Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 1, No. 3, 2026

A Dynamic Grain and Semantic-FCM Model for Long-Term Multivariate PM2.5 Forecasting

Zhitao Jia *, Lianchao Qiu

School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China

* **Corresponding author**: (Email: jiazhitao2023@163.com)

Abstract: This paper addresses long-term multivariate time series forecasting, where feature redundancy, noise interference, and complex inter-variable relationships pose significant challenges. A ridge regression model enhanced by Dynamic Granular Computing (DGC) and Fuzzy Cognitive Map (FCM) features is proposed for long-term PM2.5 concentration prediction. The DGC module transforms raw time series into granule-level features described by statistical attributes such as mean, slope, and standard devia-tion, enabling dimensionality reduction and noise mitigation in long-horizon sequences. Since statistical granules alone cannot capture latent relationships among granule-level concepts, an FCM based on nonlinear Hebbian learning is employed to extract cognitive features. These features are combined with statistical granule features and used as inputs to the ridge regression predictor. Experimental results show that granule-based models outperform baselines using raw sequences, and that FCM-enhanced models achieve higher R^2 values and more stable error performance across multiple monitoring sta-tions, demonstrating the effectiveness of the proposed approach.

Keywords: Dynamic Granular Computing, Fuzzy Cognitive Map, Long-term Multivariate Time Series Forecasting, PM2.5 Concentration.

1. Introduction

PM2.5 concentration forecasting is a fundamental task in environmental engineering, playing a critical role in public health risk assessment, air pollution control policy formulation and implementation, and the prevention of respiratory diseases. Essentially, this task can be regarded as a typical time series forecasting problem, as current PM2.5 concentrations exhibit strong temporal dependence on their historical variations. Meanwhile, PM2.5 dynamics are jointly influenced by meteorological conditions, emission source characteristics, and regional transport processes, which further renders the problem a multivariate and nonlinearly coupled forecasting task [1-3].

In previous studies, traditional statistical models, such as ARIMA and its variants, have been widely applied to air quality prediction due to their robustness in modeling trends and seasonality. For example, Zhao et al. [4] proposed a hybrid ARIMA model based on decomposition–reconstruction and optimized parameter selection for long-term PM2.5 forecasting in Beijing, achieving significantly better performance than conventional ARIMA. Yang et al. [5] combined time series models with heterogeneous image data for long-horizon hourly forecasting, demonstrating the potential of statistical methods under data-rich conditions. Rani et al. [6] compared the performance of AR, MA, and ARIMA models for short- and medium-to-long-term forecasting across different Indian cities, showing that statistical models perform well in relatively stable regions but suffer substantial accuracy degradation in areas with strong fluctuations. Ao et al. [7] employed SARIMA to model seasonal patterns of regional PM2.5 concentrations and obtained stable long-term trend predictions. In addition, the Prophet model, which explicitly decomposes trends, seasonality, and holiday effects, has also been adopted for long-term forecasting tasks [8]. However, most statistical models rely on univariate time series and thus have limited capability in representing nonlinear couplings among multiple influencing factors, making them less effective under abrupt pollution events or complex dynamic conditions.

With the rapid development of deep learning, researchers have increasingly explored neural networks to overcome the univariate limitations of statistical models by leveraging their nonlinear mapping capabilities. Among them, the Long Short-Term Memory (LSTM) network has become a mainstream approach for PM2.5 forecasting due to its ability to capture long-term dependencies and handle multivariate inputs. LSTM-based models can simultaneously model complex interactions among meteorological and pollution-related variables while maintaining memory over extended temporal horizons, leading to superior performance in multivariate nonlinear time series forecasting. For instance, Chang et al. proposed an Aggregated LSTM model that integrates data from multiple monitoring stations and external pollution sources for PM2.5 prediction [9]. Qi et al. developed a hybrid Wavelet–LSTM framework, in which wavelet decomposition is used to extract multi-scale components before LSTM-based prediction [10]. Li et al. introduced LSTM into spatiotemporal PM2.5 forecasting by incorporating PM2.5 concentrations, meteorological variables, and neighboring station information, where temporal dependencies are captured by LSTM and spatial correlations are implicitly modeled through multi-station inputs [11]. Bai et al. constructed a hybrid CNN–LSTM architecture, in which CNN extracts local patterns from pollutant and meteorological data, followed by LSTM for temporal evolution modeling [12]. Nevertheless, despite their strong modeling capacity for multivariate sequences, deep neural networks are sensitive to high-frequency noise, which may lead to unstable feature extraction. Moreover, the iterative multi-step forecasting strategy commonly adopted by deep models tends to accumulate errors over long horizons, thereby degrading prediction accuracy. In addition, the black-box nature of deep learning models limits the interpretability of

their forecasting results.

Granular Computing (GrC) offers an effective solution for data dimensionality reduction by transforming raw time series into structured granule-level representations. By transforming low-level raw sequences into "information granules" characterized by statistical attributes, GrC enables dimensionality reduction, noise suppression, and granule-level representation of time series data [15, 16]. Unlike modeling directly in the original data space, GrC performs prediction at the granule level, allowing models to preserve essential temporal characteristics in a lower-dimensional space while reducing the number of iterative steps, thereby alleviating error accumulation in long-term forecasting.

In the prediction domain, GrC-based approaches have attracted increasing attention. Yin et al. combined granular computing with support vector machines for long-term time series forecasting, where fuzzy granulation algorithms were used to transform samples into coarser-grained representations and estimate the variation ranges of continuous time series [17]. Song et al. proposed a hybrid forecasting model that integrates autoregressive integrated moving average (ARIMA) models with granular neural networks [18]. However, most existing GrC-based prediction models remain limited in feature extraction, as they primarily rely on the statistical properties of individual granules and fail to fully exploit deeper causal relationships among granule-level features.

The Fuzzy Cognitive Map (FCM) mechanism is introduced in this work to compensate for the current lack of cognitive feature extraction capabilities. FCM is a powerful neuro-fuzzy modeling technique that represents and reasons about the dynamic behavior of complex systems through concept nodes and weighted causal connections. It has been widely applied in areas such as energy systems, environmental forecasting, and medical monitoring. For example, Stach et al. compared expert-defined FCMs with those automatically learned from data and proposed data-driven weight training strategies [19]. Wang et al. developed Deep-FCM by integrating FCMs with hierarchical structures to improve modeling capability for high-dimensional multivariate sequences while retaining a certain degree of interpretability [20]. Zhen et al. combined Axiomatic Fuzzy Sets (AFS) with FCM and introduced memory mechanisms to enhance semantic representation and multi-objective prediction performance, highlighting the importance of semantic concepts and causal relationships in environmental time series forecasting [21]. In addition, Zhen et al. proposed a causal–semantic air quality prediction framework in which FCMs were used to interpret the causal chain among pollution, meteorology, and haze for haze forecasting [22].

This study aims to address data redundancy and the difficulty of capturing complex relationships in long-term multivariate time series forecasting, while reducing cumulative prediction errors. To this end, a novel prediction framework based on Dynamic Granular Computing (DGC) and FCM enhancement is proposed and applied to long-term multivariate PM2.5 forecasting. In this framework, DGC transforms raw multivariate time series into granule-level features with statistical properties, achieving dimensionality reduction and noise suppression. Subsequently, an FCM mechanism is introduced, where cognitive weights are trained using a nonlinear Hebbian learning rule to generate cognitive feature vectors that capture latent relationships among granule-level concepts of the target series. These cognitive features are concatenated with the statistical granule features of the input series and used as inputs to the prediction model.

A ridge regression predictor with a simple structure and robust parameterization is employed to mitigate the instability of deep learning models in long-horizon forecasting while maintaining interpretability. Experimental results demonstrate that, in the granule-level feature space, the FCM-enhanced ridge regression model achieves the best performance and consistently outperforms the non-enhanced ridge regression baseline, providing strong evidence that FCM-derived cognitive features can stably improve the predictive accuracy of linear models. Moreover, an important finding is that, in the granule-level space where the sample size is reduced due to granulation, simple linear models exhibit better generalization ability and robustness than complex nonlinear models such as LSTM.

This work provides an efficient, stable, and interpretable solution for modeling and forecasting long-term complex time series. The main contributions and innovations are summarized as follows:

(1) A slope-threshold–based Dynamic Granular Computing (DGC) mechanism is proposed, which achieves adaptive granularity partitioning through local trend information. This mechanism effectively suppresses noise while preserving the dominant trend characteristics of the original time series.

(2) A ridge regression prediction model integrating DGC and Fuzzy Cognitive Map (FCM) semantic features, termed DGC-FCM-Ridge, is designed. By employing FCM to extract causal relationships from the target series, the input feature space is enhanced, leading to improved forecasting accuracy while maintaining a certain degree of interpretability.

(3) Experimental results demonstrate that the proposed model consistently outperforms multiple baseline methods in terms of prediction accuracy, thereby validating the effectiveness of the DGC framework and the FCM-based semantic cognitive feature enhancement.

2. Technical Methodology

2.1. Dynamic Granulation and Numerical Feature Extraction

This module serves as the front-end processing stage of the prediction framework. Its primary objective is to transform high-dimensional, noisy raw time series data into a set of low-dimensional, high-information-density granular statistical features. This process is divided into two steps: an adaptive dynamic granulation mechanism and granular numerical feature extraction.

2.1.1. Dynamic Granulation Mechanism

Traditional time-series processing methods employ fixed-size time windows for sampling. In such cases, the partition boundaries are independent of the sequence trends, which often results in inconsistent trends within a single information granule, thereby undermining the representativeness of the extracted features.

An adaptive granulation mechanism driven by the local trends of the target sequence is developed to enable flexible and data-dependent granularity partitioning. The core concept is to dynamically determine the boundaries of information granules according to the local variation characteristics of the sequence. This approach ensures trend consistency within each information granule. Suppose the original multidimensional time-series matrix is denoted as $X \in \mathbb{R}^{N \times D}$,

The prediction target vector is $Y \in \mathbb{R}^{N\times 1}$, where N denotes the number of time steps and D represents the dimension of the input features. An information granule boundary is marked when the absolute value of the local slope of the target sequence Y exceeds a preset threshold, or when consecutive slopes have opposite signs. The sequence slope is denoted by β, and the partitioning criteria are as follows:

$$\beta(t-\Delta t, t|Y) \bullet \beta(t, t+\Delta t|Y) < 0 \ \ or \ \ |\beta(\Delta t, t|Y)| > \delta \quad (1)$$

Where, $\beta(t\text{-}\Delta t,\ t|Y)$ is the slope of the target sequence Y within the time interval from $t\text{-}\Delta t$ to t, $\beta(t,t+\Delta t\ |Y)$ is the slope of the target sequence Y within the time interval from t to $t+\Delta t$, $\beta(\Delta t,\ t|Y)$ is the local slope of the target sequence Y near time t, estimated through a local window Δt, and δ is the preset slope threshold.

Through dynamic partitioning, the original multidimensional sequence is compressed into M multidimensional information granules G_1, …, G_i, …, G_M (where, $i\in[1,\ M]$, and $M \ll N$), thereby achieving dynamic granulation. Figure 1 illustrates the dynamic granulation mechanism, where L_i represents the length of the i-th information granule:

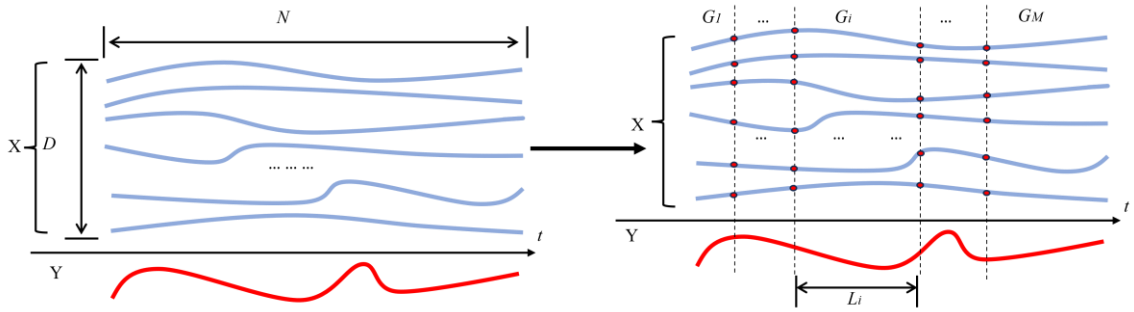

Figure 1. Schematic diagram of dynamic granular computing

This mechanism adaptively generates information granules of varying lengths: longer granules are formed during stable segments of the sequence to compress information, while shorter granules are created during periods of intense fluctuation to capture fine-grained details. This achieves a balance between noise suppression and the preservation of key trends.

2.1.2. Granular Numerical Feature Extraction

Following the completion of dynamic granulation, it is necessary to extract numerical features from the information granules that reflect the overall level, direction of change, and stability. This enables the model to learn and predict within a more compact and robust granular space. Therefore, for each dimension d of every information granule G_i, three core statistics are extracted: the granular mean $\bar{x}_{i,d}$, the granular slope $\beta_{i,d}$, and the granular standard deviation $\sigma_{i,d}$.

The granular mean $\bar{x}_{i,d}$ reflects the average level of the information granule, and its calculation formula is as follows:

$$\bar{x}_{i,d} = \frac{1}{L_i}\sum_{j=k}^{k+L_i-1} x_{j,d} \quad (2)$$

Where $x_{j,d}$ is the value of the d-th dimension feature at time j, and k is the starting index of the information granule.

The granular slope $\beta_{i,d}$ reflects the linear change trend within the information granule, calculated by fitting the data points within the granule using the ordinary least squares method:

$$\beta_{i,d} = \frac{L_i\sum_{t=1}^{L_i} t\cdot x_{t,d} - (\sum_{t=1}^{L_i} t)(\sum_{t=1}^{L_i} x_{t,d})}{L_i\sum_{t=1}^{L_i} t^2 - (\sum_{t=1}^{L_i} t)^2} \quad (3)$$

Where t is the relative time variable corresponding to the time index j, defined as $t=j-k+1 (t\in[1,\ L_i])$, and $x_{t,d}$ is the value of the d-th dimension feature at the relative time t within the information granule.

The granular standard deviation $\sigma_{i,d}$ reflects the volatility and uncertainty of the data within the information granule. It is defined as the sample standard deviation, and its calculation formula is as follows:

$$\sigma_{i,d} = \frac{1}{L_i-1}\sum_{j=k}^{k+L_i-1}(x_{j,d} - \bar{x}_{i,d})^2 \quad (4)$$

Ultimately, each information granule G_i corresponds to a statistical feature vector $S_i=\{\bar{x}_1,\ \beta_i,\ \sigma_i\}$. The collection of features from all information granules, $\{S_1, \ldots, S_i, \ldots, S_M\}$, constitutes the granular numerical feature space for subsequent processing.

2.2. Semantic Cognitive Feature Extraction Based on FCM

Numerical features only reflect the statistical properties of information granules and struggle to characterize the semantic associations and causal relationships between them. To address this, Fuzzy Cognitive Maps—which possess causal modeling capabilities and excellent interpretability—are utilized to extract higher-level granular cognitive features.

2.2.1. Semantic Fuzzification

The construction of semantic features for the FCM requires an initial transformation of numerical features into semantic descriptions. This paper employs trapezoidal membership functions to map each dimension of the granular numerical features (mean, slope, and standard deviation) onto three semantic levels: "Low", "Medium", and "High".

The formula for the trapezoidal membership function $\mu(x;\ a,\ b,\ c,\ d)$ is defined as follows:

$$\mu(x;a,b,c,d) = \begin{cases} 0, x < a \\ \frac{x-a}{b-a}, a \le x < b \\ 1, b \le x < c \\ \frac{d-x}{d-c}, c \le x < d \\ 0, x \ge d \end{cases} \quad (5)$$

The boundary parameters $a,\ b,\ c,\ d$ of the trapezoidal membership function $\mu(x;\ a,\ b,\ c,\ d)$ are adaptively determined based on the distribution of the feature values to ensure a reasonable semantic partition.

Remark. The granular semantic feature vector S_x of the input sequence is fed into the prediction model as an enhanced feature; meanwhile, the granular semantic features of the target sequence are used to drive FCM learning, generating a cognitive relationship feature vector S_{cog} that reflects granular-level interactions to further improve the model's

predictive capability.

2.2.2. Causal Reasoning and Cognitive Feature Extraction

(1) Fuzzy Cognitive Maps (FCM)

FCM is a cognitive modeling tool used to describe and analyze causal relationships between concepts within complex systems. An FCM consists of a set of concept nodes $C=\{C_1, \ldots C_n\}$, a weight matrix W, an activation function $f(\cdot)$, and an activation vector A(t). Within this framework, concept C represents a key semantic state of the system. The weight $w_{pq} \in [-1, 1]$ describes the strength and direction of the causal influence of concept C_p on C_q: a positive value indicates a promotional relationship, a negative value indicates an inhibitory relationship, and the absolute value of the weight reflects the significance of the causal effect.

The system state of the FCM is described by the activation vector $A(t)=(A_1(t), \ldots, A_p(t), \ldots, A_n(t))$, where $A_p(t) \in [0, 1]$ $(1 \le p \le n)$ represents the activation level of the p-th node at time t. The state is updated via Equation (6), where $f(\cdot)$ is typically a Sigmoid function:

$$A_p(t+1) = f\left(\sum_{q=1}^{n} w_{pq} A_q(t)\right), p = 1,2,\ldots,n \quad (6)$$

(2) Weight Learning: Nonlinear Hebbian Rule

The weight matrix W is updated via the Nonlinear Hebbian Rule (NHL). Unlike traditional unsupervised Hebbian rules, the nonlinear version aims to minimize the error between the predicted activation vector $A(t)\cdot W$ and the actual activation vector A(t+1). The formula for calculating the weight increment $\Delta W_{pq}(t)$ is as follows:

$$\Delta W_{pq}(t) = \eta \cdot \left[A_p(t) \cdot A_q(t+1) - W_{pq}(t) \cdot A_p(t) \cdot A_q(t) \right] \quad (7)$$

Where $W_{pq}(t)$ is the connection weight from node p to node q at time t, and η is the learning rate.

After calculating $\Delta W_{pq}(t)$, an L_2 regularization term is applied to prevent excessive weight growth:

$$W_{pq}(t+1) = W_{pq}(t) + \Delta W_{pq}(t) - \tau W_{pq}(t) \quad (8)$$

Where τ is the weight regularization coefficient ($\tau > 0$), used to constrain the magnitude of the FCM weights W.

To ensure the convergence of the FCM network, the weights W_{pq} are restricted to the range [-1, 1] after each update, and the diagonal elements are always maintained at zero.

c. Cognitive Feature Output

Ultimately, the weight matrix W* is obtained after updates via the Nonlinear Hebbian Rule. W* serves as the causal relationship matrix learned by the FCM. The cognitive feature vector S_{cog} is then derived using Equation (9):

$$S_{cog} = W^* \bullet \mu \quad (9)$$

The product of the weight matrix and the membership degrees enables the propagation of semantic activation through the causal network. This results in a cognitive feature representation that integrates the interaction strengths between concepts, providing semantic support for the subsequent Ridge Regression prediction.

2.2.3. Feature Fusion Strategy

Feature fusion is performed by integrating granular-level statistical features with semantic cognitive features extracted by the FCM, thereby enhancing the model's ability to capture long-term dependencies and complex structural information. Statistical features characterize the numerical structure of information granules in terms of mean, slope, and volatility intensity, while semantic cognitive features reflect the latent causal dependencies and conceptual associations between different granules. The two types of features are complementary: if the model relies solely on a single feature type, it would either lack the ability to characterize semantic associations or fail to effectively describe the granular statistical structure, both of which could limit predictive performance.

Therefore, this paper adopts vector-level concatenation to construct the fused features. Assuming $S_x \in \mathbb{R}^{d_1}$, $S_{cog} \in \mathbb{R}^{d_2}$, the fused enhanced feature vector S_E is defined as:

$$S_E = [S_x, S_{cog}] \in \mathbb{R}^{d_1+d_2} \quad (10)$$

Where $[\cdot,\cdot]$ represents the concatenation operation along the feature dimension. This processing method preserves both numerical statistical information and semantic cognitive information without disrupting their respective internal structures. This allows the prediction model to complete parameter learning and forecasting within a unified fused feature space. Consequently, by supplementing the granular statistical modeling with semantic causal information, the robustness of the long-term prediction process is significantly improved.

2.3. Ridge Regression Prediction and De-granulation Reconstruction

Ridge regression can effectively mitigate multi-collinearity and noise interference within high-dimensional features while preserving the interpretability of a linear model. This facilitates the analysis of the relative contributions of granular statistical features and semantic cognitive features. Therefore, this paper employs Ridge Regression as the predictor. Figure 2 provides a schematic diagram of the Ridge Regression prediction and the de-granulation process:

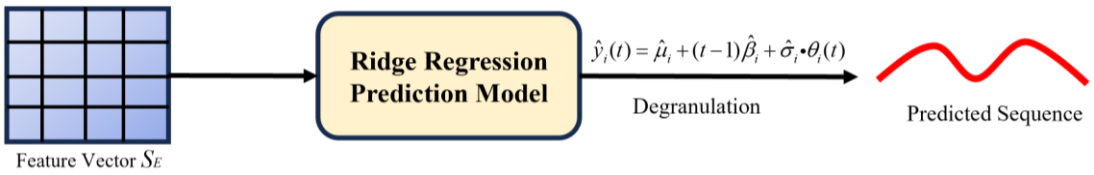

Figure 2. Schematic of Ridge Regression Prediction and De-granulation

The model takes SE as input to directly predict several future target information granules. A de-granulation operation is then performed on these predicted granules to obtain the final predicted target sequence.

The Ridge Regression prediction weights W_{ridge} are obtained using a closed-form least-squares solution to ensure the stability of the model:

$$W_{ridge} = (S_E^T S_E + \alpha I)^{-1} (S_E^T Y) \quad (11)$$

Where α is the regularization coefficient of the Ridge Regression ($\alpha > 0$), and I is the identity matrix.

For a given input S_E, the Ridge Regression outputs the predicted information granules:

$$\hat{G} = S_E \bullet W_{ridge} \tag{12}$$

This approach avoids point-by-point recursion on the original time scale by predicting the mean, slope, and standard deviation of the target information granules. This significantly reduces the accumulation of long-term iterative errors.

Once the predicted information granules are obtained, a de-granulation operation is used to restore them to a continuous sequence on the original time scale. For the *i*-th predicted information granule, let the predicted mean, slope, and standard deviation be denoted as $\hat{x}_i$, $\hat{\beta}_i$, and $\hat{\sigma}_i$, respectively. The reconstructed value for the *t*-th time point within that granule can then be generated via a linear model with a stochastic perturbation:

$$\hat{y}_i(t) = \hat{\mu}_i + (t-1)\hat{\beta}_i + \hat{\sigma}_i \bullet \theta_i(t), \quad t = 1,2,...,L_i; \ \ i = 1,2,...,H \tag{13}$$

Where θ_i(t) is a stochastic perturbation term used to enhance the naturalness and smoothness of the sequence, and H is the number of predicted information granules.

2.4. Summary of the Method

Ultimately, the complete predicted target sequence is obtained by sequentially concatenating the de-granulation results of all information granules. Figure 3 illustrates the overall flowchart of the proposed model:

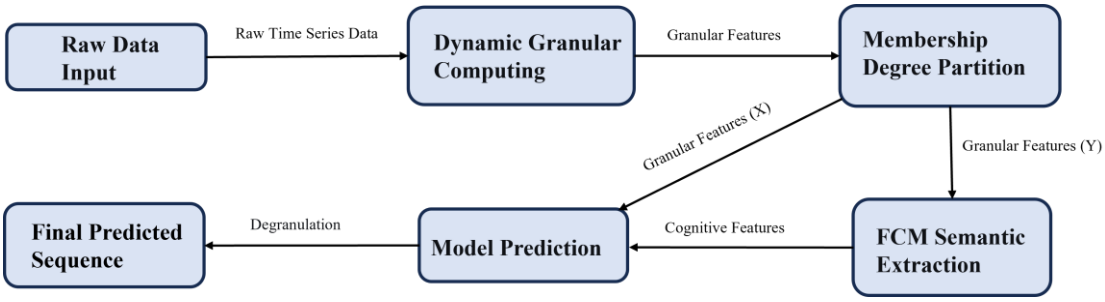

Figure 3. Overall Flowchart of the Proposed Model

The overall workflow can be summarized into the following five key steps:

Step 1: Information Granulation Input the multivariate time series and apply Dynamic Granular Computing to granulate the raw sequence. This forms granular-level statistical features consisting of the mean, slope, and standard deviation, reducing data complexity and noise while preserving trend information.

Step 2: Semantic Mapping Map the granular statistical features into a Low–Medium–High three-state semantic space to obtain a more compact granular semantic state representation with clear linguistic meaning.

Step 3: Cognitive Feature Extraction Introduce a FCM targeted at the semantic states of the target sequence. Using the Nonlinear Hebbian Learning rule to infer causal relationships between nodes, the model trains a weight matrix and generates a cognitive feature vector that reflects granular semantic associations.

Step 4: Feature Fusion and Prediction Concatenate the granular statistical features with the semantic cognitive features to construct an enhanced feature vector. This is fed into the Ridge Regression prediction module to obtain the predicted statistics for future information granules.

Step 5: De-granulation Reconstruction Perform a de-granulation operation based on the predicted granular parameters. This restores the granular results into a continuous sequence on the original time scale to produce the final prediction output.

This workflow effectively mitigates the problem of error accumulation in multi-step forecasting. Simultaneously, it leverages semantic cognitive features to enhance the model's interpretability regarding the modeling of relationships between granules.

3. Experimental Results and Analysis

The purpose of this chapter is to compare the proposed DGC-FCM-Ridge model with baseline prediction models through a series of comparative experiments, proving the superiority of this model in long-term multidimensional time series prediction tasks, and verifying the effectiveness of each key module in the model through ablation experiments.

3.1. Experimental Setup

3.1.1. Dataset Description

This paper conducts empirical research using the Beijing multi-site air quality public dataset provided by the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/501/beijing). This dataset is a typical high-frequency time series data, containing hourly observation records of several national monitoring stations in Beijing. In this experiment, empirical research was carried out at four stations in the dataset: Aotizhongxin, Changping, Dingling, and Dongsi. The time span is from March 1, 2013, to December 31, 2013 (a total of 7,345 data points), and the original data resolution is an hourly high-frequency sequence. This paper selects PM2.5 concentration as the prediction target sequence, and selects 9 other concentration indicators (fine particulate matter concentration, sulfur dioxide concentration, nitrogen dioxide concentration, carbon monoxide concentration, ozone concentration, temperature, dew point, air pressure, wind speed) as the input features of the model. To address the problem of missing data values, linear interpolation is used for filling to ensure the integrity and continuity of the sequence. In addition, to eliminate the impact of dimensional differences on model training, all feature data are processed by Min-Max normalization to the range of [0, 1], and the dataset is divided into a training set (the first 80%) and a test set (the last 20%) according to chronological order.

3.1.2. Evaluation Metrics

In order to comprehensively evaluate the performance of the proposed model in time series prediction tasks, the prediction metrics used in this experiment are: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination(R^2).

RMSE measures the square root of the ratio of the square

of the deviation between the predicted value and the true value to the number of samples, and is sensitive to large errors. Let T be the prediction horizon length, $Y_{true,t}$ be the true value, and $Y_{rec,t}$ be the predicted value. Its calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_{true,t} - Y_{rec,t})^2} \quad (14)$$

MAE measures the average absolute error of the predicted values, and its calculation formula is as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|y_{true,t} - y_{rec,t}\right| \quad (15)$$

R^2 measures the goodness of fit of the model to the observed values; the closer it is to 1, the better the fitting effect. Its calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(Y_{true,t} - Y_{rec,t})^2}{\sum_{t=1}^{T}(Y_{true,t} - \overline{Y_{true}})^2} \quad (16)$$

3.2. Performance Analysis

The proposed model is evaluated against four widely used baseline methods in time series analysis and machine learning, namely ARIMA, SVR, Ridge, and LSTM, to assess its effectiveness in long-term multivariate time series forecasting. Meanwhile, to verify the effectiveness of each module, ablation experiments were designed for DGC-Ridge, DGC-LSTM, and DGC-FCM-LSTM.

3.2.1. Baseline Experiment Comparison

For a fair comparison, the prediction horizons of all baseline models are aligned with the effective forecasting span of the DGC-based models. In the dynamic granulation model, the values of the future H information granules are predicted; therefore, in this experiment, the prediction horizon of the classic baseline models is set to the average length of the information granules in the training set. This setting ensures that all models have the same effective prediction horizon on the original high-frequency time scale, thereby achieving a fair evaluation of the models.

Furthermore, to prevent bias caused by hyperparameter selection, the regularization parameters for the Ridge regression model and SVR, as well as the learning rate and number of hidden units for LSTM, were determined through empirical tuning and cross-validation. The ARIMA model adopts a fixed order of (2, 1, 2) to maintain training stability and prevent overfitting. All parameters of the DGC-FCM-Ridge model (including granulation windows, slope thresholds, and FCM learning rates, etc.) were obtained through systematic search on the validation set. All models are compared under the same inputs, outputs, sample partitioning, and evaluation metrics, ensuring the objectivity and reproducibility of the experimental conclusions. The experimental results are shown in Table 1-4:

Table 1. Comparison of baseline experimental results for the Aotizhongxin station (1,457 hourly samples)

Models	R^2	MAE	RMSE
ARIMA	-0.0239	60.5874	95.2201
SVR	-0.0125	57.3361	93.5749
LSTM	0.0025	55.5503	78.4270
Ridge	-0.0052	56.9272	79.9308
DGC-FCM-Ridge	0.5717	37.0774	52.8173

Table 2. Comparison of baseline experimental results for the Changping station (1,457 hourly samples)

Models	R^2	MAE	RMSE
ARIMA	-0.0173	60.8845	70.6432
SVR	0.0192	45.5223	66.7633
LSTM	0.0005	44.0859	62.3036
Ridge	-0.0087	45.0130	64.8021
DGC-FCM-Ridge	0.5936	34.6450	47.8501

Table 3. Comparison of baseline experimental results for the Dingling station (1,457 hourly samples)

Models	R^2	MAE	RMSE
ARIMA	-0.0013	60.5874	95.2201
SVR	-0.0482	57.3361	93.5749
LSTM	0.0294	55.5503	78.4270
Ridge	-0.0011	56.9272	79.9308
DGC-FCM-Ridge	0.5423	37.0774	52.8173

Table 4. Comparison of baseline experimental results for the Dongsi station (1,457 hourly samples)

Models	R^2	MAE	RMSE
ARIMA	0.0038	56.4314	64.8449
SVR	0.0113	39.0943	61.6921
LSTM	0.0425	39.5331	60.6640
Ridge	0.0317	38.9723	59.9136
DGC-FCM-Ridge	0.6021	35.6312	46.1753

Based on the comparison of baseline experimental results across four different monitoring stations (Table 1 to Table 4), DGC-FCM-Ridge demonstrates a significant performance advantage in all test scenarios. Across the three metrics—R^2, MAE, and RMSE—the performance of DGC-FCM-Ridge substantially outperforms traditional baseline models (ARIMA, SVR, LSTM, and Ridge). This indicates its superior ability to capture complex, nonlinear features within the raw temporal data. Furthermore, this performance improvement validates the effectiveness of the DGC and FCM semantic extraction mechanisms, which are capable of transforming raw data into more cognitively meaningful granular features, thereby enhancing both the prediction accuracy and the interpretability of the model.

3.2.2. Ablation Study Analysis

To verify the effectiveness of each module in DGC-FCM-Ridge, this section conducts an ablation study. Using the Aotizhongxin dataset and the same parameter settings as the baseline experiments, three model variants were constructed to verify the independent and synergistic contributions of the dynamic granular computing mechanism and the FCM semantic enhancement mechanism:

DGC-LSTM: A deep learning model based solely on dynamic granular computing;

DGC-Ridge: A ridge regression model based solely on dynamic granular computing;

DGC-FCM-LSTM: A model that introduces FCM semantic features on top of DGC-LSTM.

Through comparison with these three variants, the mechanism of action of granular abstraction and semantic enhancement on prediction accuracy and stability can be clearly analyzed. The numerical comparison of the four models across R^2, RMSE, and MAE metrics is shown in Table 5:

Table 5. Ablation Study Analysis (1,457 hourly samples)

Models	R^2	MAE	RMSE
DGC-LSTM	0.4132	40.1083	61.8208
DGC-FCM-LSTM	0.3838	40.7891	63.3526
DGC-Ridge	0.5557	37.6678	53.7961
DGC-FCM-Ridge	0.5717	37.0774	52.8173

The results of the ablation study in Table 5 demonstrate the impact of different module combinations on model performance. The base model, DGC-LSTM, shows mediocre performance. When FCM semantic features are added to this structure (DGC-FCM-LSTM), the overall metrics do not improve and actually show a slight decline. However, when the prediction model is replaced with Ridge Regression (DGC-Ridge), performance improves significantly. This indicates that within the granular feature space, linear models can more stably utilize statistical features—such as mean, slope, and standard deviation—compared to deep learning models. With the further introduction of FCM semantic feature extraction (DGC-FCM-Ridge), all three evaluation metrics reach their optimal levels. Specifically, the R^2 increases to 0.5717, while MAE and RMSE decrease to 37.0774 and 52.8173, respectively. This demonstrates that semantic features provide additional trend information for the Ridge Regression model, enabling it to more effectively characterize relationships across information granules, thereby enhancing the accuracy and stability of multi-step prediction.

Figures 5–8 present comparison curves between predicted and observed PM2.5 sequences for the four models, based on 1,457 hourly samples, to illustrate their prediction behaviors. It can be observed that the enhanced model shows improved smoothness and trend consistency during volatile periods:

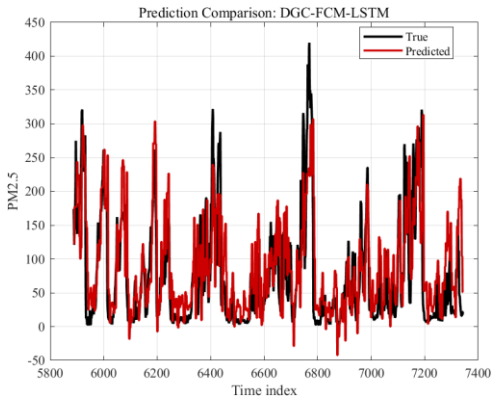

Figure 4. Predicted vs. Actual Values for DGC-FCM-LSTM

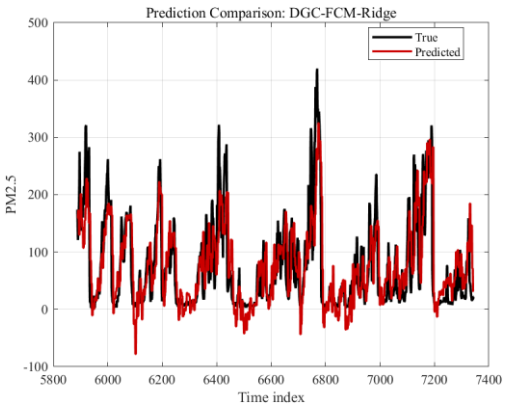

Figure 5. Predicted vs. Actual Values for DGC-FCM-Ridge

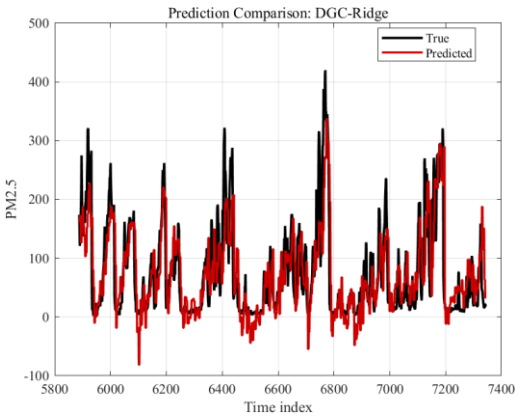

Figure 6. Predicted vs. Actual Values for DGC- Ridge

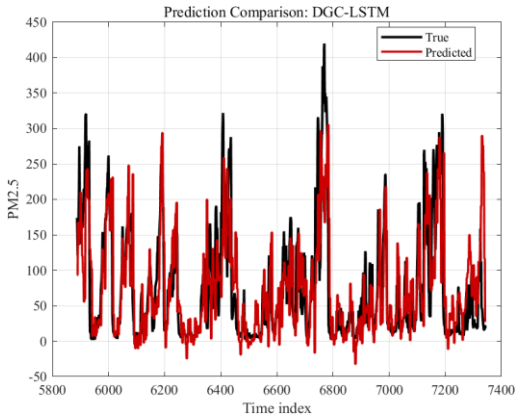

Figure 7. Predicted vs. Actual Values for DGC- LSTM

Meanwhile, Figures 9–11 display the error boxplots for the feature layers (Mean, Slope, and Std). These reflect that the proposed model significantly reduces the error standard deviation and the proportion of outliers through semantic regularization.

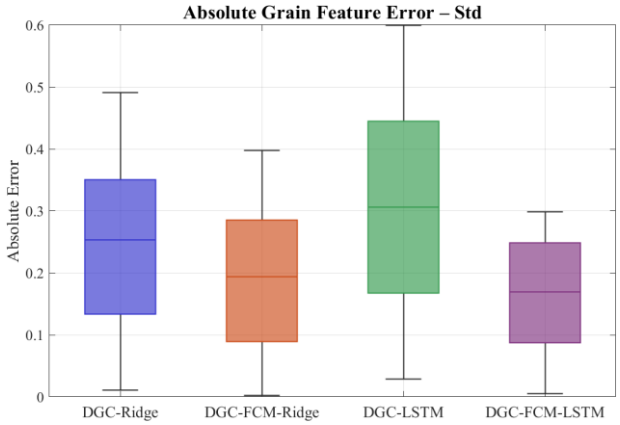

Figure 8. Absolute Error Comparison of Granular Features: Standard Deviation

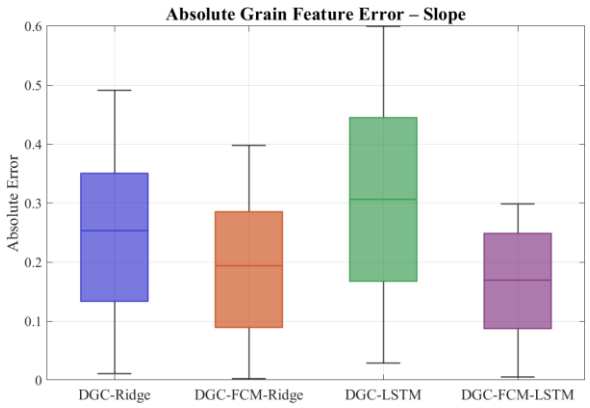

Figure 9. Absolute Error Comparison of Granular Features: Slope

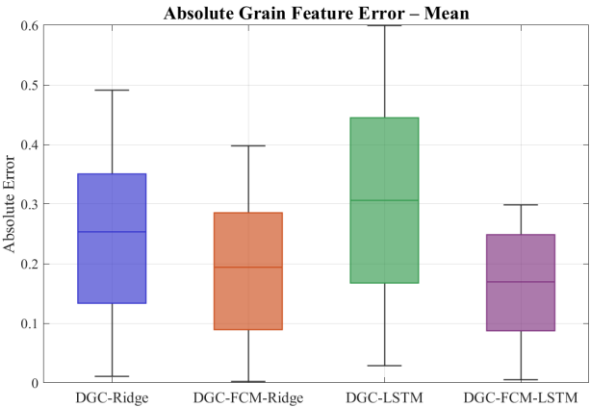

Figure 10. Absolute Error Comparison of Granular Features: Mean

Figures 8–10 illustrate the prediction error distributions across the granular layers (Mean, Slope, and Std). In these boxplots, the DGC-FCM-LSTM model exhibits a more concentrated error distribution with fewer outliers, indicating higher overall stability. However, this model reacts weakly to peak and valley fluctuations, showing a degree of underfitting, which results in a lower R^2 compared to DGC-FCM-Ridge. In contrast, DGC-FCM-Ridge better tracks the trends of the actual sequences, achieving the highest R^2, though its error is slightly larger in highly volatile regions. This suggests that error stability and trend-fitting capability are complementary metrics; DGC-FCM-Ridge is better suited for long-term trend prediction tasks.

3.3. Sensitivity Analysis of Key Parameters

This section explores the impact of the slope threshold δ on the model's predictive performance. As shown in Figure 11, the experiment tests the effect on R^2 within the range of $\delta \in [0.1, 0.6]$, while keeping the minimum window size $W_{min}=3$ and the prediction horizon $H=2$ fixed.

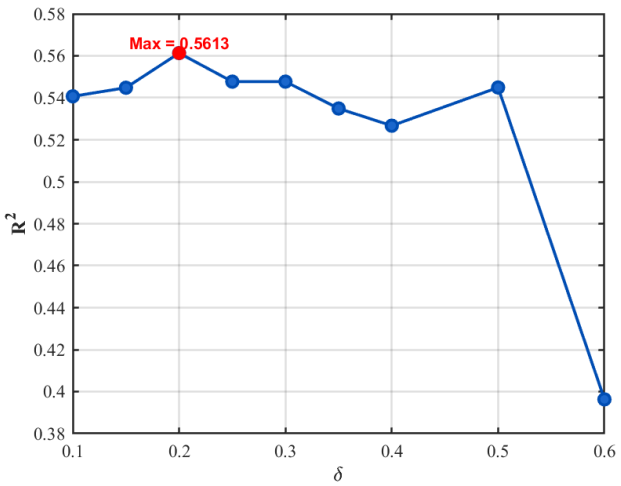

Figure 11. R^2 as a function of the slope threshold δ

Figure 11 shows that the slope threshold δ significantly impacts the model's R^2 performance. When δ is too small (e.g., 0.1), granulation is overly aggressive, resulting in too many short information granules and poor performance. As δ increases toward 0.2, R^2 reaches its peak, indicating an optimal balance between information compression and preservation. Further increasing δ (e.g., beyond 0.5) leads to overly conservative granulation, where the granules are too few and too long to capture critical turning points, causing R2 to decline rapidly.

3.4. Analysis of Model Error Characteristics

A histogram of the differences between predicted and actual values is constructed (see Figure 12) to evaluate the overall error distribution during temporal reconstruction and the prediction accuracy at the information granule level. Additionally, boxplots were generated for the prediction errors of the mean, slope, and standard deviation of the information granules (see Figure 13).

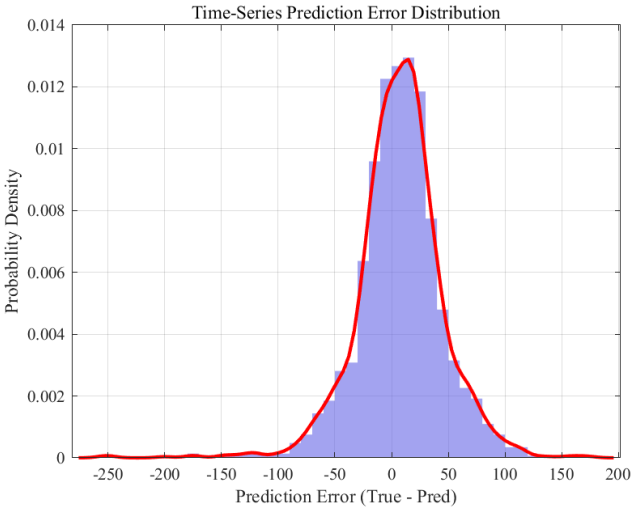

Figure 12. Error Distribution Plot

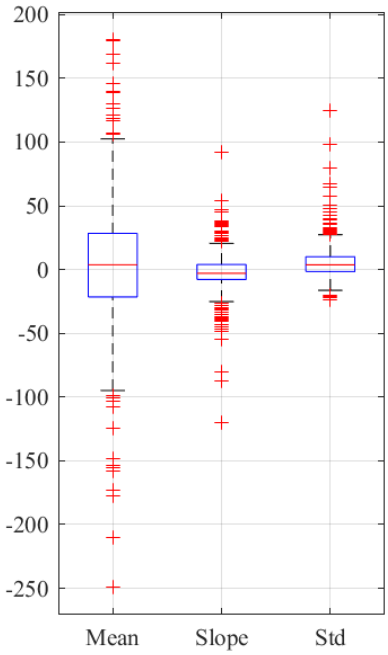

Figure 13. Boxplot of Granular Prediction Errors

Figure 12 shows that the error approximately follows a zero-mean symmetric distribution and is primarily concentrated within the ±100 range, indicating that the model has no systemic bias and maintains stable, consistent statistical characteristics. Figure 13 presents the prediction error distribution for the granular outputs, where the Mean dimension exhibits a taller box and a wider range of outliers, suggesting some volatility in mean prediction. In contrast, the Slope and Standard Deviation dimensions have narrower boxes and lower outlier magnitudes, showing the model's superior stability in capturing local trends and fluctuations. All three error types have medians near zero and symmetric distributions, implying that errors arise mainly from the local dynamic differences of the samples.

3.5. Discussion on Model Interpretability

An interpretability analysis is conducted from two perspectives—semantic structure and feature contribution—to examine the internal mechanisms and feature rationality of the proposed model. The FCM module describes the causal relationships between semantic nodes via a weight matrix W; Therefore, a heatmap of this weight matrix is plotted to demonstrate semantic structural characteristics. Meanwhile, the linear weights of the Ridge Regression module reflect the degree of contribution of each input feature to the prediction results; thus, a feature contribution bar chart is plotted for quantitative analysis. Together, these two types of visualizations reveal the model's interpretability from both the structural and feature levels.

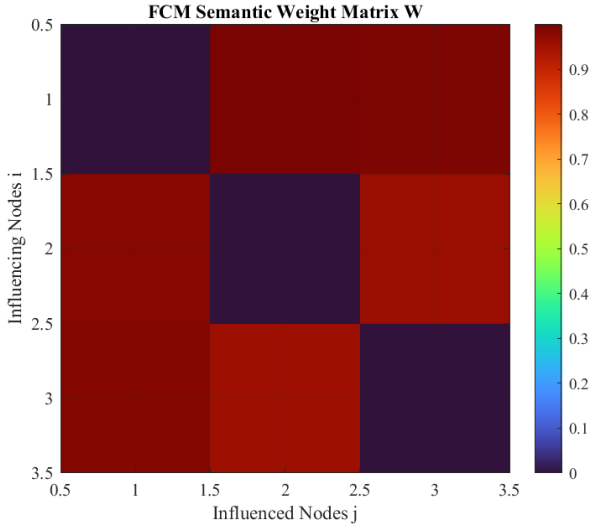

Figure 14. Heatmap of FCM Weight Matrix

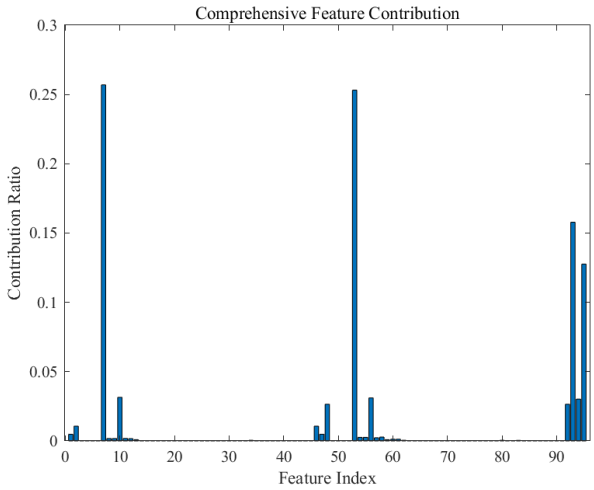

Figure 15. Distribution of Feature Contributions

The decision-making mechanism of DGC-FCM-Ridge can be explained through the visualization of the FCM semantic weight matrix and the Ridge Regression feature contributions. The FCM weight matrix heatmap (Figure 14) shows high values in non-diagonal elements, reflecting significant coupling between semantic features; this indicates the model's ability to capture the mutual influence of internal granular features, providing a more interpretable representation for the Ridge Regression. The feature contribution distribution (Figure 15) exhibits a clear sparse structure, showing that the prediction process primarily relies on a few key granular statistical features. In the semantic-enhanced feature region (indices beyond 90), the higher weights demonstrate that semantic features extracted by the FCM provide complementary cognitive information, helping to capture nonlinear associations and enhance model robustness. Overall, the model achieves a sparse feature representation while maintaining semantic awareness, effectively fusing numerical patterns with cognitive characteristics.

4. Conclusion and Future Work

(1) To address error accumulation and limited interpretability in long-term multivariate time series forecasting, this study proposes the DGC-FCM-Ridge model. A trend-driven dynamic granulation strategy is introduced to extract robust granule-level statistical features, which

effectively suppress noise and enhance the stability of long-horizon prediction.

(2) A fuzzy cognitive map is employed to model causal relationships among granules. The resulting cognitive features complement the statistical granule features, enabling the model to capture latent inter-granular dependencies that are difficult to represent explicitly. By integrating granule-level statistical features and FCM-based cognitive features within a regression framework, the proposed model achieves stable and consistent forecasting performance, while avoiding the prediction instability and overfitting risks commonly observed in deep learning–based long-term forecasting models.

(3) Extensive experimental results demonstrate that the proposed approach outperforms multiple baseline methods, including ARIMA and LSTM, across various evaluation metrics in terms of both prediction accuracy and reliability. Furthermore, in the reduced feature space obtained through granulation, simple linear predictors exhibit superior generalization performance compared with complex nonlinear models, highlighting the effectiveness of combining fine-grained representations with interpretable modeling structures for long-term forecasting.

Future work can be extended in three directions. First, global optimization strategies, such as hybrid particle swarm optimization algorithms, can be incorporated to improve the convergence behavior of FCM weight learning. Second, the proposed framework can be extended to multi-objective forecasting to better capture interactions among multiple variables. Third, by integrating online learning mechanisms with visual causal analysis, an intelligent forecasting system with real-time updating and dynamic monitoring capabilities can be developed, providing coherent and interpretable solutions for complex tasks in environmental monitoring and industrial process control.

Acknowledgments

This work was supported by the Key Research Projects Plan of Henan Provincial Colleges and Universities [13250045]

Natural Science Foundation of Henan Province [252300423319]

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Zhitao Jia: Conceptualization, Investigation, Methodology, Software, Writing – original draft.

Lianchao Qiu: Funding acquisition, Supervision, Writing – review & editing.

References

[1] L. Zhao, Z. Li and L. Qu, "Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition," Heliyon, vol. 8, no. 12, e12239, Dec. 2022.

[2] F. Xiao, M. Yang, H. Fan, G. Fan and M. A. A. Al-qaness, "An improved deep learning model for predicting daily PM2.5 concentration," Scientific Reports, vol. 10, Art. no. 20988, Dec. 2020.

[3] Z. Liu, D. Ji and L. Wang, "PM2.5 concentration prediction based on EEMD-ALSTM," Scientific Reports, vol. 14, Art. no. 12636, Jun. 2024.

[4] Zhao, B., Wang, J., & Yu, X. Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on optimized decomposition and parameter selection. Heliyon, 2022, 8(5): e09456.

[5] Yang, H., Zhou, J., Li, Y., & Liu, W. Time-Series Forecasting of PM2.5 and PM10 Concentrations Based on the Integration of Surveillance Images. Sensors, 2025, 25(3): 8374.

[6] Rani, M., Rajasekar, P., & Sinha, P. Forecasting PM2.5 concentrations using statistical modeling for Bengaluru and Delhi regions. Environmental Monitoring and Assessment, 2023, 195(4): 456.

[7] X. Ao, J. Zhang, and L. Chen, "Analysis and prediction of PM2.5 concentration based on seasonal time series models," Journal of Environmental Sciences, vol. 41, no. 7, pp. 2874–2883, 2021.

[8] Kim,S.,&Park,J.Long-term air pollution forecasting using Facebook Prophet model: A case study of Seoul.Atmospheric Environment, 2020, 224: 117307.

[9] Chang, L., Zhang, Z., Zhao, Y., & Chen, W. (2020). An LSTM-based aggregated model for air pollution forecasting. Environmental Science and Pollution Research, 27, 38827–38839.

[10] Qi Y, Li Q, Karimian H, Liu D. A hybrid model for spatiotemporal forecasting of PM2.5 based on wavelet transform and long short-term memory network. Science of The Total Environment, 2019, 664: 1–10.

[11] Li, X., Liu, Y., & Wang, H. (2021). Research on PM2.5 spatiotemporal forecasting model based on LSTM neural network. Environmental Monitoring and Assessment, 193, 1–13.

[12] Bai, Y., Zhuang, Y., & Wang, Y. (2024). Prediction of PM2.5 concentration based on a CNN-LSTM neural network algorithm. Scientific Reports, 14, 1123.

[13] Pedrycz W. Granular computing: An introduction. Human-Centric Information Processing Through Granular Modelling, Springer, 2008: 1–17.

[14] Yao Y, Lin T Y. Granular computing: Basic issues and possible solutions. Rough Sets and Current Trends in Computing, Springer, 1999: 46–57.

[15] Artiemjew, P. (2020). About Granular Rough Computing — Overview of Decision System Approximation Techniques and Future Perspectives. Algorithms, 13(4), 79.

[16] Pedrycz W, Homenda W. Building the fundamentals of granular computing: A principle of justifiable granularity. Applied Soft Computing, 2013, 13(10): 4209–4218.

[17] Yin Y. Prediction and analysis of time series data based on granular computing and support vector machines[J]. Frontiers in Computational Neuroscience, 2023, 17: 1192876.

[18] Song M., Wang R., Li Y. Hybrid time series interval prediction by granular neural network and ARIMA[J]. Granular Computing, 2023, 9(3): 1–12.

[19] Stach W, Kurgan L A, Pedrycz W. Expert-based and computational methods for developing fuzzy cognitive maps. International Journal of Approximate Reasoning, 2005, 39(2–3): 149–192.

[20] Wang, X., et al. (2020). Deep Fuzzy Cognitive Maps for Interpretable Multivariate Time Series Forecasting. IEEE Trans. Fuzzy Systems (2020).

[21] Z. Peng, W. Liu and S. -K. Oh, "AFS-FCM With Memory: A Model for Air Quality Multi-Dimensional Prediction With Interpretability," in IEEE Transactions on Big Data, vol. 11, no. 2, pp. 810-820, April 2025.

[22] Z. Peng, W. Liu, and S. An, “Haze pollution causality mining and prediction with PS-FCM,” Information Sciences, vol. 523, pp. 307–317, 2020.