



# A Novel Electromagnetic Spectrum Prediction Model Based upon Multi-Dimensional Feature Fusion

Yu Han<sup>1</sup>, Jiayi Song<sup>1</sup>, Zhiqi Wang<sup>1</sup>, Wei Xiang<sup>2</sup>, P. Takis Mathiopoulos<sup>3</sup>, Guan Gui<sup>4,\*</sup> and Yun Lin<sup>1,\*</sup>

<sup>1</sup> College of Information and Communication Engineering, Harbin Engineering University, Harbin 150000, China

<sup>2</sup> School of Engineering and Mathematic Sciences, La Trobe University, Melbourne VIC 3086, Australia

<sup>3</sup> Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens 15784, Greece

<sup>4</sup> School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

## Abstract

In the era of increasingly scarce spectrum resources, electromagnetic spectrum (EMS) prediction has emerged as a critical means for enhancing spectrum utilization efficiency. However, most of the existing EMS methods primarily exploit low-dimensional features such as temporal, frequency, or spatial characteristics in an individual fashion, which limits their ability to fully capture the inherent complexity of spectrum dynamics. To improve the performance, this paper proposes a novel EMS prediction model, which involving three operations, namely multi-dimensional decoupling, feature fusion and temporal prediction. Firstly, for multi-dimensional decoupling operation, we propose a Multi-dimensional Feature Extraction (MFE) module, which characterizes the complex temporal-frequency-spatial variations of EMS data

by leveraging both single-dimensional features and cross-dimension dependencies (i.e., temporal-frequency, temporal-spatial, and frequency-spatial relationships). By explicitly modeling these correlations, the MFE module enhances the prediction performance of the proposed model. Secondly, to reduce redundancy between these decoupled multi-path features, we introduce a Tensor-Feature-Fused (TF) module. Through a bidirectional cross-attention mechanism, the proposed TF module enables symmetric information exchange between multi-path features and the original spectrum data, by selectively integrating both inter-path features and intra-feature information. Finally, by employing a Temporal Convolutional Network (TCN), the data obtained by the TF module are processed to capture multi-scale dependencies so that the accuracy of spectrum prediction is enhanced. The performance of the proposed model, termed as MFE-TFTCN, has been extensively evaluated by means of computer simulations. Various experimental results obtained through the use of a publicly European multi-location dataset have demonstrated that, compared to state-of-the-art EMS prediction methods, the proposed model achieves superior prediction performance by effectively capturing temporal-frequency-spatial interdependencies.



Academic Editor:

Xinde Li

Submitted: 10 September 2025

Accepted: 09 November 2025

Published: 20 November 2025

Vol. 3, No. 1, 2026.

10.62762/CJIF.2025.747641

\*Corresponding authors:

✉ Guan Gui

guiguan@njupt.edu.cn

✉ Yun Lin

linyun@hrbeu.edu.cn

## Citation

Han, Y., Song, J., Wang, Z., Xiang, W., Mathiopoulos, P. T., Gui, G., & Lin, Y. (2025). A Novel Electromagnetic Spectrum Prediction Model Based upon Multi-Dimensional Feature Fusion. *Chinese Journal of Information Fusion*, 3(1), 1–16.



© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** electromagnetic spectrum prediction, multi-dimensional decoupling, bidirectional cross-attention mechanism, temporal convolutional network.

## 1 Introduction

With the continuous growth in user frequency demand, wireless communication technologies are developing and evolving rapidly. Consequently, the explosive increase in communication signals is intensifying the problem of limited and non-renewable EMS resources, so that spectrum scarcity is an ever growing concern. To deal with this problem, EMS prediction technology plays a crucial role as it forecasts future signal trends, warns of sudden spectrum changes, and assists in identifying spectrum holes. These capabilities are fundamental for improving spectrum utilization, optimizing frequency allocation strategies, and ensuring reliable communication under congested conditions. They also support the future development of intelligent, fully connected communication systems [1].

However, as spectrum collection and storage technologies improve, spectrum data has become increasingly high-dimensional. Research focusing solely on low-dimensional characteristics, such as temporal features, is no longer effective, and nowadays multi-dimensional spectrum prediction is receiving significant attention. To enhance prediction accuracy, models must go beyond individual dimensions like time, frequency, or space. In fact it is necessary to jointly consider and analyze high-dimensional interactions, especially in scenarios like 5G/6G IoT, vehicle networks, and smart cities where a very large number of devices operate in coordinated clusters.

Feature extraction from multi-dimensional data method can be categorized into i) Statistical methods; ii) Neural Networks(NNs)-based extraction methods; iii) Hybrid methods combining statistics and deep learning. Statistical methods are the earliest and simplest, offering strong interpretability. In general for all these methods, their main objective is to extract statistical features or map data into a new feature space using techniques like principal component analysis (PCA) [2] or change vector analysis (CVA) [3]. For instance, Chen et al. [4] have proposed a multi-dimensional feature analysis framework using time-domain hurst exponent, information entropy, frequency-domain variance-to-mean ratio, and short-time Fourier transform-based time-frequency representations. However, these methods are not effective in dealing with high-dimensional data and

often lack dynamic feature updating. On the other hand, NNs can deeply mine nonlinear relationships in high-dimensional data but usually lack interpretability and involve rigid module designs [5]. Ji et al. [6] have proposed a single encoder-multiple decoder structure to address complex task-specific features, which unfortunately lacks interpretability from a physical perspective. To balance interpretability and learning capacity, hybrid models combining statistical and deep learning methods have become popular [7]. These models aim to retain interpretability while capturing complex, nonlinear features. However, by converting to one-dimensional symbolic sequences, important information might be lost. Yan et al. [8] have addressed issues of non-stationarity and spatiotemporal correlation by decomposing target sequences into intrinsic mode functions (IMFs), then reconstructing them using fuzzy entropy. Yet, many domain-specific approaches often rely on the intrinsic physical properties of high-dimensional data, which limits their scalability and generalization.

EMS prediction is typically categorized by dimensionality into: one-dimensional time-series prediction, two-dimensional time-frequency prediction and higher-dimensional predictions [9]. As the number of dimensions increases, this leads to higher computational demands and a greater need for prior knowledge. Early studies on this issue have mainly focused on temporal variations of spectrum, neglecting other dimensional information, which results in the extracted features offered limited interpretability [10]. Earlier studies have considered traditional signal analysis and statistical models, such as deep Gaussian processes [11] and Holt's exponentially weighted moving average [12]. Deep learning models, like long short-term memory networks(LSTM) [13] and gated recurrent unit networks (GRU) [14], have shown improved performance in one-dimensional time-series prediction. Other models have considered statistical and physical features with deep learning for two-dimensional time-frequency prediction although they have often relied on frequency features, which have limited their scalability [15]. For example, Ren et al. [16] have used CNN-ResNet-based models with nearest-neighbor interpolation to fill sensing gaps and reduce sensing costs. Basak et al. [17] have considered the joint operation of CNN and LSTM for automated spectrum prediction, using spectrogram matrices and transfer learning to predict time-frequency sequences under unknown rates and patterns. Li et al. [18] have

used graph convolutional networks (GCNs) to extract hidden time-frequency knowledge. Nevertheless, as the electromagnetic environment is becoming more and more complex, researchers have turned their attention to spatial and other physical-domain features beyond time and frequency [19]. However, LSTM-based models still tend to overlook spatial propagation characteristics [20]. Although the models extract higher-level spatial features, they are limited by predefined physical structures and fixed dimensions [21]. To address this limitation, Li et al. [22] by using a 3<sup>rd</sup> Hankelized tensor, have proposed to model temporal-frequency-spatial data, framing spectrum prediction as a tensor completion task. However, such tensor completion methods requires high computational complexity. On the other hand, Deng et al. [23] have considered multi-way normalization on tensor time-series data to isolate heterogeneous low-dimensional substructures by using rather simple statistical models with limited feature extraction capability.

Motivated by the above, in this paper we present a novel highly-accurate and interpretable multi-dimensional prediction model for EMS. The proposed model fully exploits high-dimensional spectrum features and thoroughly explores cross-dimensional coupling interactions leading to a comprehensive understanding and precise prediction of spectrum dynamics. Furthermore, it performs multi-dimensional feature extraction (MFE), tensor-feature fusion (TF), and temporal-frequency-spatial prediction with temporal convolutional network (TCN) to support efficient spectrum utilization and coordinated allocation, and thus it will be termed as MFE-TFTCN. Within this framework, the main contributions of this paper can be summarized as follows:

- We propose a novel MFE module based on multi-dimensional decoupling. By modeling not only temporal, frequency, and spatial features but also their mutual correlations, the module exploits the latent regularities of EMS variations, thereby improving prediction performance.
- We introduce a temporal convolutional NN enhanced with a bidirectional cross-attention mechanism to improve the accuracy and robustness of spectrum prediction. This mechanism enables symmetric information exchange between features and original data, and assigns attention weights to both salient features

and intra-feature information. It effectively fuses parallel multi-path features and reduce redundancy among these decoupled features and learns multi-scale temporal dependencies of multi-dimensional sequences.

- Extensive simulations on a real-world multi-location dataset evaluated with multiple metrics, demonstrate that the proposed model outperforms state-of-the-art multi-dimensional prediction methods. Comparative and ablation studies further confirm its advantage in effectively capturing temporal–frequency–spatial interdependencies.

## 2 Problem Statement

As common in EMS studies, all elements of the multi-dimensional spectrum data utilized in this paper are power spectral density (PSD) functions having location, time, and frequency as parameters representing the ratio of the signal power to the resolution bandwidth measured by a spectrum analyzer at a specific location. Essentially these PSD functions reflect the electromagnetic energy present in a given environment with its value expressed in dBm. This paper adopts PSD rather than spectrum occupancy to describe spectrum states. The main reason for this is that, unlike occupancy data, which involves both signal energy and noise, PSD avoids errors introduced during noise extraction and occupancy calculation.

For a three-dimensional spectrum tensor  $\mathcal{X} \in \mathbb{R}^{T \times F \times S}$  with time, frequency, and space dimensions, each element  $\mathcal{X}_{t,f,s}$ ,  $t \in \{1, \dots, T\}$ ,  $f \in \{1, \dots, F\}$ ,  $s \in \{1, \dots, S\}$  represents the PSD value measured at time index  $t$ , frequency index  $f$ , and spatial location  $s$ . The description of the tensor and its subarrays is shown in Table 1.

With the advancement of spectrum collection and storage technologies, EMS data is evolving toward higher dimensionality. Accurate prediction of three-dimensional or even higher-dimensional spectrum data requires more observation perspectives and higher-quality multi-dimensional features, a problem which is considered in this paper.

## 3 Methodology

As illustrated in Figure 1, multi-dimensional EMS prediction involves a three-dimensional tensor  $\mathcal{X} \in \mathbb{R}^{T \times F \times S}$ , where  $T$  is the total number of time steps in the time dimension,  $F$  is the total bandwidth in

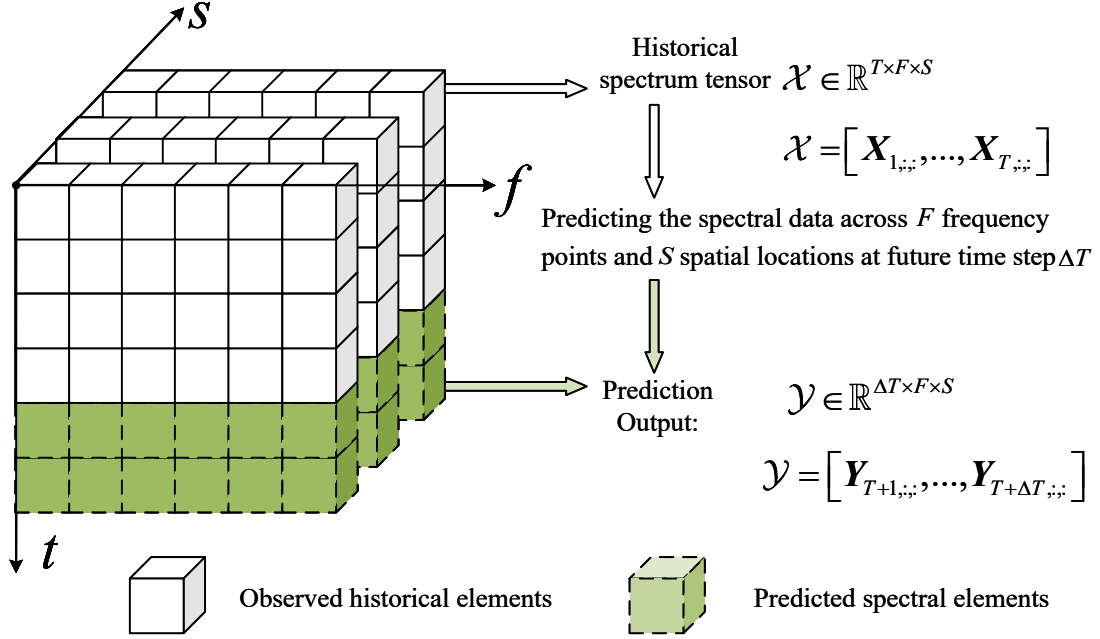


Figure 1. Illustration of the tensor-based data structure for multi-dimensional EMS prediction.

Table 1. Tensor and subarray representation.

| Symbol                                                      | Definition                                                                           |
|-------------------------------------------------------------|--------------------------------------------------------------------------------------|
| $x_{t,f,s}$                                                 | PSD value at time $t$ , frequency $f$ , and spatial locations $s$                    |
| $\mathbf{x}_{:,f,s}, \mathbf{x}_{t,:s}, \mathbf{x}_{t,f,:}$ | Tensor fiber with one varying dimension while holding the other two dimensions fixed |
| $\mathbf{x}_{:,f,s}, \mathbf{x}_{t,:s}, \mathbf{x}_{t,f,:}$ | Tensor slice with two varying dimensions while keeping one dimension fixed           |
| $\mathcal{X} \in \mathbb{R}^{T \times F \times S}$          | The three-dimensional spectrum tensor                                                |
| $\mathcal{Z} \in \mathbb{R}^{T \times F \times S \times H}$ | The potential four-dimensional representation of the spectrum tensor                 |

the frequency dimension, and  $S$  is the number of discrete spatial locations. By decomposing along the time axis, we obtain  $\mathcal{X} = [\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(T)}]$ , where each slice  $\mathbf{X}^{(t)}, t \in \{1, \dots, T\}$  is a signal power matrix across all frequencies and locations at time. Its evolution is not only time-dependent, but also

affected in a nonlinear manner by the other dimensions. This work aims to predict a future spectrum tensor  $\hat{\mathcal{Y}} = [\mathbf{Y}^{(T+1)}, \dots, \mathbf{Y}^{(T+\Delta T)}]$ , which represents the spectrum values at future steps across all frequency and spatial positions, as also illustrated in Figure 1.

### 3.1 System Model

This paper proposes the MFE-TFTCN model, which constructs a “multi-dimensional decoupling-feature fusion-temporal-frequency-spatial prediction” model. The proposed model integrates statistical methods with deep learning to enable high-precision, interpretable, and scalable multi-dimensional EMS prediction and its overall architecture of MFE-TFTCN is illustrated in Figure 2.

As illustrated in this architecture, firstly the input three-dimensional historical spectrum tensor  $\mathcal{X} \in \mathbb{R}^{T \times F \times S}$  is preprocessed using Z-Score normalization and a fully-connected (FC) layer to produce a four-dimensional latent representation  $\mathcal{Z} \in \mathbb{R}^{T \times F \times S \times H}$ , where  $H$  denotes the number of hidden channels. Next, the two MFE modules perform independent and joint feature extraction, respectively. They extract adaptive features that capture the independent influence of each dimension and the interactions across dimensions from  $\mathcal{Z}$ .

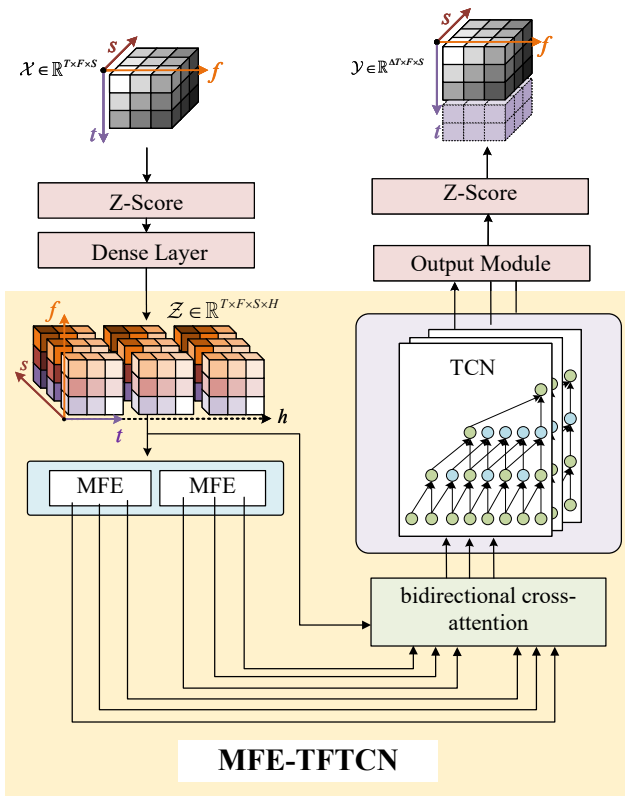


Figure 2. Block diagram of the MFE-TFTCN model.

These extracted features, together with  $Z$ , are passed through a bidirectional cross-attention module to enable symmetric information exchange. Then, TCN is used to learn multiscale temporal patterns. Finally, the output module, composed of skip connections and FC layers, transforms the feature dimensions to produce the predicted spectrum tensor  $\hat{Y} = [Y^{(T+1)}, \dots, Y^{(T+\Delta T)}]$ , which consists of  $\Delta T$  time steps and  $F \times S$  predicted elements.

### 3.2 Data Preprocessing

To reduce scale differences across the various dimensions of the spectrum data, we first apply Z-Score normalization to the original data, as follows:

$$\text{Z-Score}(x_{t,f,s}) = \frac{x_{t,f,s} - \mu}{\sigma}, \quad \begin{matrix} t=1, \dots, T; \\ f=1, \dots, F; \\ s=1, \dots, S \end{matrix} \quad (1)$$

where  $x_{t,f,s}$  is an element of the spectrum tensor,  $\mu$  is the sample mean, and  $\sigma$  is the sample standard deviation. To unify units and stabilize the data distribution, the commonly used global Z-Score normalization method is applied to the mean and standard deviation of all elements in the tensor:

$$\mu = \frac{1}{T \cdot F \cdot S} \sum_{s=1}^S \sum_{f=1}^F \sum_{t=1}^T x_{t,f,s}, \quad x_{t,f,s} \in \mathcal{X}, \quad (2)$$

$$\sigma = \sqrt{\frac{1}{T \cdot F \cdot S} \sum_{s=1}^S \sum_{f=1}^F \sum_{t=1}^T (x_{t,f,s} - \mu)^2}. \quad (3)$$

It is noted that the Z-Score normalization transforms the data through a linear operation to set the mean to 0 and the standard deviation to 1, which reduces the scale differences between multi-dimensional features. It also preserves the relative distances between samples, making it more aligned with statistical assumptions [24].

To avoid samples with insufficient statistics caused by too few values in the considered dimension, we employ a FC layer to map the normalized spectrum data into a four-dimensional space, resulting in a latent representation of the spectrum data, denoted as  $Z$ . Compared to the original tensor, a hidden channel dimension is added, which increases the number of elements within each tensor fiber and tensor slice, making the data better suited for subsequent statistical methods.

In the next two sections (3.3 and 3.4), we describe how the model deeply extracts and analyzes the statistical features of the data to better identify the variation patterns and trends across dimensions. It is noted that the elements in the latent representation are used for internal computation and learning within the network.

### 3.3 Multi-dimensional Feature Extraction

To improve the accuracy of temporal-frequency-spatial spectrum prediction, the model not only captures the trends of variation in temporal, frequency, and spatial dimensions, but also extracts high-dimensional features and the interactions between different dimensions in the spectrum data. Therefore, the complex multi-dimensional variations in spectrum data are decomposed into two parts: i) Independent changes within each dimension; and ii) Interactions across dimensions. To capture these two types of variations, the feature extraction module consists of an independent feature extraction module and a joint feature extraction module. The independent feature extraction module separately captures the independent variations in time, frequency, and space. The joint feature extraction module focuses on interactions between pairs of dimensions, including time-frequency, frequency-space, and time-space. Each module is composed of multiple homogeneous and parallel components. Each component is designed to extract either an independent variation within a single dimension or an interaction between two dimensions. Consequently, the multi-dimensional

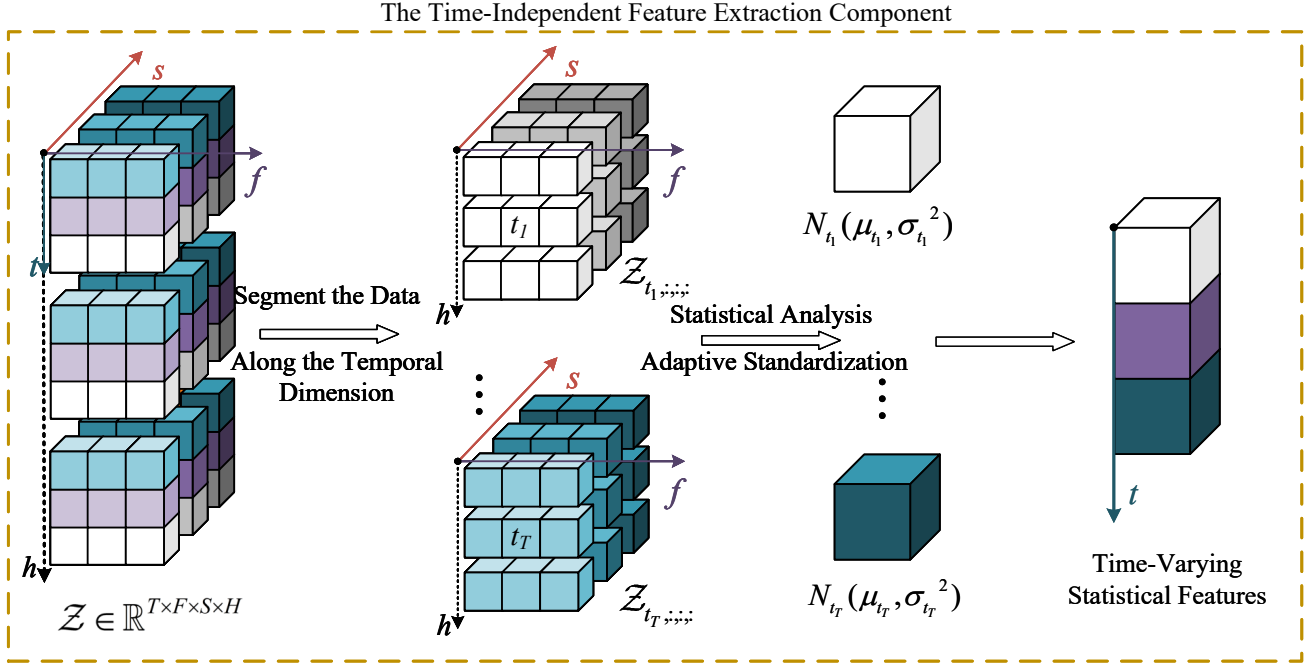


Figure 3. Time-independent feature extraction within the tensor structure of the multi-dimensional EMS.

feature extraction module can adapt to variations in both the number and types of input dimensions.

Any element in the normalized latent representation  $Z \in \mathbb{R}^{T \times F \times S \times H}$  of the spectrum data can be indexed using a four-tuple corresponding to time, frequency, space, and hidden channel. From a probabilistic statistical perspective, the variation of spectrum data along a specific dimension can be described by the joint distribution its elements. As an example, let us take the time-independent feature extraction component (TIFE-T) to illustrate the principle of extracting independent features in a specific dimension. Its operation within the context of the tensor structure of the multi-dimensional EMS is shown in Figure 3.

More specifically, the latent representation  $Z \in \mathbb{R}^{T \times F \times S \times H}$  is split along the time dimension into  $T$  three-dimensional sub-tensors  $Z^{(t_1)}, t_1 \in \{1, \dots, T\}$ , each containing  $F \times S \times H$  elements. Together, these sub-tensors represent the potential variation patterns of all frequencies, locations, and hidden channels at time step  $t_1$ . The collection  $\{z_{t_1, f, s, h} \mid f = 1, \dots, F; s = 1, \dots, S; h = 1, \dots, H\}$  is treated as a statistical sample for feature estimation:

$$\mu_{t_1} = \mathbb{E} [Z^{(t_1)}] \approx \frac{1}{F \cdot S \cdot H} \sum_{f=1}^F \sum_{s=1}^S \sum_{h=1}^H z_{t_1, f, s, h}, \quad (4)$$

$$\sigma_{t_1} = \sqrt{\mathbb{E} [(Z^{(t_1)} - \mu_{t_1})^2]} \approx \sqrt{\frac{1}{F \cdot S \cdot H} \sum_{f=1}^F \sum_{s=1}^S \sum_{h=1}^H (z_{t_1, f, s, h} - \mu_{t_1})^2}, \quad (5)$$

where  $\mu_{t_1}$  and  $\sigma_{t_1}$  represent the mean and standard deviation of  $Z^{(t_1)}$  respectively. The variations along the frequency and spatial dimensions are treated as samples for statistical estimation, and all elements involved in the computation belong to the same time step. In Eqs. 4 and 5, the mean and standard deviation reflect the average signal strength and fluctuation range at each time step, representing the independent distribution of the data over time.

In this way, the TIFE-T component isolates and ignores the frequency and spatial variations, ensuring that the extracted features are independent along the time dimension. Given the higher level abstraction of the data processed by TIFE-T, and the nonlinear and dynamic nature of these features, each sub-tensor  $Z^{(t_1)}$  is further processed by a learnable normalization layer to obtain the time-independent feature tensor:

$$Z_{\text{TIFE-T}} = [\hat{Z}^{(1)}, \dots, \hat{Z}^{(T)}], \quad (6)$$

$$\hat{Z}^{(t)} = \left\{ \hat{z}_{t, f, s, h} \mid f = 1, \dots, F; s = 1, \dots, S; h = 1, \dots, H \right\}, \quad (7)$$

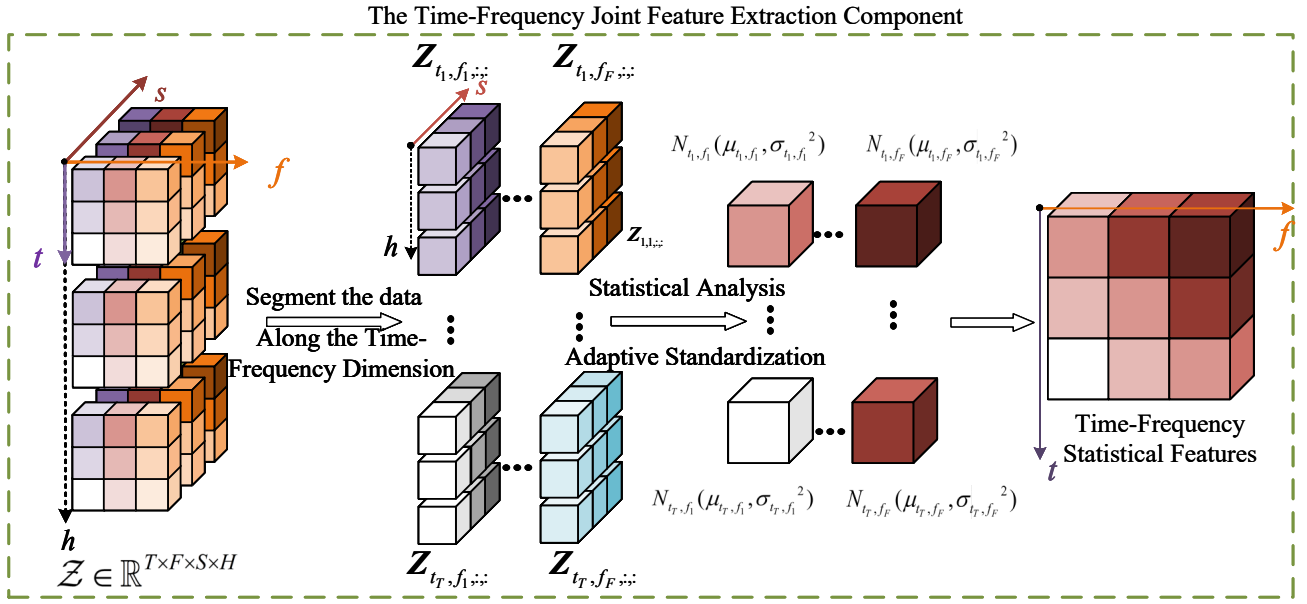


Figure 4. Time-frequency joint feature extraction within the tensor structure of the multi-dimensional EMS.

$$\hat{z}_{t,f,s,h} = \frac{z_{t,f,s,h} - \mu_t}{\sigma_t + \varepsilon} + \beta_t \cdot z_{t,f,s,h}, \quad z_{t,f,s,h} \in \mathcal{Z}^{(t)}. \quad (8)$$

In the above equation  $\varepsilon$  is a small constant added to avoid division by zero,  $\gamma_t$  and  $\beta_t$  are learnable parameters, where  $\gamma_t$  represents data-dependent dynamic weights and  $\beta_t$  is a static bias term independent of the data. Clearly, the TIFE-T component outputs the time-independent feature tensor  $\mathcal{Z}_{\text{TIFE-T}}$ . This statistical approach approximates the marginal distribution  $P(\mathcal{Z}_{t,f,s,h} | t \in \{1, \dots, T\})$ , capturing and quantifying the independent impact of temporal variations. Similarly, the frequency- and spatial-independent feature extraction components follow the same procedure as TIFE-T and output the corresponding feature tensors  $\mathcal{Z}_{\text{TIFE-F}}$  and  $\mathcal{Z}_{\text{TIFE-S}}$  respectively.

In real scenarios, due to dynamic spectrum access and other multi-dimensional collaborative communication technologies, spectrum data no longer varies independently along a single dimension. Instead, it exhibits complex, nonlinear couplings across dimensions. To model such cross-dimensional variations, it is necessary to introduce pairwise joint feature representations. Taking the time-frequency joint feature extraction (TFJFE-TF) component as an example, the principle of time-frequency joint feature extraction within the tensor structure of the multi-dimensional EMS is illustrated in Figure 4.

The four-dimensional tensor at the input is decomposed simultaneously along the time

and frequency dimensions to generate a set of two-dimensional sub-matrices extending across both dimensions:

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}^{(1,1)} & \dots & \mathcal{Z}^{(1,T)} \\ \vdots & \ddots & \vdots \\ \mathcal{Z}^{(F,1)} & \dots & \mathcal{Z}^{(T,F)} \end{bmatrix}. \quad (9)$$

Each sub-matrix is then subjected to statistical processing to generate a joint feature tensor that  $\mathcal{Z}$  varies along these two dimensions, i.e.:

$$\begin{aligned} \mathcal{Z}_{\text{TFJFE-TF}} &= \text{TFJFE-TF}(\mathcal{Z}) \\ &= \begin{bmatrix} \hat{\mathcal{Z}}^{(1,1)} & \dots & \hat{\mathcal{Z}}^{(1,T)} \\ \vdots & \ddots & \vdots \\ \hat{\mathcal{Z}}^{(F,1)} & \dots & \hat{\mathcal{Z}}^{(T,F)} \end{bmatrix}, \end{aligned} \quad (10)$$

where  $\mathcal{Z}_{\text{TFJFE-TF}}$  denotes the joint time-frequency feature tensor, and  $\text{TFJFE-TF}(\cdot)$  represents the operation of the time-frequency joint feature extraction component. Furthermore,  $\hat{\mathcal{Z}}^{(t,f)} \in \mathbb{R}^{S \times H}$ ,  $t \in \{1, \dots, T\}$ ,  $f \in \{1, \dots, F\}$  refers to the two-dimensional feature sub-matrix extracted at time step  $t$  and frequency point  $f$ , with the set of all  $\hat{\mathcal{Z}}^{(t,f)}$ ,  $\{\hat{z}_{t,f,s,h} | s = \{1, \dots, S\}; h = \{1, \dots, H\}\}$  used as samples for statistical feature estimation:

$$\mu_{t,f_1} = \mathbb{E}[\mathcal{Z}^{(t,f_1)}] \approx \frac{1}{S \cdot H} \sum_{s=1}^S \sum_{h=1}^H \hat{z}_{t,f_1,s,h}, \quad (11)$$

$$\sigma_{t,f_1} = \sqrt{\mathbb{E} \left[ \left( \mathbf{Z}^{(t_1,f_1)} - \mu_{t_1,f_1} \right)^2 \right]} \quad (12)$$

$$\approx \sqrt{\frac{1}{S \cdot H} \sum_{s=1}^S \sum_{h=1}^H (\hat{z}_{t_1,f_1,s,h} - \mu_{t_1,f_1})^2},$$

where each element  $\hat{z}_{t_1,f_1,s,h}$  in the latent representation is a scalar.  $\mu_{t_1,f_1}$  and  $\sigma_{t_1,f_1}$  represent the mean and variance of  $\mathbf{Z}^{(t_1,f_1)}$  at time step  $t_1$  and frequency point  $f_1$ , respectively. In this case, it is clear that computed  $\mu_{t_1,f_1}$  and  $\sigma_{t_1,f_1}$  vary only with the combination of time and frequency. The mean and standard deviation characterize the joint variation pattern of the spectral data across the time-frequency dimensions, thereby decoupling the influence of the spatial dimension. We further apply learnable parameter-based normalization to the corresponding two-dimensional sub-matrix  $\mathbf{Z}^{(t_1,f_1)}$  to obtain the joint time-frequency feature tensor:

$$\mathbf{Z}_{\text{TFJFE-TF}} = \begin{bmatrix} \hat{\mathbf{Z}}^{(1,1)} & \dots & \hat{\mathbf{Z}}^{(1,T)} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{Z}}^{(F,1)} & \dots & \hat{\mathbf{Z}}^{(T,F)} \end{bmatrix}, \quad (13)$$

$$\hat{\mathbf{Z}}^{(t_1,f_1)} = \left\{ \hat{z}_{t_1,f_1,s,h} \mid s = \{1, \dots, S\}; h = \{1, \dots, H\} \right\}, \quad (14)$$

$$\hat{z}_{t_1,f_1,s,h} = \gamma_{t_1,f_1} \frac{\hat{z}_{t_1,f_1,s,h} - \mu_{t_1,f_1}}{\sigma_{t_1,f_1} + \varepsilon} + \beta_{t_1,f_1}, \quad (15)$$

where  $\gamma_{t_1,f_1}$  and  $\beta_{t_1,f_1}$  are the learnable parameters at time step  $t_1$  and frequency point  $f_1$ ,  $\varepsilon$  is a small constant added to avoid division by zero. Thus,  $\mathbf{Z}_{\text{TFJFE-TF}}$  is output as the joint time-frequency feature tensor. Through this approach, the derived statistics approximate the joint distribution  $P(\mathbf{Z}_{t,f,s,h} \mid t \in \{1, \dots, T\}; f \in \{1, \dots, F\})$  of the two dimensions, extracting and quantifying the interaction between time and frequency. Similarly, the frequency-space and time-space joint feature extraction modules undergo analogous processing, providing the feature tensors  $\mathbf{Z}_{\text{TFJFE-FS}}$  and  $\mathbf{Z}_{\text{TFJFE-TS}}$ .

The next step is to combine adaptive learnable parameters with underlying statistical features to complete the feature extraction for three-dimensional sub-tensors at each time step. This hybrid method, integrating statistical and deep learning-based feature extraction, leverages the strengths of both approaches. On one hand, the time-varying mean and variance derived from statistical methods fully utilize the physical significance of measured data, enhancing

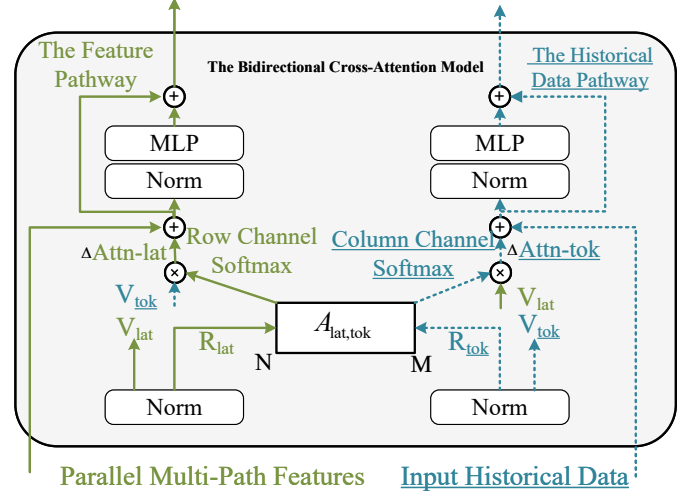


Figure 5. Tensor feature fusion module based on a multi-dimensional bidirectional cross-attention mechanism.

model interpretability. On the other hand, the end-to-end loss minimization dynamically adjusts the learnable parameters, capturing complex patterns in the data and improving nonlinear feature extraction capability of the model, thereby compensating for the computational limitations of purely statistical methods.

### 3.4 Feature Fusion and Prediction

When describing an object, two key questions arise. The first one is "what" the object contains and the second one is "where" the object is located. By focusing on the "what," the position of the object can be quickly identified, while "what" defines its internal structure, enabling rapid localization.

Extending this concept to NNs, the feature pathway describing "where" and the historical data pathway describing "what" exhibit a symmetric and interdependent relationship, leading to nearly symmetrical attention patterns. Motivated by this analogy, we introduce a bidirectional cross-attention mechanism for cross-dimensional feature fusion [25]. This mechanism interactively learns the mutual dependencies between historical data and features, enhancing cross-dimensional representations while reducing redundant information and shrinking the scale of input features for the prediction model. The operation of the tensor feature fusion module, based on a multi-dimensional bidirectional cross-attention, is illustrated in Figure 5.

First, normalization is applied to refine the two data

pathways. Reference values are generated using learnable linear projections of the input data: the feature pathway creates reference values  $\mathcal{R}_{\text{tok}} \in \mathbb{R}^{N \times D}$  and  $\mathcal{V}_{\text{tok}} \in \mathbb{R}^{N \times D}$ , while the historical data pathway creates reference values  $\mathcal{R}_{\text{lat}} \in \mathbb{R}^{N \times D}$  and  $\mathcal{V}_{\text{lat}} \in \mathbb{R}^{N \times D}$ :

$$\mathcal{R}_{\text{tok}} = \mathcal{Z}_{\text{tok}} \cdot \mathbf{W}_{\mathcal{R}_{\text{tok}}}, \mathbf{W}_{\mathcal{R}_{\text{tok}}} \in \mathbb{R}^{D_{\text{tok}} \times D}, \quad (16)$$

$$\mathcal{R}_{\text{lat}} = \mathcal{Z}_{\text{in}} \cdot \mathbf{W}_{\mathcal{R}_{\text{lat}}}, \mathbf{W}_{\mathcal{R}_{\text{lat}}} \in \mathbb{R}^{D_{\text{lat}} \times D}, \quad (17)$$

$$\mathcal{V}_{\text{tok}} = \mathcal{Z}_{\text{tok}} \cdot \mathbf{W}_{\mathcal{V}_{\text{tok}}}, \mathbf{W}_{\mathcal{V}_{\text{tok}}} \in \mathbb{R}^{D_{\text{tok}} \times D}, \quad (18)$$

$$\mathcal{V}_{\text{lat}} = \mathcal{Z}_{\text{in}} \cdot \mathbf{W}_{\mathcal{V}_{\text{lat}}}, \mathbf{W}_{\mathcal{V}_{\text{lat}}} \in \mathbb{R}^{D_{\text{lat}} \times D}, \quad (19)$$

where, a set of  $N = 6$  learnable independent and joint feature tensors serve as the input for the feature pathway, processed in parallel as:

$$\mathcal{Z}_{\text{tok}} = \text{Concat} \left( \mathcal{Z}_{\text{TIFE-T}}, \mathcal{Z}_{\text{TIFE-F}}, \mathcal{Z}_{\text{TIFE-S}}, \right. \\ \left. \mathcal{Z}_{\text{TFJFE-TF}}, \mathcal{Z}_{\text{TFJFE-TS}}, \mathcal{Z}_{\text{TFJFE-FS}} \right). \quad (20)$$

Meanwhile, the latent representation of the original spectral data, containing raw information, is used as the input  $\mathcal{Z}_{\text{lat}} = \mathcal{Z} \in \mathbb{R}^{T \times F \times S \times H}$  for the historical data pathway. Thus, the feature pathway continuously extracts contextual information from the historical data pathway through iterative updates of  $(\mathcal{R}_{\text{lat}}, \mathcal{V}_{\text{lat}})$ , while, the historical data pathway progressively captures abstract information from the feature pathway via iterative updates of  $(\mathcal{R}_{\text{tok}}, \mathcal{V}_{\text{tok}})$ . Following the multi-head attention mechanism, the correlation between each feature in  $\mathcal{Z}_{\text{in}}$  and  $\mathcal{Z}_{\text{tok}}$  is computed, producing a cross-correlation matrix  $\bar{\mathbf{A}}$ :

$$\bar{\mathbf{A}}_{\text{lat,tok}} = \left( \frac{\mathcal{R}_{\text{lat}} \mathcal{R}_{\text{tok}}^\top}{\sqrt{D}} \right) = \bar{\mathbf{A}}_{\text{tok,lat}}^\top, \quad (21)$$

where  $\bar{\mathbf{A}}_{\text{lat,tok}}$  and  $\bar{\mathbf{A}}_{\text{tok,lat}}^\top$  represent the correlation matrices from the feature pathway to the historical data pathway and vice versa. Through cross-learning, the two pathways capture each other's inherent symmetric tendencies. Using these matrices and value tensors, the module computes attention weights for both pathways:

$$\Delta_{\text{lat}}^{\text{attn}} = \text{softmax}(\bar{\mathbf{A}}_{\text{lat,tok}}) \cdot \mathcal{V}_{\text{tok}}, \quad (22)$$

$$\Delta_{\text{tok}}^{\text{attn}} = \text{softmax}(\bar{\mathbf{A}}_{\text{tok,lat}}) \cdot \mathcal{V}_{\text{lat}}, \quad (23)$$

where  $\Delta_{\text{lat}}^{\text{attn}}$  and  $\Delta_{\text{tok}}^{\text{attn}}$  denote the attention-weighted outputs. The historical data pathway's attention weights  $\Delta_{\text{lat}}^{\text{attn}}$  incorporate information from the feature pathway's value tensor  $\mathcal{V}_{\text{tok}}$ , iteratively infusing feature-derived knowledge into the historical data

pathway. Similarly, historical data insights are progressively integrated into the feature pathway.

Finally, on the feature pathway, the symmetrically attended weights  $\Delta_{\text{lat}}^{\text{attn}}$ , which are now enriched with historical data information, are combined with the historical pathway's reference values  $\mathcal{V}_{\text{lat}}$  to refine the feature data:

$$\hat{\mathcal{Z}}_{\text{lat}} = \text{MLP}(\text{Norm}(\Delta_{\text{lat}}^{\text{attn}} + \mathcal{Z}_{\text{in}})) + \Delta_{\text{lat}}^{\text{attn}} + \mathcal{Z}_{\text{in}}, \quad (24)$$

where a linear layer normalizes the attention-weighted input:  $\mathcal{Z}_{\text{Norm}} = \text{Norm}(\mathcal{Z}_{\text{in}}) = \gamma \odot \frac{\Delta_{\text{tok}}^{\text{attn}} + \mathcal{Z}_{\text{in}} - \mu}{\sigma + \varepsilon} + \beta$ , where  $\odot$  denotes element-wise multiplication,  $\gamma$  denotes scaling parameter and  $\beta$  denotes offset parameter. Finally, a multilayer perceptron  $\text{MLP}(\cdot)$  further adjusts the normalized features locally to enhance representation.

### 3.5 Temporal Convolutional Network and Output Module

For the above mentioned fused features  $\hat{\mathcal{Z}}_{\text{lat}}$ , a TCN is employed to extract temporal causality and long-term dependencies in the data, enabling predictions of future wideband spectrum measurements across multiple locations.

The TCN consists of residual blocks with varying spans, each comprising dilated causal convolution, weight normalization, activation functions, and  $1 \times 1$  convolution. Through parameter configuration, the dilated causal convolution learns temporal dependencies at different receptive fields. Weight normalization stabilizes training by reparameterizing convolutional filters to improve gradient flow, while activation functions introduce nonlinearity to enable the extraction of complex temporal dependencies. The  $1 \times 1$  convolution adjusts feature dimensions to capture evolving temporal patterns.

The output module incorporates skip connections [26] to preserve rich structural features, feeding multiple TCN layers into FC layers for multi-step predictions. Specifically, the output block contains two FC layers with ReLU activation, transforming hidden representations into the target output space. Mathematically, this can be expressed as:

$$\hat{\mathcal{Y}} = \phi \left( \phi \left( \hat{\mathcal{Z}} \right) * \mathcal{W}_{o1} + \mathbf{b}_{o1} \right) * \mathcal{W}_{o2} + \mathbf{b}_{o2}, \quad (25)$$

where  $\mathcal{W}_{o1} \in \mathbb{R}^{H \times H}$ ,  $\mathbf{b}_{o1} \in \mathbb{R}^H$ ,  $\mathcal{W}_{o2} \in \mathbb{R}^{H \times \Delta T}$ ,  $\mathbf{b}_{o2} \in \mathbb{R}^{\Delta T}$  represent the filter weights and biases.  $*$  denotes matrix multiplication.  $\phi(\cdot)$  denotes the element-wise ReLU activation function.  $\hat{\mathcal{Y}} \in \mathbb{R}^{\Delta T \times F \times S}$

represents the predicted results for  $F$  positions and  $S$  locations at time step size  $\Delta T$ . The FC layers integrate the sequential features learned by previous layers and map them to the target output, thus completing the transformation from sequential features to final predictions. For the loss function design, an end-to-end Mean Squared Error (MSE) criterion is adopted to train the model, using it as the evaluation metric, i.e.:

$$\text{MSE} = \frac{1}{\Delta T \cdot F \cdot S} \sum_{k=1}^S \sum_{j=1}^F \sum_{i=1}^{\Delta T} (y_{i,j,k} - \hat{y}_{i,j,k})^2, \quad (26)$$

where  $y_{i,j,k}$ ,  $\hat{y}_{i,j,k}$  denote ground truth and predictions respectively. The MSE loss aligns with our MFE principle by assuming normal distribution of errors. Compared to alternatives, e.g. Mean Absolute Error (MAE) Loss Function, its adaptive gradient magnitude enables stable convergence even with fixed learning rates. The model continuously updates hidden layers through backpropagation to minimize MSE between predictions and ground truth, ultimately achieving optimal multi-dimensional forecasting performance.

## 4 Experiments and Performance

### 4.1 Experimental Setup

#### 4.1.1 Dataset Description

The experiments used the publicly available European dataset, and its key characteristics are summarized in Table 2. The original dataset has a temporal resolution of 10 seconds and a frequency resolution of 20kHz [27]. Downsampling is applied along the time and frequency dimensions using averaging [28], resulting in a tensor of shape (1440, 30, 3), where the number 1440 represents the time dimension (6-minute resolution, aggregating 6 days of data); the number 30 denotes the frequency dimension (10 MHz resolution, covering 300 MHz) and the number 3 corresponds to the spatial dimension i.e. three measurement locations.

**Table 2.** European dataset overview.

| Dataset              | ElectroSense                             |
|----------------------|------------------------------------------|
| Site                 | Switzerland, Netherlands, United Kingdom |
| Collection Time      | 2021.5.2 7:00-2021.5.8 7:00              |
| Frequency Band       | 2400-2700MHz                             |
| Temporal Resolution  | 10s                                      |
| Frequency Resolution | 20kHz                                    |

All experiments have been implemented using Python

3.6 and Matlab R2022b, with CUDA 11.1 for GPU acceleration. The hardware platform employs an AMD Ryzen 7 5800H with Radeon Graphics and an NVIDIA GeForce RTX 3060 Laptop GPU.

#### 4.1.2 Evaluation Metrics

To quantitatively assess the performance of structural missing data reconstruction, two widely used metrics in data imputation and regression have been adopted. The first one is the Root Mean Squared Error (RMSE), which measures the square root of the average squared deviations between predicted and true values, i.e.:

$$\text{RMSE} = \sqrt{\frac{1}{T \cdot F \cdot S} \sum_{s=1}^S \sum_{f=1}^F \sum_{t=1}^T (y_{t,f,s} - \hat{y}_{t,f,s})^2}. \quad (27)$$

The second one was the Mean Absolute Percentage Error (MAPE) which computes the relative percentage error between predictions and ground truth, providing a clear indication of reconstruction accuracy. Mathematically, this can be expressed as:

$$\text{MAPE} = \frac{1}{T \cdot F \cdot S} \sum_{s=1}^S \sum_{f=1}^F \sum_{t=1}^T \left| \frac{y_{t,f,s} - \hat{y}_{t,f,s}}{y_{t,f,s}} \right| \times 100\%, \quad (28)$$

where  $\hat{y}_{t,f,s}$  and  $y_{t,f,s}$  are the true and predicted values, respectively, and  $t \in \{1, \dots, T\}$ ,  $f \in \{1, \dots, F\}$  index the time, frequency, and spatial dimensions. Additionally, the required running time was used to evaluate computational complexity.

#### 4.1.3 Parameter Configuration

The input data have been partitioned into training, validation, and test sets at an 8:1:1 ratio. The model took 15 previous time steps as input to predict the subsequent 1-3 steps. During training and testing, the batch size was fixed at 8 with a learning rate of 0.0001. The fully-connected layers have been configured with 16 and 20 hidden channels for the European dataset.

The feature extraction module employed three parallel independent extractors and three joint extractors to process the three-dimensional input data, whose six output feature tensors are weight-fused and normalized via Sigmoid function. The dimension-aware TCN architecture incorporated 16 attention heads, where each causal dilated convolution uses a kernel size of 2 with dilation factor 2. The entire model is trained end-to-end using accumulated mean squared error loss.

**Table 3.** Performance comparisons for different models in multi-steps prediction on the European dataset.

|                         | location       | steps | BHT-ARIMA | ROSP   | CNN    | LSTM     | DeepGLO-TCN | TTS-Norm | MFE-TFTCN     |
|-------------------------|----------------|-------|-----------|--------|--------|----------|-------------|----------|---------------|
| R<br>M<br>S<br>E        | Switzerland    | 1     | 1.2758    | 1.1422 | 0.8456 | 0.736443 | 0.8001      | 0.7086   | <b>0.4686</b> |
|                         |                | 2     | 1.3144    | 1.3886 | 0.8864 | 0.7593   | 0.8256      | 0.7472   | <b>0.4715</b> |
|                         |                | 3     | 1.3454    | 1.5183 | 0.8562 | 0.7954   | 0.8612      | 0.7810   | <b>0.4744</b> |
|                         | United Kingdom | 1     | 1.0636    | 0.9927 | 0.6846 | 0.5696   | 0.6286      | 0.5819   | <b>0.4399</b> |
|                         |                | 2     | 1.0763    | 1.1200 | 0.7445 | 0.5745   | 0.6992      | 0.6235   | <b>0.4390</b> |
|                         |                | 3     | 1.1095    | 1.3192 | 0.7559 | 0.5893   | 0.7163      | 0.6456   | <b>0.4451</b> |
|                         | Netherlands    | 1     | 1.1861    | 1.0846 | 0.8268 | 0.5921   | 0.6409      | 0.6219   | <b>0.4427</b> |
|                         |                | 2     | 1.2015    | 1.2976 | 0.8700 | 0.6052   | 0.8228      | 0.6406   | <b>0.4457</b> |
|                         |                | 3     | 1.2391    | 1.4342 | 0.8467 | 0.6249   | 0.8408      | 0.6518   | <b>0.4456</b> |
| M<br>A<br>P<br>E<br>(%) | Switzerland    | 1     | 1.2104    | 1.1696 | 0.8324 | 0.5236   | 0.8291      | 0.5087   | <b>0.3204</b> |
|                         |                | 2     | 1.2610    | 1.2731 | 0.9442 | 0.5345   | 0.9373      | 0.5643   | <b>0.3228</b> |
|                         |                | 3     | 1.3065    | 1.4944 | 0.9441 | 0.5549   | 0.9850      | 0.5985   | <b>0.3276</b> |
|                         | United Kingdom | 1     | 1.0098    | 0.9263 | 0.6265 | 0.3222   | 0.6455      | 0.4334   | <b>0.2565</b> |
|                         |                | 2     | 1.0302    | 1.0998 | 0.8405 | 0.3454   | 0.8111      | 0.4504   | <b>0.2559</b> |
|                         |                | 3     | 1.0807    | 1.2108 | 0.8346 | 0.3580   | 0.8613      | 0.4621   | <b>0.2565</b> |
|                         | Netherlands    | 1     | 1.1200    | 1.0289 | 0.7047 | 0.3773   | 0.7168      | 0.4568   | <b>0.2784</b> |
|                         |                | 2     | 1.2083    | 1.2600 | 0.8354 | 0.4007   | 0.7819      | 0.4726   | <b>0.2817</b> |
|                         |                | 3     | 1.2300    | 1.3214 | 0.8276 | 0.4199   | 0.7977      | 0.4760   | <b>0.2805</b> |

## 4.2 Comparative Performance Experiments

To quantitatively evaluate the predictive performance of the proposed model, six previously used multi-dimensional forecasting methods were selected as benchmarks, namely BHT-ARIMA [29], ROSP [30], CNN [31], LSTM [32], DeepGLO-TCN [33] and TTSNorm [23]. To ensure fair comparisons, all models underwent hyperparameter optimization to ensure peak performance. To mitigate randomness, non-deep learning models were evaluated using 10-trial averages, while deep learning models used the average test errors from 10 post-training runs.

The obtained performance results have been summarized in Table 3, where it can be seen that the MFE-TFTCN model achieves the lowest prediction errors in the prediction task of temporal-frequency-spatial EMS. Compared to the second-best model, MFE-TFTCN reduces the average RMSE by 23.57% and MAPE by 25.03%. The superior performance stems from its ability to capture multi-dimensional interactions through bidirectional cross-attention mechanisms, enabling selective feature learning across frequencies and locations.

Among all the considered baseline models, BHT-ARIMA shows the worst single-step prediction accuracy due to its statistical nature being vulnerable to spectrum data randomness; ROSP performs poorly

in multi-step forecasting as its tensor completion framework lacks computational capacity for handling missing data bursts; DeepGLO-TCN and similar deep learning models exhibit 24.30% higher RMSE variance across locations compared to MFE-TFTCN, as they primarily focus on temporal-scale interactions while neglecting frequency-space dimension couplings.

## 4.3 Ablation Experiments

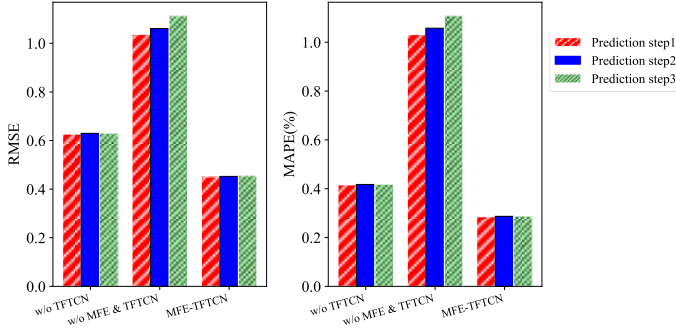
To validate core modules, we conducted the following ablation studies by modifying MFE-TFTCN:

a)w/o TFTCN: Removes the bidirectional cross-attention fusion module

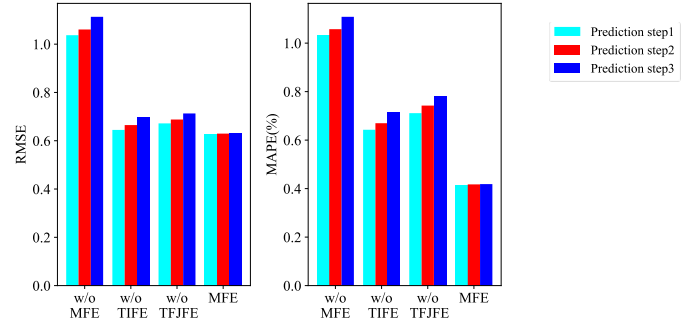
b)w/o MFE & TFTCN: Retains only the TCN and output layers, with both feature extraction and fusion modules removed

All model variants are trained with identical parameters and under the same conditions. Results averaged over 10 test runs and 3 locations are visualized in Figure 6, demonstrating the contribution of each module to prediction accuracy.

Compared to the removal of TFTCN variant, the MFE-TFTCN model achieves 28.03% and 31.28% reduction in RMSE and MAPE, respectively. This improvement stems from the bidirectional cross-attention mechanism the ability to dynamically



**Figure 6.** Average prediction error performance of the considered variant models based on MFE-TFTCN on the European dataset with different prediction steps.



**Figure 7.** Average prediction error performance of the considered variant models based on MFE module on the European dataset with different prediction steps.

allocate cross-dimensional attention weights, effectively mitigating the adverse effects of spatial heterogeneity and uneven element distribution in multi-dimensional forecasting. This result confirms that the dual attention mechanism of TFTCN, which simultaneously weights cross-dimensional relationships and single-dimensional features, becomes increasingly effective as the dataset size increases.

The significant performance gap between MFE-TFTCN and the two ablated models (w/o TFTCN and w/o MFE & TFTCN) validates two critical findings: First, the TFTCN module substantially enhances the prediction accuracy by mitigating spatial-temporal irregularities through adaptive attention weighting. Second, spatial heterogeneity fundamentally impacts multi-location spectrum forecasting, which TFTCN effectively addresses via selective feature learning across geographical locations.

To further evaluate the contributions of the TIFE and TFJFE modules, we developed three additional variants on the MFE module:

- a) w/o TIFE: Removes the independent feature extraction module
- b) w/o TFJFE: Eliminates the joint feature extraction module
- c) w/o MFE: Uses only stacked convolutions without any specialized feature extraction

To prevent the attention mechanism from introducing additional variables, all three variant models and the MFE model are evaluated without the TFTCN module. Each variant model is trained using identical parameter settings under the same experimental conditions. The final results represent the average errors from 10 test runs, with visualization showing the mean errors

across three measurement locations, as presented in Figure 7.

When trained under identical conditions (10-run average across 3 locations), the complete MFE model achieves the lowest errors, proving its multi-dimensional feature extraction capability. The w/o MFE variant performs worst due to its failure to capture cross-dimensional dependencies, processing each temporal sequence in isolation.

Both w/o TFJFE and w/o TIFE show intermediate performance, confirming that spectrum data exhibits both dimension-specific patterns (e.g., tidal effects in time, antenna configurations in space) and cross-dimensional interactions (e.g., time-frequency characteristics for dynamic spectrum sharing, frequency-space transmission loss patterns). The w/o TFJFE model retains dimension-isolated features like signal bandwidth determinants and spatial distribution patterns, while w/o TIFE preserves critical cross-dimensional relationships essential for modern spectrum sharing systems.

#### 4.4 Different Normalization Experiments

To investigate the effect of different normalization approaches prior to feature extraction, we evaluated three variants of the MFE-TFTCN model with alternative standardization methods:

- a) Global normalization: Applies Z-score standardization using the mean and standard deviation computed across all elements in the spectral tensor and is used in performance comparisons and ablation studies.
- b) Spatial normalization: Performs Z-score standardization separately for each geographical location using location-specific statistics.

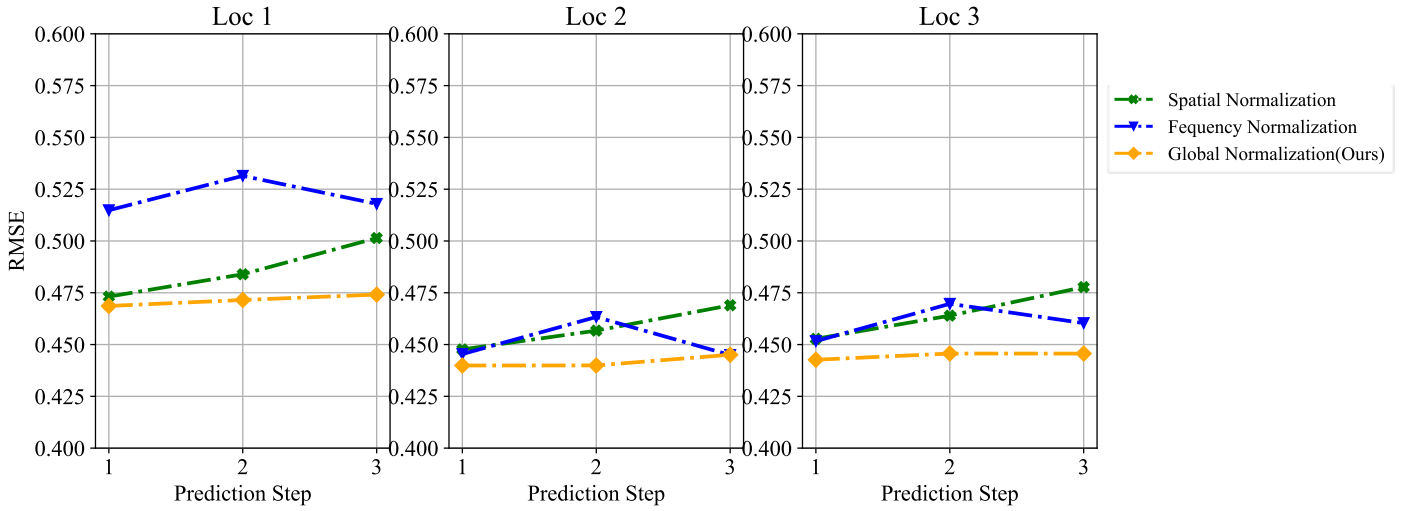


Figure 8. RMSE performance vs. prediction step for each location.

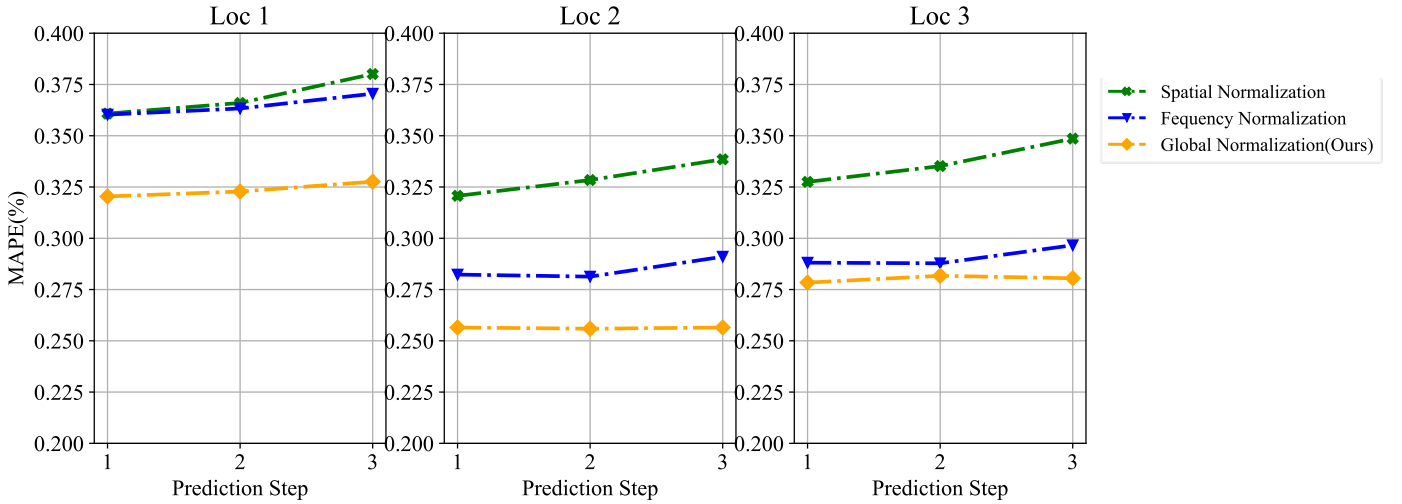


Figure 9. MAPE performance vs. prediction step for each location.

c) Frequency normalization: Conducts Z-score standardization independently for each frequency point along the frequency dimension.

All normalization variants were trained under identical hyperparameters and environmental conditions. The results represent average metrics from 10 test iterations, with location-wise mean errors visualized in Figure 8 and Figure 9.

The global normalization method demonstrated marginally superior performance, reducing RMSE by 1.70% and MAPE by 3.72%, compared to the second-best approach. This method applies uniform standardization across all elements of the multi-dimensional spectrum data, offering low computational complexity while achieving optimal

results on the homogeneous European dataset, indicating strong generalization capability.

## 5 Conclusion

This paper has proposed a novel multi-dimensional feature extraction and fusion method (MFE-TFTCN) that addressed the challenges of high-dimensional and cross-dimensional feature learning in spectrum prediction through a dimension-decoupling approach. The MFE-TFTCN architecture employed parallel homogeneous modules combining statistical and deep learning approaches, making it inherently adaptable to higher-dimensional prediction tasks beyond EMS applications. Experimental results on the real-world multi-location dataset have demonstrated superior performance compared to the state-of-the-art

multi-dimensional forecasting models, achieving the average reductions of 23.57% in RMSE and 25.03% in MAPE. The proposed model maintains stable prediction accuracy across varying forecasting horizons and geographical locations.

## Data Availability Statement

Data will be made available on request.

## Funding

This work was supported by the National Natural Science Foundation of China under Grant 62201172.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Ethical Approval and Consent to Participate

Not applicable.

## References

- [1] Wang, C. X., Di Renzo, M., Stanczak, S., Wang, S., & Larsson, E. G. (2020). Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges. *IEEE Wireless Communications*, 27(1), 16-23. [CrossRef]
- [2] Zhou, Z., Lyu, G., Huang, Y., Wang, Z., Jia, Z., & Yang, Z. (2024, August). Sdformer: transformer with spectral filter and dynamic attention for multivariate time series long-term forecasting. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, Jeju, Republic of Korea (pp. 3-9).
- [3] Li, T. B., Su, Y. T., Song, D., Li, W. H., Wei, Z. Q., & Liu, A. A. (2024). Multi-scale spatial-temporal transformer for meteorological variable forecasting. *IEEE Transactions on Circuits and Systems for Video Technology*. [CrossRef]
- [4] Chen, X., Zhang, H., Zhang, S., Feng, J., Xia, H., Rao, P., & Ai, J. (2024). A space infrared dim target recognition algorithm based on improved DS theory and multi-dimensional feature decision level fusion ensemble classifier. *Remote Sensing*, 16(3), 510. [CrossRef]
- [5] Ya, T. U., Yun, L. I. N., Haoran, Z. H. A., Yu, W. A. N. G., Guan, G. U. I., & Shiwen, M. A. O. (2022). Large-scale real-world radio signal recognition with deep learning. *Chinese Journal of Aeronautics*, 35(9), 35-48. [CrossRef]
- [6] Ji, A., Li, D., Dai, Z., Cui, M., Yu, L., & Duan, Z. (2025). A hybrid graph memory network approach with multi-level feature representation for traffic flow forecast. *Expert Systems with Applications*, 130316. [CrossRef]
- [7] Yazdanpanah, O., Chang, M., & Ali Bakhshi, E. (2025). Attention-based hybrid convolutional-long short-term memory network for bridge pier hysteresis and backbone curves prediction. *Integrated Computer-Aided Engineering*, 32(2), 176-195. [CrossRef]
- [8] Yan, J., Peng, Y., Li, Q., Shao, H., Bin, G., Zheng, J., ... & Zhou, J. (2025). Riesz Time-frequency Spectrum Transform and Its Application in Rolling Bearing Fault Diagnosis. *IEEE Sensors Journal*. [CrossRef]
- [9] Lin, Y., Tu, Y., & Dou, Z. (2020). An improved neural network pruning technology for automatic modulation classification in edge devices. *IEEE Transactions on Vehicular Technology*, 69(5), 5703-5706. [CrossRef]
- [10] Lin, Y., Zhao, H., Ma, X., Tu, Y., & Wang, M. (2020). Adversarial attacks in modulation recognition with convolutional neural networks. *IEEE Transactions on Reliability*, 70(1), 389-401. [CrossRef]
- [11] Cao, D., Zhao, J., Hu, W., Zhang, Y., Liao, Q., Chen, Z., & Blaabjerg, F. (2021). Robust deep Gaussian process-based probabilistic electrical load forecasting against anomalous events. *IEEE Transactions on Industrial Informatics*, 18(2), 1142-1153. [CrossRef]
- [12] Rahim, V. A., & Prema, S. C. (2021, December). A combined spectrum prediction and sensing approach for cognitive radios. In *2021 IEEE 18th India Council International Conference (INDICON)* (pp. 1-6). IEEE. [CrossRef]
- [13] Radhakrishnan, N., Kandeepan, S., Yu, X., & Baldini, G. (2024, December). Multi step temporal spectrum occupancy prediction using deep learning. In *2024 17th International Conference on Signal Processing and Communication System (ICSPCS)* (pp. 1-6). IEEE. [CrossRef]
- [14] Brown, C., & Ghasemi, A. (2023). Evolution toward data-driven spectrum sharing: Opportunities and challenges. *IEEE Access*, 11, 99680-99692. [CrossRef]
- [15] Gao, Y., Zhao, C., & Fu, N. (2021, September). Joint multi-channel multi-step spectrum prediction algorithm. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)* (pp. 1-5). IEEE. [CrossRef]
- [16] Ren, X., Mosavat-Jahromi, H., Cai, L., & Kidston, D. (2021). Spatio-temporal spectrum load prediction using convolutional neural network and ResNet. *IEEE Transactions on Cognitive Communications and Networking*, 8(2), 502-513. [CrossRef]
- [17] Basak, S., Rajendran, S., Pollin, S., & Scheers, B. (2023). Spectrum prediction for protocol-aware RF jamming. *IEEE Transactions on Cognitive Communications and Networking*, 10(2), 363-373. [CrossRef]
- [18] Li, S., Sun, Y., Han, Y., Zhang, Z., Yao, M., Chen, J., ... & Lin, Y. (2024). CL-MFGCN: Graph Structure Contrastive Learning and Multi-Scale Feature Fusion Graph Convolutional Network for Spectrum Prediction. *IEEE Internet of Things Journal*. [CrossRef]

- [19] Naikwadi, M. H., & Patil, K. P. (2022). A multi-dimensional real world spectrum occupancy data measurement and analysis for spectrum inference in cognitive radio network. *International Journal of Communication Networks and Information Security*, 14(2), 244-260. [CrossRef]
- [20] Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6), 7833-7848. [CrossRef]
- [21] Zhang, H., Tian, Q., & Han, Y. (2022, September). Multi channel spectrum prediction algorithm based on GCN and LSTM. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)* (pp. 1-5). IEEE. [CrossRef]
- [22] Li, X., Liu, Z., Xu, Y., Wang, X., & Song, T. (2020, August). Recovering missing values from corrupted historical spectrum observations for dependable spectrum prediction. In *2020 IEEE/CIC International Conference on Communications in China (ICCC)* (pp. 941-946). IEEE. [CrossRef]
- [23] Deng, J., Deng, J., Yin, D., Jiang, R., & Song, X. (2023). Tts-norm: Forecasting tensor time series via multi-way normalization. *ACM Transactions on Knowledge Discovery from Data*, 18(1), 1-25. [CrossRef]
- [24] Jebur, R. S., & Thaher, R. H. (2022, October). Multiple FBG wavelength peaks detection using Z-score algorithm. In *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 223-226). IEEE. [CrossRef]
- [25] Tian, G., Zhang, C., Shi, Y., & Li, X. (2024). MultiWaveNet: A long time series forecasting framework based on multi-scale analysis and multi-channel feature fusion. *Expert Systems with Applications*, 251, 124088. [CrossRef]
- [26] He, N., Fang, L., Li, S., Plaza, J., & Plaza, A. (2019). Skip-connected covariance network for remote sensing scene classification. *IEEE transactions on neural networks and learning systems*, 31(5), 1461-1474. [CrossRef]
- [27] Rajendran, S., Calvo-Palomino, R., Fuchs, M., Van den Bergh, B., Cordobés, H., Giustiniano, D., ... & Lenders, V. (2017). Electrosense: Open and big spectrum data. *IEEE Communications Magazine*, 56(1), 210-217. [CrossRef]
- [28] Yang, M., Yang, J., Xiao, Z., & Huang, M. (2021, January). A modular spectrum sensing node for Resources-Oriented Radio Monitoring. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-8). IEEE. [CrossRef]
- [29] Shi, Q., Yin, J., Cai, J., Cichocki, A., Yokota, T., Chen, L., ... & Zeng, J. (2020, April). Block Hankel tensor ARIMA for multiple short time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 5758-5766). [CrossRef]
- [30] Li, X., Wang, X., Song, T., & Hu, J. (2021). Robust online prediction of spectrum map with incomplete and corrupted observations. *IEEE Transactions on Mobile Computing*, 21(12), 4583-4594. [CrossRef]
- [31] Bohara, B., Fernandez, R. I., Gollapudi, V., & Li, X. (2022, November). Short-term aggregated residential load forecasting using BiLSTM and CNN-BiLSTM. In *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)* (pp. 37-43). IEEE. [CrossRef]
- [32] Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28.
- [33] Sen, R., Yu, H. F., & Dhillon, I. S. (2019). Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. *Advances in neural information processing systems*, 32.



**Yu Han** received the B.S. and M.E. degrees in information and communication engineering from Harbin Engineering University (HEU), Harbin, China, in 2009 and 2013, respectively, and Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2021. From 2017 and 2019, he was with the James Cook University, Australia, as a Visiting Scholar. He has been with the College of Information and Communication Engineering, HEU, as a Lecturer since 2021. His current research interests include complex electromagnetic environment perception, the Internet of Things and massive information access. (Email: yu.han1001@hrbeu.edu.cn)



**Jiayi Song** received the B.S. degree in Communication Engineering from Hebei University of Technology, Tianjin, China, in 2024. (Email: songjiayi@hrbeu.edu.cn)



**Zhiqi Wang** received the M.S. degree in Electronic Information Engineering from Harbin Engineering University, Harbin, China, in 2025. (Email: 22heuwzq@hrbeu.edu.cn)



**Wei Xiang** received the Ph.D. degree in Telecommunications Engineering from the University of South Australia, Adelaide, Australia, in 2004. His research interests include the Internet of Things, coding and signal processing for multimedia communication systems. He served as an Editor for IEEE Communications Letters and an Associate Editor for Springer's Telecommunications Systems. (Email:

w.xiang@latrobe.edu.au)



**Guan Gui** received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. His research interests include intelligent sensing and recognition, intelligent signal processing, and physical layer security. He is the Editor-in-Chief of KSII Transactions on Internet and Information Systems, and an Editorial Board Member of journals such as IEEE Transactions on Information

Forensics and Security, IEEE Internet of Things Journal, and IEEE Transactions on Vehicular Technology. (Email: guiguan@njupt.edu.cn)



**P. Takis Mathiopoulos** received the Ph.D. degree in Digital Communications from the University of Ottawa, Ottawa, Ontario, Canada, in 1989. His research interests include wireless terrestrial and satellite communication systems and networks, as well as remote sensing; lidar systems; and information technologies, including blockchain systems. He has served on the editorial boards of several archival journals, including acting as a Regional Editor

for IET Communications, and a Specialty Chief Editor for IEEE Transactions on Communications, Remote Sensing Journal, and Frontiers in Aerospace and Space Communications. (Email: mathio@ieee.org)



**Yun Lin** received the Ph.D. degree in Communication and Information Systems from Harbin Engineering University, Harbin, China, in 2010. His research interests include machine learning and data analysis over wireless networks, signal processing and analysis, cognitive radio and software-defined radio, as well as artificial intelligence and pattern recognition. He serves as an Associate Editor for IEEE Internet of Things, IEEE

Transactions on Reliability, Digital Communications and Networks, Wireless Networks, and KSII Transactions on Internet and Information Systems, and a China Regional Editor for International Journal of Performability Engineering. (Email: linyun@hrbeu.edu.cn)