6.11 Test specimen and its sensor system

The scaled model (1: 20 ratio, dimensions: $2.4m \times 2.2m \times 6.5m$) was tested on a six-DoF shaking table ($5m \times 5m$ platform) to capture vibration responses under varying damage levels. Four seismic waves of increasing intensity, representing minor, moderate, major, and super major earthquakes, were sequentially applied. Prior to seismic excitation, a 10-minute free-vibration signal was recorded. Similarly, after each seismic event, structural responses under white noise excitation of identical duration were measured. To ensure damage occurrence, the excitation was aligned with the weak axis of the building's plan, while state assessments along the strong axis followed technically identical procedures and are not presented here. For the same rationale, only simulated soft soil site conditions were replicated in the shaking table tests. The seismic input adopted the north-south component of the 1940 El Centro earthquake (30 s duration), with peak

ground accelerations adjusted to 0.02 g, 0.08 g, 0.15 g, and 0.25 g to match the four intensity levels. In this study, accelerations at critical floors were selected as the primary measurement and evaluation variables. Nine accelerometers, arranged as shown in Fig. 6.11, recorded data at a sampling rate of 100 Hz, with detailed test protocols summarized in Table 6.5.

Table 6.5 Shaking table test scenario

Test content	Excitation	Duration	Dataset
Free vibration	White noise	10 min	$\boldsymbol{X^c, Y^c}$
Minor earthquake	El Centro wave ($\mathbf{0.008}\ \boldsymbol{g}$)	30 s	/
Free vibration	White noise	10 min	$\boldsymbol{X^{mi}, Y^{mi}}$
Moderate earthquake	El Centro wave ($\mathbf{0.02}\ \boldsymbol{g}$)	30 s	/
Free vibration	White noise	10 min	$\boldsymbol{X^{mo}, Y^{mo}}$
Major earthquake	El Centro wave ($\mathbf{0.15}\ \boldsymbol{g}$)	30 s	/
Free vibration	White noise	10 min	$\boldsymbol{X^{ma}, Y^{ma}}$
Super major earthquake	El Centro wave ($\mathbf{0.25}\ \boldsymbol{g}$)	30 s	/
Free vibration	White noise	10 min	$\boldsymbol{X^{sm}, Y^{sm}}$

To demonstrate the assessment capability of the DT model throughout the entire service life of a structure, the DT is constructed starting from the structural design phase. At this stage, input-output data generated by a FEM (denoted as $\{\boldsymbol{X^{pre}, Y^{pre}}\}$; in practical applications, this would correspond to data from scaled mechanical experiments) are required to build and initialize the model. The input $\boldsymbol{X^{pre}}$ consists of white noise excitation with a duration of 60 s and a time step of 0.01 s, consistent with measured data.

$\{\boldsymbol{X^{pre}}, \boldsymbol{Y^{pre}}\}$ is partitioned into training and testing sets in a $5:1$ ratio. The datasets presented in Table 6.5, representing structural data at different stages, are primarily used for state assessment and model updating.

6.3.4 State assessment results and discussion

This section sequentially presents the workflow of the MFRC-based DT model. The process encompasses initialization using simulated or experimental data, calibration with SHM data from healthy structures, state assessment after routine disasters, model updating, and guidance for structural repair and reinforcement following severe disasters. The superiority of the MFRC model will be demonstrated. Since its network architecture has been optimized, hyperparameter selection will be revisited in Section 6.3.4.1, with specific values listed in Table 6.6. Both pre-training and retraining of the model employ the SGD optimizer. During pre-training, a learning rate $lr = 0.00005$ and $\eta = 0.001$ are applied over 100 epochs. Each retraining phase adopts a reduced learning rate $lr = 0.00001$ for 50 epochs. The convergence behavior for each training phase will be elaborated in subsequent discussions.

Table 6.6 Hyperparameters of MFRC

$\boldsymbol{N_{res}}$	$\boldsymbol{\alpha}$	$\boldsymbol{N_{head}}$	$\boldsymbol{\rho(W_{res})}$	$\boldsymbol{spar_f}$
4000	0.85	6	0.9	3000

6.3.4.1 DT initialization and calibration

After the MFRC-based model was established and trained using the aforementioned methodology, it was validated using the test set from $\{\boldsymbol{X^{pre}}, \boldsymbol{Y^{pre}}\}$ (1,000 data points), with results illustrated in Fig. 6.12. The MSE shown in Fig. 6.12 indicates that the proposed MFRC-based virtual model achieves at least 90% accuracy in predicting structural responses at nine critical floors under white-noise ground motion excitation. Since high-rise buildings can be idealized as cantilever beams fixed to the ground, vibration amplitudes increase with elevation. Consequently, even with consistent model

accuracy, the MSE of signals measured by accelerometers at higher elevations naturally grows larger.

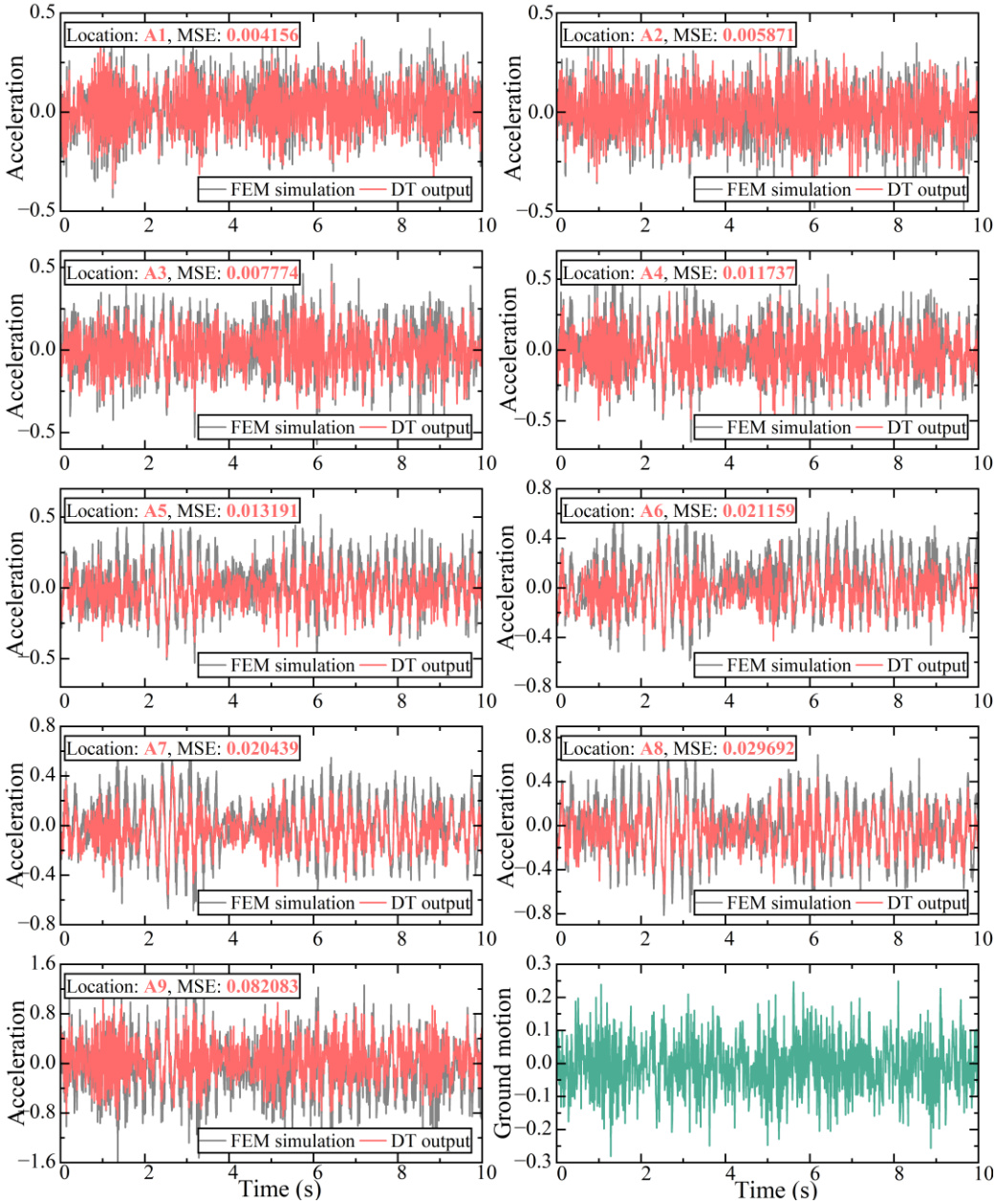


Fig. 6.12 Structural responses output by initial MFRC

Additionally, the computational time required to obtain a 1000-step response sequence was reduced from 20 minutes (iterative solving with the FEM) to 0.005 s (output from the MFRC-based DT), which is shorter than the actual duration of the time series (10 s). This demonstrates the computational efficiency of the proposed DT in enabling real-time monitoring and response determination. FFT was further applied to both simulated

response data and MFRC outputs, with signal spectra presented in Fig. 6.13. Notably, the structure behaves like a filter under seismic excitation, progressively eliminating high-frequency components from its responses. With increasing elevation, the vibration frequencies of the structure converge toward the natural frequency obtained from finite element eigenvalue analysis, except for accelerometer A9 located at the equipment tower. The MFRC model is shown to effectively capture this behavior.

Upon completion of structural construction and activation of the SHM system, the virtual entity of the DT requires calibration to mitigate errors from FEM and measurement noise. Taking accelerometer A1 as an example, Fig. 6.14 reveals that directly inputting a 1,000-step excitation from $\boldsymbol{X^c}$ into the MFRC model trained on finite element data results in a larger discrepancy between the predicted response $\widehat{\boldsymbol{Y}}^{\boldsymbol{c}}$ and the sensor-measured response $\boldsymbol{Y^c}$. The MSE increases from 0.0041 to 0.0131, which can be attributed to distributional differences between the measured dataset ($\{\boldsymbol{X^c}, \boldsymbol{Y^c}\}$) and the finite element dataset ($\{\boldsymbol{X^{pre}}, \boldsymbol{Y^{pre}}\}$). Specifically, the measured dataset exhibits greater variance and higher-frequency components due to noise contamination.

Therefore, in addition to the pre-training phase spanning 100 epochs, a retraining phase of 50 epochs (applied exclusively to the last fully connected layer) is conducted to calibrate the MFRC-based virtual entity. The training data used for this purpose comprises the first 5,000 steps of excitation-response data from $\{\boldsymbol{X^c}, \boldsymbol{Y^c}\}$ (measured from the physical entity). To avoid overfitting and reduce training time during calibration, an early stopping technique is adopted. The early stopping strategy is defined as follows: a 1000-step validation set is selected from $\{\boldsymbol{X^c}, \boldsymbol{Y^c}\}$, independent of the training and test sets; the MSE of the MFRC's predictions on this validation set is monitored; the retraining process terminates if the validation MSE fails to decrease by more than 1% of its current value for five consecutive epochs. During both pre-training and retraining, the shapes of the training loss versus epoch curves are similar, as shown in Fig. 6.14. The retraining phase achieves convergence in approximately 30 epochs, as the model's capability as a dynamic solver is retained despite changes in the mapping relationship to be learned. After

retraining, the test MSE for channel A1 decreases to 0.0047. This MSE indicates that the accuracy of the transfer-learned model is preserved even as the volatility of the target responses increases. Similar phenomena and accuracy levels are observed in the MFRC predictions at the other eight locations (A2~A8).

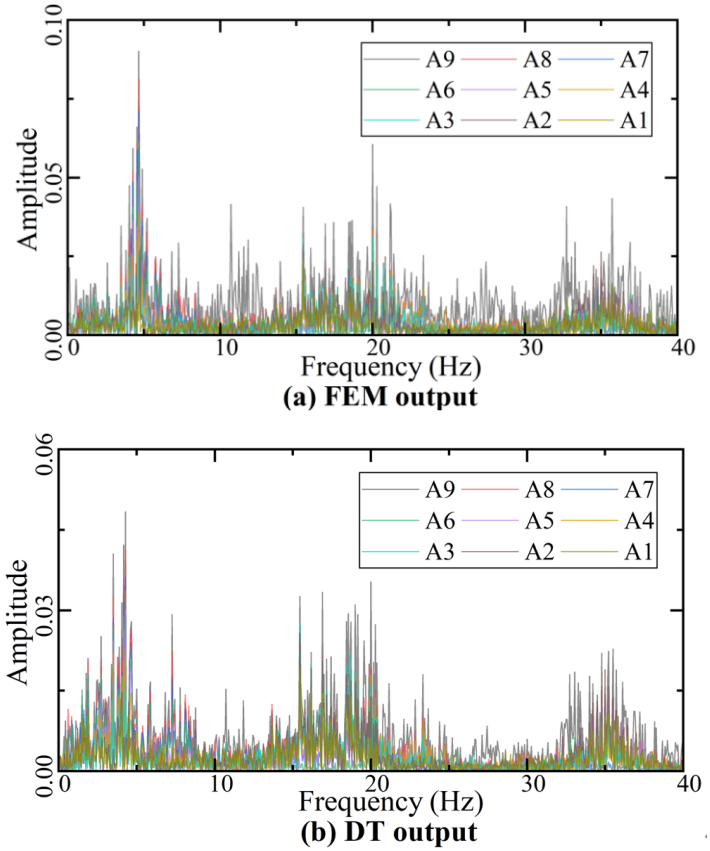


Fig. 6.13 Frequency spectrum comparison of the structural responses

Following calibration, the MFRC-based DT can be deployed for daily structural monitoring. In the absence of damaging excitations (e.g., typhoons, earthquakes, or impact loads), the structure undergoes free vibrations under minor random excitations, during which the virtual entity of the DT maintains relatively high and stable accuracy in response prediction. This phase may persist for extended periods until the first damaging event occurs. The unused portion of $\{\boldsymbol{X^c}, \boldsymbol{Y^c}\}$ (54,000 data points) is utilized to simulate the daily monitoring process. The dataset is segmented into 1,000-step sequences, and the prediction accuracy of the model for 54 segments of A1 response segments is illustrated in Fig. 6.15. The MSEs of the calibrated MFRC model for response estimation at the other

eight locations are summarized in Table 6.7.

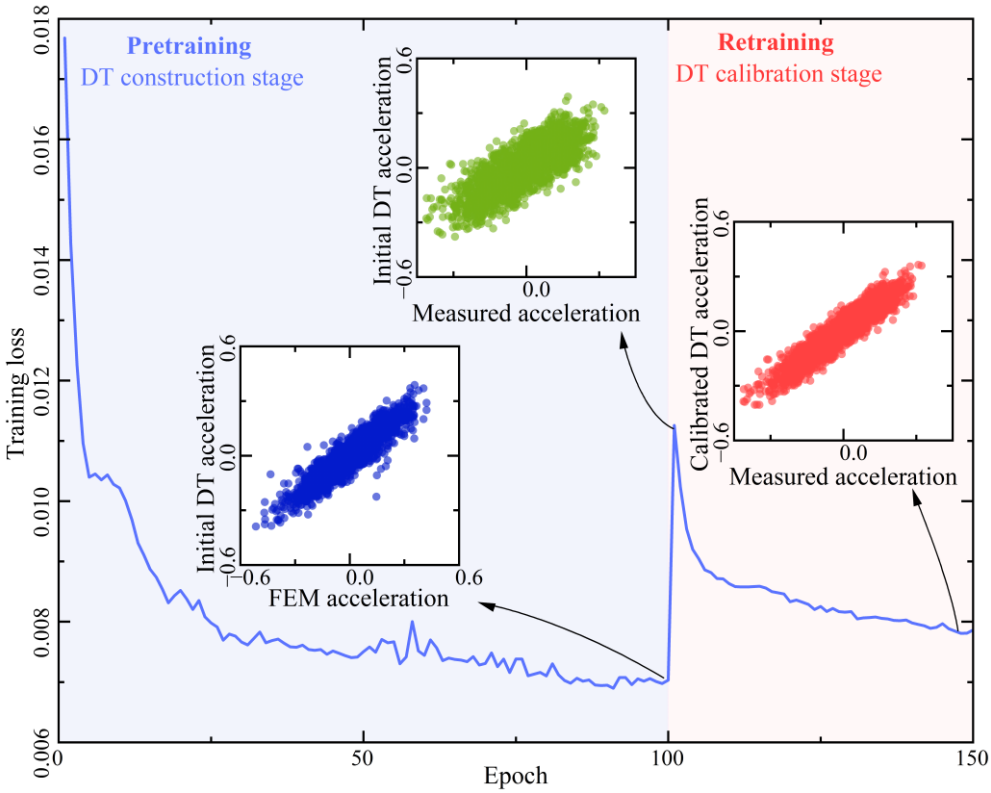


Fig. 6.14 Training loss-epoch curves of MFRC during initialization and calibration process

As shown in Fig. 6.15, the MSE between the model-predicted and measured A1 responses fluctuates around 0.0052, with maximum and minimum segment errors of approximately 0.007 and 0.004, respectively. During this phase, the excitations approximate white noise, with no correlation between sequences. However, as long as the structural properties remain unchanged, the DT consistently demonstrates stable response prediction capabilities. This confirms the suitability of the proposed MFRC network for long-term monitoring tasks, requiring only short-term excitation-response datasets for initial training. Similar prediction performance is observed at the other eight accelerometer locations. The MSEs listed in Table 6.7 are slightly higher than those in Fig. 6.12, attributed to discrepancies between field-measured and simulated data domains. These discrepancies, primarily caused by noise, make it more challenging for the calibrated MFRC to estimate noisy structural vibrations from noisy excitation inputs.

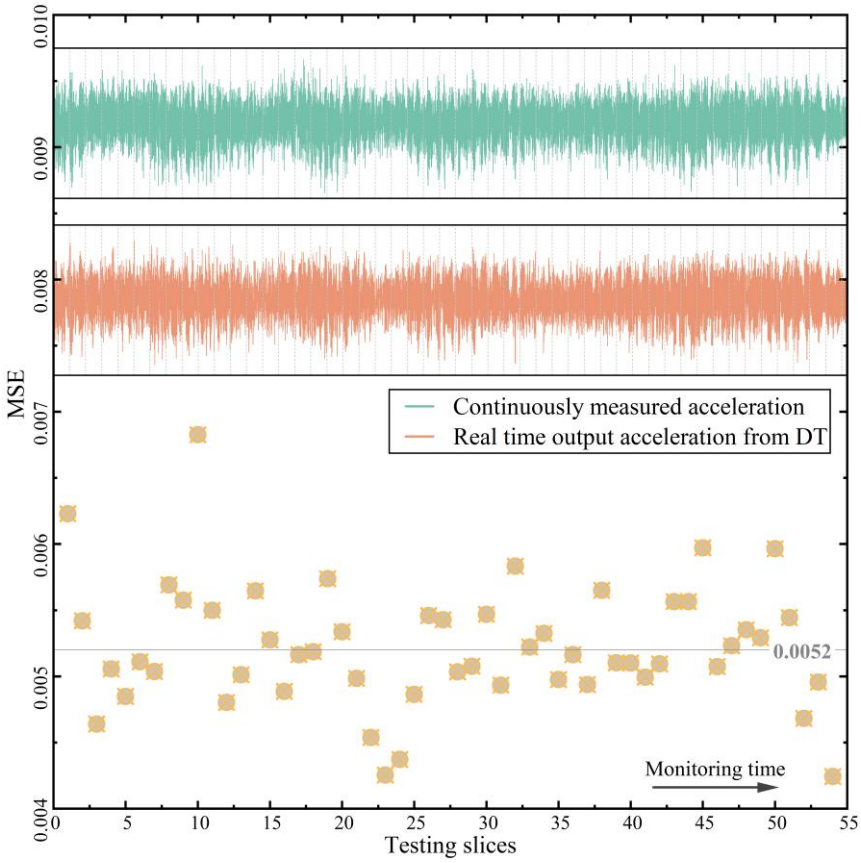


Fig. 6.15 MFRC A1 solving accuracy during long term monitoring

Table 6.7 MFRC solving accuracy of other channels during long term monitoring

Sensor No.	Solving MSE	Sensor No.	Solving MSE
A2	0.006639	A6	0.024324
A3	0.007678	A7	0.030358
A4	0.009482	A8	0.042567
A5	0.017112	A9	0.101475

To ensure the reliability of the results in this section and subsequent analyses, the superiority of the MFRC model over other Seq2Seq models in structural response prediction is validated, focusing on the effectiveness of its multi-head structure and full-rank criterion. Additionally, the hyperparameter selection for the ESN under the multi-

head, full-rank configuration is further discussed. For performance comparison, benchmark models including a standard ESN, multi-head ESN, full-rank ESN, and LSTM are selected. The hyperparameters for all ESN variants align with those listed in Table 6.6. The LSTM comprises two hidden layers, each with 50 neurons. All models are trained using identical strategies and datasets, with a segment of channel A1 from $\{\boldsymbol{X^c}, \boldsymbol{Y^c}\}$ selected as the test set. Comparative results are visualized in Fig. 6.16 (for clarity, only 500 steps within the segment are shown).

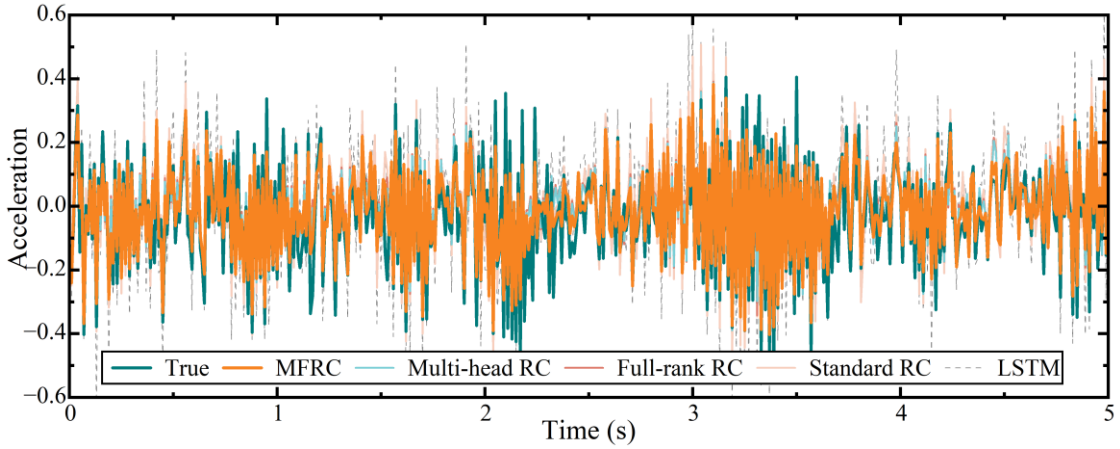


Fig. 6.16 Model comparison on structural response solving

The results demonstrate that all models successfully learned the mapping between excitation and response, particularly in capturing frequency characteristics. However, ESN-based models generally achieve higher accuracy and faster inference speeds compared to traditional models with trainable recurrent layers (e.g., the LSTM, which exhibits an MSE of 0.012). Among the two proposed mechanisms, the multi-head structure contributes more significantly by expanding reservoir dimensions and providing diverse encoding patterns for input sequences. ESN models with multi-head structures yield MSE values below 0.0065. While the full-rank criterion exhibits a smaller effect, it enhances the utilization of reservoir neurons, reducing the MSE from 0.0061 (multi-head ESN) to 0.0053 (MFRC). Overall, the proposed MFRC model improves accuracy by approximately 20% compared to conventional models in predicting structural responses under seismic excitation.

In the MFRC network, the original hyperparameter N_{res} is expanded into two parameters: the number of heads (N_{head}) and the number of neurons per head (N_{res}, assuming uniform head sizes). Additionally, the sparsity is redefined as an additional sparsity parameter (denoted as $spar_f$, calculated as the number of non-zero elements in $\boldsymbol{W^{res}}$ minus N_{res}), constrained by the full-rank criterion. The impact of these hyperparameters on prediction accuracy is investigated via grid search, using the same dataset as in the previous model comparison tasks. The search ranges, intervals, and results are summarized in Fig. 6.17.

When fixing the number of neurons per head to 4,000 (as per Table 5.6), the estimation MSE for acceleration decreases continuously with increasing N_{head}. However, as shown in Fig. 6.17(a), the MSE approaches a lower bound, indicating diminishing returns from further increasing the number of parallel heads. Thus, N_{head} is set to 6 for this study, resulting in a total reservoir size of 24,000 neurons. To evaluate the trade-offs under a fixed total reservoir size (24,000 neurons), the MSE is compared when allocating neurons to different numbers of heads. Notably, increasing the number of heads under this constraint degrades accuracy (Fig. 6.17(b)), as the multi-head structure restricts the reservoir to selecting diagonal blocks of varying sizes within an $N_h N_{res} \times N_h N_{res}$ matrix, which introduces prediction errors. Nevertheless, distributing neurons across multiple heads significantly reduces computational costs during encoding, scaling from ${N_h}^2$ to N_h. However, excessive heads reduce the dimensionality per head below the threshold required to maintain the "echo state" property in ESNs (Jaeger & Haas 2004).

As shown in Fig. 6.17(c), when the full-rank criterion is unmet, the MSE rises sharply with decreasing sparsity. Conversely, when $spar_f$ exceeds 3,000, the MSE gradually rebounds to 0.01. This occurs because excessive sparsity complicates the state signal patterns in the reservoir matrix $\boldsymbol{R}$, leading to confusion in the output layer training. Based on these observations, $spar_f$ is optimized to 2,000, corresponding to an average of 1.5 non-zero elements per row in $\boldsymbol{W^{res}}$.

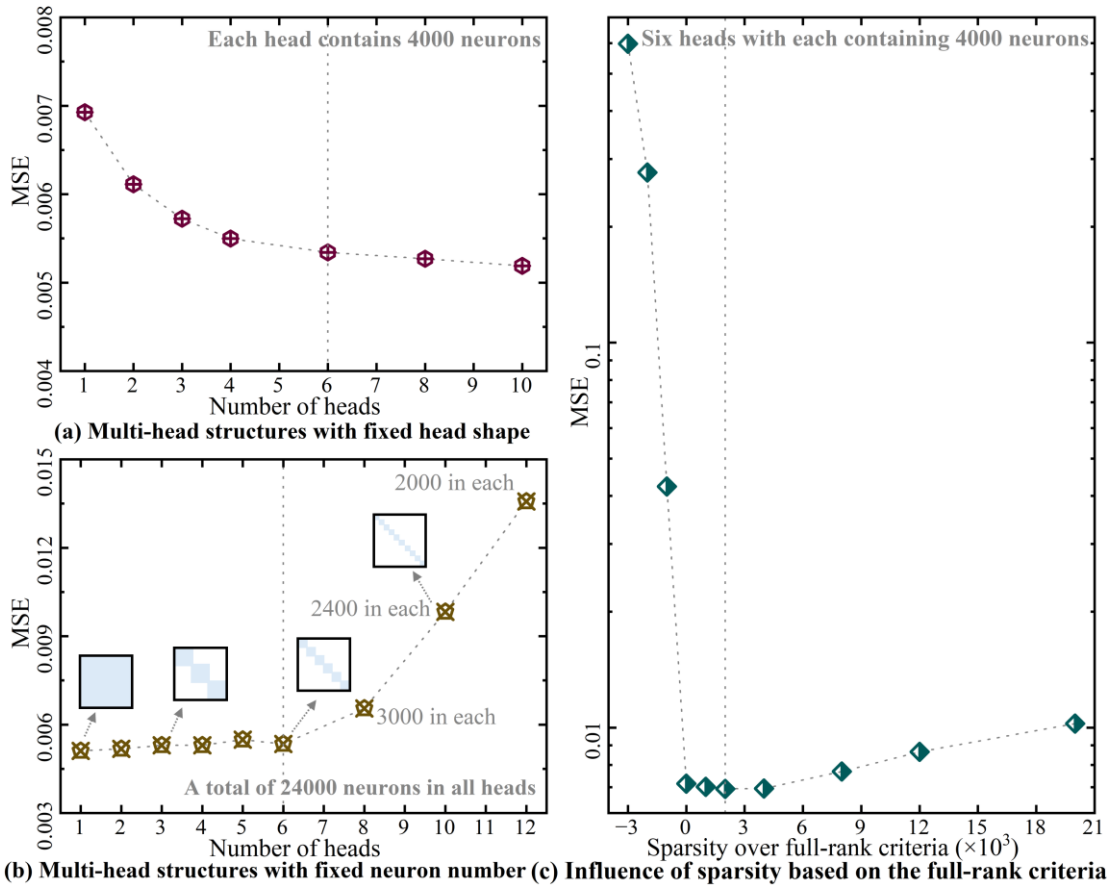


Fig. 6.17 MFRC hyperparameter searching results

6.3.4.2 DT guided model updating

During the service life of structures, they are often subjected to non-white-noise external excitations. For building structures, most of these excitations are low-intensity hazards such as tropical storms or minor earthquakes, as opposed to extreme events. According to design codes, high-rise buildings are designed to withstand such low-intensity hazards with sufficient strength. A hypothesis follows: if a structure remains undamaged after enduring such events, its response patterns to white-noise excitation should remain unchanged, allowing the pre-event DT to maintain consistent accuracy in vibration estimation. To validate this, a minor seismic excitation (0.008 g) was applied to the structure, as illustrated in Fig. 6.18. Visual inspection of the high-rise building model revealed no apparent damage (e.g., spalling or cracks). Post-earthquake, the excitation was restored to white noise, and the calibrated MFRC was used for structural state assessment. The post-event white-noise excitation ($\boldsymbol{X^{mi}}$) was input to the network, and

the predicted accelerations at nine locations were compared with measured post-event accelerations (Fig. 6.18, bottom: a 1000-step segment of $\boldsymbol{Y^{mi}}$ and $\boldsymbol{\hat{Y}^{mi}}$ for A1). For reference, pre-event white-noise-induced acceleration signals are shown in Fig. 6.18 (top: a 1000-step segment of $\boldsymbol{Y^{mi}}$ and $\boldsymbol{\hat{Y}^{mi}}$ for A1). FFT results of both response sets indicate that the structure's first natural frequency remained stable at approximately 4.5 Hz, suggesting negligible changes in dynamic properties caused by the minor earthquake. Furthermore, the MSE for A1 accelerations fluctuated slightly from 0.0052 (pre-event) to 0.0057 (post-event), while A2's MSE shifted from 0.0066 to 0.0064. All other output channels exhibited MSE variations within 10% of their original values.

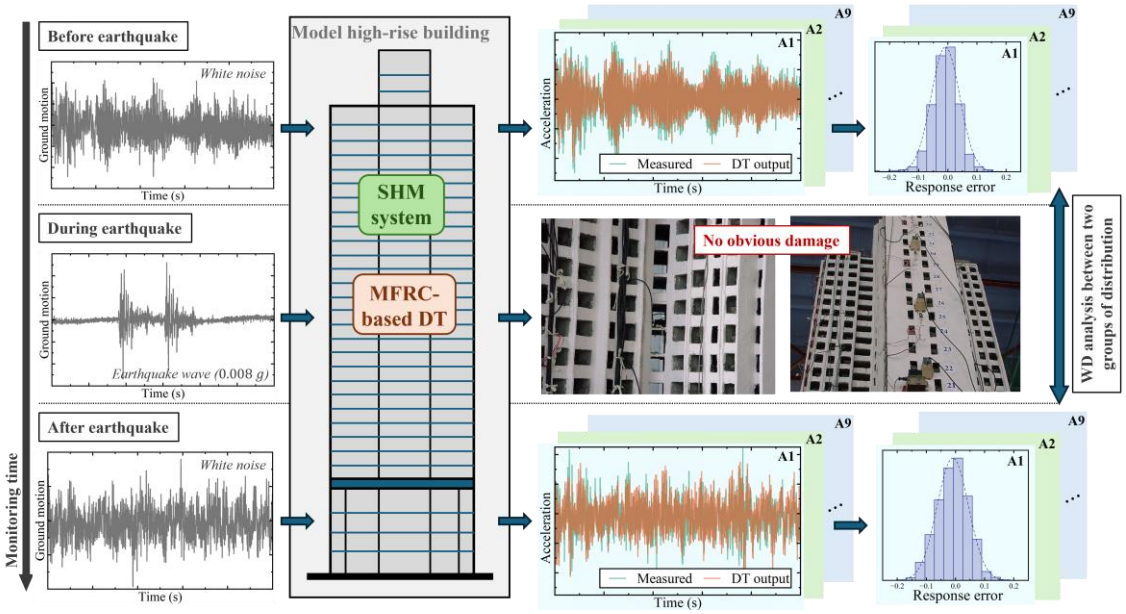


Fig. 6.18 Structural state assessment from DT after minor earthquake

On the right side of Fig. 6.18, histograms of prediction errors under pre- and post-event white-noise excitations are presented. Using the pre-event error distribution as a baseline, the WD between pre- and post-event error distributions was calculated as a damage indicator (Table 6.8), following the method in Section 6.3.2. The low WD values, comparable to those between pre-event error distributions from 54 segments, confirm that measurement noise, rather than model bias, dominates the WD variations. This indicator, sensitive to error distributions rather than acceleration magnitudes, demonstrates robust generalization. Except for A9 (with distinct vibration amplitudes), the WDs for the

remaining eight channels remained similar regardless of acceleration scales, indicating no abrupt changes in mechanical properties across any floor. The stable and low WDs between pre- and post-event MFRC prediction errors validate its effectiveness as a damage indicator for SHM.

Table 6.8 WD between MFRC error distributions before and after minor earthquake

Sensor No.	$d_W(\boldsymbol{D}_{o,i}, \boldsymbol{D}_{n,i})$	Sensor No.	$d_W(\boldsymbol{D}_{o,i}, \boldsymbol{D}_{n,i})$
A1	0.007898	A6	0.004258
A2	0.002982	A7	0.003811
A3	0.004159	A8	0.007337
A4	0.001879	A9	0.024981
A5	0.004518	/	/

Additionally, during operational phases, structures inevitably experience degradation or minor damage due to harsh environmental conditions and sudden impacts. Degradation typically reduces structural stiffness, making the system more flexible and altering its excitation-response relationship. However, most early-stage degradation or damage poses minimal risk and is challenging to localize. Thus, updating the DT to reflect the degraded state is often sufficient, eliminating the need for physical repairs. For example, after applying a ground motion with a peak acceleration of 0.02 g (simulating a moderate earthquake) to the model structure (Fig. 6.19), minor cracks emerged primarily at beam-column joints. These cracks were short in length and width, with limited quantity. Such slight damage altered the structural state, manifested as divergent responses under identical excitation distributions. As shown in the first row of Fig. 6.19, the MFRC-based DT estimated the MSE of the last 1,000-step segment (A1 channel of $\boldsymbol{Y^{mi}}$) as 0.0049. However, when predicting the first 1000-step segment of $\boldsymbol{Y^{mo}}$, the MSE increased to 0.0105, a growth exceeding 100% (Fig. 6.19, third row). This discrepancy highlights the

structural state shift caused by the damage.

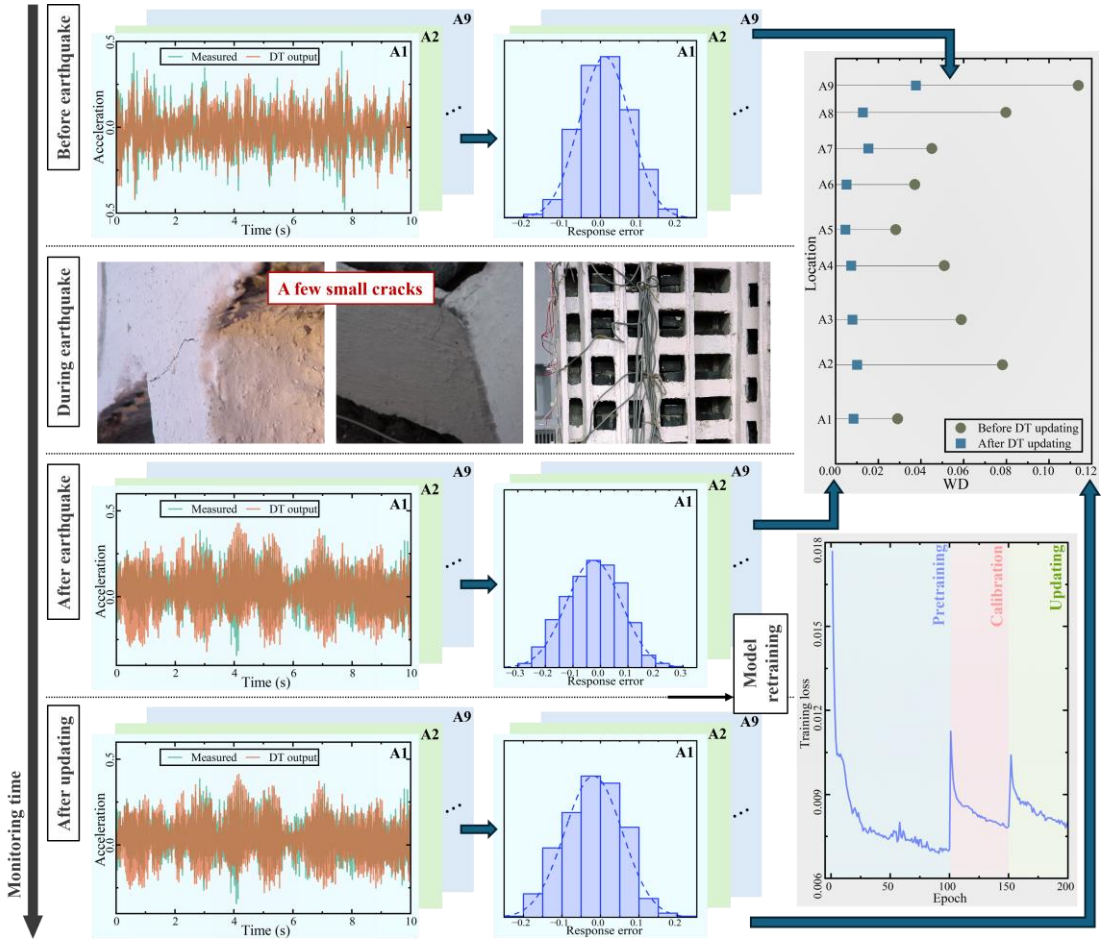


Fig. 6.19 Structural state assessment and model updating from DT after moderate earthquake

Similar or even higher growth rates were observed in the other eight channels. Specifically, the standard deviation of the error distribution in the DT's estimated responses increased from 0.066 to 0.109, resulting in an expanded distance between pre- and post-earthquake error distributions. Consequently, the WDs shown in the upper-right panel of Fig. 6.19 are significantly higher than those listed in Table 6.8. A WD threshold can be empirically defined based on structural criticality and inspection costs. When the WD exceeds this threshold, the structure is deemed to have degraded sufficiently from its initial state to warrant a new round of on-site inspection. In practice, most structural degradation may not manifest visible symptoms, and the structure can continue operating safely despite minor damage. Typically, the WD between error distributions from pre- and post-event responses estimated by an outdated DT is at least three times higher (often

more) than that of an updated model. Therefore, a threshold can be set as twice the average WD between pre-event error distributions across all segments in $\{\boldsymbol{X^{mi}}, \boldsymbol{Y^{mi}}\}$ to guide model updates. In summary, during routine monitoring, if the WD between the error distribution of MFRC predictions for a white-noise excitation segment and historical distributions exceeds this threshold, it indicates new structural degradation, necessitating model retraining or structural reinforcement. The location of degradation likely corresponds to the channel exhibiting the most significant WD shift in the joint distribution.

For structures with degradation not requiring repairs or newly reinforced systems, the DT must be updated using data from the "updated" structure to ensure reliable dynamic response predictions. Here, a 5,000-step segment from the post-earthquake measurement dataset ($\{\boldsymbol{X^{mo}}, \boldsymbol{Y^{mo}}\}$) was used to fine-tune the final layer parameters of the MFRC. The training loss versus epoch curve is shown in the lower-right panel of Fig. 6.19. After 50 additional training epochs, the loss converged to a level comparable to that achieved during the initial calibration. Following the update, the MSE for A1 acceleration estimation decreased to 0.0059, with the error distribution's standard deviation dropping to 0.074. The WDs in Fig. 6.19 also returned to magnitudes consistent with those in Table 5.8, indicating that no new damage had occurred if subsequent monitoring showed no significant MSE increase relative to this baseline. Notably, even after model updates, the MFRC's prediction errors gradually increased slightly as the structure underwent minor and moderate earthquakes. This trend may stem from degraded structural vibration modes becoming less "pure" compared to the healthy state or potential loosening or damage in the SHM system caused by seismic events.

6.3.4.3 DT guided risk alarm and structural repairing

As a critical function of the DT model, damage detection and early warning should be performed after severe disasters, such as extreme typhoons or high-intensity earthquakes. To investigate this, two seismic waves with peak accelerations of 0.15 g and 0.25 g were applied to the high-rise building model to induce damage of varying severities. Fig. 6.20(a)

illustrates the observed damage. After the major earthquake (0.15 g), multiple cracks formed, including partial column fractures and cross-shaped cracks in the core tube shear walls. Following the super major earthquake (0.25 g), horizontal and vertical cracks interconnected and penetrated structural components, causing more severe damage. Notably, concrete spalling and exposed reinforcement occurred near transfer slabs and at connections between the foundation and superstructure.

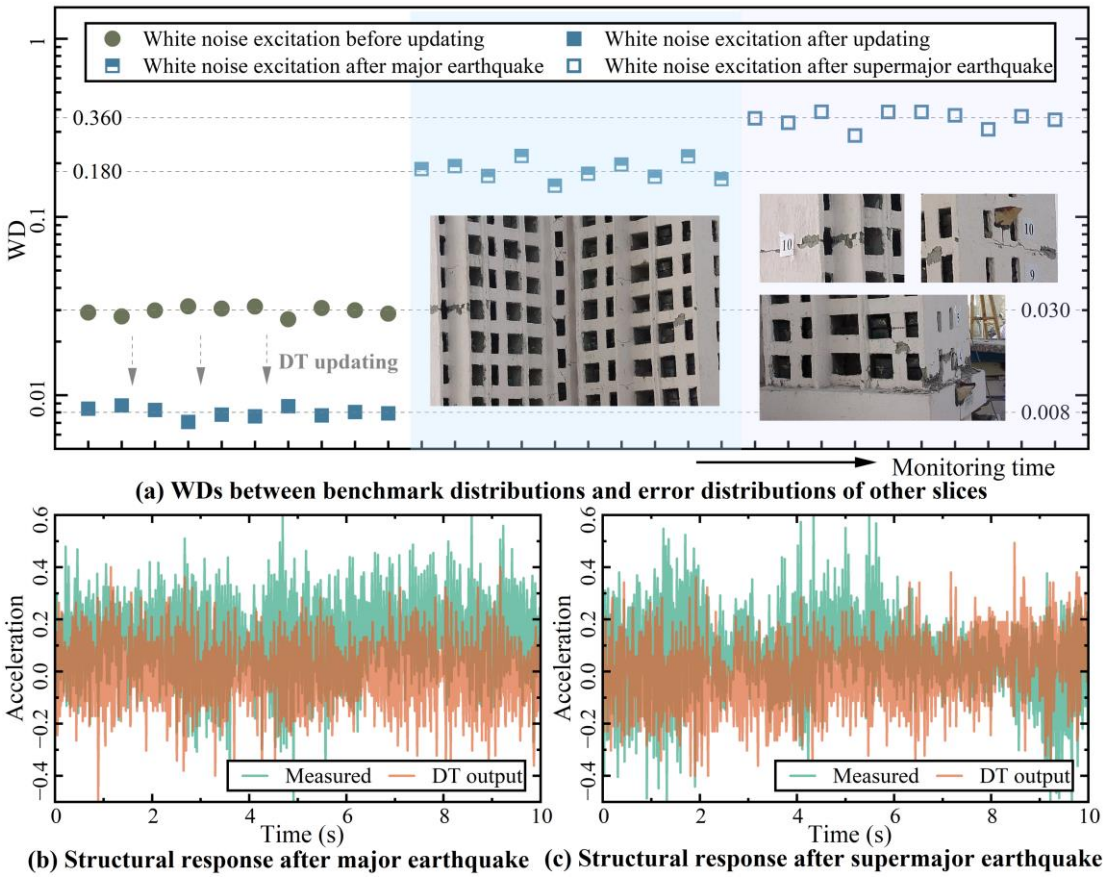


Fig. 6.20 Structural state assessment from DT after major and super major earthquake

Similar to degradation, structural damage alters the excitation-response relationship. As shown in Figs. 6.20(b) and 6.20(c), the MSEs of accelerations at nine locations estimated by the MFRC increased significantly after both earthquakes. The MFRC used here had been updated after a moderate earthquake, reflecting that most buildings subjected to severe disasters have already been in service for extended periods. For example, the MSE of A1 accelerations determined by the previously updated MFRC was approximately 0.0060 before severe damage, rising to 0.0318 and 0.0519 after the two earthquakes,

respectively. The error distribution between the first segment of $\boldsymbol{Y^{mo}}$ and the MFRC-predicted $\widehat{\boldsymbol{Y}}^{\boldsymbol{mo}}$ was used as the baseline. Fig. 6.20(a) plots the WDs between this baseline distribution and error distributions from ten random segments in $\{\boldsymbol{X^{mo}}, \boldsymbol{Y^{mo}}\}$, $\{\boldsymbol{X^{ma}}, \boldsymbol{Y^{ma}}\}$, and $\{\boldsymbol{X^{sm}}, \boldsymbol{Y^{sm}}\}$ for A1.

Table 6.9 Comparison of WD after major and super major earthquake

Sensor No.	$d_W(\boldsymbol{D}_{o,i}, \boldsymbol{D}_{n,i})$			
	Moderate earthquake	**Model updating**	**Major earthquake**	**Super major earthquake**
A2	0.078202	0.010052	0.247258	0.471771
A3	0.058911	0.007922	0.233123	0.531928
A4	0.050918	0.007292	0.272332	0.515947
A5	0.028235	0.004605	0.145571	0.291244
A6	0.037127	0.005116	0.213861	0.323937
A7	0.045124	0.015327	0.255515	0.498773
A8	0.079809	0.012787	0.274565	0.572103
A9	0.113687	0.037626	0.453348	0.850743

Structural damage markedly amplified the WDs between the baseline and post-event error distributions. Prior to the major earthquake, the WD between baseline and error distributions from white-noise excitations fluctuated around 0.008 for A1, indicating stable MFRC performance. However, this metric surged to 0.180 and 0.360 after the two earthquakes, corresponding to different damage levels. Similar trends were observed for the other eight channels (Table 6.9), with WD increases proportional to damage severity. This confirms that WD can serve as a reliable structural state indicator.

The superiority of this indicator over traditional metrics lies in three aspects: Firstly, WD is entirely data-driven, eliminating the need for comparisons with baseline numerical models (e.g., FEMs), which are time-consuming and error-prone. Secondly, WD is independent of modal parameters, global dynamic properties often insensitive to early-stage damage. Thirdly, the substantial gap between WDs of damaged versus healthy (or degraded) structures enables straightforward threshold-based damage alarms. In contrast, traditional physics-based metrics (e.g., stiffness reduction, frequency shifts) lack unlimited scalability under severe damage, reducing their discriminative power and potentially leading to misjudgments of structural risk.

6.4 Conclusions

This chapter conducts final structural damage detection and condition assessment using SHM system data acquired and processed in prior chapters. The results will guide operational decisions and, in extreme cases, serve directly as risk warnings, fulfilling the core technical objectives of SHM system deployment. The methodology operates at two levels: first, a supervised MDSC-embedded ESN classifier is proposed for time-domain signals to achieve precise damage localization and severity quantification; second, an updatable, interactive DT framework based on MFRC and WD is developed for rapid condition evaluation using long-term operational data. Key findings are summarized below:

(1) The primary innovation lies in integrating MDSC layers to extract spatiotemporal correlations and critical signal features, coupled with Eros-PCA-based data selection to autonomously identify valuable segments from massive SHM datasets. Validation using a three-span model bridge with simulated damage demonstrated that the MDSC-ESN model achieved approximately 1% error in severity estimation across 17 damage patterns at various locations. Robustness analysis revealed that diagnostic accuracy heavily depends on the diversity of damage modes in the training set, while severity intervals can be sparse, the training data must comprehensively cover all potential damage

locations. The model maintained reliable damage identification under noise levels below 10%, but performance degraded significantly with further reductions in SNR.

(2) The proposed DT framework targets sustainable operation of urban high-rise buildings. It can be established at any lifecycle stage and activated using limited data (from FEM or scaled experiments) for real-time condition monitoring and rehabilitation guidance. The framework comprises an MFRC-based updatable virtual entity for structural vibration prediction and a WD-based unsupervised damage detection module linking physical and virtual entities. Validation via shaking table tests confirmed that MFRC predictions matched measured data with over 20% higher accuracy than conventional methods in both time and frequency domains. The MFRC demonstrated stable generalization for free vibration responses when structural properties remained unchanged. Transfer learning-based retraining enabled model calibration and updates post-degradation, keeping MSE increases within 10% after updates. WD proved highly sensitive, showing at least 100% increments between error distributions from adjacent damage levels. Hyperparameters (N_{head}, $spar_f$) were optimized via grid search to balance computational efficiency and accuracy.

CONCLUSIONS AND FUTURE WORKS

7.1 Conclusions

The SHM and its data processing serve as effective means and critical components of intelligent structural operation and maintenance. When implemented in conjunction with related technologies (e.g., anomaly detection, structural rehabilitation, and DTs), they can significantly enhance the sustainability and risk resilience of building structures, infrastructure, and even urban clusters, while reducing post-aging and disaster losses. Achieving these objectives requires comprehensive understanding and mining of SHM-acquired data. This study employs novel NNs for time-series data processing to deliver higher-performance solutions for SHM system tasks across its full lifecycle. Key conclusions are summarized as follows:

(1) Sensor deployment phase: Sensor placement primarily relies on structural design models and relevant information. First, the optimization criteria for sensor placement were refined by proposing the FIA optimization objective, which demonstrates higher sensitivity to local information compared to conventional indicators like MAC. The HEOA search algorithm was modified to address combinatorial optimization problems. Considering the critical dependence of search efficiency and solution quality on initial candidate point sets, a PageRank-based measurement point ranking algorithm utilizing structural characteristic matrices was developed for preliminary screening of effective points. This novel sensor placement theory was validated through FEM analysis of Shanghai Rafael Cloud Corridor, proving that the FIA criterion not only ensures accurate time-domain signal inference across locations but also achieves high modal independence. The PageRank ranking algorithm essentially replaces manual screening by eliminating most invalid measurement points from massive model degrees of freedom.

(2) Data preprocessing phase: This phase focuses on filtering raw data from deployed sensor systems. For conventional noise in SHM data, an IKF method based on ARIMAX state transition matrices was adopted. Unlike traditional low-pass filters, its key parameters are determined by sensor specifications and transition model performance, eliminating unreliable manual assumptions while better preserving effective frequency-domain features. For missing data, a Bi-ESN-R network was designed for high-precision reconstruction, outperforming conventional RNN methods in long-term missing scenarios and handling extreme cases (multi-channel or endpoint missing). For outliers, an FD-ESN generator combined with LOF algorithm was developed, achieving precise outlier detection (anomaly segment less than 25%) through frequency-specific fitting and anomaly index amplification, though limitations exist in trend accumulation and zero-oscillation regions. Experimental validation on Canton Tower and Shanghai Tower monitoring data demonstrated strong adaptability of these preprocessing methods.

(3) Data inference phase: This phase infers unmeasured information through spatiotemporal correlations using preprocessed SHM data. For prediction tasks, a memory-enhanced dESN-eLSTM predictor was proposed and applied to wind field prediction for Lanzhou-Xinjiang HSR WSs, enabling high-precision reconstruction of in-plane or out-of-plane wake fields behind barriers while maintaining functionality under extreme layouts (unilateral, sparse, and long-distance sensors). For synchronized mapping (estimating unmeasured but valuable variables), the Echoformer network with random reservoir replacing positional encoding was developed. Applied to Shenzhen Metro noise inference, it outperformed conventional models in accuracy and robustness for near and far-field noise estimation. For physical system solving (governed by control equations with variable-dependent mappings), the GPES method was proposed by enhancing Seq2Seq generalization. It successfully solved PD system

counterforces through sub-model fitting and generated the continuous ExESN model, achieving accuracy of over 80% in estimating input-output mappings under unfamiliar excitation frequencies, providing reliable references for damper performance calibration.

(4) Evaluation and decision-making phase: This phase conducts structural damage detection and condition assessment based on SHM measurements and inferred data to inform maintenance decisions and risk warnings. For damage information in time-domain signals, an ESN-based supervised classifier embedded with MDSC was developed. By extracting multi-dimensional key features, it achieved precise damage localization and severity quantification. Validated on a three-span model bridge dataset, the MDSC-ESN model exhibited approximately 1% error across 17 damage patterns and maintained reliability under 10% noise. For long-term structural state evolution, an updatable interactive DT architecture (WD guided MFRC framework) was proposed for sustainable operation. Activated by minimal data, it enables real-time state evaluation guiding repairs. The MFRC module demonstrated accuracy improvement of over 20% and faster computation in shake-table tests with strong generalization. The WD-based unsupervised anomaly detection module effectively distinguished post-disaster structural states. Model updates for gradual degradation and rehabilitation were achieved through transfer learning, ensuring long-term applicability.

This systematic framework demonstrates significant advancements in SHM system performance across deployment, preprocessing, inference, and decision-making stages, providing robust technical support for intelligent infrastructure management.

7.2 Future works

The limitations of the research findings presented in this paper and corresponding improvement strategies are outlined as follows:

(1) The sensor deployment scheme derived from the proposed theory lacks validation through measured data. Verification of the FIA criterion requires

dense measurement points, which proves challenging for full-scale structural implementations. A scaled-down dynamic experiment targeting specific structures could be considered to validate this methodology.

(2) The proposed DT framework relies on calculating structural responses through measured excitations, whereas complete measurement of actual structural excitations (particularly wind loads) remains impractical. To address this, the input-output configuration of the virtual entity could be modified to establish partial channel responses as inputs and remaining channel responses as outputs. This adjustment enables the model to learn health-state mapping relationships between response channels, with the inherent requirement that mapping fluctuations intensify as structural degradation progresses.

(3) While post-damage diagnosis and warning systems demonstrate value, pre-disaster estimation of potential structural damage holds greater engineering significance. Current NNs lack inherent capacity for progressive self-degradation modeling. A viable solution involves organizing long-term monitoring data into WD sequences correlated with historical disaster events, followed by application of the proposed prediction algorithm to forecast WD sequence trajectories. This approach could ultimately enable pre-disaster estimation of structural damage evolution.

REFERENCES

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. Gpt-4 technical report. arXiv preprint arXiv, 2023: 2303.08774.

Akbar, M. A., Qidwai, U., & Jahanshahi, M. R. An evaluation of image-based structural health monitoring using integrated unmanned aerial vehicle platform. Structural Control and Health Monitoring, 2019, 26(1): e2276.

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., & Muneer, A. LSTM inefficiency in long-term dependencies regression problems. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2023, 30(3): 16-31.

Albdour, N., & Zanoon, N. A steganographic method based on roberts operator. Jordan Journal of Electrical Engineering, 2020, 6(3): 266.

Alghushairy, O., Alsini, R., Soule, T., & Ma, X. A review of local outlier factor algorithms for outlier detection in big data streams. Big Data and Cognitive Computing, 2020, 5(1): 1.

Ali, A. A. I. M., Jamaludin, S., Imran, M. M. H., Ayob, A. F. M., Ahmad, S. Z. A. S., Akhbar, M. F. A., ... & Ramli, M. R. Computer vision and image processing approaches for corrosion detection. Journal of Marine Science and Engineering, 2023, 11(10): 1954.

Antonelo E, Camponogara E, & Foss B. Echo state networks for data-driven downhole pressure estimation in gas-lift oil wells. Neural Networks, 2017, 85: 106-117.

AS, R. A., & Gopalan, S. Comparative analysis of eight direction Sobel edge detection algorithm for brain tumor MRI images. Procedia Computer Science, 2022, 201: 487-494.

Assirati, L., Silva, N. R. D., Berton, L., Lopes, A. D. A., & Bruno, O. M. Performing edge detection by Difference of Gaussians using q-Gaussian kernels. In Journal of Physics: Conference Series, 2014, 490(1): 012020.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. A

review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical Systems and Signal Processing, 2021, 147: 107077.

Azhar, A. S., Kudus, S. A., Jamadin, A., Mustaffa, N. K., & Sugiura, K. Recent vibration-based structural health monitoring on steel bridges: Systematic literature review. Ain Shams Engineering Journal, 2024, 15(3): 102501.

Bai, S., Kolter, J. Z., & Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv, 2018: 1803.01271, 10.

Balochian, S., & Baloochian, H. Edge detection on noisy images using Prewitt operator and fractional order differentiation. Multimedia Tools and Applications, 2022, 81(7): 9759-9770.

Bararnia, H., & Esmaeilpour, M. On the application of physics informed neural networks (PINN) to solve boundary layer thermal-fluid problems. International Communications in Heat and Mass Transfer, 2022, 132: 105890.

Behnia, A., Chai, H. K., & Shiotani, T. Advanced structural health monitoring of concrete structures with the aid of acoustic emission. Construction and Building Materials, 2014, 65: 282-302.

Berkhin, P. A survey on PageRank computing. Internet Mathematics, 2005, 2(1): 73-120.

Bezas, K., Komianos, V., Koufoudakis, G., Tsoumanis, G., Kabassi, K., & Oikonomou, K. Structural health monitoring in historical buildings: a network approach. Heritage, 2020, 3(3): 796-818.

Bisheh, H. B., & Amiri, G. G. Structural damage detection based on variational mode decomposition and kernel PCA-based support vector machine. Engineering Structures, 2023, 278: 115565.

Burchi, M., & Vielzeuf, V. Efficient conformer: Progressive downsampling and grouped attention for automatic speech recognition. In 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021: 8-15.

Cao, J., Bu, F., Wang, J., Bao, C., Chen, W., & Dai, K. Reconstruction of full-field dynamic responses for large-scale structures using optimal sensor placement. Journal of Sound and Vibration, 2023, 554, 117693.

Chan, T. H., Yu, L., Tam, H. Y., Ni, Y. Q., Liu, S. Y., Chung, W. H., & Cheng, L. K. Fiber Bragg grating sensors for structural health monitoring of Tsing Ma bridge: Background and experimental observation. Engineering structures, 2006, 28(5): 648-659.

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. Recurrent neural networks for multivariate time series with missing values. Scientific reports, 2018, 8(1): 6085.

Chen, T., Mao, K., Huang, X., & Wang, M. Y. Dissipation mechanisms of nonobstructive particle damping using discrete element method. In Smart Structures and Materials 2001: Damping and Isolation, 2001, 4331: 294-301.

Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. Neural ordinary differential equations. Advances in Neural Information Processing Systems, 2018: 31.

Chen, J., Jiang, X., Yan, Y., Lang, Q., Wang, H., & Ai, Q. Dynamic warning method for structural health monitoring data based on ARIMA: Case study of Hong Kong–Zhuhai–Macao bridge immersed tunnel. Sensors, 2022, 22(16), 6185.

Chen, S. Y., Wang, Y. W., & Ni, Y. Q. Gross outlier removal and fault data recovery for SHM data of dynamic responses by an annihilating filter-based Hankel-structured robust PCA method. Structural Control and Health Monitoring, 2022, 29(12): e3144.

Chen, W., Dai, S., & Zheng, B. ARIMA-FEM method with prediction function to solve the stress–strain of perforated elastic metal plates. Metals, 2022, 12(2): 179.

Cheng, M., Zhang, Q., & Cao, Y. An Early Warning Model for Turbine Intermediate-Stage Flux Failure Based on an Improved HEOA Algorithm Optimizing DMSE-GRU Model. Energies, 2024, 17(15): 3629.

Chmielewski, D. J., Palmer, T., & Manousiouthakis, V. On the theory of optimal sensor placement. AIChE journal, 2002, 48(5): 1001-1012.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv, 2014: 1406.1078.

Choe, D. E., Kim, H. C., & Kim, M. H. Sequence-based modeling of deep learning with LSTM and GRU networks for structural damage detection of floating offshore wind turbine blades. Renewable Energy, 2021, 174: 218-235.

Chong, J. W., Kim, Y., & Chon, K. H. Nonlinear multiclass support vector machine–based health monitoring system for buildings employing magnetorheological dampers. Journal of Intelligent Material Systems and Structures, 2014, 25(12): 1456-1468.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv, 2014: 1412.3555.

Civera, M., Pecorelli, M. L., Ceravolo, R., Surace, C., & Zanotti Fragonara, L. A multi-objective genetic algorithm strategy for robust optimal sensor placement. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(9): 1185-1202.

Cobb, R. G., & Liebst, B. S. Sensor placement and structural damage identification from minimal sensor information. AIAA journal, 1997, 35(2): 369-374.

Dang, D. Z., Lai, C. C., Ni, Y. Q., Zhao, Q., Su, B., & Zhou, Q. F. Image classification-based defect detection of railway tracks using fiber Bragg grating ultrasonic sensors. Applied Sciences, 2022, 13(1): 384.

Deif, D. S., & Gadallah, Y. An ant colony optimization approach for the deployment of reliable wireless sensor networks. IEEE Access, 2017, 5: 10744-10756.

Deng, Z., Huang, M., Wan, N., & Zhang, J. The current development of structural health monitoring for bridges: A review. Buildings, 2023, 13(6): 1360.

Deng, E., Yue, H., Ni, Y. Q., He, X. H., Yang, W. C., & Chen, Z. W. Wake dynamic characteristics of windproof structures in embankment–bridge sections along a high-speed railway under natural strong crosswinds. Physics of Fluids, 2023, 35(5).

Deng, Y., Zhao, Y., Ju, H., Yi, T. H., & Li, A. Abnormal data detection for structural health monitoring: State-of-the-art review. Developments in the Built Environment, 2024,

17: 100337.

Dezhkam, A., Manzuri, M. T., Aghapour, A., Karimi, A., Rabiee, A., & Shalmani, S. M. A Bayesian-based classification framework for financial time series trend prediction The Journal of Supercomputing, 2023, 79(4): 4622-4659.

Dhillon, S. S., & Chakrabarty, K. Sensor placement for effective coverage and surveillance in distributed sensor networks. In 2003 IEEE Wireless Communications and Networking, 2003, 3: 1609-1614.

Diamanti, K., & Soutis, C. Structural health monitoring techniques for aircraft composite structures. Progress in Aerospace Sciences, 2010, 46(8): 342-352.

Ding, L. Y., Zhou, C., Deng, Q. X., Luo, H. B., Ye, X. W., Ni, Y. Q., & Guo, P. Real-time safety early warning system for cross passage construction in Yangtze Riverbed Metro Tunnel based on the internet of things. Automation in construction, 2013, 36: 25-37.

Dong, Y., Pan, Y., Wang, D., & Cheng, T. Corrosion detection and evaluation for steel wires based on a multi-vision scanning system. Construction and Building Materials, 2022, 322: 125877.

Du, L., Gao, R., Suganthan, P. N., & Wang, D. Z. Bayesian optimization based dynamic ensemble for time series forecasting. Information Sciences, 2022, 591: 155-175.

Du, W., Cote, D., & Liu, Y. Saits: Self-attention-based imputation for time series. Expert Systems with Applications, 2023, 219: 119619.

Dung, C. V. Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction, 2019, 99: 52-58.

Ercan, T., & Papadimitriou, C. Bayesian optimal sensor placement for parameter estimation under modeling and input uncertainties. Journal of Sound and Vibration, 2023, 563: 117844.

Farrar, C. R., & Worden, K. An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2007, 365(1851): 303-315.

Farrar, C. R., Dervilis, N., & Worden, K. The Past, Present and Future of Structural Health Monitoring: An Overview of Three Ages. Strain, 2025, 61(1): e12495.

Franchi, F., Marotta, A., Rinaldi, C., Graziosi, F., Fratocchi, L., & Parisse, M. What can 5g do for public safety? structural health monitoring and earthquake early warning scenarios. Sensors, 2022, 22(8): 3020.

Garriga, C., Hedlund, A., Tang, Y., & Wang, P. Rural-urban migration, structural transformation, and housing markets in China. American Economic Journal: Macroeconomics, 2023, 15(2), 413-440.

Gao, X., Ji, X., Zhang, Y., Zhuang, Y., & Cai, E. Structural displacement estimation by a hybrid computer vision approach. Mechanical Systems and Signal Processing, 2023, 204: 110754.

Giurgiutiu, V., & Yu, L. Comparison of short-time Fourier transform and wavelet transform of transient and tone burst wave propagation signals for structural health monitoring. In 4th International Workshop on Structural Health Monitoring, 2003: 1267-1274.

Gomes, G. F., & Pereira, J. V. P. Sensor placement optimization and damage identification in a fuselage structure using inverse modal problem and firefly algorithm. Evolutionary Intelligence, 2020, 13(4): 571-591.

Gong, Z., Tang, Y., & Liang, J. Patchmixer: A patch-mixing architecture for long-term time series forecasting. arXiv preprint arXiv, 2023: 2310.00655.

Graves, A., Jaitly, N., & Mohamed, A. R. Hybrid speech recognition with deep bidirectional LSTM. In 2013 IEEE workshop on automatic speech recognition and understanding, 2013: 273-278.

Gou, J., Yu, B., Maybank, S. J., & Tao, D. Knowledge distillation: A survey. International Journal of Computer Vision, 2021, 129(6): 1789-1819.

Gui, G., Pan, H., Lin, Z., Li, Y., & Yuan, Z. Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. KSCE Journal of Civil Engineering, 2017, 21: 523-534.

Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K., & Funaya, K. Robust online time series prediction with recurrent neural networks. In 2016 IEEE international conference on data science and advanced analytics (DSAA), 2016: 816-825.

Guo, D., Zhu, Q., Yang, D., Xie, Z., Dong, K., Zhang, W., ... & Liang, W. DeepSeek-Coder: When the Large Language Model Meets Programming--The Rise of Code Intelligence. arXiv preprint arXiv, 2024: 2401.14196.

Guratzsch, R. F., & Mahadevan, S. Structural health monitoring sensor placement optimization under uncertainty. AIAA journal, 2010, 48(7): 1281-1289.

Hajializadeh, D., OBrien, E. J., & O'Connor, A. J. Virtual structural health monitoring and remaining life prediction of steel bridges. Canadian Journal of Civil Engineering, 2017, 44(4): 264-273.

Hamilton, J. D. Time series analysis. Princeton university press. 2020.

Hao, H., Bi, K., Chen, W., Pham, T. M., & Li, J. Towards next generation design of sustainable, durable, multi-hazard resistant, resilient, and smart civil engineering structures. Engineering Structures, 2023, 277: 115477.

Hasani, R., Lechner, M., Amini, A., Rus, D., & Grosu, R. Liquid time-constant networks. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(9): 7657-7666.

Hassani, S., Dackermann, U., Mousavi, M., & Li, J. A systematic review of data fusion techniques for optimized structural health monitoring. Information Fusion, 2024, 103: 102136.

He, C., Xing, J., Li, J., Yang, Q., Wang, R., & Zhang, X. A new optimal sensor placement strategy based on modified modal assurance criterion and improved adaptive genetic algorithm for structural health monitoring. Mathematical Problems in Engineering, 2015, 2015(1): 626342.

He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-779.

He, Y., Zhang, L., Chen, Z., & Li, C. Y. A framework of structural damage detection for civil structures using a combined multi-scale convolutional neural network and echo state network. Engineering with Computers, 2023, 39(3): 1771-1789.

Hester, D., Brownjohn, J., Bocian, M., Xu, Y., & Quattrone, A. Using inertial measurement units originally developed for biomechanics for modal testing of civil engineering structures. Mechanical Systems and Signal Processing, 2018, 104: 776-798.

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

Hospedales, T., Antoniou, A., Micaelli, P., & Storkey, A. Meta-learning in neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(9): 5149-5169.

Hsieh, W. W. Nonlinear multivariate and time series analysis by neural network methods. Reviews of Geophysics, 2004, 42(1).

Hu, J., Hu, X., Kong, F., & Wu, H. Vulnerability analysis of super high-rise building security system based on Bayesian network and digital twin technology. Process Safety and Environmental Protection, 2024, 187: 1047-1061.

Huang, C., Xu, H., Xu, Y., Dai, P., Xia, L., Lu, M., ... & Ye, Y. Knowledge-aware coupled graph neural network for social recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence 2021, 35(5): 4115-4122.

Ibrahim, A. S., Han, Z., & Liu, K. R. Distributed energy-efficient cooperative routing in wireless networks. IEEE Transactions on Wireless Communications, 2008, 7(10): 3930-3941.

Jaeger H, and Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 2004; 304(5667): 78-80.

Jafari, B., Khaloo, A., & Lattanzi, D. Long-term monitoring of structures through point cloud analysis. In Health Monitoring of Structural and Biological Systems 2016: 9805: 637-644.

Jaseena, K. U., & Kovoor, B. C. Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. Energy Conversion and Management, 2021, 234: 113944.

Javadinasab Hormozabad, S., Gutierrez Soto, M., & Adeli, H. Integrating structural control, health monitoring, and energy harvesting for smart cities. Expert Systems, 2021, 38(8): e12845.

Jia, J., Feng, S., & Liu, W. A triaxial accelerometer monkey algorithm for optimal sensor placement in structural health monitoring. Measurement Science and Technology, 2015, 26(6): 065104.

Jiang, H. J., Lu, X. L., Liu, X. J., & He, L. S. Performance-based seismic design principles and structural analysis of Shanghai Tower. Advances in Structural Engineering, 2014, 17(4): 513-527.

Jiang, J., & Lai, Y. C. Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. Physical review research, 2019, 1(3): 033056.

Jiao, P., Ye, X., Zhang, C., Li, W., & Wang, H. Vision-based real-time marine and offshore structural health monitoring system using underwater robots. Computer-Aided Civil and Infrastructure Engineering, 2024, 39(2): 281-299.

Jin, Z. Q., Zhao, X., Zhao, T. J., & Li, J. Q. Chloride ions transportation behavior and binding capacity of concrete exposed to different marine corrosion zones. Construction and Building Materials, 2018, 177: 170-183.

Kaartinen, E., Dunphy, K., & Sadhu, A. LiDAR-based structural health monitoring: Applications in civil infrastructure systems. Sensors, 2022, 22(12): 4610.

Kammer, D. C., & Yao, L. Enhancement of on-orbit modal identification of large space structures through sensor placement. Journal of Sound and Vibration, 1994, 171(1): 119-139.

Kankanamge, Y., Hu, Y., & Shao, X. Application of wavelet transform in structural health monitoring. Earthquake Engineering and Engineering Vibration, 2020, 19: 515-532.

Khajavi, S. H., Motlagh, N. H., Jaribion, A., Werner, L. C., & Holmström, J. Digital twin: vision, benefits, boundaries, and creation for buildings. IEEE Access, 2019, 7: 147406-147419.

Khodarahmi, M., & Maihami, V. A review on Kalman filter models. Archives of Computational Methods in Engineering, 2023, 30(1): 727-747.

Khodayar, M., & Wang, J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Transactions on Sustainable Energy, 2018, 10(2): 670-681.

Kinet, D., Mégret, P., Goossen, K. W., Qiu, L., Heider, D., & Caucheteur, C. Fiber Bragg grating sensors toward structural health monitoring in composite materials: Challenges and solutions. Sensors, 2014, 14(4): 7394-7419.

Larger, L., Baylón-Fuentes, A., Martinenghi, R., Udaltsov, V. S., Chembo, Y. K., & Jacquot, M. High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification. Physical Review X, 2017, 7(1): 011015.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. Temporal convolutional networks for action segmentation and detection. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 156-165.

Lei, X., Sun, L., & Xia, Y. Lost data reconstruction for structural health monitoring using deep convolutional generative adversarial networks. Structural Health Monitoring, 2021, 20(4): 2069-2087.

Li, C. S., Lam, S. S., Zhang, M. Z., & Wong, Y. L. Shaking table test of a 1: 20 scale high-rise building with a transfer plate system. Journal of Structural Engineering, 2006, 132(11): 1732-1744.

Li, N., Zhang, X. Y., Zhou, X. T., Leng, J., Liang, Z., Zheng, C., & Sun, X. F. Introduction of structural health and safety monitoring warning systems for Shenzhen-Hong Kong Western Corridor Shenzhen Bay Bridge. In Health Monitoring of Structural and Biological Systems, 2008, 6935: 490-499.

Li, G. Y., Ma, W., Mu, Y. H., Wang, F., Fan, S. Z., & Wu, Y. H. Effects of freeze-thaw cycle on engineering properties of loess used as road fills in seasonally frozen ground regions, North China. Journal of Mountain Science, 2017, 14: 356-368.

Li, D., Chen, D., Goh, J., & Ng, S. K. Anomaly detection with generative adversarial networks for multivariate time series. arXiv preprint arXiv, 2018:1809.04758.

Li Y. T., Bao T. F., Gong J., & Zhang K. The prediction of dam displacement time series using STL, extra-trees, and stacked LSTM neural network. IEEE Access, 2020, 8: 94440-94452.

Lian, J., & Hui, G. Human evolutionary optimization algorithm. Expert Systems with Applications, 2024, 241: 122638.

Liao, J., Fukui, H., Urakami, T., & Morisaki, H. Effect of biofilm on ennoblement and localized corrosion of stainless steel in fresh dam-water. Corrosion Science, 2010, 52(4): 1393-1403.

Lin, L., Wang, F., Xie, X., & Zhong, S. Random forests-based extreme learning machine ensemble for multi-regime time series prediction. Expert Systems with Applications, 2017, 83: 164-176.

Liu, T., Yang, B., & Zhang, Q. Health monitoring system developed for Tianjin 117 high-rise building. Journal of Aerospace Engineering, 2017, 30(2): B4016004.

Liu, J., Ouyang, H., Han, X., & Liu, G. Optimal sensor placement for uncertain inverse problem of structural parameter estimation. Mechanical Systems and Signal Processing, 2021, 160: 107914.

Lorencin, I., Andelic, N., Spanjol, J., & Car, Z. Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis. Artificial intelligence in medicine, 2020, 102: 101746.

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature machine intelligence, 2021, 3(3): 218-229.

Lun, S., Zhang, Z., Li, M., & Lu, X. Parameter optimization in a leaky integrator echo

state network with an improved gravitational search algorithm. Mathematics, 2023, 11(6): 1514.

Ma, Q., Shen, L., & Cottrell, G. W. DeePr-ESN: A deep projection-encoding echo-state network. Information Sciences, 2020, 511: 152-171.

Mahmud, M. A., Bates, K., Wood, T., Abdelgawad, A., & Yelamarthi, K. A complete internet of things (IoT) platform for structural health monitoring (SHM). In 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 2018: 275-279.

Majumder, M., Gangopadhyay, T. K., Chakraborty, A. K., Dasgupta, K., & Bhattacharya, D. K. Fibre Bragg gratings in structural health monitoring—Present status and applications. Sensors and Actuators A: Physical, 2008, 147(1): 150-164.

Mander, J. B., Panthaki, F. D., & Kasalanati, A. Low-cycle fatigue behavior of reinforcing steel. Journal of Materials in Civil Engineering, 1994, 6(4): 453-468.

Mehrjoo, A., Song, M., Moaveni, B., Papadimitriou, C., & Hines, E. Optimal sensor placement for parameter estimation and virtual sensing of strains on an offshore wind turbine considering sensor installation cost. Mechanical Systems and Signal Processing, 2022, 169: 108787.

Myeong, W. C., Jung, K. Y., Jung, S. W., Jung, Y. H., & Myung, H. Drone-type wall-climbing robot platform for structural health monitoring. In Proc. Int. Conf. Advances in Experimental Structural Engineering 2015: 286-389.

Namba, S., Kuwano, S., & Kato, T. An investigation of Leq, L10, and L50 in relation to loudness. The Journal of the Acoustical Society of America, 1978, 64(S1): S58-S58.

Nguyen, T., Chan, T. H., Thambiratnam, D. P., & King, L. Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring. Mechanical Systems and Signal Processing, 2015, 64: 313-324.

Nguyen, T. T., Dang, V. H., & Pham, T. T. Development of an automatic and knowledge-infused framework for structural health monitoring based on prompt engineering. Frontiers of Structural and Civil Engineering, 2024, 18(11): 1752-1774.

Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. Technology innovation in developing the

structural health monitoring system for Guangzhou New TV Tower. Structural Control and Health Monitoring, 2009, 16(1): 73-98.

Nong, X., Luo, X., Lin, S., Ruan, Y., & Ye, X. Multimodal deep neural network-based sensor data anomaly diagnosis method for structural health monitoring. Buildings, 2023, 13(8): 1976.

Ochieng, F. X., Hancock, C. M., Roberts, G. W., & Le Kernec, J. A review of ground-based radar as a noncontact sensor for structural health monitoring of in-field wind turbines blades. Wind Energy, 2018, 21(12): 1435-1449.

Oh, B. K., Glisic, B., Kim, Y., & Park, H. S. Convolutional neural network–based data recovery method for structural health monitoring. Structural Health Monitoring, 2020, 19(6): 1821-1838.

Ozbey, B., Unal, E., Ertugrul, H., Kurc, O., Puttlitz, C. M., Erturk, V. B., ... & Demir, H. V. Wireless displacement sensing enabled by metamaterial probes for remote structural health monitoring. Sensors, 2014, 14(1): 1691-1704.

Ozkul, T., A. A finite element formulation for dynamic analysis of shells of general shape by using the Wilson-θ method. Thin-walled Structures, 2004, 42(4): 497-513.

Panaretos, V. M., & Zemel, Y. Statistical aspects of Wasserstein distances. Annual Review of Statistics and its Application, 2019, 6(1): 405-431.

Park, H. S., Lee, H. M., Adeli, H., & Lee, I. A new approach for health monitoring of structures: terrestrial laser scanning. Computer-Aided Civil and Infrastructure Engineering, 2007, 22(1): 19-30.

Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. How to construct deep recurrent neural networks. arXiv preprint arXiv, 2013: 1312.6026.

Peng, Z., Li, J., Hao, H., & Zhong, Y. Smart structural health monitoring using computer vision and edge computing. Engineering Structures, 2024, 319: 118809.

Perera, R., & Torres, R. Structural damage detection via modal data with genetic algorithms. Journal of Structural Engineering, 2006, 132(9): 1491-1501.

Perry, M., Fusiek, G., Niewczas, P., Rubert, T., & McAlorum, J. Wireless concrete

strength monitoring of wind turbine foundations. Sensors, 2017, 17(12): 2928.

Peter, D., & Silvia, P. ARIMA vs. ARIMAX-which approach is better to analyze and forecast macroeconomic time series. in: Proc. Int. Conf. Mathematical methods in Economics, 2012: 136-140.

Pines, D., & Salvino, L. Structural health monitoring using empirical mode decomposition and the Hilbert phase. Journal of Sound and Vibration, 2006, 294(1-2): 97-124.

Pooya, S. M. H., & Massumi, A. A novel damage detection method in beam-like structures based on the relation between modal kinetic energy and modal strain energy and using only damaged structure data. Journal of Sound and Vibration, 2022, 530: 116943.

Pourzeynali, S., Zhu, X., Ghari Zadeh, A., Rashidi, M., & Samali, B. Comprehensive study of moving load identification on bridge structures using the explicit form of Newmark-β method: Numerical and experimental studies. Remote Sensing, 2021, 13(12): 2291.

Prasanna, P., Dana, K. J., Gucunski, N., Basily, B. B., La, H. M., Lim, R. S., & Parvardeh, H. Automated crack detection on concrete bridges. IEEE Transactions on automation science and engineering, 2014, 13(2): 591-599.

Qin, Z., Zhang, P., Wu, F., & Li, X. Fcanet: Frequency channel attention networks. In Proceedings of the international conference on computer vision, 2021: 783-792.

Qiu, D., Fu, K., Xue, Y., Tao, Y., Kong, F., & Bai, C. TBM tunnel surrounding rock classification method and real-time identification model based on tunneling performance. International Journal of Geomechanics, 2022, 22(6), 04022070.

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

Ravanelli, M., Brakel, P., Omologo, M., & Bengio, Y. Light gated recurrent units for speech recognition. IEEE Transactions on Emerging Topics in Computational

Intelligence, 2018, 2(2): 92-102.

Ravizza, G., Ferrari, R., Rizzi, E., & Dertimanis, V. On the denoising of structural vibration response records from low-cost sensors: A critical comparison and assessment. Journal of Civil Structural Health Monitoring, 2021, 11(5): 1201-1224.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

Revach, G., Shlezinger, N., Ni, X., Escoriza, A. L., Van Sloun, R. J., & Eldar, Y. C. KalmanNet: Neural network aided Kalman filtering for partially known dynamics. IEEE Transactions on Signal Processing, 2022, 70: 1532-1547.

Riveiro, B., DeJong, M. J., & Conde, B. Automated processing of large point clouds for structural health monitoring of masonry arch bridges. Automation in Construction, 2016, 72: 258-268.

Rizvi, A. R., Khan, P. R., & Ahmad, S. Crack detection in railway track using image processing. International Journal of Advance Research, Ideas and Innovations in Technology, 2017, 3(4): 489-496.

Roy, A., Saffar, M., Vaswani, A., & Grangier, D. Efficient content-based sparse attention with routing transformers. Transactions of the Association for Computational Linguistics, 2021, 9: 53-68.

Rui, E. Z., Zeng, G. Z., Ni, Y. Q., Chen, Z. W., & Hao, S. Time-averaged flow field reconstruction based on a multifidelity model using physics-informed neural network (PINN) and nonlinear information fusion. International Journal of Numerical Methods for Heat & Fluid Flow,2024, 34(1): 131-149.

Sagheer, A., & Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing, 2019, 323: 203-213.

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. Recent advances in recurrent neural networks. arXiv preprint arXiv, 2017: 1801.01078.

Santos, A., Figueiredo, E., Silva, M. F. M., Sales, C. S., & Costa, J. C. W. A. Machine

learning algorithms for damage detection: Kernel-based approaches. Journal of Sound and Vibration, 2016, 363: 584-599.

Sapankevych, N. I., & Sankar, R. Time series prediction using support vector machines: a survey. IEEE Computational Intelligence Magazine, 2009, 4(2): 24-38.

Schmidt, R. M. Recurrent neural networks (RNNs): A gentle introduction and overview. arXiv preprint arXiv, 2019: 1912.05911.

Schrauwen, B., Wardermann, M., Verstraeten, D., Steil, J. J., & Stroobandt, D. Improving reservoirs using intrinsic plasticity. Neurocomputing, 2008, 71(7-9): 1159-1171.

Schuster, M., & Paliwal, K. K. Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.

Shaban, A. Determination of concrete properties using hyperspectral imaging technology: A review. Science Journal of Physics, 2013, 2013: 102.

Shen, Z., Zhang, M., Zhao, H., Yi, S., & Li, H. Efficient attention: Attention with linear complexities. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3531-3539.

Shi, Z. Y., Law, S. S., & Zhang, L. M. Optimum sensor placement for structural damage detection. Journal of Engineering Mechanics, 2000, 126(11): 1173-1179.

Shih, S. Y., Sun, F. K., & Lee, H. Y. Temporal pattern attention for multivariate time series forecasting. Machine Learning, 2019, 108: 1421-1441.

Shin, J. U., Kim, D., Kim, J. H., & Myung, H. Micro-aerial vehicle type wall-climbing robot mechanism for structural health monitoring. In Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2013, 8692: 363-369.

Silik, A., Noori, M., Altabey, W. A., Ghiasi, R., & Wu, Z. Comparative analysis of wavelet transform for time-frequency analysis and transient localization in structural health monitoring. Structural Durability & Health Monitoring, 2021, 15(1): 1.

Sivasuriyan, A., Vijayan, D. S., Devarajan, P., Stefańska, A., Dixit, S., Podlasek, A., ... & Koda, E. Emerging trends in the integration of smart sensor technologies in structural health monitoring: a contemporary perspective. Sensors, 2024, 24(24):

8161.

Sofi, A., Regita, J. J., Rane, B., & Lau, H. H. Structural health monitoring using wireless smart sensor network-An overview. Mechanical Systems and Signal Processing, 2022, 163: 108113.

Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W., Nadler, B. R., & Czarnecki, J. J. A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory, USA, 2003, 1(16): 10-12989.

Soman, R., Moaf, F. O., Fiborek, P., Kudela, P., & Kurpińska, M. Investigating the suitability of the matched fiber Bragg grating approach for guided wave based structural health monitoring. Measurement, 2025, 242: 115935.

Sreenath, S., Malik, H., Husnu, N., & Kalaichelavan, K. Assessment and use of unmanned aerial vehicle for civil structural health monitoring. Procedia Computer Science, 2020, 170: 656-663.

Staszewski, W. J., bin Jenal, R., Klepka, A., Szwedo, M., & Uhl, T. A review of laser Doppler vibrometry for structural health monitoring applications. Key Engineering Materials, 2012, 518: 1-15.

Stubbs, N., & Park, S. Optimal sensor placement for mode shapes via Shannon's sampling theorem. Computer-Aided Civil and Infrastructure Engineering, 1996, 11(6): 411-419.

Su, J. Z., Xia, Y., Chen, L., Zhao, X., Zhang, Q. L., Xu, Y. L., ... & Chen, A. R. Long-term structural performance monitoring system for the Shanghai Tower. Journal of civil structural health monitoring, 2013, 3: 49-61.

Sun, Z., & Chang, C. C. Structural damage assessment based on wavelet packet transform. Journal of structural engineering, 2002, 128(10): 1354-1361.

Sun, C., Song, M., Cai, D., Zhang, B., Hong, S., & Li, H. A systematic review of echo state networks from design to application. IEEE Transactions on Artificial Intelligence, 2022, 5(1): 23-37.

Sun, Z., Mahmoodian, M., Sidiq, A., Jayasinghe, S., Shahrivar, F., & Setunge, S. Optimal

sensor placement for structural health monitoring: A comprehensive review. Journal of Sensor and Actuator Networks, 2025, 14(2): 22.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence 2017, 31(1).

Tan, C., Uddin, N., OBrien, E. J., McGetrick, P. J., & Kim, C. W. Extraction of bridge modal parameters using passing vehicle response. Journal of Bridge Engineering, 2019, 24(9): 04019087.

Tan, Y., & Zhang, L. Computational methodologies for optimal sensor placement in structural health monitoring: A review. Structural Health Monitoring, 2019, 19(4): 1287-1308.

Tan, Y. K., Wang, Y. L., Ni, Y. Q., Zhang, Q. L., & Wang, Y. W. Improved bidirectional echo state network-based time series reconstruction and prediction for structural response. Structural Health Monitoring, 2024: 14759217241253082.

Tan, Y. K., Wang, Y. L., Deng, E., Ye, X., Zhang, Y., & Ni, Y. Q. Automatic damage detection and data completion for operational bridge safety using convolutional echo state networks. Automation in Construction, 2024, 166: 105606.

Tan, Y. K., Ni, Y. Q., Zhang, S. X., Zhang, Q. L., & Wang, Y. W. A fast, information-interactive, and reservoir computing-based digital twin for high-rise building operation. Expert Systems with Applications, 2025, 269, 126390.

Tibaduiza, D., Torres-Arredondo, M. Á., Vitola, J., Anaya, M., & Pozo, F. A damage classification approach for structural health monitoring using machine learning. Complexity, 2018, 2018(1): 5081283.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. Llama: Open and efficient foundation language models. arXiv preprint

arXiv, 2023: 2302.13971.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. Attention is all you need. Advances in Neural Information Processing systems, 2017: 30.

Wan, H. P., & Ni, Y. Q. Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. Structural Health Monitoring, 2019, 18(4): 1282-1309.

Wang, X., Jin, Y., & Hao, K. Evolving local plasticity rules for synergistic learning in echo state networks. IEEE transactions on neural networks and learning systems, 2019, 31(4): 1363-1374.

Wang, Y. W., Ni, Y. Q., & Wang, X. Real-time defect detection of high-speed train wheels by using Bayesian forecasting and dynamic model. Mechanical Systems and Signal Processing, 2020, 139: 106654.

Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., ... & Yu, J. Traffic flow prediction via spatial temporal graph neural network. In Proceedings of the web conference, 2020: 1082-1092.

Wang, L., Xiao, T., Liu, S., Zhang, W., Yang, B., & Chen, L. Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation. Gondwana Research, 2023, 123: 27-40.

Wang, N., Shang, L., & Song, X. A transformer-optimized deep learning network for road damage detection and tracking. Sensors, 2023, 23(17): 7395.

Wang, H., Li, B., Fan, S., Wu, Y., & Liu, X. TimeSQL: Improving multivariate time series forecasting with multi-scale patching and smooth quadratic loss. Information Sciences, 2024, 671: 120652.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. A survey of transfer learning. Journal of Big data, 2016, 3: 1-40.

Wong, K. Y., Man, K. L., & Chan, W. Y. K. Application of global positioning system to structural health monitoring of cable-supported bridges. In Health Monitoring and

Management of Civil Infrastructure Systems 2001, 4337: 390-401.

Wu, C. J., Liao, W. H., & Wang, M. Y. Modeling of granular particle damping using multiphase flow theory of gas-particle. Journal of Vibration and Acoustics, 2004, 126(2): 196-201.

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 753-763.

Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. Graph neural networks in recommender systems: a survey. ACM Computing Surveys, 2022, 55(5): 1-37.

Xie, X., Wang, Q., Shahrour, I., Li, J., & Zhou, B. A real-time interaction platform for settlement control during shield tunnelling construction. Automation in Construction, 2018, 94: 154-167.

Xu, Z. Q. J., Zhang, Y., & Xiao, Y. Training behavior of deep neural network in frequency domain. In Neural Information Processing: 26th International Conference, ICONIP 2019, 1(26): 264-274.

Yan, L., Yan, J., Xu, G., He, S., & Zhang, Z. Safety analysis of temporary anchorage system for immersed tube in Shenzhen-Zhongshan Link. Journal of Traffic and Transportation Engineering (English Edition), 2024, 11(4): 797-813.

Yang, Y., Chadha, M., Hu, Z., & Todd, M. D. An optimal sensor placement design framework for structural health monitoring using Bayes risk. Mechanical Systems and Signal Processing, 2022, 168: 108618.

Ye, X. W., Su, Y. H., & Han, J. P. A state-of-the-art review on fatigue life assessment of steel bridges. Mathematical Problems in Engineering, 2014, 2014(1): 956473.

Ye, X. W., Dong, C. Z., & Liu, T. A review of machine vision-based structural health monitoring: methodologies and applications. Journal of Sensors, 2016, 2016(1): 7103039.

Ye, X., Ni, Y. Q., Sajjadi, M., Wang, Y. W., & Lin, C. S. Physics-guided, data-refined

modeling of granular material-filled particle dampers by deep transfer learning. Mechanical Systems and Signal Processing, 2022, 180: 109437.

Yelisetti, S., Saini, V. K., Kumar, R., Lamba, R., & Saxena, A. Uncertainty aware learning model for thermal comfort in smart residential buildings. IEEE Transactions on Industry Applications, 2023, 60(1): 1909-1918.

Yi, T., Zhou, G., Li, H., & Wang, C. Optimal placement of triaxial sensors for modal identification using hierarchic wolf algorithm. Structural Control and Health Monitoring, 2017, 24(8): e1958.

Yi, Z. nmODE: neural memory ordinary differential equation. Artificial Intelligence Review, 2023, 56(12): 14403-14438.

Yu, B., Yin, H., & Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv, 2017, 1709.04875.

Yu, Y., Si, X., Hu, C., & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 2019, 31(7): 1235-1270.

Yu, T., Tang, Q., & Vinayaka, S. Identifying structural properties of a steel railway bridge for structural health monitoring using laser Doppler vibrometry. Automation in Construction, 2024, 160: 105320.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. Ts2vec: Towards universal representation of time series. In Proceedings of the AAAI conference on artificial intelligence 2022, 36(8): 8980-8987.

Yuan, L., Ni, Y. Q., Deng, X. Y., & Hao, S. A-PINN: Auxiliary physics informed neural networks for forward and inverse problems of nonlinear integro-differential equations. Journal of Computational Physics, 2022, 462: 111260.

Yuan, F., Yan, Z., Zhang, R., Yang, Y., Jian, S., & Tang, B. Research on Vibration Characteristic Analysis and Fault Diagnosis Method of Oil-Immersed Transformer Based on Multi-Physics Coupling. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 19(10): 1621-1630.

Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model.

Neurocomputing, 2003, 50: 159-175.

Zhang, X., Zhou, X., Lin, M., & Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 6848-6856.

Zhang, L., Hu, X., Xie, Z., Shi, B., Zhang, L., & Wang, R. Field measurement study on time-varying characteristics of modal parameters of super high-rise buildings during super typhoon. Journal of Wind Engineering and Industrial Aerodynamics, 2020, 200: 104139.

Zhang, S., Hou, W., Yin, J., & Lin, Z. A review of research and practice on the theory and technology of reservoir dam risk assessment. Sustainability, 2022, 14(22): 14984.

Zhang, Y., Dang, D. Z., Wang, Y. W., & Ni, Y. Q. Damage identification for railway tracks using ultrasound guided wave and hybrid probabilistic deep learning. Construction and Building Materials, 2024, 418: 135466.

Zhang, J., Huang, M., Wan, N., Deng, Z., He, Z., & Luo, J. Missing measurement data recovery methods in structural health monitoring: The state, challenges and case study. Measurement, 2024: 114528.

Zhang, S., Tan, Y., Ge, H., & Zhang, Q. Safety-Function-Environment Evaluation System for Large-Span Cable-Supported Bridges: Theory and Case Studies. Sustainability, 2024, 16(4): 1414.

Zhang, S., Tan, Y., Ge, H., & Zhang, Q. An ultra-thin bolt tension sensor and online monitoring system: For application in hydropower plant unit. Frontiers of Structural and Civil Engineering, 2024, 18(9): 1388-1400.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 2019, 115: 213-237.

Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. Do RNN and LSTM have long memory? In International Conference on Machine Learning, 2020: 11365-11375.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. A survey of large language models. arXiv preprint arXiv, 2023: 2303.18223.

Zheng, J., Su, M., Ying, W., Tong, J., & Pan, Z. Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis. Measurement, 2021, 179: 109425.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence 2021, 35(12): 11106-11115.

Zhou, X., Sun, K., Wang, J., Zhao, J., Feng, C., Yang, Y., & Zhou, W. Computer vision enabled building digital twin using building information model. IEEE Transactions on Industrial Informatics, 2022, 19(3): 2684-2692.

Zhu, L., Fu, Y., Chow, R., Spencer Jr, B. F., Park, J. W., & Mechitov, K. Development of a high-sensitivity wireless accelerometer for structural health monitoring. Sensors, 2018, 18(1): 262.