

Article

Model Validation for Multivariate Functional Responses via Autoencoder-Based Dual-Layer Feature Extraction

Dengyu Wu, Xiaodong Zhang, Daobo Sun, Haidong Lin, Jinhui Li and Baoqiang Zhang *

School of Aerospace Engineering, Xiamen University, Xiamen 361000, China; wudengyu@stu.xmu.edu.cn (D.W.); zhangxiaodong@stu.xmu.edu.cn (X.Z.); sundaobo@stu.xmu.edu.cn (D.S.); xmulhd@163.com (H.L.); 15892137005@163.com (J.L.)

* Correspondence: bqzhang@xmu.edu.cn

Abstract

Model validation for complex simulation models with multivariate functional responses poses significant challenges, as it involves the dual coupling of physical correlations among variables and field correlations in time-series data. A novel Autoencoder-based Dual-Layer Feature Extraction (AE-DLFE) method is proposed. The first layer uses joint principal component analysis to decouple physical correlations, while the second layer develops an Autoencoder-improved Feature Selective Validation (AE-FSV) method that adaptively extracts features of time-series data and measures feature discrepancies via deep representation learning. On this basis, a new validation metric named U-PCDM (Uncertainty Principal Component Difference Measure) is developed to quantify the discrepancies between simulation and experiment under uncertainty. Theoretical analysis confirms the boundedness and unique temporal permutation sensitivity of the proposed metric. Case study results demonstrate that the proposed AE-FSV enhances the evaluation accuracy of traditional FSV on transient data. Furthermore, compared to benchmark methods such as MD-pooling, the U-PCDM metric significantly improves computational efficiency—especially in high-dimensional scenarios—while maintaining consistent model rankings. This work effectively addresses the heterogeneous correlation coupling issue, offering a robust quantitative tool for model validation.

Keywords: model validation; multivariate functional responses (MVFR); autoencoder (AE); feature selective validation (FSV); deep representation learning; uncertainty quantification (UQ)

MSC: 62R10; 62P30

Academic Editor: Stoyan Tranev,
Velichka Traneva

Received: 11 January 2026
Revised: 10 February 2026
Accepted: 12 February 2026
Published: 13 February 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Modeling and simulation technologies are widely applied in fields such as aerospace, electromagnetic scattering, and structural dynamics, as engineering systems are becoming increasingly complex. As a result, accurately assessing the credibility of models in engineering scenarios has become a key factor in determining whether simulation results can be applied to engineering practice. Consequently, Uncertainty Quantification (UQ) and Verification and Validation (V&V) techniques have attracted growing attention [1,2]. Model validation is defined as the process of evaluating the extent to which a model

accurately represents the real world from the perspective of its intended use. In recent years, model validation has evolved from deterministic comparisons based on single-point errors to systematic validation frameworks integrating statistical inference and uncertainty. Within these frameworks, validation metrics serve as critical tools for quantifying discrepancies between simulations and experiments under uncertainty [3].

Validation metrics can be broadly categorized based on the dimensionality of the response into univariate and multivariate metrics. Each category further encompasses scenarios involving either single or multiple validation sites. In univariate response systems, the Area Metric [4] effectively characterizes the distributional discrepancies between two sets of samples by comparing the area difference between their Cumulative Distribution Functions (CDFs). Extensions include the normalized Area Metric based on probability density function (PDF) [5], the u-pooling method for multiple validation site scenarios [6], and the global Area Metric for spatial multi-point wind fields [7]. For multivariate response systems, it is necessary to simultaneously consider the correlations among multiple physical quantities. Li et al. [8] proposed the t-pooling method, which achieves multivariate probability integral transform via joint CDF, but suffers from high computational costs. To improve efficiency, subsequent studies combined Principal Component Analysis (PCA) [9] and Mahalanobis Distance (MD) [10] with the Area Metric, respectively, resulting in a significant reduction in computational burden. In addition, the multi-metric fusion method [11], the energy distance-based metric [12], and methods suitable for multiple validation site scenarios such as MD-pooling [10], COPMV [13], and EMD-pooling [14] have further enriched the validation metric methodologies for multivariate responses.

However, the outputs of contemporary high-performance computing simulations are rarely restricted to simple scalars or vectors. Instead, they increasingly manifest as functional data, leading to two additional categories of validation problems: Functional Responses and Multivariate Functional Responses (MVFRs). As a core concept in Functional Data Analysis (FDA) [15], MVFRs refer to data composed of multiple interrelated variables that simultaneously vary over time (or other continuous domains such as space). Model validation for such systems poses a significant mathematical challenge: the “dual coupling of correlations”. These datasets are governed by physical correlations among different physical quantities and field correlations characterizing the temporal or spatial evolution of each quantity. Existing validation metrics and discrepancy assessments have largely evolved from scalar statistics [4–14] or univariate signal processing [16–21] and often struggle to decouple these heterogeneous correlation structures. This limitation gives rise to “validation blind spots”: conflating field correlations with physical correlations or ignoring either type, resulting in the omission of key physical information carried by temporal evolution or synergistic changes in physical quantities, ultimately leading to erroneous validation and decision-making outcomes.

In the field of FDA, research on MVFR has predominantly centered on statistical modeling and pattern recognition rather than simulation model validation. For instance, Happ et al. [22] proposed Multivariate Functional Principal Component Analysis (MFPCA) to handle multivariate functional data defined on varying dimensional domains. Rigueira et al. [23] developed a machine learning framework integrating multivariate functional depth, directional outlyingness, and clustering methods for anomaly detection in water quality sensor data. Additionally, Nolan et al. [15] presented a Bayesian MFPCA method that addresses the challenge of sparse or irregular data, enabling cross-variable information fusion and rapid computation. Its biomedical interpretive value was further demonstrated in COVID-19 research. While these FDA methods effectively model functional structures and capture temporal dynamics, they generally lack the uncertainty quantification mechanisms and specific discrepancy measures required for model

validation. Nevertheless, their strategies for handling functional data offer valuable insights for developing validation metrics for MVFR.

Table 1 summarizes the characteristics of the existing validation methods. Overall, validation metrics for MVFR still face the following challenges:

Table 1. Taxonomy of validation methods.

Response Category	Representative Method	Physical Correlation	Field Correlation	Uncertainty Quantification	Discrepancy Quantification
Univariate Response	Area Metric	No	No	Yes	Yes
Multivariate Response	MD-pooling	Yes	No	Yes	Yes
Functional Response	Feature Selective Validation (FSV)	No	Yes	No	Yes
MVFR	MFPCA	Yes	Yes	No	No

1. MVFR data inherently exhibit both cross-variable physical correlations and intra-variable field correlations. These two types of correlations are significantly different in terms of mathematical structures and physical interpretations. However, existing validation metrics predominantly focus on problems where the output is a scalar (univariate response) or a vector (multivariate response). These methods prioritize physical correlations, effectively handling cross-variable dependencies, but often neglect field correlations or conflate them with physical correlations, thereby losing sensitivity to the structural evolution of the field. Conversely, while functional responses are primarily addressed within the FDA domain, these approaches typically lack the uncertainty quantification mechanisms and discrepancy measures essential for model validation.
2. The digital era has driven an increasing demand for real-time model validation and decision-making. However, given the significant data volumes characteristic of MVFR, the computational efficiency of validation metric algorithms remains a critical bottleneck.

Therefore, there is an urgent need for a systematic method capable of strictly decoupling and integrating both physical and field correlations while effectively incorporating uncertainty quantification. To achieve this, this paper proposes an Autoencoder-based Dual-Layer Feature Extraction (AE-DLFE) method, characterized by a “Dual Correlation Decoupling” strategy. The first layer employs the Joint PCA (JPCA) to decouple physical correlations; the second layer develops an Autoencoder-improved Feature Selective Validation (AE-FSV) method, which adaptively extracts field features and assesses discrepancies via data-driven deep representation learning. On this basis, this paper constructs the Uncertainty Principal Component Difference Measure (U-PCDM) validation metric and a decision criterion. These contributions are designed to resolve the validation challenges posed by the dual coupling of correlations, ultimately enabling efficient and robust model validation and decision-making under uncertainty.

The remainder of this paper is organized as follows: Section 2 details the theoretical framework of the AE-DLFE method and the construction of the U-PCDM validation metric, along with its mathematical properties. Section 3 conducts a comprehensive performance evaluation and validation on both numerical benchmarks and engineering simulation examples. Section 4 provides an in-depth discussion on the proposed method and the obtained results. Section 5 summarizes the main contributions and conclusions of this paper.

2. Theory and Methodology

2.1. Autoencoder-Based Dual-Layer Feature Extraction Method

For complex simulation systems generating MVFR, let the simulation dataset be defined as $\mathcal{D}_{\text{sim}} = \{\mathbf{Y}_{\text{sim}}^{(i)}\}_{i=1}^{N_{\text{sim}}}$, and the corresponding experimental dataset as $\mathcal{D}_{\text{exp}} = \{\mathbf{Y}_{\text{exp}}^{(j)}\}_{j=1}^{N_{\text{exp}}}$. Here, N_{sim} and N_{exp} denote the sample sizes for the simulation and experiment, respectively. For an arbitrary sample, its response can be represented as a temporal matrix $\mathbf{Y} \in \mathbb{R}^{L \times V}$, as shown in Equation (1):

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_V] = \begin{bmatrix} y_1(t_1) & y_2(t_1) & \cdots & y_V(t_1) \\ y_1(t_2) & y_2(t_2) & \cdots & y_V(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(t_L) & y_2(t_L) & \cdots & y_V(t_L) \end{bmatrix} \quad (1)$$

where L is the field length and V is the number of coupled physical quantities.

Equation (1) reveals the inherent “dual coupling of correlations” in MVFR: physical correlations exist among physical variables (columns), while field correlations govern the evolution within each variable's time series (rows). To address this challenge, this paper proposes the AE-DLFE method, the framework of which is illustrated in Figure 1. This method consists of two layers of feature extraction: the first layer employs the JPCA to decouple physical correlations. In the second layer, this paper proposes the AE-FSV method to adaptively extract field features and quantify discrepancies. A detailed description of the AE-DLFE method is provided below.

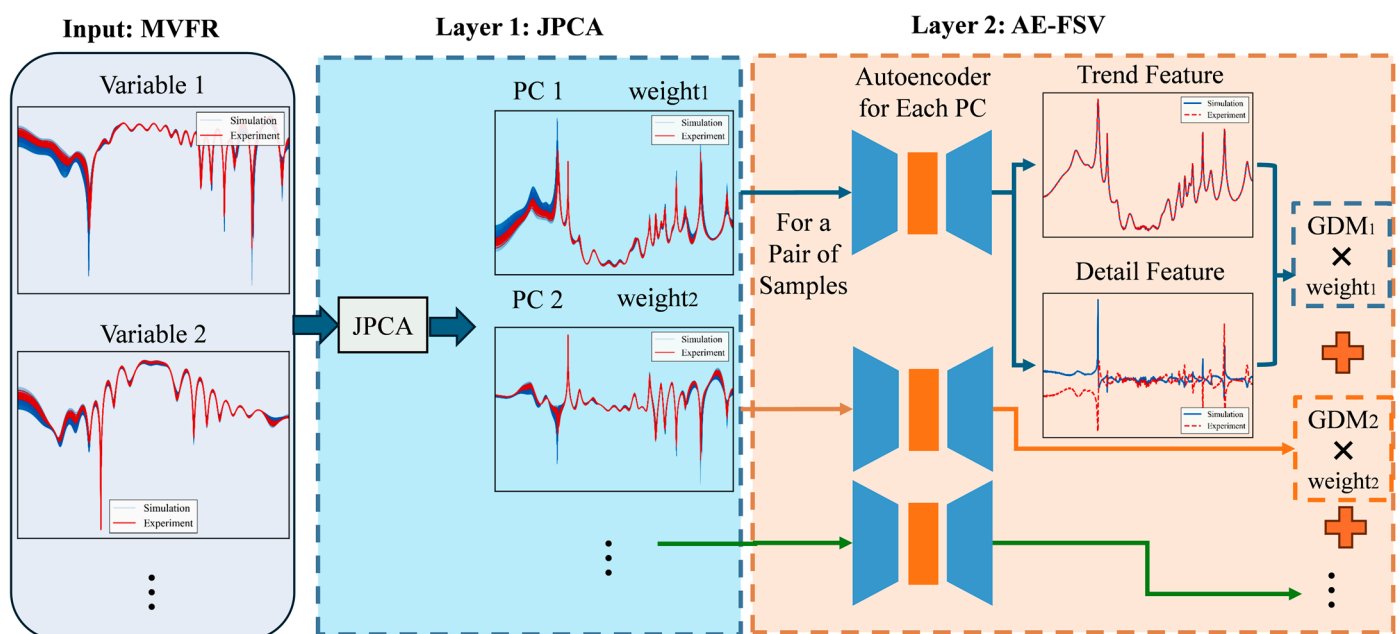


Figure 1. Framework of AE-DLFE method. Layer 1 employs JPCA to decouple physical correlations, transforming the MVFR data into multiple independent temporal sequences. In Layer 2, an individual autoencoder is trained for each principal component channel to extract field features (specifically, trend features and detail features). Finally, the global discrepancy of the MVFR data is derived by computing the weighted sum of feature discrepancies across all channels.

2.1.1. Layer 1: JPCA-Based Physical Correlation Decoupling

Physical correlations stem from the intrinsic relationships among different physical quantities governed by physical mechanisms. In the model validation process for multivariate response systems with physical correlations, the primary focus is not on the variation trends or details of the transition from physical quantity A to B, but rather on the interdependency of each physical quantity and its mutual coupling with others. Consequently, the validation results should remain invariant to the permutation of physical quantities within the dataset.

Given these characteristics, PCA technology has great application potential in addressing such correlations. As a classic feature extraction and dimensionality reduction technique, PCA is widely used in fields such as data science, statistics, and machine learning [24,25]. Its core objective is to eliminate redundant correlations in the original data by identifying orthogonal principal components in the data, thereby laying the foundation for further evaluations.

To ensure that all samples from simulations and experiments under uncertainty undergo the same transformation, this paper employs the classic JPCA method for the first layer of feature extraction.

First, all samples from the simulation dataset $\mathcal{D}_{\text{sim}}$ and experimental dataset $\mathcal{D}_{\text{exp}}$ are stacked vertically to obtain the joint data matrix $\mathbf{Y}_{\text{joint}} \in \mathbb{R}^{(N_{\text{sim}}+N_{\text{exp}}) \times V}$:

$$\mathbf{Y}_{\text{joint}} = \begin{bmatrix} (\mathbf{Y}_{\text{sim}}^{(1)})^T & \cdots & (\mathbf{Y}_{\text{sim}}^{(N_{\text{sim}})})^T & (\mathbf{Y}_{\text{exp}}^{(1)})^T & \cdots & (\mathbf{Y}_{\text{exp}}^{(N_{\text{exp}})})^T \end{bmatrix}^T \quad (2)$$

Since different physical quantities usually exhibit varying dimensions and scales, it is necessary to standardize the joint data matrix $\mathbf{Y}_{\text{joint}}$ column-wise. Let $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_V]$ and $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_V]$ denote the mean vector and standard deviation vector of $\mathbf{Y}_{\text{joint}}$, respectively. The standardized joint data matrix $\bar{\mathbf{Y}}_{\text{joint}}$ can be calculated by Equation (3):

$$\bar{y}_{k,v} = \frac{y_{k,v} - \mu_v}{\sigma_v} \quad (3)$$

where $\bar{y}_{k,v}$ and $y_{k,v}$ are the elements in the k -th row and v -th column of $\bar{\mathbf{Y}}_{\text{joint}}$ and $\mathbf{Y}_{\text{joint}}$ matrices, respectively. $\bar{\mathbf{Y}}_{\text{joint}}$ can be denoted as

$$\bar{\mathbf{Y}}_{\text{joint}} = \begin{bmatrix} (\bar{\mathbf{Y}}_{\text{sim}}^{(1)})^T & \cdots & (\bar{\mathbf{Y}}_{\text{sim}}^{(N_{\text{sim}})})^T & (\bar{\mathbf{Y}}_{\text{exp}}^{(1)})^T & \cdots & (\bar{\mathbf{Y}}_{\text{exp}}^{(N_{\text{exp}})})^T \end{bmatrix}^T \quad (4)$$

Subsequently, the covariance matrix $\mathbf{C} \in \mathbb{R}^{V \times V}$ of $\bar{\mathbf{Y}}_{\text{joint}}$ is computed via Equation (5) to capture the global correlation structure among physical variables:

$$\mathbf{C} = \frac{1}{(N_{\text{sim}} + N_{\text{exp}})L - 1} \bar{\mathbf{Y}}_{\text{joint}}^T \bar{\mathbf{Y}}_{\text{joint}} \quad (5)$$

Through the eigenvalue decomposition of matrix $\mathbf{C}$, eigenvalues ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_V$) and their corresponding orthogonal eigenvectors ($\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_V$) are obtained. It is well known that $\mathbf{C}$ is a positive semi-definite matrix, and all its eigenvalues are non-negative. A larger eigenvalue indicates greater variance (information) contained in the direction of the corresponding eigenvector. Therefore, a cumulative

contribution threshold $\eta = 0.9$ is established. When the cumulative proportion of the first m eigenvalues exceeds the threshold (as shown in Equation (6)), it means that the projection of data onto these eigenvectors has retained sufficient information:

$$\frac{\sum_{k=1}^m \lambda_k}{\sum_{j=1}^V \lambda_j} \geq \eta \quad (6)$$

Let $\mathbf{W}_m = [\mathbf{w}_1, \dots, \mathbf{w}_m] \in \mathbb{R}^{V \times m}$ be the projection matrix composed of the first m orthogonal principal components; then the simulation projection data $\mathbf{P}_{\text{sim}}$ and experimental projection data $\mathbf{P}_{\text{exp}}$ with decoupled physical correlations are given by

$$\mathbf{P}_{\text{sim}} = \begin{bmatrix} \bar{\mathbf{Y}}_{\text{sim}}^{(1)} \\ \vdots \\ \bar{\mathbf{Y}}_{\text{sim}}^{(N_{\text{sim}})} \end{bmatrix} \mathbf{W}_m, \mathbf{P}_{\text{exp}} = \begin{bmatrix} \bar{\mathbf{Y}}_{\text{exp}}^{(1)} \\ \vdots \\ \bar{\mathbf{Y}}_{\text{exp}}^{(N_{\text{exp}})} \end{bmatrix} \mathbf{W}_m \quad (7)$$

In addition, the normalized eigenvalues can serve as weights $r_k (k=1, 2, \dots, m)$ for the subsequent metric calculation:

$$r_k = \frac{\lambda_k}{\sum_{j=1}^m \lambda_j} \quad (8)$$

2.1.2. Layer 2: Autoencoder-Improved Feature Selective Validation

Having obtained the projection data $\mathbf{P}_{\text{sim}}$ and $\mathbf{P}_{\text{exp}}$ with decoupled physical correlations from Layer 1, the subsequent challenge lies in evaluating the discrepancies between these two sets of projected data. The hypothesis that the dependency structure of the MVFR system is primarily characterized by second-order statistics (i.e., covariance) is adopted. Consequently, the decorrelation provided by JPCA serves as a robust approximation for statistical independence, allowing for the complex multivariate PDF to be effectively factored into marginal densities. This factorization property is critical for information preservation. It implies that the global distributional discrepancy between simulation and experiment can be effectively decomposed into an aggregation of discrepancies within individual principal component channels, thereby satisfying additivity. Therefore, independently evaluating feature discrepancies for each channel in Layer 2 and subsequently computing a weighted sum does not result in the loss of validation information regarding the system's dependency structure. The problem is thus transformed into quantifying the discrepancies between functional responses within single principal component channels.

Unlike physical correlations, field correlations manifest in the dynamic evolution process of a single physical quantity changing over time (or space). In the validation of functional response models, researchers focus not only on amplitude discrepancies, but also on the consistency of variation trends and detail features between adjacent field points. Consequently, the sequential order of field points will affect the validation metric results, and traditional permutation-invariant metrics are no longer applicable. This necessitates the proposed dual-layer feature extraction method to decouple and process the two types of correlations in separate layers.

FSV is a discrepancy evaluation method for functional response data developed from the field of Electromagnetic Compatibility (EMC) [26]. By decomposing and comparing

features such as amplitude, trend, and details of two sets of data, it comprehensively quantifies the global and local discrepancies of the data. Traditional FSV is based on the IEEE Std 1597.1 standard [26] and employs Fast Fourier Transform (FFT) to decompose signals into low-frequency “trend features” and high-frequency “detail features”. However, FFT is a global transform that lacks time-frequency localization capability and can lead to distorted results for complex responses characterized by abrupt transients and high-frequency fluctuations [27]. To address this, Shi et al. [28] and Niu et al. [29] introduced the Smoothed Pseudo-Wigner-Ville Distribution (SPWVD) and Wavelet Transform (WT) respectively to improve FSV. Although these methods have made progress in specific scenarios, they rely on fixed mathematical basis functions and hyperparameters predefined by expert experience, and thus cannot adaptively match specific physical systems.

To overcome the aforementioned drawbacks, this paper introduces autoencoders (AE) [30,31] and proposes the AE-FSV method. This method captures the latent physical dynamics inherent in simulation data through data-driven deep representation learning, thereby achieving adaptive decomposition and discrepancy evaluation of functional responses.

Specifically, for the simulation projection data $\mathbf{P}_{\text{sim}}$ and experimental projection data $\mathbf{P}_{\text{exp}}$ with physical correlations decoupled in Layer 1, each sample is decomposed into m principal component channels column-wise. For each channel sequence $\mathbf{x}$, an independent undercomplete autoencoder network is constructed based on all simulation sample curves. The network adopts a compression-expansion symmetric structure, as illustrated in Figure 2.

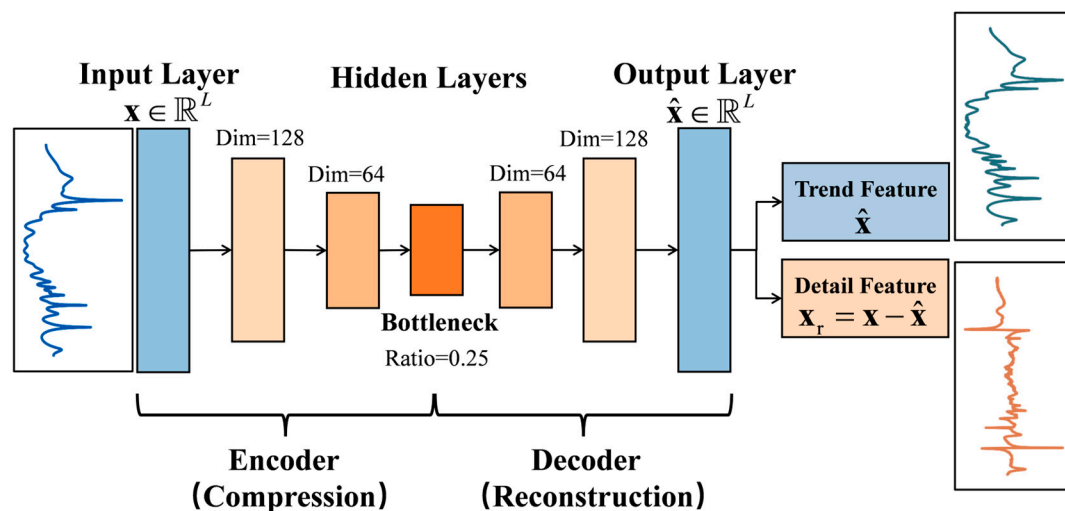


Figure 2. Structure of the autoencoder network for a single principal component channel.

The network consists of an encoder $f_{\theta}(\cdot)$ and a decoder $z_{\phi}(\cdot)$. The encoder maps the input high-dimensional sequence $\mathbf{x} \in \mathbb{R}^L$ (where L is the field length) to a low-dimensional latent manifold $\xi \in \mathbb{R}^{n_b}$ (where $n_b = \alpha_b L$ denotes the dimension of the bottleneck layer with the bottleneck ratio $\alpha_b = 0.25$) through a progressive compression path ($L \rightarrow 128 \rightarrow 64 \rightarrow n_b$). Mathematically, this process $f_{\theta}(\cdot)$ is defined as:

$$\mathbf{h}_1 = \tanh(\mathbf{\Omega}_1 \mathbf{x} + \mathbf{b}_1) \quad (9)$$

$$\mathbf{h}_2 = \tanh(\mathbf{\Omega}_2 \mathbf{h}_1 + \mathbf{b}_2) \quad (10)$$

$$\xi = \mathbf{\Omega}_\xi \mathbf{h}_2 + \mathbf{b}_\xi \quad (11)$$

where $\mathbf{\Omega}$ and $\mathbf{b}$ represent the weight matrices and bias vectors, respectively.

Subsequently, the corresponding decoder $z_\phi(\cdot)$ reconstructs the data from the latent features via a symmetric path ($n_b \rightarrow 64 \rightarrow 128 \rightarrow L$):

$$\hat{\mathbf{x}} = \tanh(\mathbf{\Omega}_{\text{out}}(\tanh(\mathbf{\Omega}_4(\tanh(\mathbf{\Omega}_3 \xi + \mathbf{b}_3)) + \mathbf{b}_4)) + \mathbf{b}_{\text{out}}) \quad (12)$$

The training process of the network is an unsupervised deep learning task. To ensure robustness and rigor, the following specific training scheme is adopted:

1. Loss Function: The Mean Squared Error (MSE) is selected as the loss function:

$$\mathcal{L}(\theta, \phi) = \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \left\| \mathbf{x}_{\text{sim},i} - z_\phi \left(f_\theta \left(\mathbf{x}_{\text{sim},i} \right) \right) \right\|^2 \quad (13)$$

where $\mathbf{x}_{\text{sim},i}$ denotes the i -th input simulation sample curve and θ and ϕ are the network parameters of the encoder and decoder, respectively.

2. Initialization: The network weights are initialized using the Kaiming Uniform method. Specifically, weights and biases are sampled from a uniform distribution $U(-\sqrt{k_u}, \sqrt{k_u})$, where $k_u = 1/n_f$, and n_f is the number of input nodes from the previous layer.
3. Regularization: Explicit Dropout is not employed. Instead, the bottleneck layer acts as the primary structural regularizer. By forcing data through this compressed representation, the network is compelled to learn the dominant physical manifold (trend features) while filtering out high-frequency variations (detail features) that cannot be effectively compressed.
4. Stopping Criterion: Unlike supervised learning, Early Stopping is not utilized here. The objective is to fully capture the physical manifold of the simulation training set, and the bottleneck layer already provides sufficient regularization to prevent overfitting. Consequently, a fixed-epoch strategy (500 epochs) is adopted.
5. Training Set: The AE is trained on the complete simulation dataset of the corresponding model to construct a comprehensive reference manifold. In scenarios involving the ranking of multiple competing models, a dedicated AE is trained for each principal component of each model to ensure fair evaluation.

The specific parameter settings for the network architecture and training process are detailed in Table 2.

Table 2. Training parameters of the autoencoder network.

Parameters	Settings
Bottleneck layer ratio	0.25
Activation function	Tanh
Optimization algorithm	Adam
Learning rate	0.001
Batch size	Full batch
Number of training epochs	500
Initialization method	Kaiming Uniform

Assuming that the response of a physical system is governed by a set of deterministic physical equations, the data mathematically reside on a low-dimensional manifold characterized by high structure and correlation. In this context, the trained autoencoder serves as a data-driven nonlinear filter capable of adaptively identifying and preserving signal features consistent with the underlying simulation model mechanisms. Deep networks tend to prioritize learning low-frequency functions over high-frequency details [31]. Thus, for any sample curve $\mathbf{x} \in \mathbb{R}^L$ in the principal component channel, the reconstructed signal $\hat{\mathbf{x}}$ encodes the amplitude and dominant evolutionary trend conforming to the model's physical principles. In contrast, the residual $\mathbf{x}_r$ between the original signal and the reconstructed signal captures the high-frequency noise, measurement interference, or microscopic dynamic characteristics not captured by the AE. Therefore, $\hat{\mathbf{x}}$ and $\mathbf{x}_r$ can be referred to as the trend feature and detail feature, respectively. The $\hat{\mathbf{x}}$ and $\mathbf{x}_r$ of curve $\mathbf{x} \in \mathbb{R}^L$ can be defined as follows:

$$\hat{\mathbf{x}} = z_{\phi} \left(f_{\theta} (\mathbf{x}) \right) \quad (14)$$

$$\mathbf{x}_r = \mathbf{x} - \hat{\mathbf{x}} \quad (15)$$

By replacing the FFT and simple filters used in traditional FSV with this autoencoder-based feature extraction method, the Amplitude Difference Measure (ADM) and Feature Difference Measure (FDM) can be computed more effectively at the k -th field point between any two sets of signals (distinguished by superscripts $\bullet^{\text{set1}}$ and $\bullet^{\text{set2}}$):

$$ADM(k) = \left| \frac{|\hat{\mathbf{x}}^{\text{set1}}(k)| - |\hat{\mathbf{x}}^{\text{set2}}(k)|}{\frac{1}{L} \sum_{i=1}^L (|\hat{\mathbf{x}}^{\text{set1}}(i)| + |\hat{\mathbf{x}}^{\text{set2}}(i)|)} \right| \quad (16)$$

$$FDM_I(k) = \left| \frac{|\hat{\mathbf{x}}^{\text{set1}'}(k)| - |\hat{\mathbf{x}}^{\text{set2}'}(k)|}{\frac{2}{L} \sum_{i=1}^L (|\hat{\mathbf{x}}^{\text{set1}'}(i)| + |\hat{\mathbf{x}}^{\text{set2}'}(i)|)} \right| \quad (17)$$

$$FDM_{II}(k) = \left| \frac{|\mathbf{x}_r^{\text{set1}'}(k)| - |\mathbf{x}_r^{\text{set2}'}(k)|}{\frac{6}{L} \sum_{i=1}^L (|\mathbf{x}_r^{\text{set1}'}(i)| + |\mathbf{x}_r^{\text{set2}'}(i)|)} \right| \quad (18)$$

$$FDM_{III}(k) = \left| \frac{\left| \mathbf{x}_r^{\text{set1}''}(k) \right| - \left| \mathbf{x}_r^{\text{set2}''}(k) \right|}{\frac{7.2}{L} \sum_{i=1}^L \left(\left| \mathbf{x}_r^{\text{set1}''}(i) \right| + \left| \mathbf{x}_r^{\text{set2}''}(i) \right| \right)} \right| \quad (19)$$

$$FDM(k) = 2(FDM_I(k) + FDM_{II}(k) + FDM_{III}(k)) \quad (20)$$

where the superscripts $\bullet'$ and $\bullet''$ denote the first-order and second-order derivatives calculated via the central difference method, respectively, and L is the field length. Notably, since the $\hat{\mathbf{x}}$ extracted by the proposed method naturally retains the Direct Component (DC) of the signal, there is no need to separately set a DC discrepancy index in Equation (16), as is required in traditional FSV.

To obtain a comprehensive metric, the Global Difference Measure (GDM) is calculated via Equation (21):

$$GDM = \frac{1}{L} \sum_{i=1}^L \sqrt{K_{ADM} \cdot ADM^2(i) + K_{FDM} \cdot FDM^2(i)} \quad (21)$$

where K_{ADM} and K_{FDM} are weight coefficients, both set to 1 in this paper.

Unlike traditional FSV, which assigns high-frequency components to the detail feature, AE-FSV preserves physically significant high-frequency features within the trend feature. Consequently, the comparison of these features shifts from the first-order and second-order derivatives of high-frequency features (Equations (18) and (19)) to the direct comparison of amplitudes (Equation (16)) and first-order derivatives (Equation (17)). This is methodologically sound, as these high-frequency components contain essential physical information, making their explicit amplitude comparison more critical than the comparison of second-order derivatives.

2.2. Construction of the U-PCDM Validation Metric

After obtaining the GDM for each principal component, a global metric is required to evaluate the discrepancies between single samples. However, considering the prevalent uncertainties in engineering practice (such as parameter fluctuations and measurement noise), the validation metric should not merely be limited to the deterministic discrepancy evaluation between single samples, but rather should rigorously assess the model's capacity to characterize real experimental results under the influence of uncertainty.

To this end, based on the proposed AE-DLFE method, this section further constructs a new validation metric: U-PCDM. The computational framework of this metric is illustrated in Figure 3, which consists of two parts: the uncertainty quantification of the Principal Component Difference Measure (PCDM) and the subsequent calculation of the U-PCDM metric along with decision-making criteria.

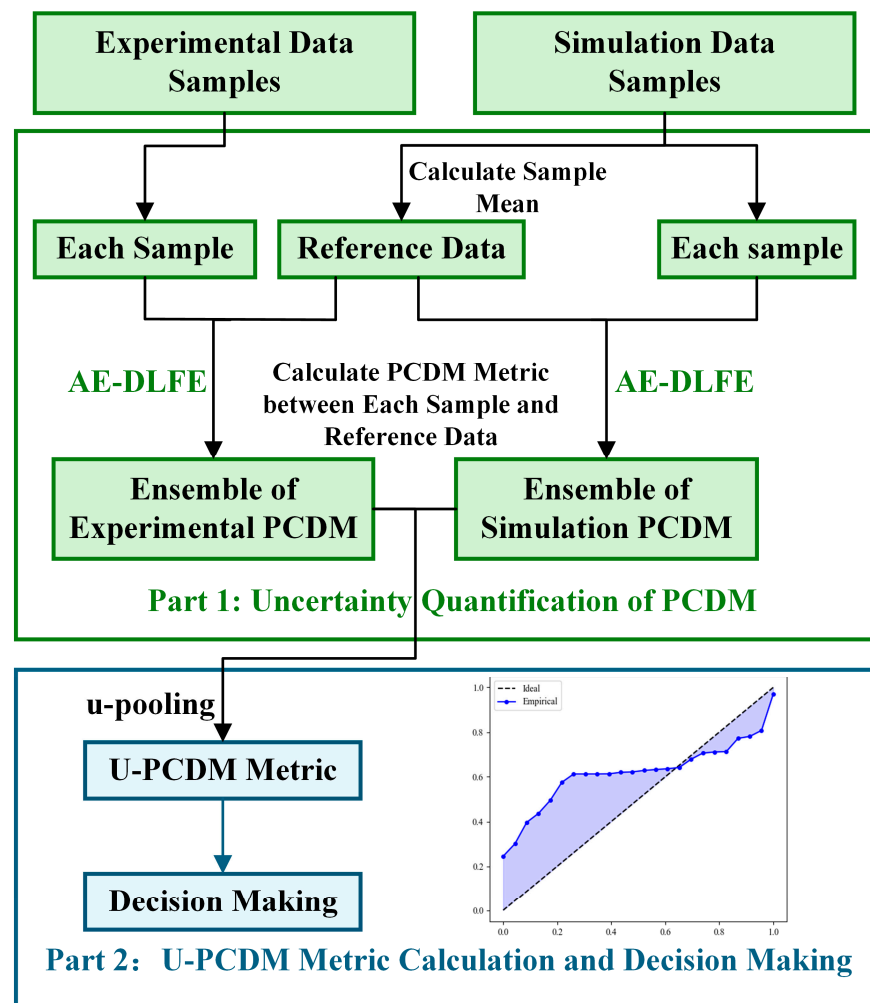


Figure 3. Computational framework of U-PCDM metric.

2.2.1. Uncertainty Quantification of PCDM

For any arbitrary pair of samples $\mathbf{Y}^{(j)}$ and $\mathbf{Y}^{(k)}$ from the simulation dataset $\mathcal{D}_{\text{sim}} = \{\mathbf{Y}_{\text{sim}}^{(i)}\}_{i=1}^{N_{\text{sim}}}$ and experimental dataset $\mathcal{D}_{\text{exp}} = \{\mathbf{Y}_{\text{exp}}^{(i)}\}_{i=1}^{N_{\text{exp}}}$ of the multivariate functional response system, applying the proposed AE-DLFE method yields the GDM metric of each principal component channel. The PCDM metric, reflecting the deterministic discrepancies between the two specific MVFR samples, is derived via a weighted summation using weights r_k derived from Equation (8), as shown in Equation (22):

$$D_{\text{PCDM}}(\mathbf{Y}^{(j)}, \mathbf{Y}^{(k)}) = \sum_{i=1}^m r_i \cdot \text{GDM}_i \quad (22)$$

Based on the aforementioned process, the PCDM metric not only inherits the benefits of decoupled correlations and feature discrepancy evaluation, but also effectively aggregates the complex high-dimensional information into a single scalar.

To quantify uncertainty, we adopt the simulation mean as the reference data $\mathbf{Y}_{\text{ref}}$, and the PCDM metric for every sample in both datasets is calculated relative to this reference. The resulting simulation PCDM set $\mathcal{S}_{\text{sim}} = \{D_{\text{PCDM}}(\mathbf{Y}_{\text{sim}}^{(i)}, \mathbf{Y}_{\text{ref}})\}_{i=1}^{N_{\text{sim}}}$ and experimental PCDM set $\mathcal{S}_{\text{exp}} = \{D_{\text{PCDM}}(\mathbf{Y}_{\text{exp}}^{(i)}, \mathbf{Y}_{\text{ref}})\}_{i=1}^{N_{\text{exp}}}$ effectively capture the global feature

uncertainty of the original datasets $\mathcal{D}_{\text{sim}}$ and $\mathcal{D}_{\text{exp}}$. Therefore, utilizing the PCDM metric—acting as a representative feature scalar—in place of the original MVFR data is scientifically justified. This strategy has effectively converted the complex validation metric problem of MVFR systems into a more manageable validation problem of univariate response systems.

2.2.2. U-PCDM Metric Calculation and Decision Making

To quantify the distributional discrepancy between $\mathcal{S}_{\text{sim}}$ and $\mathcal{S}_{\text{exp}}$, and thereby derive the final U-PCDM metric, this paper employs the u-pooling method [5], the computational principle of which is illustrated in Figure 4.

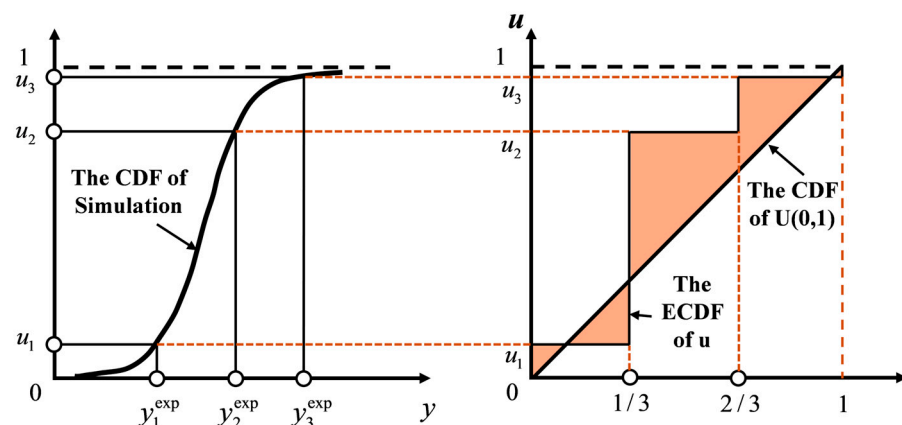


Figure 4. Computational principle of the u-pooling method. The shaded region represents the area difference between the empirical and theoretical CDFs, which corresponds to the validation metric value.

First, Kernel Density Estimation (KDE) is employed to fit the PDF of the simulation PCDM set $\mathcal{S}_{\text{sim}}$, from which the CDF is derived, denoted as $F_{\text{sim}}(\cdot)$. The bandwidth of KDE h_{KDE} is set by multiplying Scott's Rule by a bandwidth scaling factor α_{KDE} :

$$h_{\text{KDE}} = 1.06 \cdot \hat{\sigma} \cdot N^{-1/5} \cdot \alpha_{\text{KDE}} \quad (23)$$

where $\alpha_{\text{KDE}} = 1$ in this paper.

Subsequently, each sample in $\mathcal{S}_{\text{exp}}$ is transformed into the probability space of $[0,1]$ via $F_{\text{sim}}(\cdot)$, denoted as $u_j, j = 1, 2, \dots, N_{\text{exp}}$:

$$u_j = F_{\text{sim}}\left(D_{\text{PCDM}}\left(\mathbf{Y}_{\text{exp}}^{(j)}, \mathbf{Y}_{\text{ref}}\right)\right) \quad (24)$$

Under the assumption that $F_{\text{sim}}(\cdot)$ is estimated accurately, and according to the Probability Integral Transformation (PIT) theorem, if $\mathcal{S}_{\text{sim}}$ and $\mathcal{S}_{\text{exp}}$ are continuous and statistically consistent, the transformed set $\{u_j\}_{j=1}^{N_{\text{exp}}}$ should follow a standard uniform distribution $U(0,1)$. Thus, the proposed new validation metric U-PCDM is defined as the area enclosed between the Empirical Cumulative Distribution Function (ECDF) $F_u(u)$ of the set $\{u_j\}_{j=1}^{N_{\text{exp}}}$ and the CDF of the standard uniform distribution, scaled by a normalization factor of 2:

$$D_{\text{U-PCDM}}(\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{exp}}) = 2 \int_0^1 |F_u(u) - u| du \quad (25)$$

Mathematically, this u-pooling process establishes a theoretical connection between our method and the classic Area Metric framework. While the traditional Area Metric directly quantifies the distance between two univariate CDFs, U-PCDM transforms the MVFR into a univariate variable containing model validation information via AE-DLFE. This effectively converts the MVFR validation problem first into multiple independent functional response metrics and subsequently into a simple univariate area metric problem. This strategy of “dimensionality reduction + area metric” aligns with existing validation methodologies (such as MD-pooling), while U-PCDM extends this logic to scenarios where each response variable is a function.

2.2.3. Mathematical Properties of U-PCDM

To ensure the theoretical rigor of U-PCDM as a validation metric, this subsection establishes its fundamental mathematical properties.

1. Boundedness

Proposition 1 (Boundedness). *For any simulation dataset and experimental dataset formatted as in Equation (1), the U-PCDM metric satisfies the boundedness property: $0 \leq D_{\text{U-PCDM}} \leq 1$.*

Proof of Proposition 1. Since the value range of any CDF is $[0, 1]$, it follows that $D_{\text{U-PCDM}} = 2 \int_0^1 |F_u(u) - u| du \geq 2 \int_0^1 0 du = 0$. The maximum possible area difference between two CDFs is 0.5 (occurring when the distributions are entirely non-overlapping). With the normalization factor of 2, the maximum value becomes 1. Therefore, $0 \leq D_{\text{U-PCDM}} \leq 1$. $\square$

Proposition 1 proves the boundedness of the U-PCDM metric, which is a desirable property as it eliminates the influence of varying physical dimensions and orders of magnitude, thereby facilitating standardized comparisons. Referring to the normalized metric grade classification proposed by Suo et al. [32], the decision criterion based on the U-PCDM metric is formulated as shown in Table 3. Note that the decision criteria presented in Table 3 are empirical reference values derived from the normalized nature of the U-PCDM metric. However, model validation is intrinsically tied to the intended use and risk tolerance of the specific application. Therefore, the interval $[0.1, 0.25)$ should be interpreted as a “discretionary transition zone”. When the U-PCDM metric falls within this range, the model should be cautiously accepted or rejected, strictly contingent upon specific application requirements.

Table 3. Decision criterion based on the U-PCDM metric.

Grade	U-PCDM Value	Decision
Excellent	$[0, 0.1)$	Accept the model
Good	$[0.1, 0.25)$	Conditional Acceptance ¹
Fair	$[0.25, 0.5)$	Calibration Required
Poor	$[0.5, 1]$	Reject the model

¹ Cautiously accept or reject the model based on application requirements.

2. Consistency

Proposition 2 (Consistency). Let F_{sim} and F_{exp} be the CDFs of the PCDM values for the simulation and experimental datasets, respectively. The U-PCDM metric satisfies the following properties: $D_{\text{U-PCDM}} = 0$ if and only if $F_{\text{sim}} = F_{\text{exp}}$ almost everywhere.

Proof of Proposition 2. The proof is derived based on the PIT theorem and the non-negativity of the metric norm.

Sufficiency: Assume $F_{\text{sim}}(p) = F_{\text{exp}}(p)$ for all p . Let P be a random variable representing the experimental data. The transformed variable is defined as $u = F_{\text{sim}}(P)$. Since $P \sim F_{\text{exp}}$ and $F_{\text{sim}}(p) \equiv F_{\text{exp}}(p)$, by the PIT theorem, the variable $u \sim U(0,1)$. Consequently, $D_{\text{U-PCDM}}(\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{exp}}) = 2 \int_0^1 |F_u(v) - v| dv = 2 \int_0^1 |v - v| dv = 0$.

Necessity: Assume $D_{\text{U-PCDM}} = 0$, since the integrand $|F_u(v) - v|$ is non-negative and continuous on $[0,1]$, the integral being zero implies that the integrand must be zero almost everywhere. This implies that u follows a standard uniform distribution. According to the Inverse PIT theorem, if $u = F_{\text{sim}}(P) \sim U(0,1)$, then the random variable P must follow the distribution F_{sim} . Since P is drawn from F_{exp} , it follows that $F_{\text{sim}} = F_{\text{exp}}$. The proposition holds. $\square$

3. Behavior in Limiting Cases

Proposition 3 (Limiting Behavior). In the extreme case where the support of the simulation distribution Ω_{sim} and the experimental distribution Ω_{exp} are disjoint (i.e., $\Omega_{\text{sim}} \cap \Omega_{\text{exp}} = \emptyset$), the metric approaches its upper bound: $D_{\text{U-PCDM}} \rightarrow 1$.

Proof of Proposition 3. Consider the disjoint case where all experimental samples are strictly smaller (or larger) than the simulation samples. The transformed values $u = F_{\text{sim}}(P)$ will converge to the boundaries of the unit interval: $u \rightarrow 0$ if $P < \Omega_{\text{sim}}$ or $u \rightarrow 1$ if $P > \Omega_{\text{sim}}$. Consequently, $\lim_{\Omega_{\text{sim}} \cap \Omega_{\text{exp}} = \emptyset} D_{\text{U-PCDM}}(\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{exp}}) = \lim_{\Omega_{\text{sim}} \cap \Omega_{\text{exp}} = \emptyset} 2 \int_0^1 |F_u(v) - v| dv = 2 \int_0^1 |1 - v| dv$ or $2 \int_0^1 |0 - v| dv = 1$. This confirms that the metric effectively quantifies the worst-cases discrepancy. $\square$

4. Invariance to Physical Variable Permutation

Proposition 4 (Invariance to Physical Variable Permutation). Let $\mathbf{Y} \in \mathbb{R}^{L \times V}$ be the response matrix. The U-PCDM metric is invariant under arbitrary variable permutation. Formally, for any transformation $\mathbf{Y}' = \mathbf{Y}\mathbf{P}$, where $\mathbf{P}$ is a $V \times V$ permutation matrix, the condition $D_{\text{U-PCDM}}(\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{exp}}) = D_{\text{U-PCDM}}(\mathcal{D}'_{\text{sim}}, \mathcal{D}'_{\text{exp}})$ holds.

Proof of Proposition 4. Let $\bar{\mathbf{Y}}$ be the standardized joint data matrix. The permuted matrix is $\bar{\mathbf{Y}}' = \bar{\mathbf{Y}}\mathbf{P}$. The covariance matrix $\mathbf{C}'$ is derived as

$$\mathbf{C}' = \frac{1}{N-1} (\bar{\mathbf{Y}}\mathbf{P})^\top (\bar{\mathbf{Y}}\mathbf{P}) = \mathbf{P}^\top \left(\frac{1}{N-1} \bar{\mathbf{Y}}^\top \bar{\mathbf{Y}} \right) \mathbf{P} = \mathbf{P}^\top \mathbf{C} \mathbf{P}.$$

Since $\mathbf{P}$ is orthogonal, ($\mathbf{P}^\top = \mathbf{P}^{-1}$), $\mathbf{C}'$ and $\mathbf{C}$ are similar matrices. According to the spectral theorem, similar matrices share the same characteristic polynomial and, consequently, the same set of eigenvalues. Since the weighting factors r_k (Equation (8)) depend solely on the eigenvalues, and the principal component projection represents the same subspace (up to axis reordering which is aggregated by summation), the final scalar metric remains invariant. $\square$

It is worth noting that while Proposition 4 guarantees invariance from a strict linear algebra perspective, the practical implementation of the AE-FSV layer involves neural network training, which introduces inherent stochasticity (e.g., random weight initialization). Consequently, minor fluctuations in the final U-PCDM value may be observed across different training runs even under invariant transformations.

5. Temporal Permutation Sensitivity

Unlike traditional validation metrics that rely solely on statistical distribution information (and are thus permutation-invariant), the U-PCDM metric is sensitive to the sequential order of field points (stemming from the sequential feature extraction mechanism of the AE-FSV layer). Mathematically, for any $L \times L$ non-identity permutation matrix $\mathbf{Q}$, the metric generally satisfies

$$D_{\text{U-PCDM}} \left(\left\{ \mathbf{Q}\mathbf{Y}_{\text{sim}}^{(i)} \right\}_{i=1}^{N_{\text{sim}}}, \left\{ \mathbf{Q}\mathbf{Y}_{\text{exp}}^{(j)} \right\}_{j=1}^{N_{\text{exp}}} \right) \neq D_{\text{U-PCDM}} (\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{exp}}) \quad (26)$$

This sensitivity is a deliberate design choice based on the nature of the data. Admittedly, permutation invariance is a virtue when the validation scope is restricted to global statistical attributes (e.g., confirming the normality of observation noise), where the temporal sequence is irrelevant. However, for MVFR, the temporal topology encodes critical physical information such as causality and transient evolution. A metric satisfying temporal invariance would inherently fail to distinguish between a correct signal and its “temporally permuted” version, leading to false acceptance. Therefore, U-PCDM strictly penalizes discrepancies in dynamic evolution, preserving the physical integrity of the validation.

3. Case Studies

3.1. Numerical Cases

3.1.1. Validation of the AE-FSV on Transient Data

Traditional FSV methods tend to yield distorted results when analyzing transient response data. To validate the enhanced performance of the AE-FSV method on such problems, this paper employs three representative transient response benchmark cases (Case 1, Case 2, Case 3) derived from Reference [33], as shown in Figure 5. These three datasets cover non-stationary characteristics of transient response data such as abrupt transients and high-frequency fluctuations, which can effectively verify the robustness of the AE-FSV method.

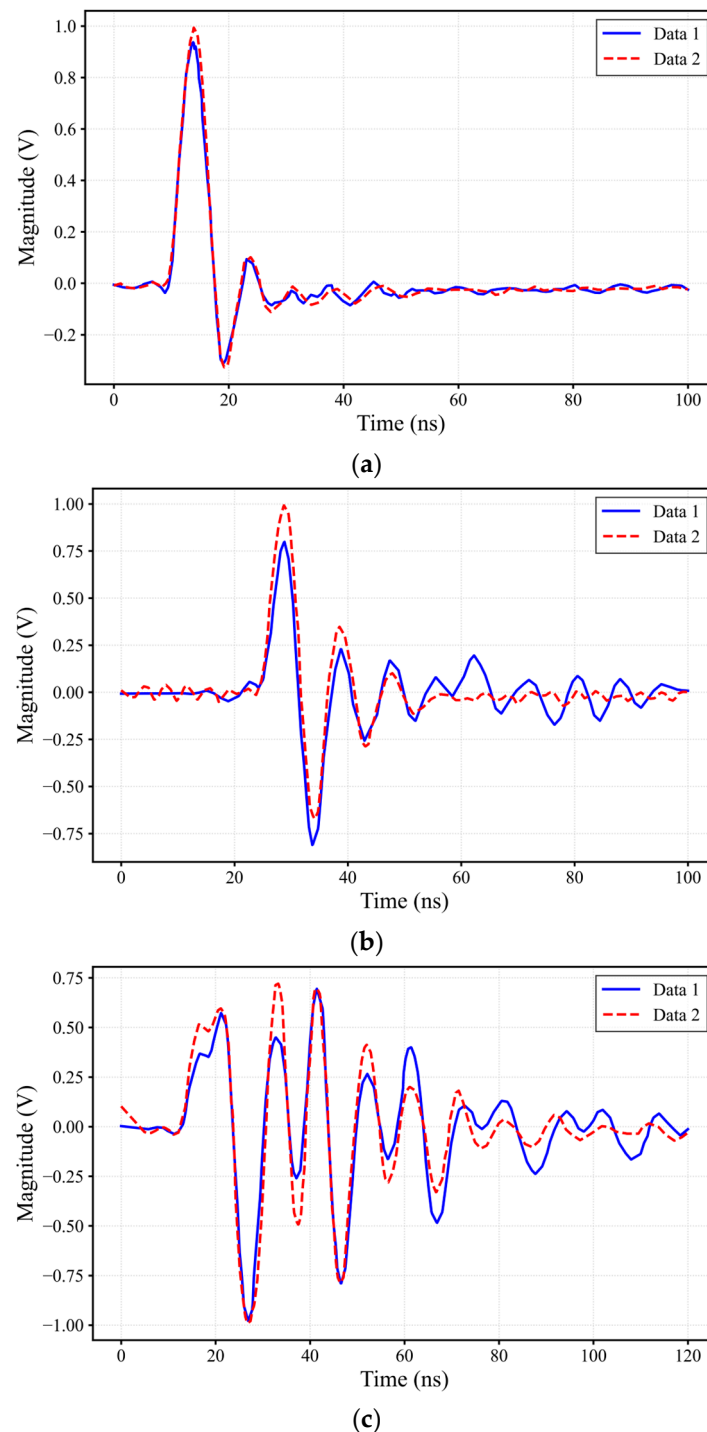


Figure 5. Three representative transient response benchmark cases cited from Reference [33]: (a) Case 1; (b) Case 2; (c) Case 3.

In each case, “Data 1” is treated as experimental data, while “Data 2” is regarded as the simulation data to be evaluated. To train the neural network of AE-FSV, amplitude fluctuations and random noise were added to the original data to construct a synthetic simulation dataset, which was used as the training set. Figure 6 presents the time-frequency analysis of the signal decomposition results of the experimental data in Case 1 obtained by AE-FSV. In the time domain (Figure 6a), the extracted trend feature accurately captures the dominant transient dynamic characteristics, including the steep rising edge, main peak, and subsequent damped oscillation. The detail feature fluctuates with low amplitude around 0, which is visually similar to random noise. In the spectral analysis

(Figure 6b), the Power Spectral Density (PSD) of the trend feature almost overlaps with that of the original signal at low frequencies (0–0.05 Hz), confirming that the AE preserves the core physical energy. At higher frequency bands (>0.2 Hz), the energy of the trend feature attenuates naturally, while the detail feature exhibits a broadband energy distribution without obvious harmonic peaks. This spectral characteristic confirms that the detail feature is mainly composed of random noise without physical regularity or incoherent high-frequency errors, rather than dynamically meaningful features with physical significance. This verifies the rationality of signal decomposition by the AE.

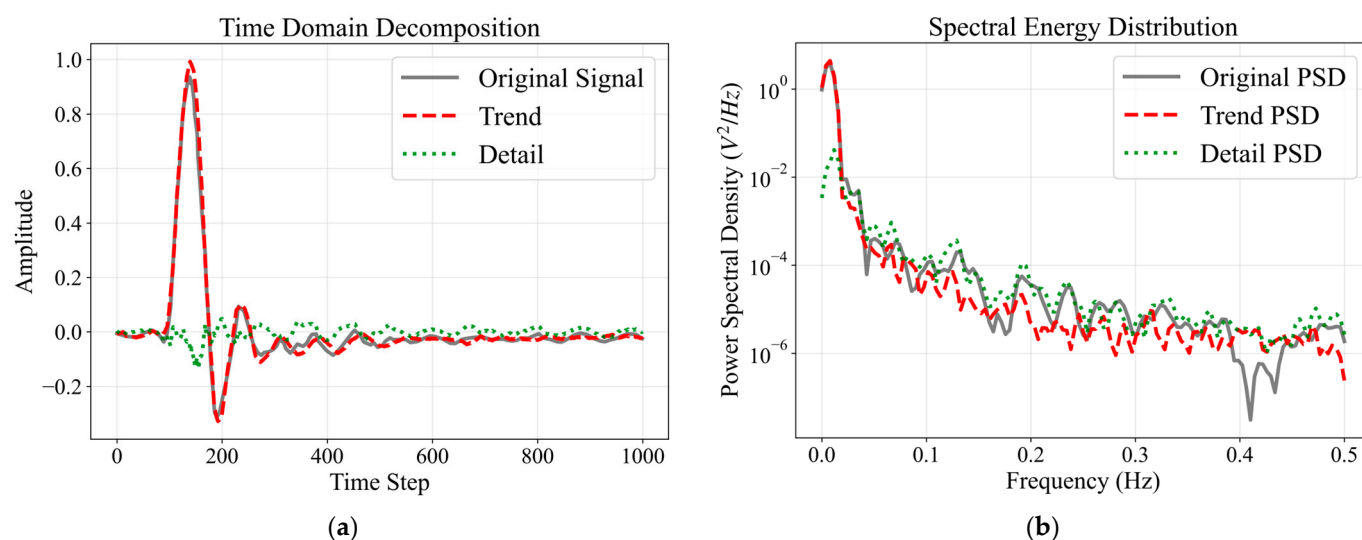


Figure 6. Time-frequency analysis of the signal decomposition results of the experimental data in Case 1 obtained by AE-FSV. **(a)** Time domain: The extracted “Trend” feature accurately captures the dominant transient dynamics (e.g., the main peak), while the “Detail” feature consists of residual fluctuations. **(b)** Spectral domain: The “Trend” feature retains the core low-frequency physical energy (<0.2 Hz), whereas the “Detail” feature shows a broadband distribution characteristic of noise.

Table 4 compares the evaluation results of these datasets across industry experts, traditional FSV, Dynamic Time Warping (DTW) [34,35], and the proposed AE-FSV. Among them, DTW, as a classical time-series similarity measurement method, is also included in the comparison. It can be seen that, in all three cases, the evaluation results of the AE-FSV method are highly consistent with those of industry experts, while the traditional FSV method exhibits significant deviations from expert consensus in Case 2 and Case 3. Meanwhile, the results of DTW suggest that the curve consistency of Case 3 is superior to that of Case 2, which also contradicts the conclusions of industry experts. This indicates that the AE-FSV method exhibits superior robustness when dealing with transient response problems where the traditional FSV and DTW fail.

Table 4. Comparison of evaluation results with expert assessment as the ground truth. The proposed AE-FSV aligns consistently with expert consensus across all transient cases, whereas Traditional FSV and DTW exhibit significant deviations in Cases 2 and 3.

Case	Expert Evaluation ¹	Traditional FSV ¹	DTW	AE-FSV
1	Very Good	Very Good ²	0.227	0.138 (Very Good)
2	Good	Fair	1.628	0.256 (Good)
3	Very Good	Fair	1.983	0.181 (Very Good)

¹ Data source: Reference [33]. ² According to the IEEE Std 1597.1 standard [26], FSV metric values have clear grade classifications: “Excellent” (<0.1), “Very Good” (0.1–0.2), “Good” (0.2–0.4), “Fair” (0.4–0.8), “Poor” (0.8–1.6), and “Very Poor” (>1.6).

3.1.2. Hyperparameter Sensitivity Analysis of U-PCDM

To evaluate the robustness of the U-PCDM metric, a sensitivity analysis of four key hyperparameters is conducted in this subsection. First, an MVFR system with five coupled physical variables ($V = 5$) and 201 time steps ($L = 201$) was constructed, as shown in Equation (27).

$$\begin{cases} y_1 = x_1 e^{-0.2t} \cdot \cos(x_2 t) + \varepsilon \\ y_2 = x_3 \sin(2t) \cdot [1 + 0.5 \cos(x_4 t)] + \varepsilon \\ y_3 = 0.2x_1 t + 0.3 \sin(10x_4 t) + \varepsilon \\ y_4 = \frac{x_2}{1 + e^{-2(t-5)}} + 0.1x_3 \cos(0.5t) + \varepsilon \\ y_5 = (x_1 + x_3) \sin t + (x_2 + x_4) \cos(3t) \end{cases} \quad (27)$$

where responses $y_1(t)$, $y_2(t)$, $y_3(t)$, $y_4(t)$ and $y_5(t)$ are all functions of time, with a domain of $t \in [0, 10]$ and a sampling step of $\Delta t = 0.05$. Model parameters x_1 , x_2 , x_3 and x_4 are subject to random uncertainties and follow specific probability distributions. The ε represents observation noises, governed by a Gaussian distribution with zero mean and a standard deviation of σ_ε .

Two sets of test models (P1 and P2) with model parameters following different probability distributions were designed (as shown in Table 5). Among them, P1 is used to generate synthetic experimental data and P2 is the model to be validated.

Table 5. Uncertainty information of model parameters in Model P1 and P2.

Model No.	x_1	x_2	x_3	x_4
P1	$N(2, 0.1^2)$	$N(5, 0.2^2)$	$N(1, 0.05^2)$	$N(0.5, 0.02^2)$
P2	$N(1.9, 0.1^2)$	$N(5.1, 0.2^2)$	$N(0.95, 0.05^2)$	$N(0.52, 0.02^2)$

Figures 7–10 show the hyperparameter sensitivity analysis of U-PCDM under four different scenarios. These scenarios vary in terms of sample size, noise level, and dimensionality (controlled by selecting specific response variables). The analysis focuses on four hyperparameters: the cumulative contribution rate threshold η , the bottleneck layer ratio α_b , the number of training epochs E_{train} and KDE bandwidth scaling factor α_{KDE} .

The analysis yields the following insights:

- **Cumulative Contribution Rate Threshold (η):** It can be seen from panels (a) of Figures 7–10 that, whether in the scenarios of large samples or small samples, large noise or small noise, or high-dimensional or low-dimensional responses, U-PCDM shows a step-like change with increasing η and converges around $\eta = 0.9$. This indicates that when $\eta = 0.9$, sufficient information can be retained while reducing the data dimension and computational load.
- **Bottleneck Layer Ratio (α_b):** It can be seen from panels (b) of Figures 7–10 that, in all cases, U-PCDM remains basically stable when α_b is less than 0.45. Therefore, setting α_b to 0.25 is reasonable.
- **Number of Training Epochs (E_{train}):** It can be seen from panels (c) of Figures 7–10 that U-PCDM is greatly affected by E_{train} . When E_{train} is less than 300, U-PCDM decreases or increases with the increase of E_{train} and, at this time, the AE is in an

underfitting state, leading to an incorrect underestimation of the discrepancy between simulation and experimental data by U-PCDM. When E_{train} is greater than 300, the rate of change slows down, and U-PCDM reaches a phased stability around $E_{\text{train}} = 500$. When E_{train} is greater than 550, U-PCDM increases slightly, and there is a risk of overfitting at this time. Consequently, taking $E_{\text{train}} = 500$ is a relatively reasonable choice.

- KDE Bandwidth Scaling Factor (α_{KDE}): It can be seen from panels (d) of Figures 7–10 that, in all cases, the results of U-PCDM are insensitive to variations in α_{KDE} .

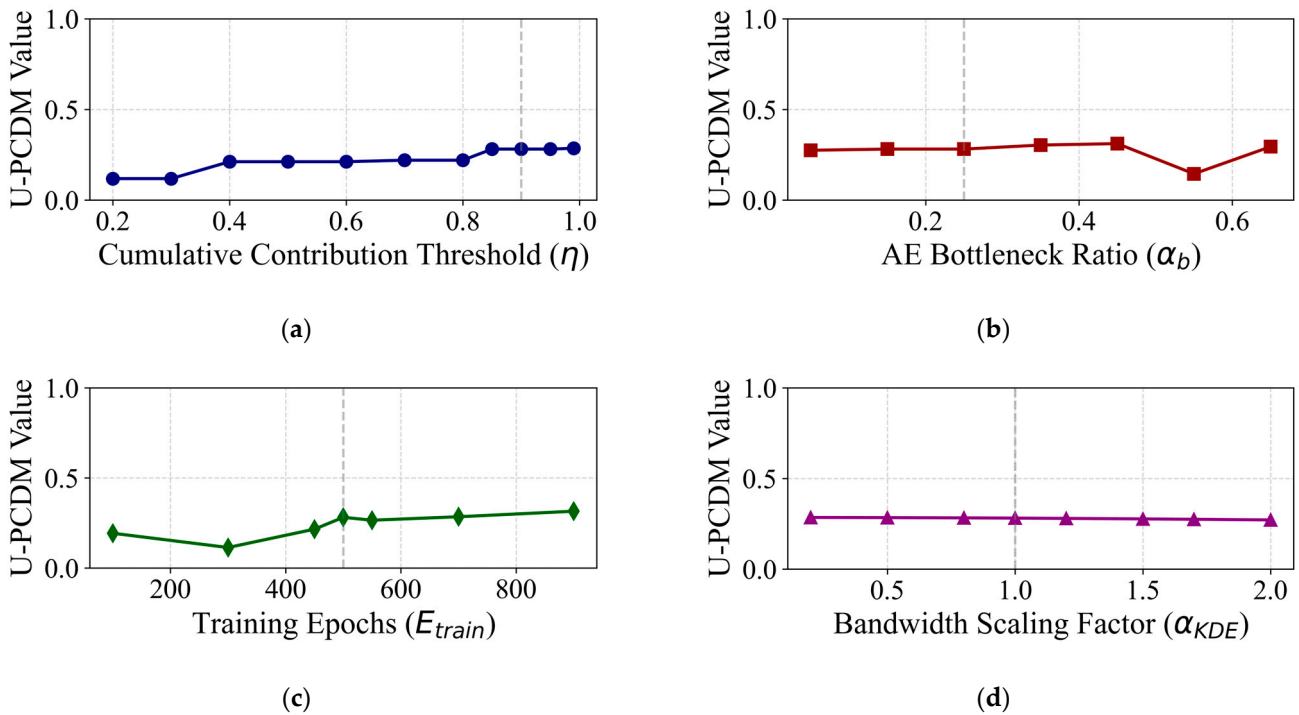
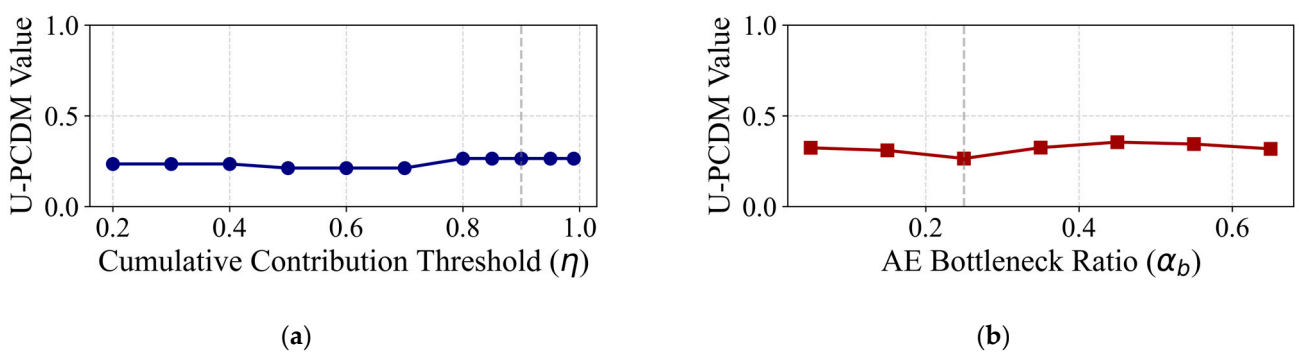


Figure 7. Hyperparameter sensitivity analysis of U-PCDM: scenario of large samples, small noise, and high dimensions ($N_{\text{sim}} = 2000$, $N_{\text{exp}} = 200$, $\sigma_{\varepsilon} = 0.001$, $V = 5$): (a) cumulative contribution rate threshold η ; (b) bottleneck layer ratio α_b ; (c) number of training epochs E_{train} ; (d) KDE bandwidth scaling factor α_{KDE} .



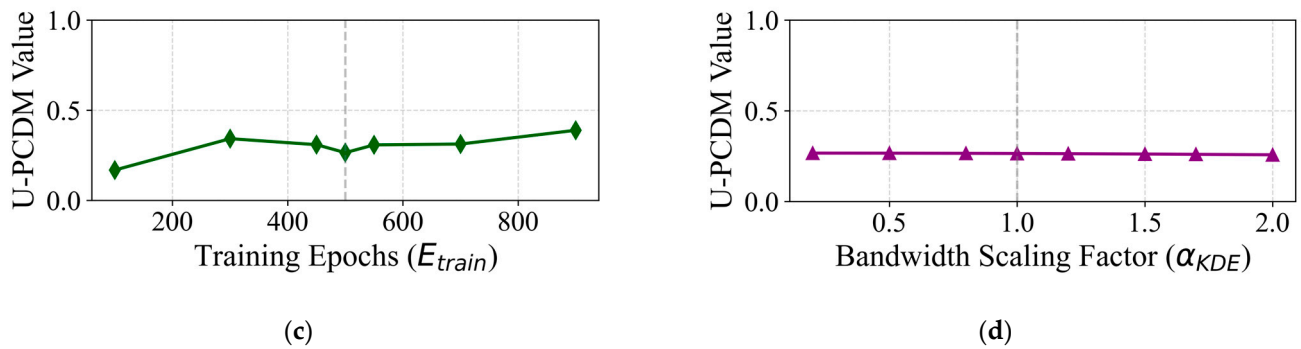


Figure 8. Hyperparameter sensitivity analysis of U-PCDM: scenario of large samples, small noise, and low dimensions ($N_{sim} = 2000$, $N_{exp} = 200$, $\sigma_\varepsilon = 0.001$, $V = 3$): (a) cumulative contribution rate threshold η ; (b) bottleneck layer ratio α_b ; (c) number of training epochs E_{train} ; (d) KDE bandwidth scaling factor α_{KDE} .

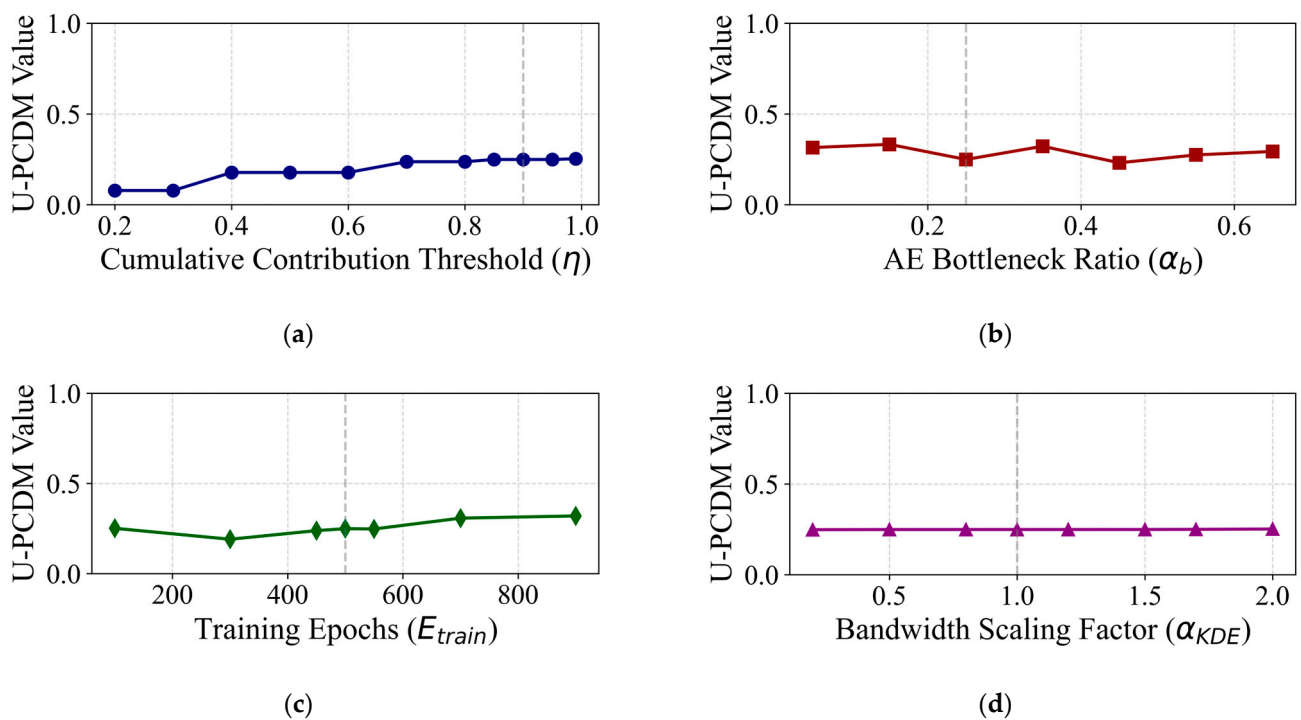
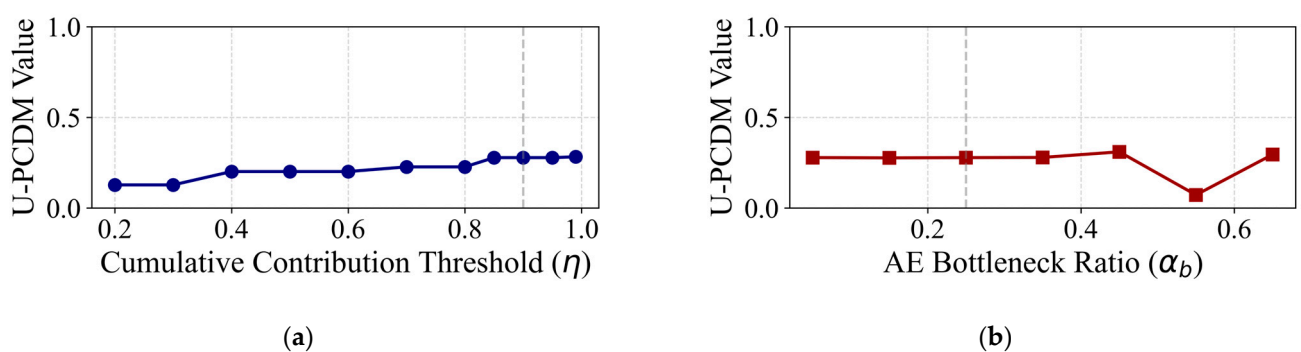


Figure 9. Hyperparameter sensitivity analysis of U-PCDM: scenario of small samples, small noise, and high dimensions ($N_{sim} = 200$, $N_{exp} = 20$, $\sigma_\varepsilon = 0.001$, $V = 5$): (a) cumulative contribution rate threshold η ; (b) bottleneck layer ratio α_b ; (c) number of training epochs E_{train} ; (d) KDE bandwidth scaling factor α_{KDE} .



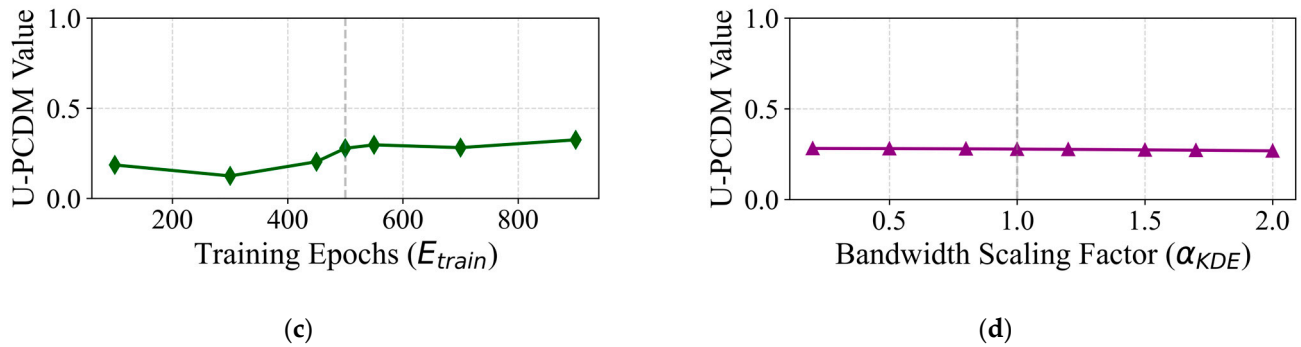


Figure 10. Hyperparameter sensitivity analysis of U-PCDM: scenario of large samples, large noise, and high dimensions ($N_{sim} = 2000$, $N_{exp} = 200$, $\sigma_{\varepsilon} = 0.005$, $V = 5$): (a) cumulative contribution rate threshold η ; (b) bottleneck layer ratio α_b ; (c) number of training epochs E_{train} ; (d) KDE bandwidth scaling factor α_{KDE} .

3.1.3. Effectiveness of U-PCDM in Numerical Examples

This section establishes an MVFR system, as defined in Equation (28):

$$\begin{cases} y_6(t) = x_5 \sin 2t + x_6 \cos 3t + \varepsilon \\ y_7(t) = x_5 \cos 2t + x_6 \sin 3t + 0.5x_6 + \varepsilon \end{cases} \quad (28)$$

where responses $y_6(t)$ and $y_7(t)$ are both functions of time, with a domain of $t \in [0, 10]$ and a sampling step of $\Delta t = 0.1$. Model parameters x_5 and x_6 are subject to random uncertainties and follow specific probability distributions. The ε represents observation noises, governed by a Gaussian distribution with zero mean and a standard deviation of $\sigma_{\varepsilon} = 0.001$.

To evaluate the effectiveness of the U-PCDM metric, six sets of test models (M1–M6) with model parameters following different probability distributions were designed. Specifically, M1 serves as the benchmark model, and its responses are used to generate synthetic experimental data for calculating the validation metrics of each model. During the validation process, the sample sizes of the experimental and simulation data are set to 200 and 2000, respectively. The point cloud visualization of the sample distribution for M5 versus the experimental data is illustrated in Figure 11.

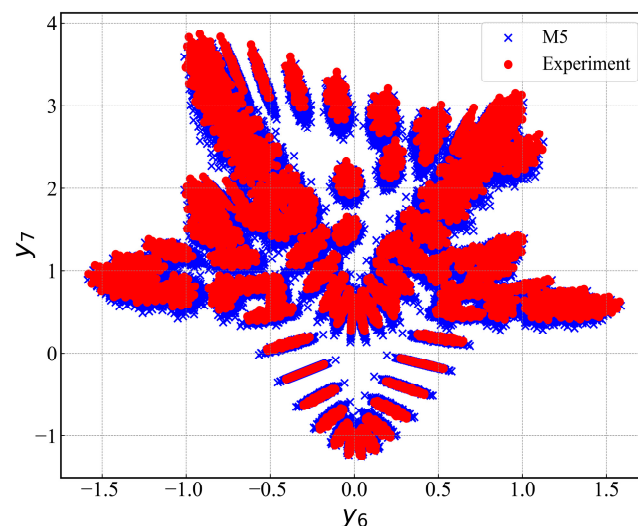


Figure 11. Point cloud visualization of the sample distribution for Model M5 and experimental data.

Table 6 details the parameter distributions of each model and compares the performance of the proposed U-PCDM against MD-pooling [10], COPMV [13], the Energy-Distance-based metric (ED) [12] and a variant of U-PCDM where the AE-FSV layer is replaced by DTW (denoted as U-PCDM-DTW).

As shown in Table 6, MD-pooling, COPMV and the proposed U-PCDM method yield the same ranking of the six models: $M1 < M2 < M5 < M4 < M3 < M6$. This consistency with the established MD-pooling benchmark confirms the accuracy of the U-PCDM metric in validating MVFR models. However, ED and the U-PCDM-DTW method provide divergent rankings, which are $M2 < M1 < M5 < M3 < M4 < M6$ and $M1 < M2 < M5 < M3 < M4 < M6$, respectively. Since $M1$ is the correct model, the conclusion from the ED results that $M2$ is superior to $M1$ is incorrect. It can also be observed that the ED and U-PCDM-DTW methods disagree with the other methods in the ranking of $M3$ and $M4$. From the perspective of ranking consistency with the benchmark method MD-pooling, the effectiveness of ED and the U-PCDM-DTW in this MVFR system case appears limited.

From the average computation time of the two methods, the time consumption of the U-PCDM is approximately 9% of that of the MD-pooling and COPMV method and approximately 20% of that of the ED and U-PCDM-DTW, demonstrating a significant improvement in computational efficiency. This marked reduction in runtime is primarily attributed to the efficient dimensionality reduction achieved by the JPCA layer and the compact feature extraction of the autoencoder within the AE-DLFE framework.

Table 6. Comparison of parameter distributions and metric results across different models ($V = 2$, $L = 101$, $N_{\text{sim}} = 2000$, $N_{\text{exp}} = 200$, $\sigma_{\varepsilon} = 0.001$). The proposed U-PCDM yields a model ranking identical to the benchmark MD-pooling method and shows superior efficiency.

Model No.	Model Parameters	MD-Pooling (Normalized to [0, 1])	COPMV	ED	U-PCDM-DTW	U-PCDM (Proposed)
M1	$\begin{cases} x_1 \sim N(1, 0.05^2) \\ x_2 \sim N(0.5, 0.02^2) \end{cases}$	0.042	0.013	0.0022	0.038	0.041 (Excellent)
M2	$\begin{cases} x_1 \sim N(1, 0.053^2) \\ x_2 \sim N(0.5, 0.022^2) \end{cases}$	0.063	0.017	0.0020	0.043	0.044 (Excellent)
M3	$\begin{cases} x_1 \sim N(1, 0.07^2) \\ x_2 \sim N(0.5, 0.03^2) \end{cases}$	0.358	0.093	0.017	0.285	0.348 (Fair)
M4	$\begin{cases} x_1 \sim N(0.96, 0.05^2) \\ x_2 \sim N(0.48, 0.02^2) \end{cases}$	0.326	0.063	0.165	0.317	0.293 (Fair)
M5	$\begin{cases} x_1 \sim N(0.96, 0.053^2) \\ x_2 \sim N(0.48, 0.022^2) \end{cases}$	0.240	0.039	0.152	0.244	0.291 (Fair)
M6	$\begin{cases} x_1 \sim N(0.9, 0.05^2) \\ x_2 \sim N(0.4, 0.02^2) \end{cases}$	0.991	1.521	0.602	0.990	0.935 (Poor)
Average computation time		66.93 s	66.19 s	29.07 s	29.91 s	5.86 s

Furthermore, to further study the sensitivity of the validation metrics to data structure, both block-wise time-series shuffling (as illustrated in Figure 12) and variable-order shuffling operations were performed on the data from Model M5 and the experimental data. The results are presented in Table 7. As can be seen, the values of the three metrics (MD-pooling, COPMV, and ED) remain unchanged under both perturbations. This

confirms that they primarily rely on the statistical distribution of the data and are insensitive to the evolutionary logic of the time series, which is a significant limitation for assessing temporal data. On the other hand, both the U-PCDM-DTW and U-PCDM metrics exhibited changes in their values after time-series shuffling. This temporal sensitivity indicates their capability to capture time-structure information. The value of U-PCDM-DTW remains unchanged after variable-order shuffling, demonstrating the permutation invariance property of the JPCA layer. Although theoretically, U-PCDM shares the same variable-order permutation invariance as U-PCDM-DTW, its value changed after variable-order shuffling in this instance. This minor discrepancy is attributed to the stochasticity inherent in the AE training process, rather than a structural flaw in the metric itself.

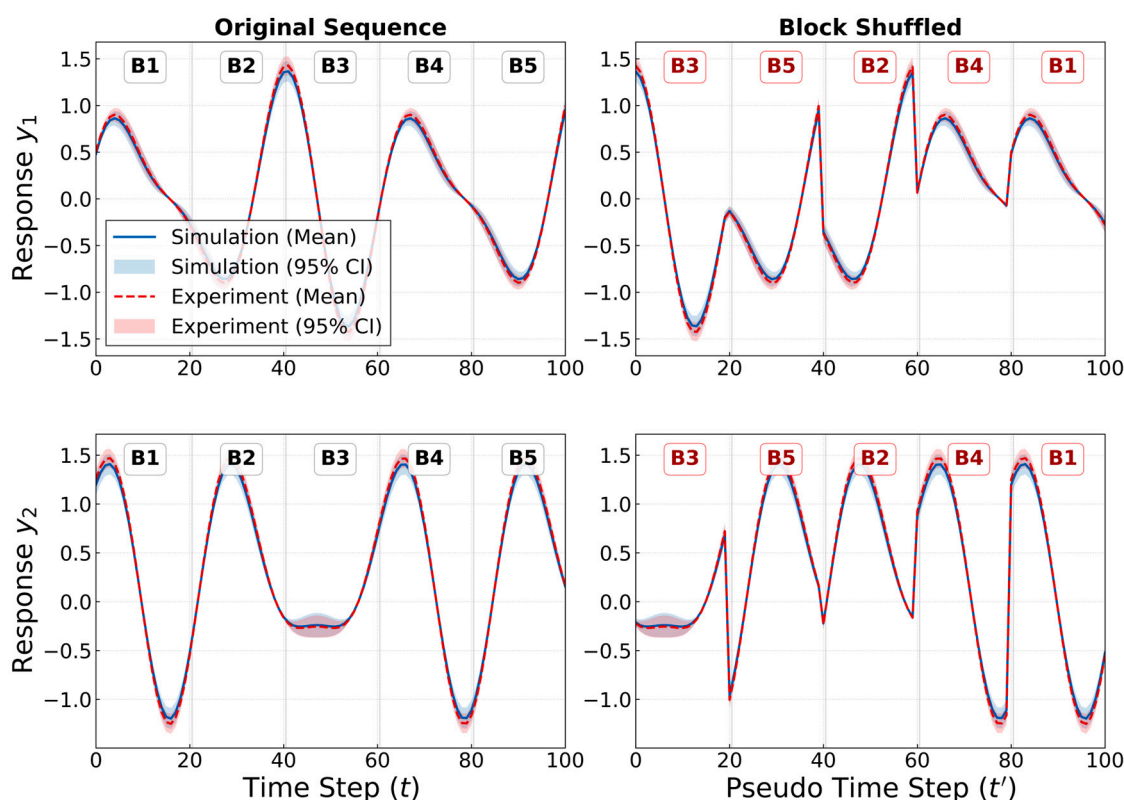


Figure 12. Comparison of samples for M5 and experimental data before and after block-wise time series shuffling.

Table 7. Comparison of metric values before and after time-series and variable-order permutations. Note that U-PCDM exhibits sensitivity to temporal shuffling (changing from 0.291 to 0.271), demonstrating its capability to capture dynamic evolution. In contrast, benchmark metrics like MD-pooling remain invariant, indicating an inability to detect temporal structural discrepancies.

	MD-Pooling	COPMV	ED	U-PCDM-DTW	U-PCDM (Proposed)
Original Value	0.240	0.039	0.152	0.244	0.291
After Temporal Shuffling	0.240	0.039	0.152	0.234	0.271
After Variable Shuffling	0.240	0.039	0.152	0.244	0.307

3.1.4. Study on the Small-Sample Stability of U-PCDM

To study the impact of random uncertainty introduced by limited experimental data on validation results, a convergence study was conducted using M5. Figure 13 depicts the distribution of validation results of the U-PCDM, MD-pooling, COPMV, ED and U-PCDM-DTW methods under different experimental sample sizes. The boxplots generated

by 200 independent Monte Carlo repeated trials for each sample size scenario intuitively reflect the statistical dispersion of the metrics.

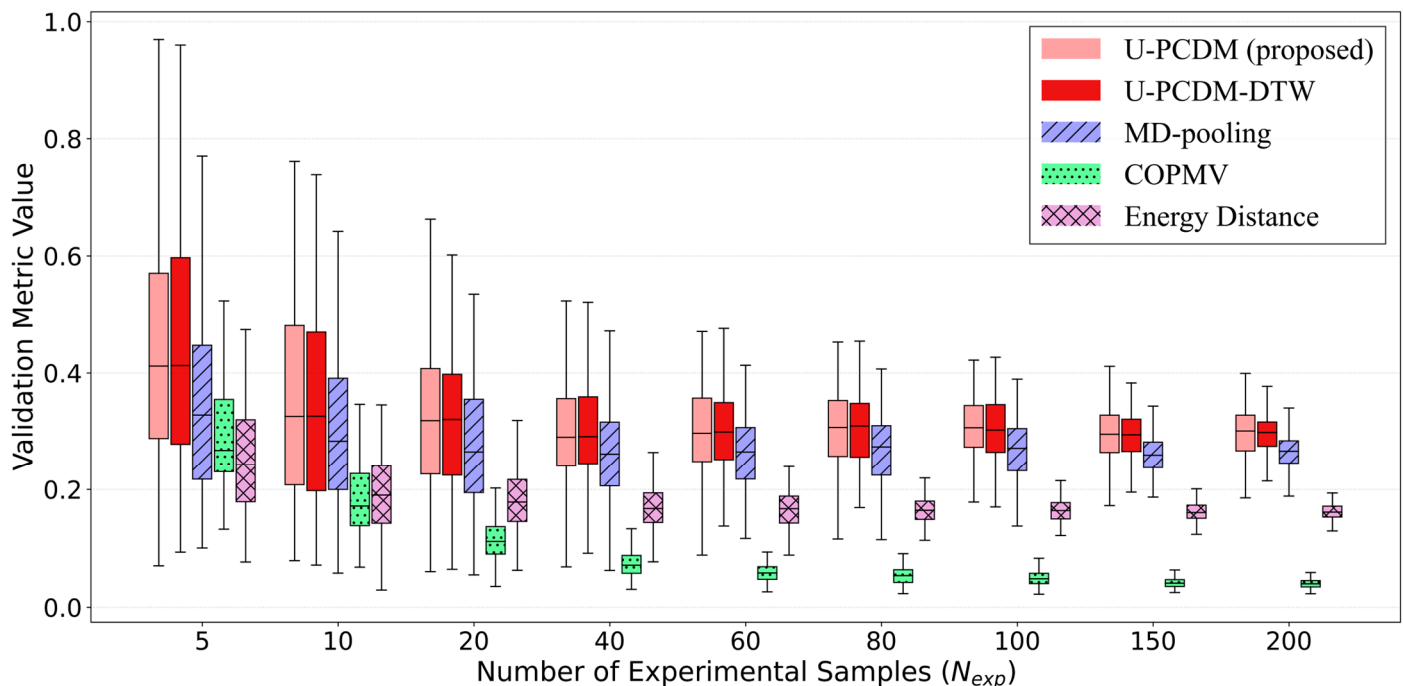


Figure 13. Distribution of metrics results under different experimental sample sizes. The diminishing dispersion (box height) indicates that U-PCDM achieves convergence stability comparable to the benchmark MD-pooling when experimental samples reach at least 40.

From the results, when experimental data is scarce (fewer than 20 samples), all metrics exhibit large interquartile ranges (high box heights) and significant fluctuations, indicating that sampling uncertainty dominates the validation results rather than the model errors. When the experimental sample size exceeds 40, the dispersion for all metrics diminishes significantly and converges to a stable state. Therefore, for the model validation of MVFR studied in this paper, it is recommended that the experimental sample size be at least 40 to obtain robust validation results. Admittedly, the dispersion of the U-PCDM is slightly higher than that of other methods in the small-sample regime; however, this is a reasonable trade-off for its superior computational efficiency and enhanced capability in identifying temporal structural discrepancies, without compromising stability once the sample size sufficiency is met.

3.1.5. Computational Complexity Analysis

To rigorously assess computational efficiency, an asymptotic complexity analysis is performed for U-PCDM and the benchmark MD-pooling method. The analysis considers the following key parameters: the number of simulation samples N_{sim} , experimental samples N_{exp} , field length L , number of physical variables V , and the number of retained principal components in the JPCA layer m (where $m \leq V$).

The computational cost of U-PCDM is governed by its dual-layer feature extraction architecture, with the primary load concentrated in the second layer (Autoencoder training). Layer 1 (JPCA) involves computing the covariance matrix of the joint dataset and performing eigen-decomposition. While the theoretical complexity is $O((N_{\text{sim}} + N_{\text{exp}})LV^2)$, the computational overhead is negligible in typical engineering contexts where the number of variables V is significantly smaller than the field length L . Layer 2 (AE-FSV) is the computational bottleneck. For the m retained principal

component channels, m independent autoencoders are trained. Given the specific network architecture, the asymptotic complexity for this phase is $O(m \cdot N_{\text{sim}} \cdot E_{\text{train}} \cdot L)$. A distinct advantage of U-PCDM is that the Kernel Density Estimation (KDE) is performed only once on the final aggregated scalar metric, adding a minimal complexity term of $O(N_{\text{sim}} \log N_{\text{sim}})$. Consequently, the total complexity of U-PCDM scales linearly with the field length L , and the computational cost is driven more by N_{sim} than by L .

In contrast, the MD-pooling method requires calculating the Mahalanobis distance and performing KDE independently at each of the L time points. The asymptotic complexity includes terms for distance calculation, $O((N_{\text{sim}} + N_{\text{exp}}) \cdot L \cdot V^2)$, and, crucially, repetitive statistical estimation, $O(L \cdot N_{\text{sim}} \log N_{\text{sim}})$. As a result, the computational burden amplifies significantly when both the field length and sample size are large.

Table 8 presents the average computation time across different complexity scenarios (where $V = 2$ corresponds to the model in Section 3.1.3, and $V = 5$ corresponds to Section 3.1.2). Comparing Case 1 and Case 4, it can be observed that the computation times for MD-pooling, COPMV, ED, and U-PCDM decrease as the sample size drops. This indicates that, for these methods, computational cost is more sensitive to N_{sim} than to L . Comparing Case 1 and Case 5, the computation times for MD-pooling, COPMV, ED, and U-PCDM show a linear increase, whereas U-PCDM-DTW remains computationally prohibitive.

Across all the high-dimensional scenarios (Case 1, Case 3 and Case 5), U-PCDM (proposed) consistently maintains the lowest computation time with minimal fluctuation. This demonstrates that U-PCDM is the most robust against increases in system complexity, achieving optimal computational efficiency for high-dimensional multivariate functional responses.

Table 8. Average computation time comparison under different complexity scenarios. Notably, U-PCDM consistently maintains the lowest computation time across high-dimensional scenarios (Cases 1, 3, and 5), demonstrating its superior computational efficiency and scalability compared to the benchmark MD-pooling.

	MD-Pooling	COPMV	ED	U-PCDM-DTW	U-PCDM (Proposed)
Case 1: $\begin{cases} V = 2, L = 101, \\ N_{\text{sim}} = 2000, N_{\text{exp}} = 200 \end{cases}$	66.93 s	66.19 s	29.07 s	29.91 s	5.86 s
Case 2: $\begin{cases} V = 2, L = 101, \\ N_{\text{sim}} = 200, N_{\text{exp}} = 20 \end{cases}$	1.04 s	1.04 s	0.33 s	3.05 s	2.07 s
Case 3: $\begin{cases} V = 2, L = 1001, \\ N_{\text{sim}} = 2000, N_{\text{exp}} = 200 \end{cases}$	677.95 s	675.63 s	45.91 s	2914.17 s	28.57 s
Case 4: $\begin{cases} V = 2, L = 1001, \\ N_{\text{sim}} = 200, N_{\text{exp}} = 20 \end{cases}$	10.25 s	10.26 s	0.45 s	295.03 s	4.14 s
Case 5: $\begin{cases} V = 5, L = 101, \\ N_{\text{sim}} = 2000, N_{\text{exp}} = 200 \end{cases}$	103.55 s	103.39 s	40.14 s	240.53 s	14.95 s

3.2. Complex Model Validation Case: NASA Almond RCS Model

This section applies the proposed U-PCDM metric to a complex model validation scenario in the field of computational electromagnetics. The validation object is the NASA Almond model [36] for Radar Cross-Section (RCS) simulation, whose geometric structure is shown in Figure 14.

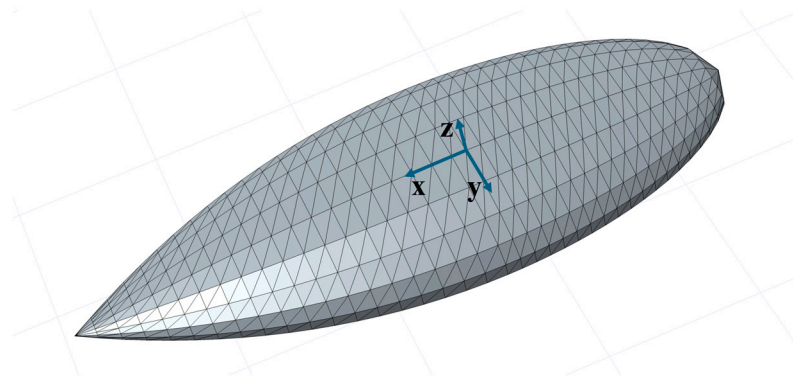


Figure 14. Geometric structure of the NASA Almond model.

The model response is an MVFR with two variables and a field length of 901 sampling points. Regarding the specific simulation settings, the operating frequency is fixed at 7 GHz. The excitation source is defined as a plane wave incident in the xy plane (elevation angle $\theta = 90^\circ$). To capture comprehensive scattering characteristics, the simulation employs dual-polarization modes—Horizontal–Horizontal (HH) and Vertical–Vertical (VV). Consequently, for each polarization, the RCS response curves are obtained by scanning the azimuth angle φ from 0° to 180° with a step size of 0.2° .

In this case study, the material parameters of the model—relative permittivity ϵ_r , relative permeability μ_r , and coating thickness h —are treated as uncertainty variables. To evaluate the performance of the validation metric, two sets of models were constructed, as detailed in Table 9. Among them, the “Accurate Model” is used to generate synthetic experimental data and the “Biased Model” is regarded as the simulation model to be validated. Simulations were performed using the solver of Altair Feko 2022 (Altair Engineering Inc., Troy, MI, USA), yielding 24 response samples of the accurate model and 256 response samples of the biased model, as shown in Figure 15.

Table 9. Uncertainty information of model parameters in the RCS case.

Model Name	ϵ_r	μ_r	h (mm)
Accurate Model	U (19.00, 23.00)	U (1.00, 1.20)	U (0.60, 0.70)
Biased Model	U (16.34, 24.51)	U (0.86, 1.29)	U (0.40, 0.70)

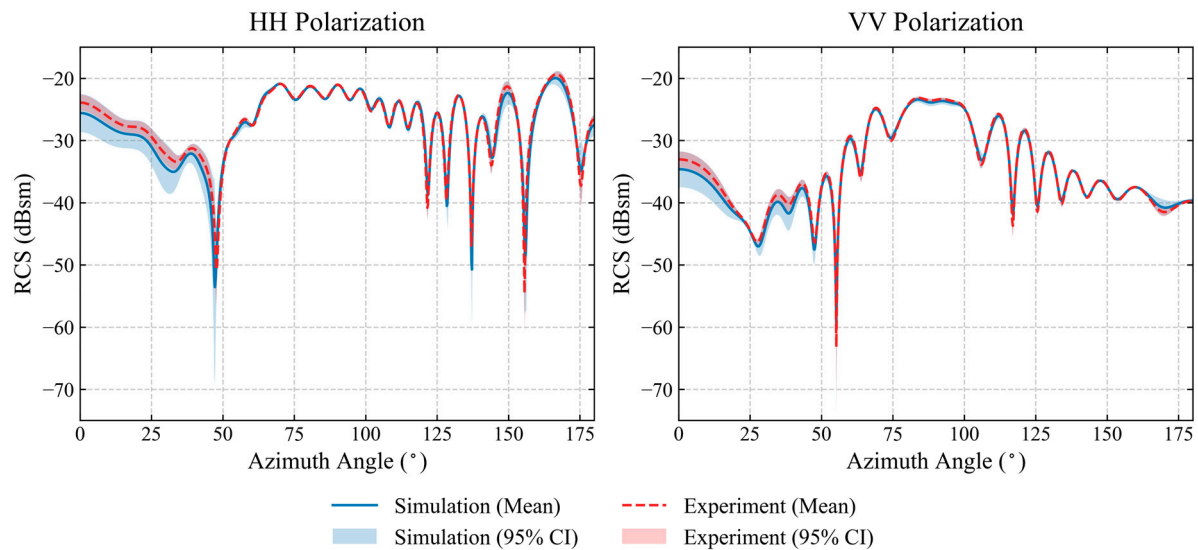


Figure 15. RCS response samples of the accurate model and biased model.

The validation results are listed in Table 10. It can be seen that the calculation result of the proposed U-PCDM metric is 0.340, which is highly consistent with that of the benchmark method MD-pooling (0.317). According to the decision criterion defined in Table 3, this result belongs to the “Fair” grade, indicating that the biased model has significant deviations and should not be directly accepted. Instead, model calibration should be performed first before re-validation. In terms of computation time, the calculation time of the proposed U-PCDM metric is approximately 5.88 s, only 48% of that of MD-pooling (12.32 s). This result further corroborates the conclusion drawn in the numerical examples: the proposed method significantly enhances computational efficiency while maintaining high validation accuracy in complex engineering applications.

Table 10. Comparison of validation results for the NASA Almond RCS case. Note that the U-PCDM metric yields a value highly consistent with the benchmark MD-pooling (0.340 vs. 0.317), confirming its accuracy. Moreover, it reduces the computation time by approximately 52% (5.88 s vs. 12.32 s), demonstrating its efficiency for complex engineering applications.

	MD-Pooling (Normalized)	U-PCDM
Validation result	0.317	0.340 (Fair)
Computation time	12.32 s	5.88 s

4. Discussion

In this paper, we have systematically explored model validation methods for MVFR. A critical insight identified is that for complex simulation systems with MVFR outputs, model validation faces the challenge of coupled physical and field correlations. The proposed AE-DLFE method and the U-PCDM validation metric effectively address these issues. This section further elaborates on the theoretical properties, performance advantages, and methodological characteristics of the proposed framework.

4.1. Superiority of AE-FSV

For functional responses, traditional FSV methods usually rely on FFT and fixed filters for signal decomposition. As shown in the transient case of AE-FSV (Table 4), when processing non-stationary signals containing abrupt transients or high-frequency fluctuations, traditional methods often produce results inconsistent with expert judgments. Although some researchers have introduced theories such as Wavelet Transform [29] and

Reassigned Smoothed Pseudo Wigner-Ville Distribution (SPWVD) [28] for improvements and achieved certain results, their performance is highly dependent on expert prior knowledge for hyperparameter selection, lacking generalization ability across different physical domains.

In contrast, the proposed AE-FSV method constructs a data-driven nonlinear filter using an autoencoder. By capturing latent physical dynamics from the simulation data itself, the autoencoder can adaptively separate trend features and detail features. As demonstrated in Table 4, this intelligent algorithm enables the AE-FSV method to overcome the limitations of traditional methods, adaptively apply to various physical problems, and provide evaluation results highly consistent with experts' cognition, thereby significantly enhancing the robustness of feature extraction.

4.2. Necessity of Dual-Layer Decoupling

For MVFR, a core finding of this study is the necessity of processing physical and field correlations separately. Existing validation metrics often conflate these correlations or neglect one entirely. As shown in the time-series and variable-order permutations experiment (Table 7), the benchmark MD-pooling method, along with COPMV and ED, is permutation-invariant because these methods primarily rely on global statistical distributions. However, from the perspective of model validation, the temporal evolution of responses carries critical physical information. The proposed U-PCDM metric effectively identifies the temporal structural discrepancies caused by shuffling, demonstrating its sensitivity to field correlations.

Furthermore, distinct from ensemble learning approaches (e.g., [37]), which aggregate multiple homogeneous models to enhance prediction accuracy, the U-PCDM framework integrates distinct functional modules—JPCA, AE-FSV, and u-pooling—to systematically address the dual-coupling challenge. This integration is designed based on the mathematical properties of each component, ensuring that both multivariate physical correlations and temporal field discrepancies are decoupled and quantified under uncertainty. Therefore, such a hierarchical combination is not merely an algorithmic aggregation, but also a structural necessity for valid MVFR assessment.

4.3. Computational Efficiency for High-Dimensional Data

In complex engineering scenarios involving uncertainty, validation metrics are frequently embedded within iterative loops for model calibration or optimization, making computational cost a decisive constraint. The U-PCDM numerical case and the NASA Almond RCS case demonstrate that the computation time of the U-PCDM metric is significantly reduced compared to the benchmark MD-pooling method. The asymptotic complexity analysis (Section 3.1.5) provides a theoretical basis for this improvement. While methods like U-PCDM-DTW and MD-pooling scale drastically with field length and sample size, U-PCDM maintains a linear scaling relationship.

Specifically, in the high-dimensional scenario (Case 1, Case 3 and Case 5), U-PCDM (proposed) consistently maintains the lowest computation time. This efficiency improvement is attributed, on the one hand, to the dimensionality reduction capability of PCA, i.e., by setting a threshold, the number of principal component channels is minimized while ensuring sufficient information. On the other hand, it benefits from the dimensionality reduction capability of the autoencoder, i.e., compressing high-dimensional functional data into compact feature representations before curve discrepancy evaluation and uncertainty quantification.

4.4. Noise Suppression Capability

The calculation process of U-PCDM relies primarily on the AE-DLFE framework. On the one hand, in the JPCA layer of AE-DLFE, minor feature vectors are truncated through the cumulative contribution threshold, which reduces the influence of non-dominant noise components to a certain extent. On the other hand, in the AE-FSV layer of AE-DLFE, the autoencoder isolates high-frequency random fluctuations as part of the detail feature, and mitigates the influence of noise to a certain extent by weighting the impact of the detail feature via Equations (17)–(21). The consistent performance of U-PCDM across all noise-injected numerical examples (Table 6) empirically verifies this capability.

4.5. Conservativeness of U-PCDM

In the calculation of U-PCDM, the entire simulation dataset is used not only for training, but also for establishing the reference distribution S_{sim} . Theoretically, this might introduce a bias in the simulation representation. However, as observed in both numerical and engineering case studies, such bias is acceptable. Therefore, U-PCDM can be regarded as a conservative upper bound of the model error. From a safety-critical engineering perspective, a metric that tends to overestimate (rather than underestimate) discrepancies is preferable, as it reduces the risk of accepting a flawed model (Type II error).

4.6. Limitations and Future Work

Although the proposed method has significant advantages, certain limitations remain:

1. **Data Dependency:** The performance of U-PCDM is partially contingent upon the quality and quantity of simulation data used to train the neural network during dual-layer feature extraction, and its stability is also limited by the number of experimental samples. Although the convergence study (Figure 13) shows that the metric tends to stabilize with a sufficient number of experimental samples, transfer learning techniques may need to be introduced in the future for scenarios with extremely sparse data.
2. **Training Strategy:** The training strategy for the autoencoder employed in this study is relatively basic. Currently, a standard reconstruction loss and fixed architectural parameters are used. While this configuration has proven effective for the cases presented, there is potential for improvement. Future research could investigate advanced training protocols, such as automated hyperparameter optimization or the adoption of Variational Autoencoders (VAEs), to further enhance the robustness of feature extraction for highly complex nonlinear systems.

Furthermore, future work will focus on extending this method to higher-dimensional spatiotemporal fields or image problems, and exploring the integration of Bayesian neural networks to further quantify the uncertainty of the feature extraction process itself.

5. Conclusions

To address the challenges of coupled heterogeneous correlations in multivariate functional response systems, this paper proposes an autoencoder-based dual-layer feature extraction method. On this basis, a new validation metric (U-PCDM) and a decision criterion are established. The main contributions and conclusions of this paper are summarized as follows:

1. **Proposal of the AE-DLFE method:** This method effectively decouples heterogeneous physical and field correlations. The first layer (JPCA) addresses physical correlations and achieves dimensionality reduction. For field correlations (the second layer), the

proposed AE-FSV method adaptively extracts the trend and detail features of functional responses and evaluates discrepancies. It demonstrates superior robustness in handling transient signals compared to traditional methods.

2. Construction of the U-PCDM metric: Based on the AE-DLFE method, a novel validation metric (U-PCDM) for multivariate functional responses is proposed. Integrating the u-pooling method, this metric enables a comprehensive evaluation of the distributional discrepancy between simulation and experimental data. Compared with existing methods, U-PCDM not only has a more rigorous mathematical foundation and physical interpretability (specifically regarding temporal structure sensitivity), but also achieves an improvement in computational efficiency on high-dimensional data, providing an efficient and robust tool for model validation under uncertain environments.
3. Establishment of a Decision Criterion based on U-PCDM results: A standardized decision criterion based on U-PCDM results is established. Validated through the computational electromagnetics (NASA Almond RCS) case, the framework successfully assigns a “Fair” grade to the biased model and provides a clear decision path for model acceptance or rejection. The results demonstrate that the method has the potential to support real-time or near-real-time decision-making in engineering applications.

Author Contributions: Conceptualization, D.W. and B.Z.; methodology, D.W.; software, D.W. and D.S.; validation, X.Z., D.S. and J.L.; formal analysis, D.W. and H.L.; investigation, D.W. and H.L.; resources, D.W.; data curation, X.Z.; writing—original draft preparation, D.W.; writing—review and editing, B.Z.; visualization, X.Z. and D.S.; supervision, B.Z.; project administration, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.18218839> (accessed on 10 February 2026). The source code is available on request from the corresponding author due to intellectual property restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Battula, S.; Alla, S.N.; Ramana, E.V.; Kumar, N.K.; Murthy, S.B. Uncertainty Quantification for Digital Twins in Smart Manufacturing and Robotics: A Review. *J. Phys. Conf. Ser.* **2024**, *2837*, 012059. <https://doi.org/10.1088/1742-6596/2837/1/012059>.
2. Xiao, W.; Zhao, J.; Lv, L.; Chen, J.; Zhang, P.; Wu, X. UQ4CFD: An Uncertainty Quantification Platform for CFD Simulation. *Aerospace* **2025**, *12*, 886. <https://doi.org/10.3390/aerospace12100886>.
3. Chen, J.T.; Xiao, W.; Zhao, W.; Zhang, P.H.; Yang, F.J.; Jin, T.; Guo, Y.Y.; Wu, X.J.; Chen, J.Q.; Wang, R.L.; et al. Advances in verification and validation in computational fluid dynamics. *Adv. Mech.* **2023**, *53*, 626–660 <https://doi.org/10.6052/1000-0992-23-012>.
4. Ferson, S.; Oberkampf, W.L. Validation of Imprecise Probability Models. *Int. J. Reliab. Saf.* **2009**, *3*, 3–22. <https://doi.org/10.1504/ijrs.2009.026832>.
5. Song, X.; Sun, D.; Liang, X. Optimal Design of Validation Experiment for Material Deterioration. *Materials* **2023**, *16*, 5854. <https://doi.org/10.3390/ma16175854>.
6. Ferson, S.; Oberkampf, W.L.; Ginzburg, L.R. Model Validation and Predictive Capability for the Thermal Challenge Problem. *Comput. Methods Appl. Mech. Eng.* **2008**, *197*, 2408–2430. <https://doi.org/10.1016/j.cma.2007.07.030>.
7. Shaukat, U.; Jakobsen, J.B.; Ikegaya, N.; Giljarhus, K.E.T. Validation of Large Eddy Simulations in Urban Wind Studies Using a New Overall Area Metric. *Build. Environ.* **2025**, *267*, 112285. <https://doi.org/10.1016/j.buildenv.2024.112285>.
8. Li, W.; Chen, W.; Jiang, Z.; Lu, Z.; Liu, Y. New Validation Metrics for Models with Multiple Correlated Responses. *Reliab. Eng. Syst. Saf.* **2014**, *127*, 1–11. <https://doi.org/10.1016/j.res.2014.02.002>.
9. Li, L.; Lu, Z. A new method for model validation with multivariate output. *Reliab. Eng. Syst. Saf.* **2018**, *169*, 124–132.

<https://doi.org/10.3390/math14040674>

10. Zhao, L.; Lu, Z.; Yun, W.; Wang, W. Validation Metric Based on Mahalanobis Distance for Models with Multiple Correlated Responses. *Reliab. Eng. Syst. Saf.* **2016**, *159*, 80–89. <https://doi.org/10.1016/j.res.2016.10.016>.
11. White, A.; Mahadevan, S.; Schmucker, J.; Karl, A. Multi-Metric Validation under Uncertainty for Multivariate Model Outputs and Limited Measurements. *J. Verif. Valid. Uncertain. Quantif.* **2022**, *7*, 041004. <https://doi.org/10.1115/1.4056548>.
12. Zhao, J.; Xiao, W.; Chen, J.; Shen, Y.; Lv, L. Validation Metric of Multi-Output Model Based on Energy Distance. *J. Phys. Conf. Ser.* **2023**, *2599*, 012043. <https://doi.org/10.1088/1742-6596/2599/1/012043>.
13. Hu, J.; Jiang, P.; Zhou, Q.; McKeand, A.; Choi, S.-K. Model Validation Methods for Multiple Correlated Responses via Covariance-Overlap Based Distance. *J. Mech. Des.* **2019**, *142*, 041401. <https://doi.org/10.1115/1.4044330>.
14. Hu, W.; Zhang, T.; Liu, Z.; Cheng, J.; Lou, Y.; Deng, X.; Tan, J. A New Validation Metric for Models with Correlated Responses Using Limited Experimental and Simulation Data. *J. Mech. Des.* **2022**, *144*, 111702. <https://doi.org/10.1115/1.4054877>.
15. Nolan, T.H.; Richardson, S.; Ruffieux, H. Efficient Bayesian Functional Principal Component Analysis of Irregularly-Observed Multivariate Curves. *Comput. Stat. Data Anal.* **2025**, *203*, 108094. <https://doi.org/10.1016/j.csda.2024.108094>.
16. Bayarri, M.J.; Berger, J.O.; Paulo, R.; Sacks, J.; Cafeo, J.A.; Cavendish, J.; Lin, C.-H.; Tu, J. A Framework for Validation of Computer Models. *Technometrics* **2007**, *49*, 138–154. <https://doi.org/10.1198/004017007000000092>.
17. Bayarri, M.J.; Walsh, D.; Berger, J.O.; Cafeo, J.; Garcia-Donato, G.; Liu, F.; Palomo, J.; Parthasarathy, R.J.; Paulo, R.; Sacks, J. Computer Model Validation with Functional Output. *Ann. Stat.* **2007**, *35*, 1874–1906. <https://doi.org/10.1214/009053607000000163>.
18. Duffy, A.P.; Martin, A.J.M.; Orlandi, A.; Antonini, G.; Benson, T.M.; Woolfson, M.S. Feature Selective Validation (FSV) for Validation of Computational Electromagnetics (CEM). Part I—The FSV Method. *IEEE Trans. Electromagn. Compat.* **2006**, *48*, 449–459. <https://doi.org/10.1109/temc.2006.879358>.
19. Wu, Y.Q.; Zhou, H.J.; Bao, X.-F.; Li, H.-Y.; Yin, W.-Y. The application of the NIPC method and the FSV method on the validation of EM simulation under HIRF environment. In *2017 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI)*; IEEE: Washington, DC, USA, 2017; pp. 561–566. <https://doi.org/10.1109/ISEMC.2017.8077932>.
20. Bongiorno, J.; Mariscotti, A. Uncertainty and sensitivity of the feature selective validation (FSV) method. *Electronics* **2022**, *11*, 2532.
21. Bai, J.; Li, X.; Niu, X. Application of the FSV method in EMC uncertainty simulation results. *IEEE Lett. Electromagn. Compat. Pract. Appl.* **2023**, *5*, 122–126.
22. Happ, C.; Greven, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.* **2018**, *113*, 649–659.
23. Rigueira, X.; Olivieri, D.; Araujo, M.; Saavedra, A.; Pazo, M. Multivariate functional data analysis and machine learning methods for anomaly detection in water quality sensor data. *Environ. Model. Softw.* **2025**, *190*, 106443.
24. Hasan, B.M.S.; Abdulazeez, A.M. A review of principal component analysis algorithm for dimensionality reduction. *J. Soft Comput. Data Min.* **2021**, *2*, 20–30.
25. Bharadiya, J.P. A Tutorial on Principal Component Analysis for Dimensionality Reduction in Machine Learning. *Int. J. Innov. Sci. Res. Technol.* **2023**, *8*, 2028–2032. <https://doi.org/10.5281/zenodo.8020810>.
26. IEEE 1597.1-2022; IEEE Standard For validation of Computational Electromagnetics Computer Modeling and Simulations. IEEE Standard Association: Piscataway, NJ, USA, 2022; pp. 1–52.
27. Jauregui, R.; Zhang, G.; Rojas-Mora, J.; Ventosa, O.; Silva, F.; Duffy, A.P.; Sasse, H. Analyzing Transient Phenomena in the Time Domain Using the Feature Selective Validation (FSV) Method. *IEEE Trans. Electromagn. Compat.* **2014**, *56*, 825–834. <https://doi.org/10.1109/temc.2013.2246167>.
28. Shi, J.; Hu, J.; Gao, Z.; Song, L.; Zhao, H.; Zhao, R.; Chen, W.; Zhang, M. Research on RCS data validation based on SPWVD. *Prog. Electromagn. Res. C* **2024**, *144*, 147–158.
29. Niu, X.; Liu, S.; Qiu, R.; Chow, X. Research on improved FSV method based on wavelet function transform. *IEEE Lett. Electromagn. Compat. Pract. Appl.* **2024**, *6*, 111–116.
30. Bourlard, H.; Kabil, S.H. Autoencoders reloaded. *Biol. Cybern.* **2022**, *116*, 389–406.
31. Nurmaini, S.; Darmawahyuni, A.; Sakti Mukti, A.N.; Rachmatullah, M.N.; Firdaus, F.; Tutuko, B. Deep Learning-Based Stacked Denoising and Autoencoder for ECG Heartbeat Classification. *Electronics* **2020**, *9*, 135. <https://doi.org/10.3390/electronics9010135>.
32. Suo, B.; Qi, Y.; Sun, K.; Xu, J. A Novel Model Validation Method Based on Area Metric Disagreement between Accelerated Storage Distributions and Natural Storage Data. *Mathematics* **2023**, *11*, 2511. <https://doi.org/10.3390/math11112511>.

33. Jauregui Tellería, R.; Rojas, J.; Silva Martínez, F. Study of Transient Phenomena with Feature Selective Validation Method. In *Progress in Electromagnetics Research Symposium Proceedings*; PIERS: Marrakesh, Morocco, 2011; pp. 1113–1117. Available online: <https://hdl.handle.net/2117/14101> (accessed on 10 February 2026).
34. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>.
35. Bagnall, A.; Lines, J.; Bostrom, A.; Large, J.; Keogh, E. The great time series classification bake off: A review and experimental evaluation of re-cent algorithmic advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660. <https://doi.org/10.1007/s10618-016-0483-9>.
36. Woo, A.C. Benchmark radar targets for the validation of computational electromagnetics programs. *IEEE Antennas Propag. Mag.* **1993**, *35*, 86.
37. Khan, W.A.; Ma, H.-L.; Chung, S.-H.; Wen, X. Hierarchical Integrated Machine Learning Model for Predicting Flight Departure Delays and Duration in Series. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103225. <https://doi.org/10.1016/j.trc.2021.103225>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.