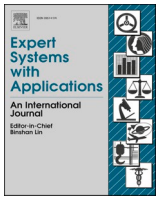
Expert Systems With Applications 245 (2024) 123008

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Application of complete ensemble empirical mode decomposition based multi-stream informer (CEEMD-MsI) in $PM_{2.5}$ concentration long-term prediction

Qinghe Zheng [a,*], Xinyu Tian [a], Zhiguo Yu [a], Bo Jin [d], Nan Jiang [b,c], Yao Ding [e], Mingqiang Yang [f], Abdussalam Elhanashi [g], Sergio Saponara [g], Kidiyo Kpalma [h]

[a] School of Intelligent Engineering, Shandong Management University, Jinan 250357, China
[b] Qingdao Research Academy of Environmental Science, Qingdao Municipal Bureau of Ecology and Environment, Qingdao 266003, China
[c] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430000, China
[d] Institute of Systems and Robotics (ISR), Department of Electrical and Computer Engineering (DEEC), University of Coimbra, Coimbra 3030-290, Portugal
[e] Key Laboratory of Optical Engineering, Xi'an Research Institute of High Technology, Xi'an 710025, China
[f] School of Information Science and Engineering, Shandong University, Qingdao 266237, China
[g] Department of Information Engineering, University of Pisa, Pisa 56122, Italy
[h] Department of Electronics and Industrial Informatics, National Institute for Applied Sciences of Rennes, Rennes F-35000, France

ARTICLE INFO

Keywords:
Environmental modelling
Pollution management
$PM_{2.5}$ prediction
Complete ensemble empirical mode decomposition (CEEMD)
Multi-stream informer (MsI)

ABSTRACT

Nowadays, air pollution has become one of the most serious environmental problems facing humanity and an inescapable obstacle limiting the sustainable development of cities and society. Although air quality sensing and management systems based on artificial intelligence and signal analysis are evolving as essential parts of intelligent cities, the mixture of local emission sources and regional transport of air pollutants still makes $PM_{2.5}$ long-term prediction challenging, especially under complex geographical and climatic conditions. In this paper, the complete ensemble empirical mode decomposition based multi-stream informer (CEEMD-MsI) is proposed to predict the hourly $PM_{2.5}$ concentration, and extensive testing and comparisons are carried out in four typical cities in Shandong, China. Firstly, CEEMD is used for signal pre-processing to construct the intrinsic mode functions (IMFs) based multi-channel representations. Then MsI is specifically designed to learn both temporal and spatial features, and complete the $PM_{2.5}$ concentration prediction. To the best of our knowledge, this is the first attempt to predict long-term $PM_{2.5}$ concentrations using a deep learning model driven by data collected from monitoring stations spanning long distances and diverse terrains. Finally, test results demonstrate that CEEMD-MsI achieves the best $PM_{2.5}$ prediction performance by comparing it with state-of-the-art methods.

1. Introduction

The sustainable development aims to maintain the natural systems' capacity that cities and society rely on (Yu et al., 2022; Han et al., 2020). Adhering to sustainable development principles can ensure the preservation of natural resources for future urban and social development while simultaneously maximizing present benefits (Teng et al., 2022). It seeks to strike the balance between environmental, social, and economic needs, promoting sustainable cities and society (Reyers et al., 2022). Despite the significant economic growth and urbanization experienced over the past few decades, there has been a corresponding increase in energy consumption and pollutant emissions, resulting in severe air pollution issues and impeding the sustainable development. Extensive studies (Lu et al., 2021; Forouzanfar et al., 2016; Yang et al., 2021) have demonstrated that atmospheric particulate matter has posed serious threats to both ecosystems and human health, even at relatively low concentrations. In particular, $PM_{2.5}$, which comprises toxic and hazardous airborne particles with an aerodynamic diameter < 2.5 μm, inevitably penetrates into the respiratory and cardiovascular systems, thereby increasing the probability of suffering from diabetes, cardiovascular disease, and lung disease (Lu et al., 2021). $PM_{2.5}$ has been ranked as the sixth leading global risk factor for death and disability by

* Corresponding author.
E-mail addresses: zqh@sdmu.edu.cn (Q. Zheng), txy@sdmu.edu.cn (X. Tian), yzg@sdmu.edu.cn (Z. Yu), jin.bo@isr.uc.pt (B. Jin), jn1992@whu.edu.cn (N. Jiang), yangmq@sdu.edu.cn (M. Yang), a.elhanashi@studenti.unipi.it (A. Elhanashi), sergio.saponara@unipi.it (S. Saponara), kidiyo.kpalma@insa-rennes.fr (K. Kpalma).

https://doi.org/10.1016/j.eswa.2023.123008
Received 3 July 2023; Received in revised form 14 December 2023; Accepted 19 December 2023
Available online 29 December 2023

World Health Organization (WHO) (Forouzanfar et al., 2016). From 1990 to 2019, the number of deaths from $PM_{2.5}$-related illnesses has surged by over 90 % worldwide (Yang et al., 2021). Approximately 7 million people die each year due to ongoing exposure to air pollution (Ma et al., 2019). Besides, the exacerbation of air pollution has facilitated the transmission of COVID-19 (Xia et al., 2019). Empirical evidence (Zoran et al., 2020) has established a significant association between $PM_{2.5}$ concentrations and the incidence and fatality rates of COVID-19. Currently, the annual average and 24-hour period $PM_{2.5}$ concentrations in many regions continue to exceed the levels of 10 μg/m^3 and 25 μg/m^3 recommended by WHO, respectively (Apte et al., 2015). In China, only less than 1 % of 500 large cities are able to meet the $PM_{2.5}$ standard and the annual concentration reaches 59 μg/m^3 (Xiao et al., 2020). This alarming situation and trend strongly indicates the urgent need for sustainable solutions to mitigate the adverse effects of economic development and urbanization on the environment and society. In addition to reducing air pollutant emissions from the source, accurate long-term prediction of $PM_{2.5}$ concentrations is one of the most effective ways to control and prevent environmental pollution. By utilizing the predicted results of $PM_{2.5}$ concentrations, targeted measures can be taken to provide practical guidance for environmental pollution control and sustainable urban planning.

Actually, the formation principle and the dissemination process of $PM_{2.5}$ are extremely complex due to its intricate properties, e.g., spatio-temporal nonlinearities (Fang et al., 2020). It is widely known that the formation of $PM_{2.5}$ is typically attributed to a diverse range of pollutants and is heavily influenced by meteorological conditions and their synergistic effects (Sun et al., 2014). In certain scenarios, the impact of meteorological factors on $PM_{2.5}$ levels surpasses that of pollutant emissions (Liang et al., 2016). The seasonal and diurnal fluctuations of $PM_{2.5}$ concentrations exhibit multiple patterns, which are subject to uncertainties in boundary and initial conditions, physical/chemical processes, and emission sources (Wu et al., 2015). Furthermore, the $PM_{2.5}$ concentrations can be influenced by the accumulation of preceding $PM_{2.5}$ thickness (Seng et al., 2021). From the geographical perspective, the $PM_{2.5}$ concentration at a specific location is influenced by both internal and external sources (Fang et al., 2020). Internal sources may comprise of direct emissions and irregular products caused by photochemical effects, which typically result in rapid increase of $PM_{2.5}$ concentrations over a short period of time. By comparison, external sources may originate from remote areas and frequently generate more subtle but persistent impacts. In particular, $PM_{2.5}$ can be easily transmitted by monsoons, leading to long-term but periodic pollution. Environmental factors including temperature, wind speed, wind direction, atmospheric pressure, and rainfall may either facilitate or hinder the diffusion of $PM_{2.5}$ (Fang et al., 2020). In general, the accurate modelling and prediction of $PM_{2.5}$ concentrations across different regions is still an extremely challenging task due to their dynamic property, i.e., the prominent non-linear and non-stationary changes in various cities.

At present, approaches for predicting $PM_{2.5}$ concentrations are primarily classified into three categories: deterministic methods, statistical modelling methods, and machine learning methods. The deterministic methods based on chemical transport modelling (CTM-DMs) attempt to explore potential complex relationships between geography, meteorology, chemistry, and $PM_{2.5}$ emissions by constructing transport and diffusion models of $PM_{2.5}$. Typical CTM-DMs include the community multi-scale air quality (CMAQ) system (Cheng et al., 2021), the weather research and forecasting method (WRF) (Cao et al., 2018), and coupled WRF-chemically model (WRF-Chem) (Hong et al., 2020). However, CTM-DMs typically require key information about the source and generation principle of pollutants, otherwise they may lead to errors. Moreover, the actual parameters involved in the diffusion mechanism, transportation, and deposition processes of pollutant sources are uncertain and difficult to measure, which affects the prediction performance. The insufficient information on the pollution sources, poor representation of physicochemical processes, and high computational costs limit the implementation of CTM-DMs. Therefore, despite offering valuable insights into $PM_{2.5}$ dispersion mechanisms and establishing the direct correlation between pollutant emissions and air pollution, the practical application of CTM-DMs remains restricted. Statistical modelling methods (SMMs) aim to establish the relationship between various types of imapct factors and $PM_{2.5}$ concentrations by designing regression models without introducing complex physical parameters. SMMs mainly include autoregressive model (ARM) (Garcia et al., 2016), weighted linear regression (WLR) (Jiang et al., 2021), autoregressive integrated moving average (ARIMA) (Zhang et al., 2018), geographically weighted regression (GWR) (Zhou and Lin, 2019), hidden Markov models (HMMs) (Xu and Wang, 2015), and so on. Given sufficient historical data observed from monitoring stations, regression models hold the advantages of being easier, faster, and more economical to implement than CTM-DMs. In addition, SMMs are capable of capturing potential site specific correlations between $PM_{2.5}$ concentrations and selected input variables (Xu and Wang, 2015). Nevertheless, SMMs driven by linear, generalized linear, or nonlinear regressions usually tend to oversimplify this relationship. HMMs have some inherent drawbacks like exorbitant training costs and sensitivity to initial conditions. In addition, a lot of nonlinear modeling approaches like heuristic modeling (Pozna et al., 2010), expert prior knowledge based modeling (Ketsarapong et al., 2012), and tensor product based model transformation (Hedrea et al., 2021) are gradually being applied. Nonlinear modelling can better capture the mapping relationship between inputs and predicted values. As methods to automatically optimize internal parameters, machine learning (ML) models based on feature extraction are capable of capturing the evolution of big data and provide valuable implicit expert knowledge, such as ensemble random forest (ERF) (Jiang et al., 2020), support vector machine (SVM) (Xu et al., 2017), and neural networks (NNs) (Li et al., 2020). In contrast, ML methods are more intuitive and easy to understand. ML based regression methods have improved the accuracy in a series of applications, including early-stage estimation (Jeong et al., 2010) and correlation analysis (Precup et al., 2022). Unfortunately, the performance of these commonly used ML models in predicting $PM_{2.5}$ concentrations is still unsatisfactory. RF models (Jiang et al., 2020) are prone to over-fitting when faced with long-period data. Since the parameters of SVR models (Xu et al., 2017) play a key role in the prediction accuracy, a large amount of computing resources are required to fine-tune them. Shallow neural networks (Li et al., 2020; Jin & Vai, 2015) are hard to be well optimized and generalized on big data. Besides, most existing ML-related work neglects the design of effective features and the key preprocessing techniques of the observation data, making it difficult to effectively capture the variation patterns of $PM_{2.5}$ time series and limiting the prediction accuracy.

As the quintessential algorithm in machine learning, end-to-end deep neural networks have shown excellent performance in sequence analysis across multiple tasks, including wireless communication (Zheng et al., 2021a), spectrum sensing (Zheng et al., 2023b; Zheng et al., 2021b), structure prediction (Zheng et al., 2022), and natural language processing (NLP) (Jin et al., 2020). Deep learning models, such as convolutional neural network (CNN) (Li et al., 2020; Zhao et al., 2023a; Zheng et al., 2018), attention-based parallel network (APNet) (Zhu et al., 2021), long short-term memory (LSTM) (Gao et al., 2021; Zheng et al., 2020), and graph attention transformer (GAT) (Tan et al., 2022), can effectively extract the crucial information and establish an accurate

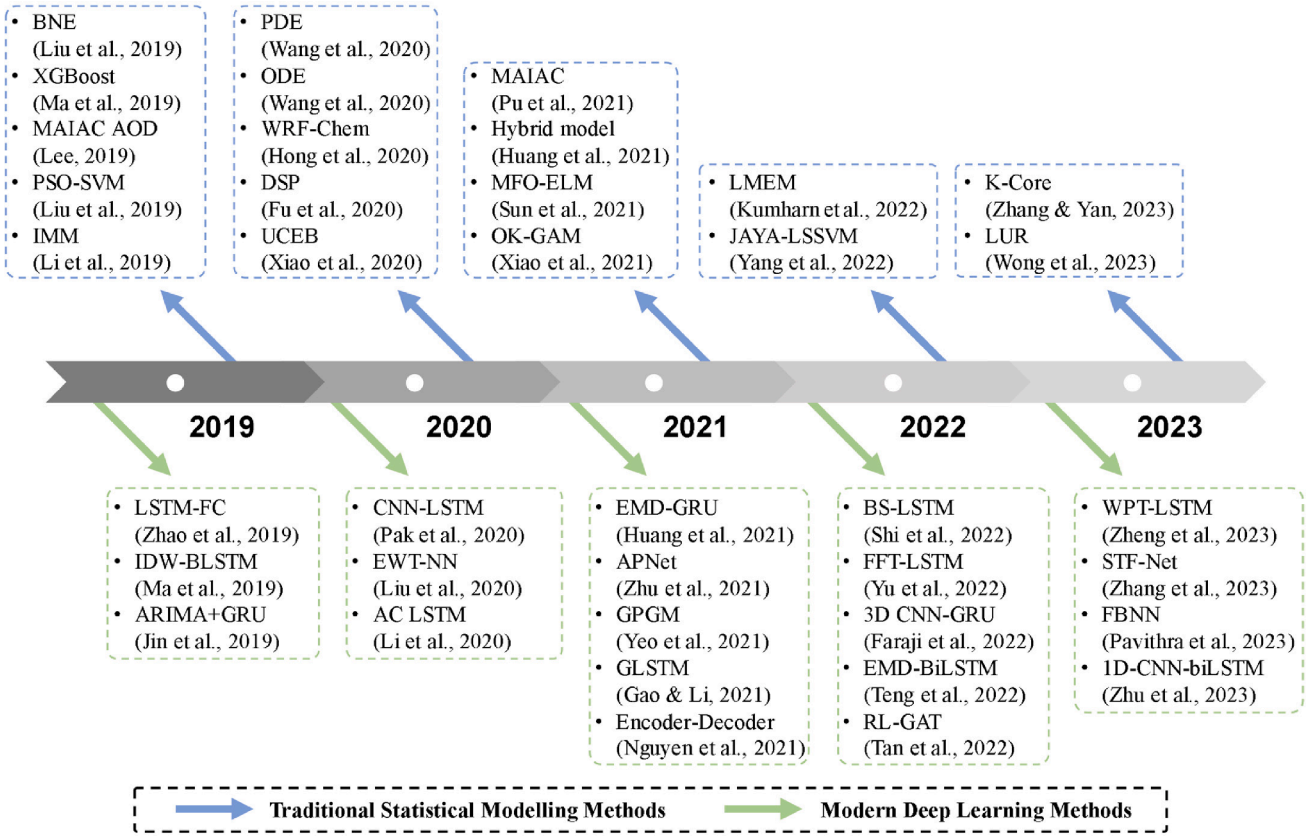


Fig. 1. The cornerstone methods for $PM_{2.5}$ concentration prediction in recent years.

mapping between input variables and $PM_{2.5}$ concentrations through the utilization of multi-layer nonlinear transformations. Consequently, deep learning approaches can substantially alleviate the requirement for conducting feature engineering, while exhibiting excellent performance in handling intricate datasets. When integrating data observed from numerous related sites into a matrix, the elements are associated via spatial correlation. CNN is suitable for extracting spatial information from multiple sites and predicting $PM_{2.5}$ concentrations, but tends to ignore the temporal information. For example, Zheng et al., (2023a) proposed using wavelet packet transform to assist in analyzing temporal information, but still faced the long-term dependence problem due to the fact that wavelet transform balances the resolution of both time and frequency domains. The LSTM overcomes the long-term dependence of time series analysis, but its efficient optimization is still a technical challenge. Some hybrid models like CNN-LSTM (Li et al., 2020), CNN-GRU (Faraji et al., 2022), and 1D-CNN-biLSTM (Zhu et al., 2023) are designed to help extract spatio-temporal features however, these models generally possess high structural complexity and correspondingly low reasoning efficiency. The compression and lightweighting of the model is worth investigating, considering the inference efficiency in practical applications (Verma et al., 2022). Furthermore, the choice of inputs for deep neural networks is a crucial but easily overlooked problem, as unrelated or noisy variables may lead to unnecessary complex loss landscapes and corresponding poor generalization (Chinatamby & Jewaratnam, 2023; Zhao et al., 2023b). Although studies have shown the impact brought by signal representations (Zheng et al., 2023b) and the enhancement of deep learning model's generalization capability by means of regularization techniques (Zheng et al., 2018), there is no targeted application for long-term prediction of $PM_{2.5}$ concentrations. Besides, some data augmentation methods like spectrum interference (Zheng et al., 2021) are challenging to directly transfer for the expansion of $PM_{2.5}$ signals. Despite recent progress in applying deep learning to $PM_{2.5}$ concentration prediction, significant efforts still need to be made to take full advantage of deep learning and break through the limitations of existing prediction methods.

To address the aforementioned issues, we propse a complete ensemble empirical mode decomposition based multi-stream informer (CEEMD-MsI) model to predicte the long-term $PM_{2.5}$ concentration. Extensive testing and comparisons are carried out in four typical cities in Shandong, China. In the proposed method, CEEMD is first used for signal pre-processing to construct the multi-channel representations based on intrinsic mode functions (IMFs). Appropriate signal representation is beneficial for the convergence of the objective function and the optimization of model parameters. Then a deep learning model integrating ProbSparse self-attention modules, i.e. MsI, is designed to learn both temporal and spatial features, and therefore to overcome the long-term dependence and to improve the generalization ability under various conditions. To the best of our knowledge, this is the first attempt to predict long-term $PM_{2.5}$ concentrations using the deep learning model driven by data collected from observation stations that span long distances and complex terrains.

Main contributions and advantages of this study are summarized as follows.

- As an efficient technique for parallel pre-processing of time series, the CEEMD is successfully adopted for frequency domain analysis of long-term $PM_{2.5}$ concentrations. It can not only solve the mode pairing problem in joint analysis of multivariate variables, but also

combine the multivariate input and component decomposition simultaneously to overcome the nonlinearity and non-smoothness of signals. Besides, the introduction of random disturbances in CEEMD suppresses the influence of noises and improves the accuracy and stability of signal decomposition.

- Instead of constructing individual models for each IMF, a single deep learning model, i.e. the multi-stream informer consisted by encoder and decoder, is specifically built to fuse $PM_{2.5}$ prediction results of multiple IMFs in the proposed method, which can significantly reduce training and deployment costs, and improve the prediction accuracy and inference efficiency. The multi-head attention mechanism enables the informer model to adapt to signals of different lengths and frequencies, thus effectively capturing long-term dependencies in temporal sequences.
- Comparative tests at 17 monitoring stations in Shandong under various evaluation metrics are conducted to prove the effectiveness of the proposed CEEMD-MsI. Comparisons with state-of-the-art statistical modelling methods (PDE (Wang et al., 2020), MFO-ELM (Sun et al., 2021), and JAYA-LSSVM (Yang et al., 2022)) and deep learning models (CNN-LSTM (Pak et al., 2020), 3D CNN-GRU (Faraji et al., 2022), and WPT-LSTM (Zheng et al., 2023a)) demonstrate that the proposed method achieves superior performance in predicting long-term $PM_{2.5}$ concentrations.
- To observe the robustness and sensitivity of CEEMD-MsI in practical applications, extensive ablation studies for various hyper-parameters are performed, including the number of averaging times and decomposition iterations in CEEMD, as well as initial learning rate, mini-batch size, weight decay, and structural complexity in MsI. Furthermore, the adaptability of CEEMD-MsI to data observation and sampling periods is also analyzed for meeting the requirements of realistic scenarios. Test results indicate that CEEMD-MsI can generalize well under different conditions.

The remainder of the paper is organized as follows. In Section 2, we briefly review the work related to traditional statistical modelling methods and modern deep learning models for $PM_{2.5}$ concentration prediction. In Section 3, the proposed CEEMD-MsI method for $PM_{2.5}$ long-term prediction is introduced in detail. In Section 4, we report the test results and comparative analysis. Finally, our work and future research plans are summarized in Section 5.

2. Related work

Urbanization has led to the perpetual deterioration of the environment, which dramatically amplifies the necessity for proficient prediction and regulation of air quality, specifically concerning $PM_{2.5}$ concentrations. To better build sustainable cities and society, researchers have proposed a large number of solutions for $PM_{2.5}$ concentration prediction in the past few years, as illustrated in Fig. 1. Evidently, the progression of $PM_{2.5}$ concentration prediction research is transitioning from statistical modelling approaches towards deep learning methods. This shift is attributable to the advancements in optimization algorithms and the development of specialized architectures that possess superior generalization ability across diverse geographies and conditions.

2.1. Traditional statistical modelling methods

Liu et al. (2019) developed a Bayesian non-parametric ensemble (BNE) based $PM_{2.5}$ prediction model, in which the adaptive and calibrated ensemble learning is used to weight the prediction result of each candidate model. BNE provided interpretable uncertainty estimates and it was applied to the Boston, outperforming existing ensemble methods by 30 %-60 % when comparing cross-validation root mean square errors. Ma et al. (Ma et al., 2020) applied the extreme gradient boosting (XGBoost) algorithm to air quality prediction and was able to predict winter heavy pollution more accurately than CTM-DMs. Utilizing of multi-source data including MAIAC and AOD, Lee (2019) computed the yearly average concentration of fine $PM_{2.5}$ at a resolution of 1 km in California. However, the missing values in observation data and the insufficient temporal resolution for individual cities are waiting to be resolved in AOD-derived approaches. Liu et al. (2019) used meteorological pattern analysis to supplement SVM for $PM_{2.5}$ class prediction. Li et al. (2019) proposed the time series and interactive multi-model (IMM) based atmospheric $PM_{2.5}$ concentration prediction algorithm, achieving better accuracy than autoregressive (AR) model and AR-Kalman prediction method. Furthermore, the partial differential equation (PDE) (Wang et al., 2020a) and the ordinary differential equation (ODE) (Wang et al., 2020b) models were designed to describe the combined influence of local emissions, cross-border transmission, and human prevention, helping improve $PM_{2.5}$ prediction accuracy. Empirical and statistical-based schemes (Hong et al., 2020) were used to optimize the estimation of aerosol initial conditions in WRF-Chem and to improve $PM_{2.5}$ prediction accuracy by integrating observation data from various sources. To incorporate the spatio-temporal correlation, Fu et al. (2020) introduced a dynamic spatial panel (DSP) approach to forecast daily serial $PM_{2.5}$ concentrations at various grid points spanning the geographic expanse of mainland China. DSP leveraged aerosol, vegetation, and meteorological remote sensing data as explanatory variables. The singular meteorological traits of urban localities situated in regions with harsh cold conditions often engender a different urban energy configuration and building arrangement compared to other urban areas, resulting in distinctive features of $PM_{2.5}$ pollution. Xiao et al. (2020) developed an urban canopy energy balance (UCEB) model for atypical regions by simulating the dispersion process and vertical concentration distribution of $PM_{2.5}$ in residential areas.

In recent years, numerous studies have been conducted to address the challenges encountered in data processing when utilizing statistical models. Pu and Yoo (2021) constructed a missing data interpolation model to improve the $PM_{2.5}$ prediction accuracy of statistical models at different locations. In (Huang et al., 2021a; Huang et al., 2021b), the bilinear interpolation method was first employed to fill in the blanks of AOD, and then the hybrid spatio-temporal modelling framework was adopted to downscale the daily $PM_{2.5}$ prediction within urban areas to a spatial resolution of 100 m. Sun & Xu (2021) attempted to filter historical data through RF and grey system approximation (GSA) model, and introduced extreme learning machine (ELM) to cope with $PM_{2.5}$ prediction, but the Moore-Penrose generalized inverse makes ELM susceptible to overfitting problems. Xiao et al. (2021) proposed to fill in the missing data by analyzing spatio-temporal trends through ordinary kriging and generalized additive mixture (OK-GAM) model. Kumharn et al. (Kumharn et al., 2022) combined monthly and daily data to fill in the gap for non-retrieval days and applied a linear mixed effects model (LMEM) as the $PM_{2.5}$ forecasting tool. Yang et al. (2022) introduced a Jaya algorithm optimized least square SVM (JAYA-LSSVM) for improving $PM_{2.5}$ concentration prediction accuracy, adopting the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and variational mode decomposition (VMD) as the predecessor steps for data preprocessing. By utilizing the K-Core idea and label distribution, Zhang & Yan (2023) obtained the weights of meteorological and environmental factors affecting $PM_{2.5}$ concentrations in each piece of data, and established multiple models for ensemble prediction. Wong et al. (2023) proposed a mixed spatial model that incorporates Kriging interpolation, land use regression (LUR), ML, and ensemble ideas, to estimate long-term variations $PM_{2.5}$ in both daytime and nighttime for nearly three decades, but its accuracy and reliability for future $PM_{2.5}$ prediction is yet to be verified.

Despite the development of statistical modelling and machine learning models for site-specific $PM_{2.5}$ concentration prediction, the robustness to the ever-changing terrain and climatic conditions, as well as their generalizability and applicability across diverse geographic areas, is yet to be empirically established. The presence of excessive

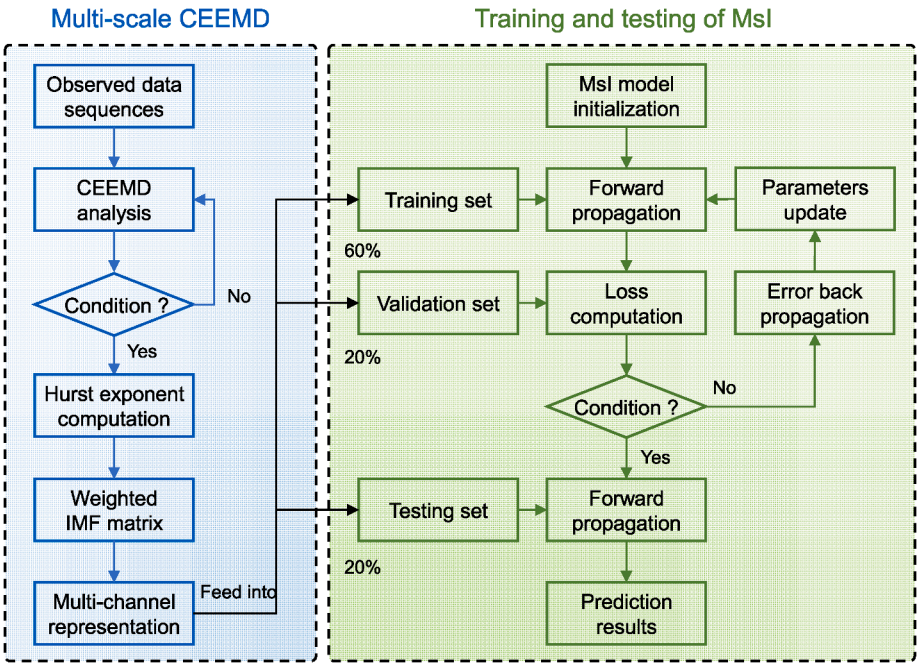


Fig. 2. The overall process of CEEMD-MsI for $PM_{2.5}$ concentration prediction.

regional differences may lead to complete failure of the original statistical model. The challenge of acquiring knowledge about long-term dependencies inherent in time series data is a major obstacle for statistical modelling methods. Typically, conventional statistical modelling techniques are hard to fulfill the practical demands of $PM_{2.5}$ concentration prediction in the context of big data-driven urban environments with growing complexity.

2.2. Modern deep learning methods

Due to excellent big data analysis and mining capabilities, a large number of deep learning methods have been proposed for $PM_{2.5}$ concentration prediction. Zhao et al. (2019) proposed a LSTM-fully connected (LSTM-FC) network that leverages previous air quality, meteorological, and weather forecast data to predict $PM_{2.5}$ concentrations at specific locations within the ensuing 48 h. Ma et al. (2019) constructed a novel framework that combines the inverse distance weighting (IDW) algorithm into the bi-directional LSTM (IDW-BLSTM) for $PM_{2.5}$ prediction at different time granularity. Jin et al. (2019) designed a comprehensive prediction factor that decomposes the original data into three components, i.e. trend, period, and residual, and then used ARIMA and two gated recurrent units (GRUs) to predict each component separately to improve the $PM_{2.5}$ prediction accuracy. Pak et al. (2020) put forth a hybrid spatiotemporal CNN-LSTM model, and used mutual information for spatiotemporal correlation analysis, taking into account both the linear and nonlinear correlations between $PM_{2.5}$ and the observation parameters. In (Liu & Chen, 2020), empirical wavelet transform (EWT) was introduced into neural networks to filter abnormal information in the original $PM_{2.5}$ concentration sequence and reduce data complexity. The utilization of the attention mechanism in the CNN-LSTM culminated in the development of AC-LSTM (Li et al., 2020), which adeptly captures the relevance of temporal states at disparate antecedent time points in predicting forthcoming $PM_{2.5}$ concentrations. In acknowledgement of the non-stationary characteristic of time series, the EMD was incorporated into GRUs (Huang et al., 2021a; Huang et al., 2021b) to enhance the performance of single GRU-based $PM_{2.5}$ concentration prediction. In (Zhu et al., 2021), an attention based parallel network (APNet) was designed to utilize the multi-layer structure of CNN-LSTM to learn both short-term and long-term knowledge to predict $PM_{2.5}$ concentrations over the next 72 h. However, existing $PM_{2.5}$ concentration prediction models still struggle to effectively capturing the complex nonlinearities of $PM_{2.5}$ variation patterns, as it is challenging for a single LSTM or simple tandem CNN-LSTM models to simultaneously characterize temporal and spatial dependencies in observed time series. Moreover, the inadequate optimization and inference efficiency of LSTM impedes its practical deployment and application.

Yeo et al. (2021) fed the combined data from 25 adjacent observation stations rather than a single station into hybrid CNN-LSTM based on the geographical polygon group method (GPGO), improving the $PM_{2.5}$ prediction accuracy by about 10 %. By introducing adjacency matrix in the LSTM cell, Gao & Li (2021) proposed a graph-based LSTM (GLSTM) to predict the $PM_{2.5}$ concentrations in Gansu, China. All the environmental observation stations are considered together as one graph, introducing spatio-temporal information while avoiding training individual models for each station. Nguyen et al. (2021) incorporated the genetic algorithm (GA) into encoder-decoder to complete the efficient feature extraction and outliers removal, thereby enhancing the $PM_{2.5}$ prediction accuracy. Drawing upon the assumption of spatial interaction, Shi et al. (2022) presented a novel balanced social LSTM (BS-LSTM) model for forecasting $PM_{2.5}$ concentration. This network possesses a distinct advantage in capturing trend information from neighboring locations. Yu et al. (2022) combined the fast Fourier transform (FFT) into LSTM to eliminate the spatio-temporal heterogeneity of $PM_{2.5}$ and impact factors, and finally achieved the long-term $PM_{2.5}$ prediction. Faraji et al. (2022) built a three-dimensional CNN by incorporating the GRU, i.e. 3D CNN-GRU. By learning and predicting all air quality monitoring stations simultaneously at varying time intervals, 3D CNN-GRU is able to maintain long-term memory and thus improve the $PM_{2.5}$ prediction accuracy. To achieve accurate short-term (within 6 h) predictions of $PM_{2.5}$ concentrations while capturing long-term (6–24 h) trends, Teng et al. (2022) developed a novel approach that combines EMD technique, sample entropy (SE) index, and bidirectional LSTM (BiLSTM). Tan et al. (2022) put forward a multi-data driven spatio-temporal $PM_{2.5}$ prediction method by using reinforcement learning based graph attention network (RL-GAT). To address issues of long-term dependence and spatial distribution description in the $PM_{2.5}$ prediction task, numerous deep learning models have incorporated specially

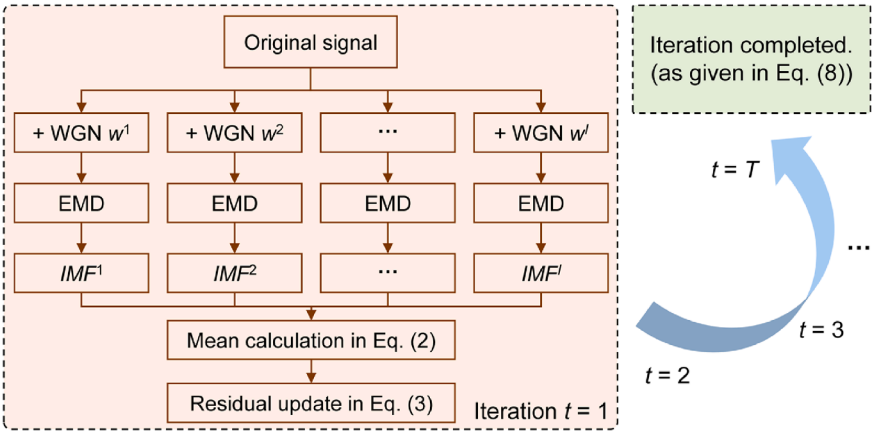


Fig. 3. Illustration of the whole CEEMD iterative process.

designed spatio-temporal information extraction modules, such as feedforward-backpropagation neural network (FBNN) (Chinatamby & Jewaratnam, 2023) and 1D-CNN-biLSTM (Zhu & Xie, 2023). On the other hand, numerous deep learning models that incorporate signal preprocessing have been developed, including the wavelet-packet transform based LSTM (WPT-LSTM) (Zheng et al., 2023a) and spatio-temporal data fusion network (STF-Net) (Zhang & Gan, 2023).

Extensive empirical studies have consistently demonstrated that the usage of multi-domain and multi-resolution signal analysis techniques for data preprocessing can enhance the cross-conditional generalization ability of deep learning models, particularly in the context of complex climate and terrain. Hence, the development of suitable preprocessing techniques is imperative to facilitate the reduction of structural complexity in deep learning models while simultaneously enhancing the capacity to extract robust features. In addition, it is essential to tailor model structures and learning strategies to unique features for $PM_{2.5}$ prediction. The applicability of model structures transferred from other tasks (e.g. image and video) to meteorological and environmental data may be limited, leading to inadequate learning of mapping from inputs to outputs, while the convergence position can be significantly influenced by learning strategies. In practical applications, the determinants influencing $PM_{2.5}$ are heterogeneous across different geographics. It is imperative to incorporate suitable analytical tools to facilitate deep learning models in attaining precise prediction of $PM_{2.5}$ concentrations in accordance with realistic scenarios.

3. CEEMD-MsI for $PM_{2.5}$ concentration long-term prediction

In this section, the proposed CEEMD-MsI for $PM_{2.5}$ long-term prediction is introduced in detail. Firstly, the meteorological and environmental time series are analyzed by CEEMD at multiple scales to form local features at different resolutions, i.e., IMFs based multi-channel representation, which are then fed into MsI to optimize the objective function by modifying all the learnable parameters. Ultimately, the converged MsI model possesses the capability to predict future $PM_{2.5}$ concentrations while overcoming the problem of long-term dependence. Fig. 2 depicts the comprehensive procedure for conducting CEEMD analysis, as well as the training and testing of MsI.

3.1. CEEMD based representation construction

3.1.1. Signal decomposition

Given a temporal data sequence $\{s_m(k)\}_{k=1,2,\ldots,K} \in \mathbb{R}^{1\times K}$ containing $m = 1, 2, \ldots, M$ types of meteorological and environmental information that has been observed for K times, the noisy signals are represented as

$$\widehat{s}^i(k) = s(k) + \varepsilon_0 w^i(k),\ i = 1,\ 2,\ \ldots,\ I \tag{1}$$

where w^i represents different realizations of white Gaussian noise (WGN), and the coefficient ε is introduced to determine the signal-to-noise ratio (SNR) at each iteration. Small coefficients ε are usually set for signals dominated by high-frequency components and vice versa. By utilizing the original EMD results, i.e. the IMF set $\mathrm{EMD}\,(s(k)) = \{IMF^j_{\mathrm{EMD}}(k)\}_{j=1,\,2,\,\ldots,\,J}$ and the individual j-th IMF component $\mathrm{EMD}^j(s(k)) = IMF^j_{\mathrm{EMD}}(k)$, EEMD can be further executed according to

$$IMF^j_{\mathrm{EEMD}}(k) = \frac{1}{I}\sum_{i=1}^{I} IMF^{i,j}_{\mathrm{EMD}}(k),\ j = 1,\ 2,\ \ldots,\ J \tag{2}$$

where $IMF^{i,j}_{\mathrm{EMD}}$ is the j-th IMF of $s^i(k)$ through EMD processing, and $IMF^j_{\mathrm{EEMD}}(k)$ indicates the j-th IMF of $s(k)$ by averaging EMD results and there are a total of J IMFs. Since each $s^i(k)$ is decomposed independently from the other realizations of WGN, the initial residual can be computed by

$$r^1(k) = s(k) - IMF^1_{\mathrm{CEEMD}}(k) \tag{3}$$

where

$$IMF^1_{\mathrm{CEEMD}}(k) = IMF^1_{\mathrm{EEMD}}(k) \tag{4}$$

Then the CEEMD results at the t-th iteration can be obtained by decomposing combinations of the residual and the first IMF component of EMD results, as given by

$$IMF^{t+1}_{\mathrm{CEEMD}}(k) = \frac{1}{I}\sum_{i=1}^{I} \mathrm{EMD}^1(r^t(k) + \varepsilon_t \mathrm{EMD}^t(w^i(k))),\ t = 1,\ 2,\ \ldots,\ T \tag{5}$$

where

$$r^t(k) = r^{t-1}(k) - IMF^t_{\mathrm{CEEMD}}(k),\ t = 2,\ 3,\ \ldots,\ T \tag{6}$$

where ε_t represents SNR controller at the t-th iteration. T denotes the total number of iterations, as well as the total number of IMFs. The operations from Eq. (5) to Eq. (7) are performed iteratively until the reconstruction residual $R(k)$ can no longer be decomposed (i.e. the residual sequence has less than two extrema), in which $R(k)$ is defined as

$$R(k) = s(k) - \sum_{t=1}^{T} IMF^t_{\mathrm{CEEMD}}(k) \tag{7}$$

If we rewrite Eq. (7), the given signal $s(k)$ can be ultimately expressed as

$$s(k) = \sum_{t=1}^{T} IMF^t_{\mathrm{CEEMD}}(k) + R(k) \tag{8}$$

The CEEMD has now been completely iterated, and it is clear that the signal can be perfectly reconstructed without any information loss. The whole iterative process is shown in Fig. 3, and the pseudo code of CEEMD analysis is summarized in Algorithm 1.

Algorithm 1 Signal decomposition based on CEEMD analysis.

Input: The observed temporal data sequence $s(k)$.

Initialization: Noise intensity ε and averaging number I.

```
Construct the noisy signal as in Eq. (1);
Compute the EMD results of the noisy signal;
Compute the EEMD results by averaging EMD results as in Eq. (2);
Compute the CEEMD results as in Eq. (4);
Compute the initial residual as in Eq. (3);
for t = 2 : T do
    for i = 1 : I do
        Construct the noisy residual as in Eq. (1);
        Compute the EMD results of the noisy residual;
    end for
    Compute the CEEMD results as in Eq. (5);
    Update the residual as in Eq. (6);
end for
```

Output: A series of IMFs and the construction residual $R(k)$.

Signal pre-processing is helpful for machine learning models to extract key features reflecting the long-term variation pattern of $PM_{2.5}$ concentration, especially CEEMD, which is adept at handling nonlinear and non-stationary signals through decomposition across various scales. CEEMD avoids the mode aliasing phenomenon encountered by the original EMD, improves the robustness of observed signals to noises and low sampling frequency, and maintains the time-varying characteristics of signals. The problem of retaining residual noises and being unable to separate components within the octave range in EEMD is also resolved during the iteration process of CEEMD. At each iteration of decompositions, CEEMD introduces the IMF component of WGN accompanied by gradually decreasing intensities ε, resulting in less residual noises in IMFs and effectively reducing the signal reconstruction error. Moreover, the global stopping criterion is set and checked after each iteration, thus resulting in efficient signal decompositions. From the perspective of $PM_{2.5}$ concentration prediction, CEEMD reduces the number of low-frequency IMF components with small amplitudes, which have little significance for expert knowledge mining of long-term meteorological and environmental data with low sampling frequency.

3.1.2. Representation construction

After obtaining IMFs of observed meteorological and environmental data at various scales (i.e., frequencies), the corresponding multi-channel representation $\mathbf{x}$ based on the correlation factor is constructed to feed into a deep learning model for mining spatio-temporal patterns and completing $PM_{2.5}$ concentration long-term predictions, as shown in Fig. 4.

The IMF matrix is first composed by concatenating IMF components at each frequency and the reconstruction residual to fully characterize the original signal, as given by

$$\begin{aligned}\mathbf{IMF} &= \mathrm{CEEMD}(s_1(k)) \oplus \mathrm{CEEMD}(s_2(k)) \oplus \ldots \oplus \mathrm{CEEMD}(s_M(k)) \\ &= [\mathbf{IMF}_1;\ \mathbf{IMF}_2;\ \ldots;\ \mathbf{IMF}_M] \in \mathbb{R}^{(T+1)\times K\times M}\end{aligned} \tag{9}$$

where

$$\begin{aligned}\mathbf{IMF}_m &= \mathrm{CEEMD}(s_m(k)) \\ &= [IMF^1_{\mathrm{CEEMD}}(k);\ IMF^2_{\mathrm{CEEMD}}(k);\ IMF^T_{\mathrm{CEEMD}}(k);\ \ldots;\ R(k)] \in \mathbb{R}^{(T+1)\times K}\end{aligned} \tag{10}$$

The obtained IMF matrix covers all intrinsic mode features of selected $PM_{2.5}$ concentrations, air indicators, and climate indicators by putting into different channels. The third dimension of size M reflects the M different types of information. Each row of **IMF** represents the signal components corresponding to a specific time scale. For different time scales, it is necessary to further improve expressiveness of **IMF** through matching weighted mechanism, as updated by

$$\mathbf{IMF}^{\dagger} = [\boldsymbol{\alpha}_1 \otimes \mathbf{IMF}_1;\ \boldsymbol{\alpha}_2 \otimes \mathbf{IMF}_2;\ \ldots;\ \boldsymbol{\alpha}_M \otimes \mathbf{IMF}_M] \in \mathbb{R}^{(T+1)\times K\times M} \tag{11}$$

where $\otimes$ denotes the Hadamard product, and $\boldsymbol{\alpha}$ represents the weighting matrix as

$$\boldsymbol{\alpha}_m = \begin{bmatrix} \alpha_{m,1} & \alpha_{m,1} & \cdots & \alpha_{m,1} \\ \alpha_{m,2} & \alpha_{m,2} & \cdots & \alpha_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m,T} & \alpha_{m,T} & \cdots & \alpha_{m,T} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{(T+1)\times K} \tag{12}$$

Correspondingly, the elements of the IMF matrix are updated to

$$\begin{aligned}\mathbf{IMF}^{\dagger}_m = [&\alpha_{m,1} IMF^1_{\mathrm{CEEMD}}(k);\ \alpha_{m,2} IMF^2_{\mathrm{CEEMD}}(k);\ \alpha_{m,T} IMF^T_{\mathrm{CEEMD}}(k);\ \ldots;\ R(k)] \\ &\in \mathbb{R}^{(T+1)\times K}\end{aligned} \tag{13}$$

It can be deduced that all elements of **IMF** are enhanced to varying

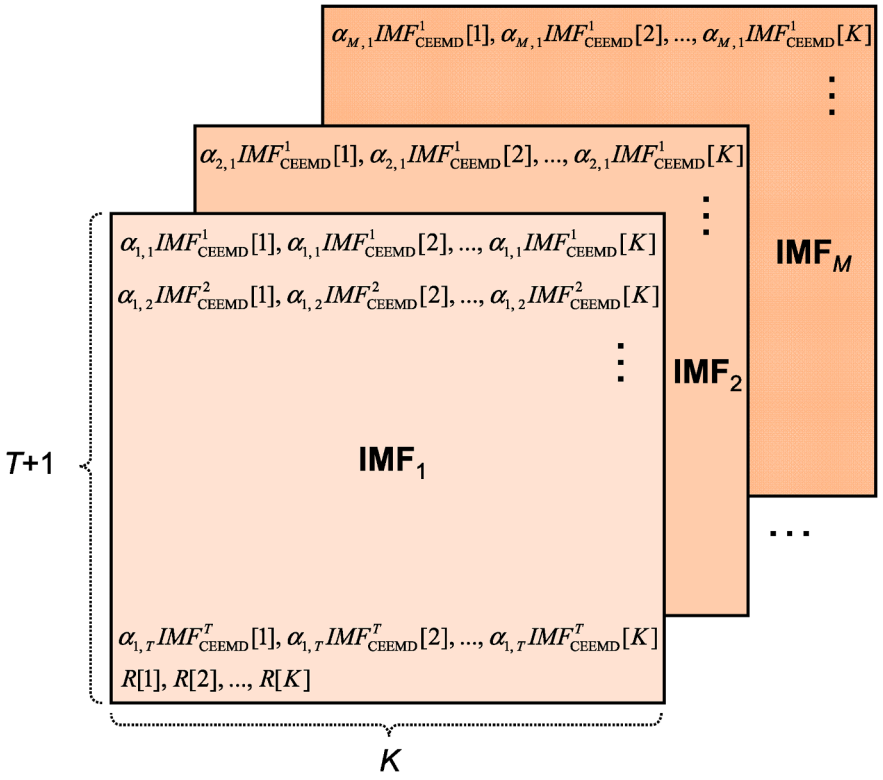


Fig. 4. Construction of the multi-channel representation of observed signals.

degrees due to $\alpha_{m,\,t} > 1$ while the reconstruction residual with an implicit fixed weight 1 remains unchanged.

The setting of the weighting matrix $\boldsymbol{\alpha}$ is computed based on the Hurst exponent H, which not only reveals the complexity of time series, but also describes their long-term correlation, as given by

$$\alpha_{m,\,t} = e^{H_{m,\,t}} \in (1,\ \ e),\ \text{if}\ H_{m,\,t} \in (0,\ 1) \tag{14}$$

where $H_{m,t}$ corresponding to $IMF^{t}_{\mathrm{CEEMD}}(k)$ can be calculated by the following steps:

Step 1: Constructing the zero-mean normalized IMF sequence $\varphi(k)$ by

$$\varphi^{t}(k) = IMF^{t}_{\mathrm{CEEMD}}(k) - \ < IMF^{t}_{\mathrm{CEEMD}}(k) > \tag{15}$$

where $<\cdot>$ is responsible for calculating the average value of a given sequence.

Step 2: Dividing the sequence $\varphi(k)$ into n non-overlapping subsequences with the same length L, ensuring that $K = nL$. Then each subsequence can be represented as

$$\phi^{t}_{i}(k) = \{\varphi^{t}(k) :\ 1 + (i-1)L \leqslant k \leqslant iL\ \},\ i = 1,\ 2,\ ...,\ n \tag{16}$$

The calculation can be repeated from both ends to eliminate boundary effects if there are redundant elements.

Step 3: Each subsequence $\phi^{t}_{i}(k)$ is fitted using a first-order polynomial function $g(\cdot)$ to obtain its fitting curve $g^{t}(k)$. In principle, the fitting function $g(\cdot)$ should exhibit linearity and low-frequency, and be equivalent to a high-pass filter for dealing with the time series.

Step 4: Computing the variance of the remaining sequence after descending convergence of each sub-sequence as

$$(F^{t}_{i}(l))^{2} = \frac{1}{L}\sum_{k=(i-1)L+1}^{il}(\phi^{t}_{i}(k) - g(k))^{2},\ l = 1,\ 2,\ ...,\ L \tag{17}$$

Step 5: Calculating the root mean square of the overall descending pulsation of the times series by

$$F^{t^2}(l) = \frac{1}{K}\sum_{k=1}^{K}(\varphi^{t}(k) - g(k))^{2} \tag{18}$$

Step 6: Calculating the pulsation $F(l)$ corresponding to different length L, and the scalar relationship between the pulsation $F(l)$ and the box length L can be described as

$$F^{t}(l) \sim L^{H_t} \tag{19}$$

where the Hurst exponent H_t can be used to characterize the memory strength of IMFs and determine the long-term correlation of IMFs with respect to its past values. Therefore, a larger H_t reflects a stronger self-similarity.

The weighted IMF matrix is now computed and then can be used as the multi-channel input (i.e. $\mathbf{x} = \mathbf{IMF}^{\dagger} \in \mathbb{R}^{(T+1)\times K\times M}$) for predicting $PM_{2.5}$ concentrations using the observed data sequences at the previous K moments. Compared with feeding original IMF sequences directly into the deep learning model, the weighted IMFs based signal representation $\mathbf{x}$ can eliminate the disturbing factors during the data acquisition process, and pay attention to the high-frequency fluctuations and profound changes hidden in the low-frequency of the data. In practical applications, the received signals are first decomposed and reconstructed by CEEMD, and then a deep learning model can be constructed to mine the implicit expert knowledge and complete the continuous prediction of future $PM_{2.5}$ concentrations.

3.2. MsI model

In this part, we introduce the proposed MsI model for feature learning of weighted IMFs obtained from CEEMD analysis of observed meteorological and environmental data, and subsequent $PM_{2.5}$ concentration prediction.

3.2.1. Model structure

The specific structure of MsI is consisted of an encoder $f(\cdot)$ and a

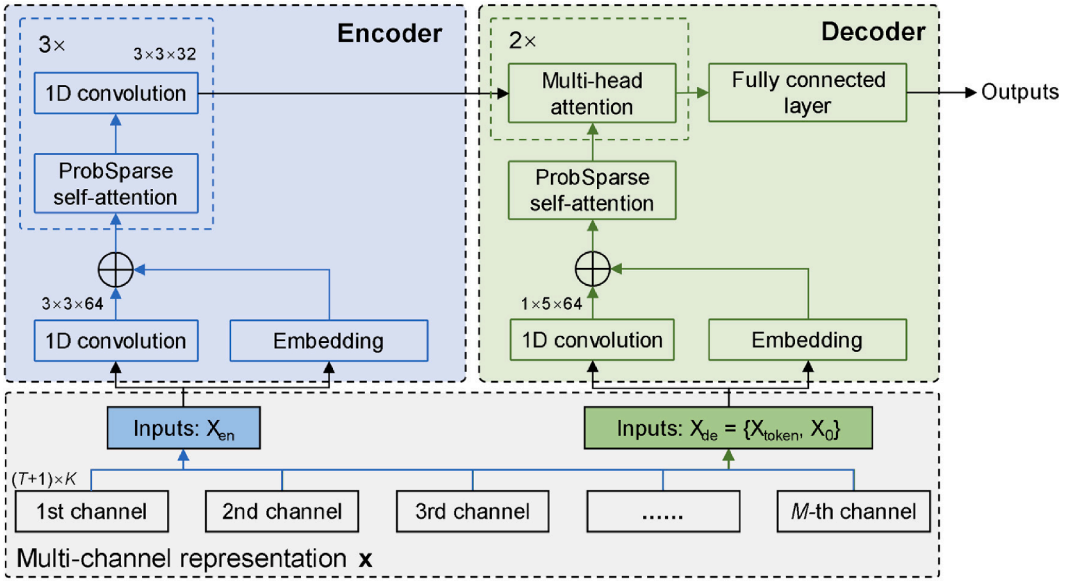


Fig. 5. Specific model structure of the proposed MsI.

decoder $h(\cdot)$, as shown in Fig. 5. The encoder composed of convolution, embedding, and attention is designed to extract robust long-range dependencies from long sequential inputs. The convolutional kernels with size 3×3 are used to mine contextual information, while the embedding is to supplement the positional information of each channel. During the forward propagation process, the i-th multi-channel representation $\mathbf{x}^i$ is first re-shaped into a two-dimensional matrix $\mathbf{X}_{\text{en}}^t$, in which each channel corresponds to an input stream and the convolutional kernels between streams are not shared. The outputs of j-th layer after 1D convolution and embedding are concatenated fed into the ProbSparse self-attention block, as given by

$$\mathbf{X}_{j+1}^t = \text{Maxpool}(\text{ELU}(\text{Conv1d}([\mathbf{X}_j^t]_{\text{AB}}))) \tag{20}$$

where

$$\text{ELU}(x) = \begin{cases} e^x - 1 & \text{if } x \leqslant 0 \\ x & \text{if } x > 0 \end{cases} \tag{21}$$

In the above equations, Maxpool$(\cdot)$ denotes the max-pooling layer with stride 2, Conv1d$(\cdot)$ performs 1D convolution operations on time dimension with kernel size of 3×3, and the exponential linear units (ELU) function is used to complete the nonlinear transform. $[\cdot]_{\text{AB}}$ represents the ProbSparse self-attention operation, which is defined as

$$[\mathbf{X}]_{\text{AB}} = \text{Softmax}(\frac{\overline{\mathbf{Q}}\mathbf{K}^T}{\sqrt{d}})\mathbf{V} \tag{22}$$

where d is the input dimension, $\mathbf{K} \in \mathbb{R}^{L_K \times d}$ and $\mathbf{V} \in \mathbb{R}^{L_V \times d}$ represent the key vector and value vector in the canonical self-attention, respectively (Gong et al., 2022). $\overline{\mathbf{Q}}$ is a sparse matrix of the same size of query vector $\mathbf{Q}$ and only contains some top queries under the sparsity measurement SM, which can be computed by

$$SM(\mathbf{q}_i, \mathbf{K}) = \ln \sum_{j=1}^{L_K} e^{\frac{\mathbf{q}_i \mathbf{k}_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{\mathbf{q}_i \mathbf{k}_j^T}{\sqrt{d}} \tag{23}$$

where $\mathbf{q}_i$ and $\mathbf{k}_j$ are the i-th and j-th row in $\mathbf{Q}$ and $\mathbf{K}$, respectively. The first term is the Log-Sum-Exp operation of $\mathbf{q}_i$ on all the keys, and the second term is their arithmetic mean. After three stacked ProbeSparse attention blocks and 1D convolution layers, the output of the encoder is fed into the decoder. Actually, the ProbeSparse attention block is able to strike a balance between capturing long-term dependencies and low computational complexity. Besides, the introduction of multi-head attention mechanism helps the model focus on factors that are critical in predicting $PM_{2.5}$ concentrations, while ignoring irrelevant factors that may have adverse effects.

The decoder $h(\cdot)$ is composed of the stacking of one ProbeSparse attention block and two identical multi-head attention layers. The input of decoder is further split into different tokens, as given by

$$\mathbf{X}_{\text{de}}^t = \{\mathbf{X}_{\text{token}}^t, \mathbf{X}_0^t\} \tag{24}$$

where $\mathbf{X}_{\text{token}}^t$ is the start token, and $\mathbf{X}_0^t$ is a placeholder for the target sequence. Through the similar 1D convolution with kernal size of 1×5, embedding operation, and ProbSparse self-attention blocks in the encoder, the outputs of the encoder and features extracted from the decoder are put through two multi-head attention layers together. The multi-head attention layer allows the model to jointly focus on the crucial information from various representation subspaces at different locations, as calculated by

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_p)\boldsymbol{W}^O \tag{25}$$

where

$$\text{head}_i = \text{Attention}(\mathbf{Q}\boldsymbol{W}_i^Q, \mathbf{K}\boldsymbol{W}_i^K, \mathbf{V}\boldsymbol{W}_i^V) = \text{Softmax}(\frac{\mathbf{Q}\boldsymbol{W}_i^Q(\mathbf{K}\boldsymbol{W}_i^K)^T}{\sqrt{d}})\mathbf{V}\boldsymbol{W}_i^V \tag{26}$$

where the projections $\boldsymbol{W}_i^Q$, $\boldsymbol{W}_i^K$, $\boldsymbol{W}_i^V$, and $\boldsymbol{W}^O$ are learnable parameter matrices, and a total of $p = 4$ parallel heads are set in the model. Finally, $PM_{2.5}$ concentration prediction results can be obtained by a fully connected layer with identity mapping.

The initialization of the MsI structure adopts empirical settings transferred from other signal analysis fields. Then we observed the optimization results of the objective function under different module combinations to select the relatively optimal structure, including the number and size of convolutional kernels, the type of activation

functions, and the location of attention mechanisms. Finally, the $PM_{2.5}$ prediction performance of the optimal model structure is comprehensively tested and compared. In fact, a well-designed model structure has little impact on $PM_{2.5}$ prediction performance.

3.2.2. Parameters optimization

Although deep learning is a highly over-parameterized non-convex model that has always been challenging to optimize, finding local optimal solutions has been widely studied. The learnable parameters $\boldsymbol{W}$ in MsI can be modified by optimizing the objective function through the error back-propagation based mini-batch stochastic gradient descent (SGD) method (Lohat et al., 2023). The objective function designed to simultaneously measure the $PM_{2.5}$ prediction accuracy and structural complexity of the deep learning model is defined as

$$\mathscr{L}(\mathbf{x}^i; \boldsymbol{W}) = (\mathbf{y}^i - \hat{\mathbf{y}}^i)^2 + \gamma \sum_j \|\boldsymbol{W}_j\|_2^2 \tag{27}$$

where $\mathbf{y}^i$ and $\hat{\mathbf{y}}^i$ represent ground-truth and forecasted $PM_{2.5}$ concentrations, respectively. $\|\boldsymbol{W}_j\|_2^2$ denotes the L2-regularization of learnable parameters in the j-th layer and γ is a hyper-parameter controlling regularization intensity. L2-regularization limits the structural complexity of the model by constraining the intensity of parameters to prevent overfitting.

Before the training of MsI, all the learnable parameters $\boldsymbol{W}_0$ are randomly initialized to follow the standard normal distribution and decorrelated. During the back-propagation process, $\boldsymbol{W}_l$ at the l-th training iteration ($l = 1, 2, \ldots, L$) can be modified as

$$\boldsymbol{W}_l = \boldsymbol{W}_{l-1} - \alpha_l \frac{\hat{\tau}_l}{\sqrt{\hat{\rho}_l} + \text{ə}} \tag{28}$$

where α_l is the dynamically adjusted learning rate at l-th iteration and determines determine the step size of descent. ə is a small constant for numerical stability. $\hat{\tau}_l$ and $\hat{\rho}_l$ denote the estimators of corrected first-order and second-order moments respectively, which are calculated by

$$\hat{\tau}_l = \frac{\tau_l}{1 - \varphi_1} \tag{29}$$

$$\hat{\rho}_l = \frac{\rho_l}{1 - \varphi_2} \tag{30}$$

where

$$\tau_l = \varphi_1 \tau_{l-1} + (1 - \varphi_1) \nabla_l \tag{31}$$

$$\rho_l = \varphi_2 \rho_{l-1} + (1 - \varphi_2) \nabla_l^2 \tag{32}$$

$$\nabla_l = \frac{1}{\omega} \sum_{i=1}^{\omega} \frac{\partial \mathscr{L}(\mathbf{x}^i; \boldsymbol{W}_{l-1})}{\partial \boldsymbol{W}_{l-1}} \tag{33}$$

In the above equations, φ_1 and φ_2 represent exponential decay rates of first-order and second-order moments respectively and belong to the interval [0, 1), ω is the batch size, and ∇_l denotes the batch gradient used to determine the descent direction of objective function at the l-th training iteration. The whole training process is summarized in Algorithm 2. Finally, we can obtain a convergent MsI model with updated parameters $\boldsymbol{W}_*$ by continuously iterating through the above steps on training samples, i.e., from Eq. (27) to Eq. (33).

After the MsI model is fully trained, i.e. the convergent model is output, it can be deployed to complete the processing of test signals. Specifically, the newly received test signals can be fed into the MsI model, and the predicted $PM_{2.5}$ value is output through the forward propagation. By comparing the predicted values with the actual $PM_{2.5}$ values using a series of metrics, the prediction performance of the model can be comprehensively evaluated.

Algorithm 2 SGD training process of MsI model.

Input: The multi-channel representation $\mathbf{x}$.

Initialization: Model structure and learnable parameters $\boldsymbol{W}_0$.

```
for l = 1 : L do
    for i = 1 : ω do
        forward propagation:
        Compute the output of the model using Eq. (20) to Eq. (26);
        Compute the objective function as in Eq. (27);
    end for
    back propagation:
    Compute the batch gradient as in Eq. (33);
    Update learnable parameters as in Eq. (28);
end for
```

Output: The convergent model with parameters $\boldsymbol{W}_*$.

4. Test results and analysis

In this section, we conduct comprehensive tests in Shandong Province, China, and compare the $PM_{2.5}$ concentration prediction accuracy and efficiency of CEEMD-MsI with state-of-the-art methods. Ablation studies are performed to verify the effectiveness of each module in CEEMD-MsI. Furthermore, the hyper-parametric sensitivity and robustness to changing observation conditions of CEEMD-MsI are observed to verify its practicality.

4.1. Test datasets

The tests are conducted using meteorological and environmental data observed in Shandong, China. Data from a total of 17 monitoring stations in four representative cities, including Jinan, Qingdao, Taian, and Zibo, were collected. The distribution of four cities and corresponding air quality monitoring stations is shown in Fig. 6. The Shandong province has an estimated land area of 155800 km^2, with a length of 721 km from east to west and 437 km from north to south. It is dominated by mountains and hills, of which the western and northern areas are plains formed by alluvial deposits of the Yellow River, the central region and the southern peninsula are gently undulating hilly areas. The mountainous area accounts for 14.59 % of the total area of Shandong, thus forming a natural barrier to airflow carrying $PM_{2.5}$. In addition, Shandong has one-fifth of the country's coastline, so ocean currents and monsoons have an undeniable impact on the spread and accumulation of $PM_{2.5}$. Actually, there are great regional differences among cities in Shandong, making it challenging for a single model to predict long-term $PM_{2.5}$ concentrations under different locations and conditions. In this case, deep learning models aided by signal preprocessing have superior generalization capability compared to traditional statistical modelling methods.

In practical applications, the main task of outdoor air quality monitoring stations is to detect and collect specific air pollutants, and upload data in real-time to cloud servers for storage and analysis. The observed data used in the tests involves $M = 12$ types of influential factors, of which the environmental data contains $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO, and the meteorological data contains temperature (AT), pressure (PRES), wind direction (WD), wind speed (WS), cloudiness (CL), and precipitation (PREC). Since there are a few missing values in the collected data, the linear interpolation within the same day is used to fill in the blanks. The meteorological and environmental data series observed over a period of time are shown in Fig. 7. The data observation period is the whole year of 2020, where the sampling interval is 1 h. Specifically, we have collected a total of 149,311 records of hourly $PM_{2.5}$ concentrations from 17 monitoring stations in Shandong, of which 60 %,

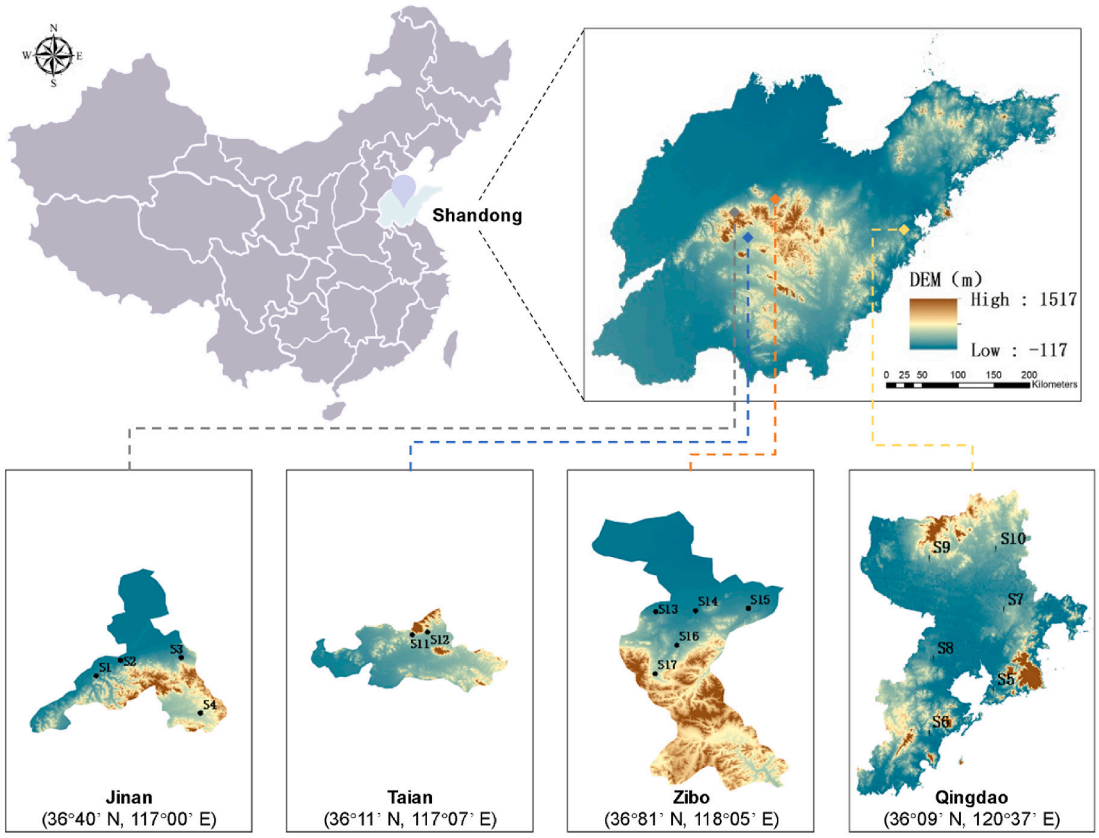


Fig. 6. The distribution of air quality monitoring stations in four cities in Shandong, China.

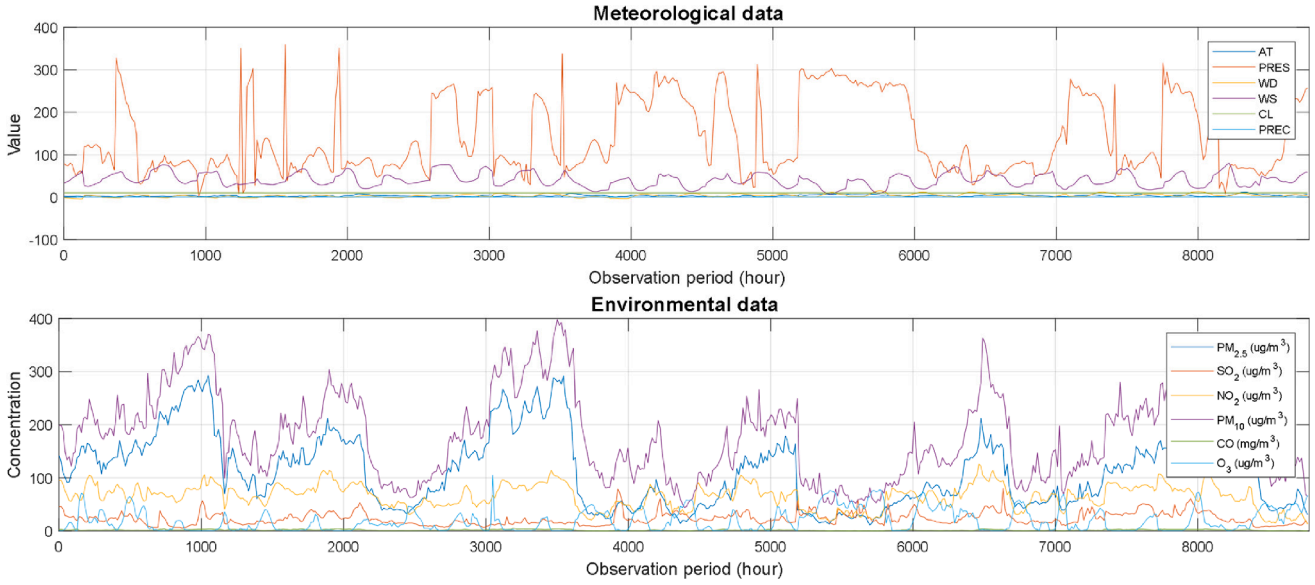


Fig. 7. Meteorological and environmental data series observed over a period of time.

20 % and 20 % are divided into training set, validation set, and test set, respectively. The training set is used to optimize model parameters, the validation set is used to select appropriate model hyper-parameters, and the test set is used to evaluate the model performance. The Pearson (linear) and Kendall (nonlinear) correlations between various factors of the observed data are analyzed in Fig. 8. The linear correlation degree between different factors and $PM_{2.5}$ concentrations is shown in (a), with historical PM_{10} accumulations being the most relevant while other factors showing joint correlation. In (b), a strong correlation is observed between meteorological data, especially between PRES, CL and PREC.

4.2. Training and test settings

The proposed CEEMD-MsI is deployed in the workstation equipped with a 64 bit Windows 10 operating system, consisting of Intel (R) Core (TM) i9-12900KS CPU, NVIDIA RTX 3090 GPU, 2 × 16 GB RAM, and 2 TB HDD. During the tests, CEEMD based signal analysis is implemented in Matlab 2022a and the deep learning model MsI is trained and tested

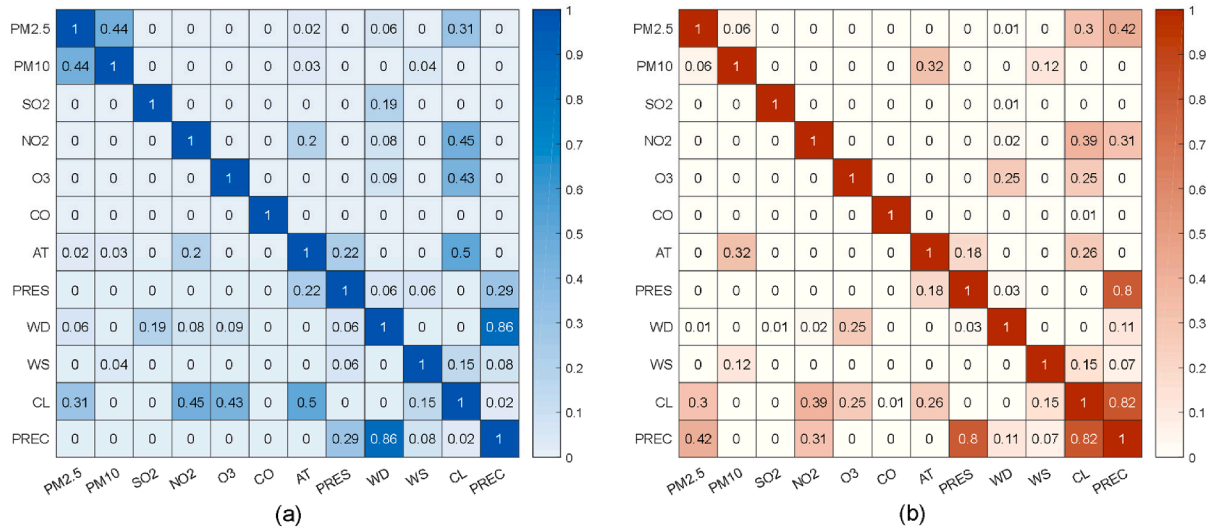


Fig. 8. Correlation analysis between various factors in meteorological and environmental data. (a) Pearson, (b) Kendall.

using Pytorch supported by Python 3.11.2. All the hyper-parameters involved in the CEEMD-MsI have been summarized in Table 1. The time series observed in the past $K = 72$ h are used to predict the $PM_{2.5}$ concentrations in the next 24 h. Initial noise intensity ε_0 in CEEMD is set to 0.5 to create an average interference. The averaging number I is set to 8 by weighing decomposition efficiency and frequency resolution. The number of decomposition iterations T for each type of influential factor is uniformly limited to 6 to ensure consistent dimensionality of the IMF matrix. During the training process of MsI, the learning rate α affects the convergence path of model loss, and is gradually reduced in a simulated annealing manner to accommodate the progressively finer loss landscapes. Batch size $\omega = 32$ is set according to the total number of training samples, which determines the gradient decent direction when optimizing the loss function by updating model parameters. Some other hyper-parameters including weight decay $\lambda = 0.01$, Dropout rate $P_D = 0.5$, decay rates $\varphi_1 = 0.9$ and $\varphi_2 = 0.999$, small constantə $= 0.00001$, and training epochs $L = 60$ (about 170,000 iterations) are set empirically. During the training process, when the objective function of the training set no longer decreases within 5 epochs, it means that the model has been trained as well as possible. When the objective function of the validation set maintains an upward trend over a period of time, the model may start to tend to overfitting.

For performance evaluation of $PM_{2.5}$ concentration prediction from multiple perspectives, five metrics including mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), determination coefficient (R^2), and Nash-Sutcliffe efficiency coefficient (NSE) are introduced as follows.

Table 1
Hyper-parameters setting in the CEEMD-MsI.

Stage	Hyper-parameter	Value
Analysis by CEEMD	observation period K	72
	noise intensity ε_0	0.5
	averaging number I	8
	decomposition iteration T	6
Training of MsI	learning rate α_0	0.01
	batch size ω	32
	Dropout rate P_D	0.5
	weight decay λ	0.0001
	decay rate φ_1	0.9
	decay rate φ_2	0.999
	small constant ə	0.00001
	training epochs L	60

$$\mathrm{MAE} = \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}|\mathbf{y}_i - \hat{\mathbf{y}}_i| \tag{34}$$

$$\mathrm{MAPE} = \frac{100}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\left|\frac{\mathbf{y}_i - \hat{\mathbf{y}}_i}{\mathbf{y}_i}\right| \tag{35}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2} \tag{36}$$

$$\mathrm{R}^2 = \frac{N_{\text{test}}\sum_{i=1}^{N_{\text{test}}}\mathbf{y}_i\hat{\mathbf{y}}_i - \sum_{i=1}^{N_{\text{test}}}\mathbf{y}_i\sum_{i=1}^{N_{\text{test}}}\hat{\mathbf{y}}_i}{\sqrt{N_{\text{test}}\sum_{i=1}^{N_{\text{test}}}\mathbf{y}_i^2 - (\sum_{i=1}^{N_{\text{test}}}\mathbf{y}_i)^2}\cdot\sqrt{N_{\text{test}}\sum_{i=1}^{N_{\text{test}}}\hat{\mathbf{y}}_i^2 - (\sum_{i=1}^{N_{\text{test}}}\hat{\mathbf{y}}_i)^2}} \tag{37}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N_{\text{test}}}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=1}^{N_{\text{test}}}(\mathbf{y}_i - \bar{\mathbf{y}})^2} \tag{38}$$

where $\bar{\mathbf{y}}$ represents the average value of observed $PM_{2.5}$ concentrations of the test data. MAE and RMSE measures the average prediction error, and MAPE describes the degree of prediction accuracy. R^2 reflects the fitting degree of regression lines to observed values, and NSE exposes the reliability of the model. It is clear from the definitions of these metrics that a method can be considered to perform better if it achieves lower MAE, MAPE, and RMSE while higher NSE and R^2 than the other comparison methods.

4.3. $PM_{2.5}$ concentration prediction performance

We first visualize the CEEMD results of 11 types of influential factors in Fig. 9. Due to the limitation of signal observation period, a total of $T = 6$ iterations have been able to completely decompose the signal at various frequencies. The main information is covered in the IMFs, while the bottom noise is concentrated in the reconstructed residuals. From a purely temporal perspective, environmental factors are more unstable than meteorological factors. However, after CEEMD preprocessing to different frequencies, the stationarity of the two types of signals gradually became consistent, alleviating the mode pairing problem. In this case, a single deep learning model is able to extract the features of multi-channel signal representations consisting of IMFs at multiple frequencies, and accomplish $PM_{2.5}$ concentration prediction.

Then we display the training process and optimization results in Fig. 10. As shown in (a), the model objective function of training set

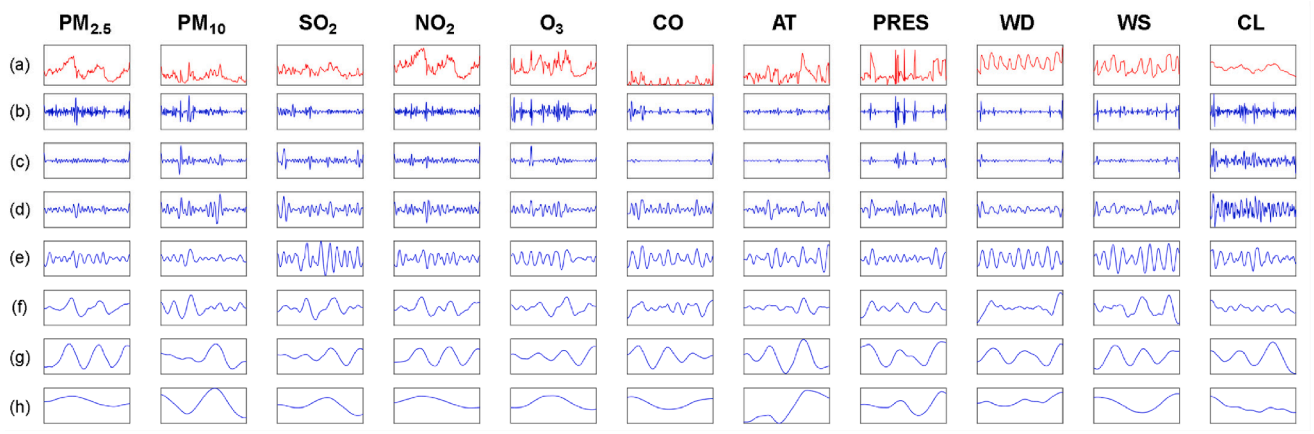


Fig. 9. Signal decomposition results of CEEMD analysis. (a): original signal, (b)-(g): IMF_1-IMF_6, (h): construction residual.

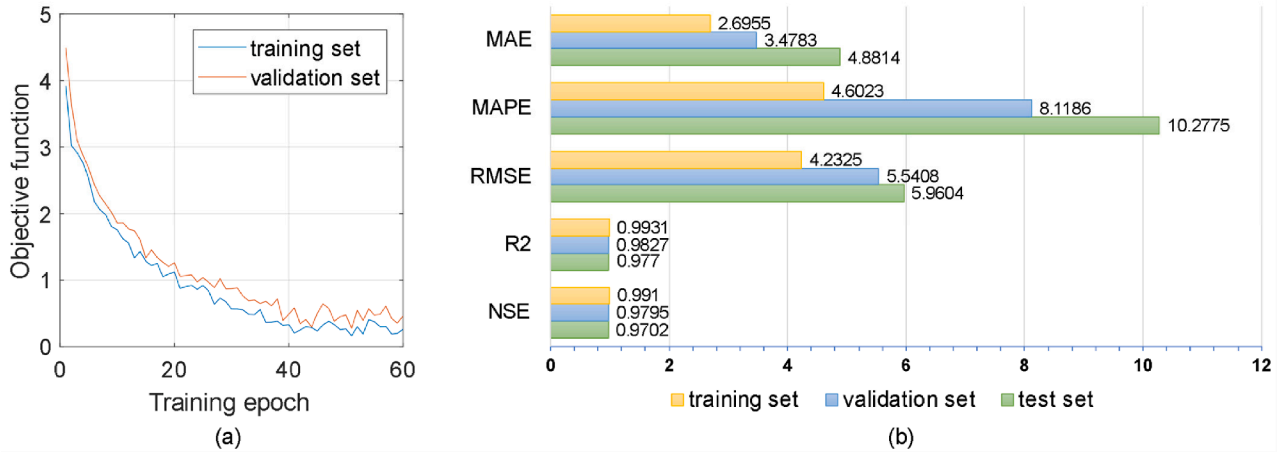


Fig. 10. Optimization results of the objective function and $PM_{2.5}$ prediction performance of CEEMD-MsI.

tends to stabilize at about 0.25 at 40 training epochs, followed by small oscillations < 0.1. By comparison, the objective function of the validation set converges to approximately 0.5 and has relatively large fluctuations. Since samples from different urban areas are used for machine learning, it is challenging for the objective function to converge stably, which is exactly beneficial to prevent overfitting and improve its generalization. The convergent model has achieved good fitting results on the training set, with MAE 2.6955, MAPE 4.6023, RMSE 4.2325, R^2 0.9931, and NSE 0.9910. The model also performs well on the test set (MAE 4.8814, MAPE 10.2775, RMSE 5.9604, R^2 0.9770, and NSE 0.9702) with acceptable generalization error (i.e. the performance gap between validation set and test set: MAE loss 1.4031, MAPE loss 2.1589, RMSE loss 0.4196, R^2 loss 0.0057, and NSE loss 0.0093).

To verify the superiority of CEEMD-MsI, a series of statistical modelling methods (PDE (Wang et al., 2020), MFO-ELM (Sun et al., 2021), and JAYA-LSSVM (Yang et al., 2022)) and deep learning models (CNN-LSTM (Pak et al., 2020), 3D CNN-GRU (Faraji et al., 2022), and WPT-LSTM (Zheng et al., 2023a)) are compared. During the tests, the test conditions and design are kept as consistent as possible, such as the period of the input signal and the predicted period. Some unique hyperparameters in different methods refer to the original settings. The $PM_{2.5}$ concentration prediction results for four cities are reported in Table 2. CEEMD-MsI achieved state-of-the-art $PM_{2.5}$ concentration prediction performance for the future 24 h in four cities, with all MAEs < 7, MAPEs < 15, RMSEs < 9, R^2 > 0.9650, and NSEs > 0.9540. Besides, deep learning models demonstrate significant advantages over statistical modelling methods, especially in this cross-geographic big data context. It is worth noting that all approaches perform the worst in Qingdao city compared to other three cities, probably due to its special geographical location. In Fig. 11, we present the comparison of average $PM_{2.5}$ concentration prediction results in four cities. It is clear that CEEMD-MsI has attained the lowest average MAE, MAPE, and RMSE, coupled with the highest average R^2 and NSE. To further evaluate the $PM_{2.5}$ long-term prediction performance, comparison results under different prediction time ranges are summarized in Table 3. $PM_{2.5}$ concentrations are predicted for up to the future 32 h. As the predicted time range increases, the $PM_{2.5}$ prediction performance of all methods decreases. The designed MsI model continues to exhibit a comparatively favorable performance, owing to its ability to capture the long-term dependency of time series data. Even predicting $PM_{2.5}$ concentrations for the future 32 h, CEEMD-MsI improved the state-of-the-art MAE, MAPE, RMSE, R^2, and NSE by 48.12 %, 36.64 %, 43.01 %, 1.42 %, and 0.99 %, respectively.

In Fig. 12, we visualize the $PM_{2.5}$ concentration prediction performance of various methods by presenting some of the results in a stair-step format. During the prediction process, both PDE and MFO-ELM displayed significant deviations from the observed values and demonstrated a high level of predictive variance. In contrast, deep learning methods, particularly CEEMD-MsI, showed smaller deviations and prediction variances. As shown in Fig. 13, about 600 points are presented in the form of a scatter plot to show the predictability of CEEMD-MsI and 6 comparison methods. Visual inspection of the observed and predicted $PM_{2.5}$ values, through the fitting of a straight line ($y = x$), reveals that CEEMD-MsI exhibits the strongest fitting capability. Notably, CEEMD-MsI shows a more pronounced advantage in predicting higher $PM_{2.5}$ concentrations above 200 $\mu g/m^3$.

Table 2
$PM_{2.5}$ concentration prediction performance of comparison methods at different cities.

City	Method	$PM_{2.5}$ concentration prediction performance				
		MAE	MAPE	RMSE	R^2	NSE
Jinan	PDE	24.9754	34.2025	31.2290	0.9111	0.8052
	MFO-ELM	21.2912	28.6699	26.0012	0.9240	0.8624
	JAYA-LSSVM	18.7872	25.2105	22.4086	0.9323	0.9001
	CNN-LSTM	14.9656	21.1094	18.5903	0.9457	0.9303
	3D CNN-GRU	12.9667	18.3582	15.9770	0.9572	0.9488
	WPT-LSTM	12.4337	17.6487	14.7477	0.9681	0.9541
	CEEMD-MsI	**4.9505**	**11.2073**	**6.0151**	**0.9754**	**0.9620**
Qingdao	PDE	25.3314	42.6425	27.1197	0.9113	0.8018
	MFO-ELM	21.8977	34.0083	23.4942	0.9182	0.8529
	JAYA-LSSVM	19.6892	33.5815	21.7792	0.9216	0.8930
	CNN-LSTM	15.2927	30.3908	16.9856	0.9398	0.9204
	3D CNN-GRU	13.4489	28.3863	16.7722	0.9410	0.9377
	WPT-LSTM	13.2356	22.7244	14.6254	0.9533	0.9438
	CEEMD-MsI	**6.2207**	**14.5419**	**8.9417**	**0.9680**	**0.9547**
Taian	PDE	21.8438	24.1115	22.8752	0.9204	0.8266
	MFO-ELM	17.6799	21.0126	19.0942	0.9411	0.8782
	JAYA-LSSVM	14.5372	15.3932	13.3320	0.9692	0.9212
	CNN-LSTM	10.7650	11.4737	10.1777	0.9725	0.9499
	3D CNN-GRU	8.5599	11.1668	10.2686	0.9733	0.9783
	WPT-LSTM	8.2004	10.2016	9.4199	0.9780	0.9800
	CEEMD-MsI	**4.0580**	**7.2701**	**4.3361**	**0.9852**	**0.9833**
Zibo	PDE	19.7317	26.4384	24.2100	0.9199	0.8251
	MFO-ELM	17.0004	23.8735	20.4351	0.9336	0.8746
	JAYA-LSSVM	12.3312	16.4445	14.7699	0.9561	0.9195
	CNN-LSTM	9.1801	12.7597	11.6879	0.9646	0.9472
	3D CNN-GRU	8.8065	12.4758	11.5577	0.9672	0.9740
	WPT-LSTM	9.0448	11.0085	10.9304	0.9698	0.9785
	CEEMD-MsI	**4.3002**	**8.0926**	**4.5514**	**0.9794**	**0.9808**

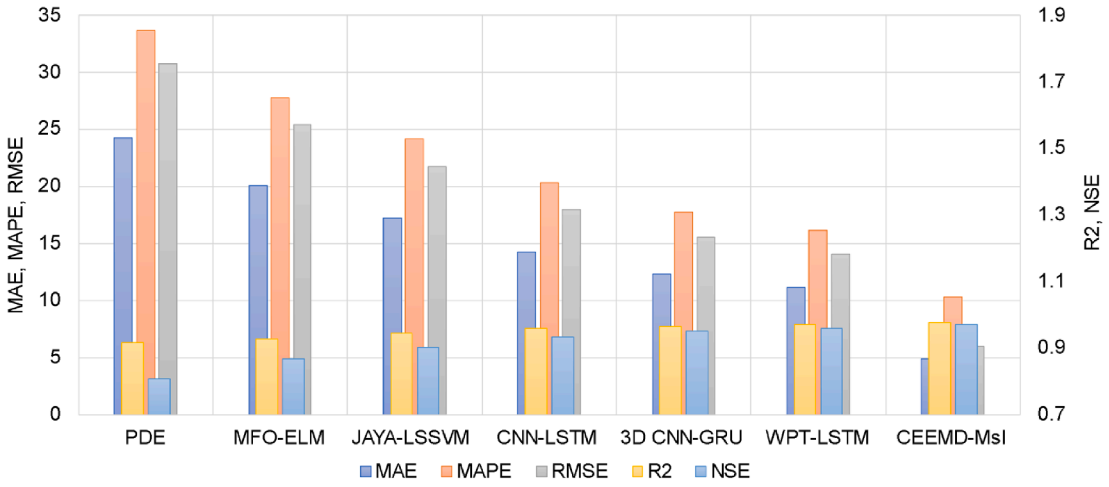


Fig. 11. Average $PM_{2.5}$ concentration prediction performance of various methods.

4.4. Ablation studies

To observe and verify the roles played by each module of CEEMD-MsI in the $PM_{2.5}$ concentration prediction task, ablation studies are conducted in this part. The predictive performance containing accuracy and efficiency of MsI driven by different input forms is reported in Table 4. MsI driven by multi-channel representations constructed based on IMFs exhibits superior prediction accuracy over the original time series, with CEEMD showing the most significant performance improvement. On the other hand, although the multi-channel representation requires longer model inference time (3.54 ms), it can still fully meet the requirements of practical applications at a signal sampling period of 1 h. In Table 5, we further observe the impact of the number of decomposition iterations of CEEMD, i.e. the number of IMFs, on the $PM_{2.5}$ prediction performance of the MsI model. It can be seen that the increase in decomposition resolution helps to improve the $PM_{2.5}$ prediction performance of the model accompanied by a negligible increase in decomposition time and inference time. In fact, the number of decomposition iterations can be considered in combination with the data observation period and the complexity of the actual $PM_{2.5}$ prediction task.

In addition to the role of CEEMD, we also judge the impact of different attention mechanisms, as shown in Table 6. Compared to the traditional multi-head attention, the adopted ProbSparse self-attention improves the average MAE, MAPE, RMSE, R^2, and NSE by 32.43 %, 23.53 %, 42.89 %, 1.09 %, and 0.94 %, respectively. The results indicate that MsI with ProbSparse self-attention is capable of extracting spatio-temporal features without introducing a large number of parameters.

4.5. Hyper-parametric sensitivity

In practical applications, the choice of hyper-parameters will affect the $PM_{2.5}$ prediction performance of deep learning models across different scenarios. In this part, we first test the sensitivity of CEEMD to

Table 3
Average $PM_{2.5}$ concentration prediction performance of comparison methods at various predicted time ranges.

Time range	Method	Average $PM_{2.5}$ concentration prediction performance of four cities				
		MAE	MAPE	RMSE	R^2	NSE
Future 12 h	PDE	15.5043	22.4895	17.3506	0.9210	0.8270
	MFO-ELM	13.5409	19.3405	15.0369	0.9501	0.8843
	JAYA-LSSVM	7.0015	13.7481	10.0518	0.9652	0.9304
	CNN-LSTM	4.4276	9.9681	6.8921	0.9805	0.9619
	3D CNN-GRU	4.2105	9.3136	6.0415	0.9878	0.9806
	WPT-LSTM	3.6260	8.3398	4.8921	0.9912	0.9836
	CEEMD-MsI	**3.0271**	**7.2853**	**3.7843**	**0.9937**	**0.9914**
Future 18 h	PDE	20.0379	28.6780	25.2847	0.9196	0.8218
	MFO-ELM	17.5229	24.0472	21.4830	0.9461	0.8712
	JAYA-LSSVM	12.9265	19.6784	15.9020	0.9588	0.9171
	CNN-LSTM	9.4350	14.1474	12.7851	0.9663	0.9446
	3D CNN-GRU	9.1266	13.8816	12.3390	0.9697	0.9622
	WPT-LSTM	9.3811	13.2950	12.8956	0.9736	0.9661
	CEEMD-MsI	**4.3244**	**9.4054**	**5.1787**	**0.9907**	**0.9799**
Future 24 h	PDE	24.2180	33.6345	30.6933	0.9151	0.8074
	MFO-ELM	20.0643	27.7107	25.4194	0.9274	0.8679
	JAYA-LSSVM	17.1944	24.1427	21.7458	0.9437	0.9033
	CNN-LSTM	14.2089	20.2795	17.9599	0.9580	0.9341
	3D CNN-GRU	12.2978	17.7348	15.5404	0.9659	0.9506
	WPT-LSTM	11.1398	16.1607	14.0739	0.9701	0.9595
	CEEMD-MsI	**4.8814**	**10.2775**	**5.9604**	**0.9770**	**0.9702**
Future 32 h	PDE	26.8645	39.1704	35.9038	0.9009	0.7946
	MFO-ELM	24.8307	31.0301	29.1432	0.9150	0.8469
	JAYA-LSSVM	23.1335	30.7763	27.0474	0.9234	0.8841
	CNN-LSTM	20.9694	26.4755	24.5298	0.9316	0.9124
	3D CNN-GRU	17.7059	24.0366	21.7087	0.9437	0.9325
	WPT-LSTM	15.8449	20.5619	17.9036	0.9442	0.9409
	CEEMD-MsI	**8.2207**	**13.0288**	**10.2024**	**0.9576**	**0.9512**

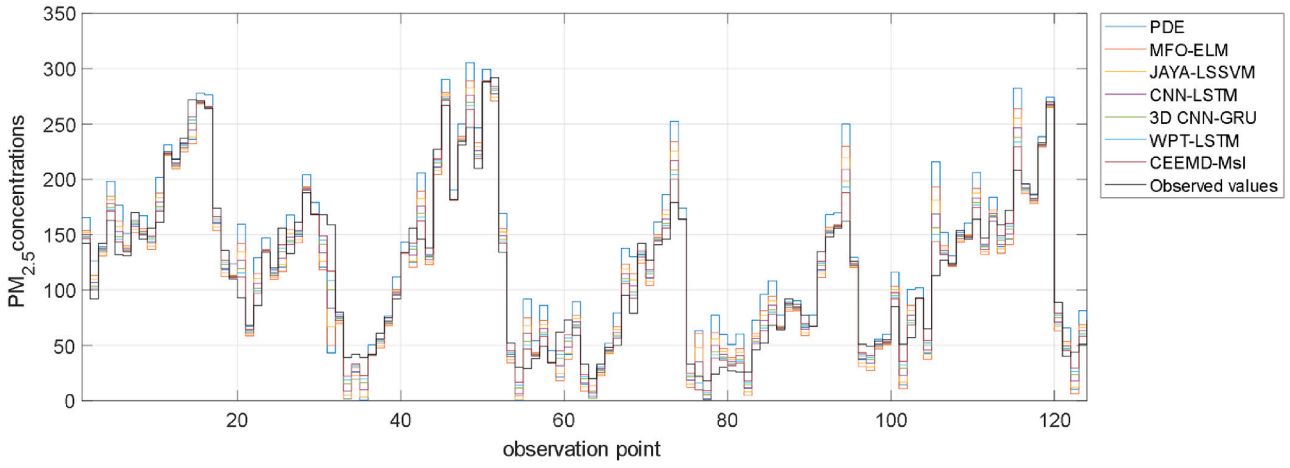


Fig. 12. Observed $PM_{2.5}$ concentrations and prediction results of various methods.

the averaging number I during the decomposition iteration process, and the results are summarized in Table 7. As the averaging number I increases, CEEMD can more effectively suppress the influence of noises and improve the accuracy and stability of decomposition results, thus leading to improved performance of MsI for feature learning and $PM_{2.5}$ prediction. The MAE, MAPE, RMSE, R^2, and NSE can be improved from 5.8987, 12.1153, 8.8220, 0.9668, and 0.9588 to 4.8770, 10.2562, 5.9479, 0.9774, and 0.9722 respectively by increasing the averaging number I from 4 to 10. Moreover, the additional decomposition time resulting from an increase in the averaging number is less than 0.1 ms, which can be negligible in both theoretical analysis and practical applications.

Then we observe the impact of hyper-paramets setting in MsI on $PM_{2.5}$ prediction performance, as shown in Fig. 14. It can be seen that the learning rate (except for a large initial value $\alpha_0 = 0.01$) has a minor impact on the final prediction performance of the proposed LSTM model, with the values of MAE, MAPE, RMSE, R^2, and NSE varying by no more than 0.45, 0.85, 1.95, 0.015, and 0.015, respectively. A large initial learning rate setting makes the model susceptible to oscillation in the late iterations and fails to converge to a stable minimum position. Moreover, even though the test data are collected from 17 observation stations at different locations in Shandong, the CEEMD-MsI still exhibits the tolerance to batch size and weight decay, and the adaptability to changing model sizes. An excessive number of parameters (140 %) can result in the model overfitting to the training samples, while an insufficient number of parameters (60 %) may lead to incomplete description of spatio-temporal features. Actually, the CEEMD-MsI driven by IMFs based multi-channel representations tends to create a flat and wide loss landscape of the objective function, making it easier to converge to the global optimal position with better generalization capability.

4.6. Robustness analysis

The standards for meteorological and environmental monitoring

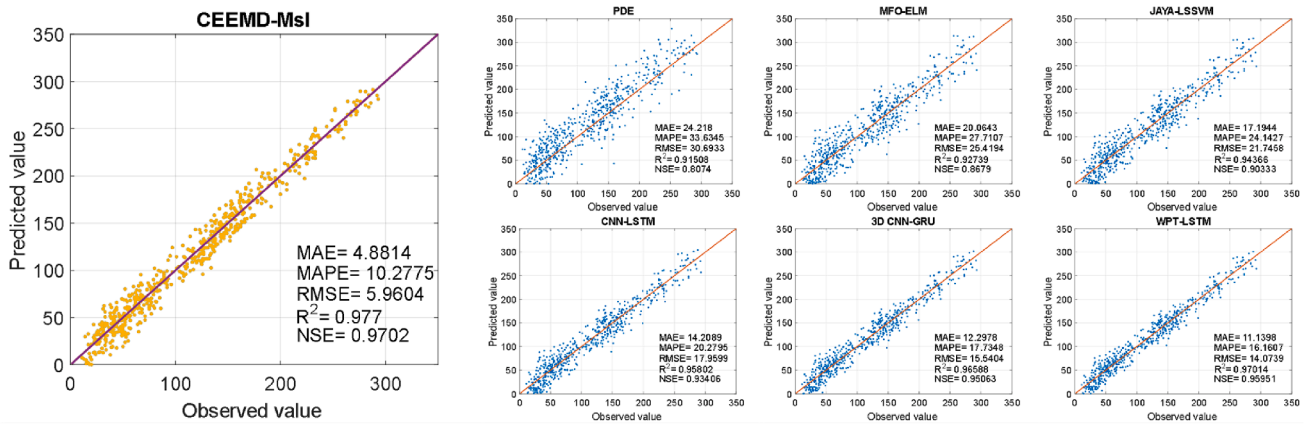


Fig. 13. Scatter plot of predicted and observed $PM_{2.5}$ concentrations of various methods.

Table 4
Comparison of average $PM_{2.5}$ prediction performance of MsI driven by different data representations.

Representation	$PM_{2.5}$ prediction performance					Inference speed (ms)	
	MAE	MAPE	RMSE	R^2	NSE	Decomposition ($T = 6$)	MsI
original time series	12.4528	17.4776	15.2798	0.9571	0.9504	–	0.89
IMFs by EMD	7.6636	12.9107	10.1755	0.9680	0.9625	0.04	3.54
IMFs by EEMD	7.1419	12.5391	9.6842	0.9704	0.9633	0.47	
IMFs by CEEMD	4.8814	10.2775	5.9604	0.9770	0.9702	0.59	

Table 5
Comparison of average $PM_{2.5}$ prediction performance of CEEMD-MsI with different decomposition iterations.

The number of IMFs	$PM_{2.5}$ prediction performance					Inference speed (ms)	
	MAE	MAPE	RMSE	R^2	NSE	CEEMD	MsI
$T = 2$	6.6177	12.4119	9.8810	0.9662	0.9580	0.17	1.24
$T = 3$	5.9574	12.0550	8.7804	0.9680	0.9602	0.22	1.35
$T = 4$	5.2461	10.9804	7.6775	0.9712	0.9624	0.31	1.76
$T = 5$	4.9648	10.3805	6.2217	0.9755	0.9679	0.44	2.30
$T = 6$	4.8814	10.2775	5.9604	0.9770	0.9702	0.59	3.54

Table 6
Comparison of average $PM_{2.5}$ prediction performance of CEEMD-MsI with different attention mechanism.

Attention mechanism	$PM_{2.5}$ prediction performance					MsI inference speed (ms)
	MAE	MAPE	RMSE	R^2	NSE	
Multi-head attention	7.2239	13.4403	10.4372	0.9665	0.9612	3.02
ProbSparse self-attention	4.8814	10.2775	5.9604	0.9770	0.9702	3.54

stations have not been established in the construction of modern sustainable cities. The monitoring stations at different locations usually have complex settings for the observation period and sampling period of the data, which places demanding requirements on $PM_{2.5}$ concentration prediction methods. Therefore, we test the robustness of CEEMD-MsI to various observation and sampling conditions, as shown in Fig. 15. The observation period from 48 h to 96 h has little effect on the predictive performance of CEEMD-MsI, and the MAE, MAPE, RMSE, R^2, and NSE are changed from 5.5483, 11.8620, 8.4554, 0.9647, and 0.9549 to 4.7226, 10.0137,5.4103, 0.98544, and 0.9735, respectively. The predictive performance of the losses is far from the level that would influence the governance policy. Therefore, the observation period can be reduced to alleviate the amount of information without losing significant $PM_{2.5}$ prediction performance, thus reducing the structural complexity of the model and increasing its inference speed.

By comparison, the manually set sampling periods from 1 h to 8 h exhibit a more pronounced impact on the CEEMD-MsI. After the data is sampled at 8 h intervals and used for model training, the $PM_{2.5}$ prediction performance of CEEMD-MsI has decreased to the MAE 22.2959, AMPE 29.7994, RMSE 25.5469, R^2 0.9217, and NSE 0.8658. Although the higher sampling period avoids the generation of bigger data, a higher sampling period results in lower temporal correlations between sampling points, posing greater challenges to the ability of capturing long-term dependencies. It is noteworthy that the majority of monitoring stations currently have the sampling period of less than 1 h, and the demand for the prediction time range is no more than 48 h in

Table 7
Comparison of average $PM_{2.5}$ prediction performance of CEEMD-MsI with different averaging number.

Averaging number	$PM_{2.5}$ prediction performance					Inference speed (ms)	
	MAE	MAPE	RMSE	R^2	NSE	CEEMD	MsI
$I = 4$	5.8987	12.1153	8.8220	0.9668	0.9588	0.50	3.54
$I = 6$	5.1158	10.9940	7.4566	0.9702	0.9639	0.54	
$I = 8$	4.8814	10.2775	5.9604	0.9770	0.9702	0.59	
$I = 10$	4.8770	10.2562	5.9479	0.9774	0.9722	0.62	

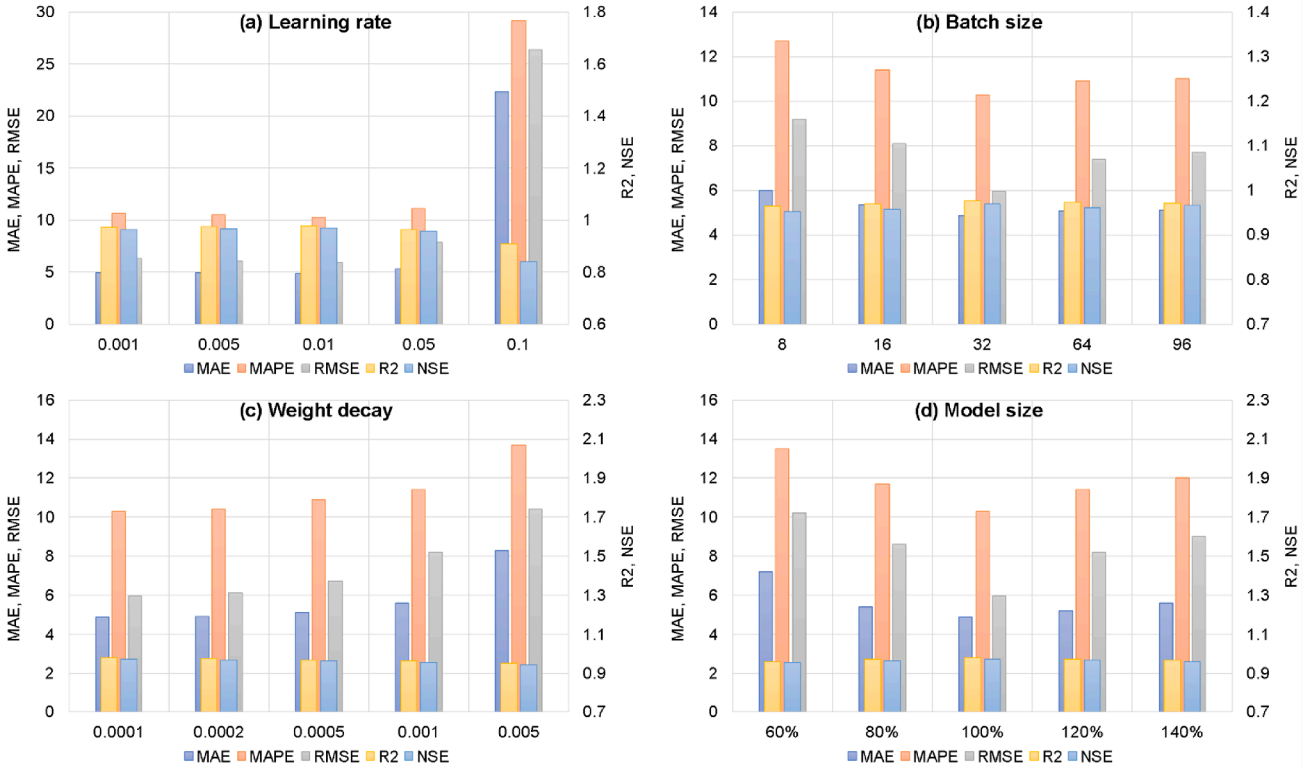


Fig. 14. Average $PM_{2.5}$ concentration prediction performance of CEEMD-MsI with a series of hyper-parameter settings.

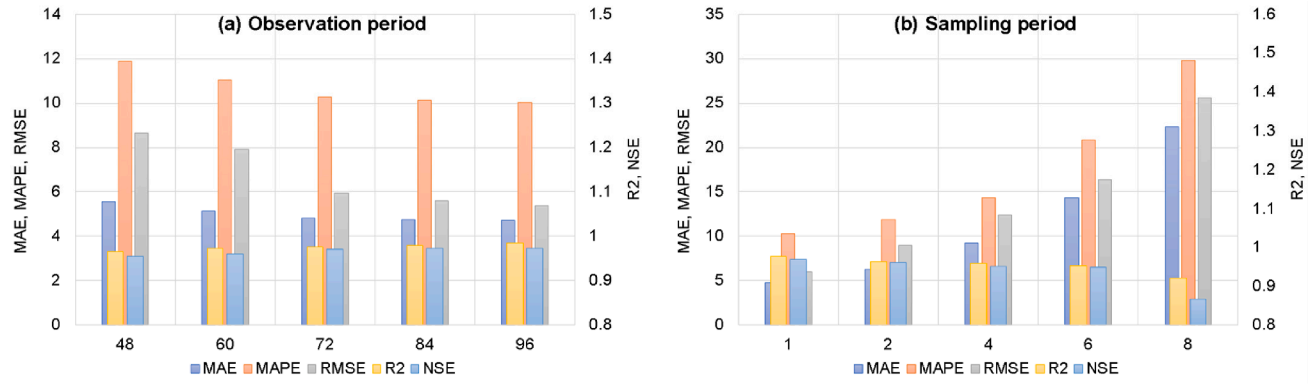


Fig. 15. Average $PM_{2.5}$ concentration prediction performance of CEEMD-MsI under different signal acquisition conditions.

practice, which can be fully met by CEEMD-MsI.

5. Conclusion

Ascertaining the precise forecast of $PM_{2.5}$ concentrations has garnered substantial interest from researchers across the globe in recent years, owing to the proven deleterious impact of $PM_{2.5}$ on human health and the sustainable development of cities and society. However, most of the existing studies focus on the selection of suitable prediction methods, while neglecting signal preprocessing and targeted model construction, resulting in unsatisfactory accuracy and efficiency. To solve this problem, we propose a MsI model driven by CEEMD based multi-channel representations to predict the $PM_{2.5}$ long-term concentrations and test its effectiveness using the meteorological and environmental data from 17 monitoring stations in Shandong, China. CEEMD can not only solve the mode pairing problem in joint analysis of multivariate variables, but also combine the multivariate input and component decomposition simultaneously to overcome the nonlinearity and non-smoothness of signals. Besides, the introduction of random disturbances in CEEMD suppresses the influence of noises and improves the accuracy and stability of signal decomposition. The MsI consisted by encoder and decoder is specifically built to fuse $PM_{2.5}$ prediction results of multiple IMFs, which can significantly reduce training and deployment costs, and improve the prediction accuracy and inference efficiency. The multi-head attention mechanism enables the informer model to adapt to signals of different lengths and frequencies, thus effectively capturing long-term dependencies in temporal sequences. To the best of our knowledge, this is the first attempt to predict long-term $PM_{2.5}$ concentrations using the deep learning model driven by data collected from various monitoring stations that span long distances and complex terrains. Through comprehensive comparisons, the effectiveness and superiority of CEEMD-MsI are proved in this study. Compared with state-of-the-art statistical modelling methods and deep learning models, the proposed CEEMD has achieved the superior $PM_{2.5}$ prediction performance with the lowest MAE, MAPE, RMSE and the highest R^2 and NSE in four typical cities. Moreover, the feasibility of deploying

CEEMD-MsI in realistic scenarios is analyzed by observing the robustness to hyper-parameters and changing conditions. Test results indicate that CEEMD-MsI can generalize well under different conditions.

At present, the intricacy and variability of terrain, climate, and environment continue to pose significant challenges to all kinds of $PM_{2.5}$ concentration prediction methodologies. In the forthcoming research, we plan to establish the federated learning based joint multi-city framework to enhance the long-term $PM_{2.5}$ concentration prediction accuracy of a single model.

CRediT authorship contribution statement

Qinghe Zheng: Conceptualization, Writing – original draft. **Xinyu Tian:** Data curation. **Zhiguo Yu:** Supervision. **Bo Jin:** Validation. **Nan Jiang:** . **Yao Ding:** Formal analysis. **Mingqiang Yang:** Methodology. **Abdussalam Elhanashi:** Investigation. **Sergio Saponara:** Project administration. **Kidiyo Kpalma:** Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by Shandong Provincial Social Science Planning Research Project (Grant Number 22CSDJ38), and Intelligent Manufacturing and Data Application Engineering Laboratory of Shandong Province.

References

Apte, J., Marshall, J. D., Cohen, A. J., & Brauer, M. (2015). Addressing global mortality from ambient PM2.5. *Environmental Science & Technology, 49*(13), 8057–8066. https://doi.org/10.1021/acs.est.5b01236

Cao, Q., Shen, L., Chen, S. C., & Pui, D. (2018). WRF modeling of PM2.5 remediation by SALSCS and its clean air flow over Beijing terrain. *Science of the Total Environment, 626*, 134–146. https://doi.org/10.1016/j.scitotenv.2018.01.062

Cheng, F. Y., Feng, C. Y., Yang, Z. M., et al. (2021). Evaluation of real-time PM2.5 forecasts with the WRF-CMAQ modeling system and weather-pattern-dependent bias-adjusted PM2.5 forecasts in Taiwan. *Atmospheric Environment, 244*, Article 117909. https://doi.org/10.1016/j.atmosenv.2020.117909

Chinatamby, P., & Jewaratnam, J. (2023). A performance comparison study on PM2.5 prediction at industrial areas using different training algorithms of feedforward-backpropagation neural network (FBNN). *Chemosphere, 317*, Article 137788. https://doi.org/10.1016/j.future.2023.02.023

Fang, B., Zhang, L., Zeng, H., et al. (2020). PM2.5-bound polycyclic aromatic hydrocarbons: Sources and health risk during non-heating and heating periods (Tangshan, China). *International Journal of Environmental Research and Public Health, 17*(2), 483. https://doi.org/10.3390/ijerph17020483

Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., & Downey, K. (2022). An integrated 3D CNN-GRU deep learning method for short-term prediction of PM2.5 concentration in urban environment. *Science of The Total Environment, 834*, Article 155324. https://doi.org/10.1016/j.scitotenv.2022.155324

Forouzanfar, M. H., Afshin, A., Alexander, L. T., et al. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet, 388*, 1659–1724. https://doi.org/10.1016/S0140-6736(16)31679-8

Fu, M., Kelly, J. A., & Clinch, J. P. (2020). Prediction of PM2.5 daily concentrations for grid points throughout a vast area using remote sensing data and an improved dynamic spatial panel model. *Atmospheric environment*, 237, 117667. Doi: 10.1016/j.atmosenv.2020.117667.

Gao, X., & Li, W. (2021). A graph-based LSTM model for PM2.5 forecasting. *Atmospheric Pollution Research, 12*(9), Article 101150. https://doi.org/10.1016/j.apr.2021.101150

Garcia, J. M., Teodoro, F., Cerdeira, R., et al. (2016). Developing a methodology to predict PM10 concentrations in urban areas using generalized linear models. *Environmental Technology, 37*(18), 2316–2325. https://doi.org/10.1080/09593330.2016. 1149228

Gong, M., Zhao, Y., Sun, J., et al. (2022). Load forecasting of district heating system based on Informer. *Energy, 253*(C), Article 124179. https://doi.org/10.1016/j.energy.2022.124179

Han, L., Zhao, J., Gao, Y., et al. (2020). Spatial distribution characteristics of PM2.5 and PM10 in Xi'an City predicted by land use regression models. *Sustainable Cities and Society, 61*, Article 102329. https://doi.org/10.1016/j.scs.2020.102329

Hedrea, E. L., Precup, R. E., Roman, R. C., et al. (2021). Tensor product-based model transformation approach to tower crane systems modeling. *Asian Journal of Control, 23*(3), 1313–1323. https://doi.org/10.1002/asjc.2494

Hong, J., Mao, F., Min, Q., et al. (2020). Improved PM2.5 predictions of WRF-Chem via the integration of Himawari-8 satellite data and ground observations. *Environmental Pollution, 263*, Article 114451. https://doi.org/10.1016/j.envpol.2020.114451

Huang, C., Hu, J., Xue, T., Xu, H., & Wang, M. (2021). High-resolution spatiotemporal modeling for ambient PM2.5 exposure assessment in China from 2013 to 2019. *Environmental Science & Technology*, 55(3), 2152–2162. Doi: 10.1021/acs.est.0c05815.

Huang, G., Li, X., Zhang, B., & Ren, J. (2021). PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. *Science of the Total Environment*, 768, 144516. Doi: 10.1016/j.scitotenv.2020.144516.

Jeong, K., Kahng, A. B., Lin, B., et al. (2010). Accurate machine-learning-based on-chip router modeling. *IEEE Embedded Systems Letters, 2*(3), 62–66. https://doi.org/10.1109/LES.2010.2051413

Jiang, N., Zheng, X., Sun, L., Zheng, H., & Zheng, Q. (2021). Long short-term memory based PM2.5 concentration prediction method. *Engineering Letters, 29*(2), 765–774. https://www.engineeringletters.com/issues_v29/issue_2/EL_29_2_46.pdf.

Jiang, N., Fu, F., Zuo, H., Zheng, X., & Zheng, Q. (2020). A municipal PM2.5 forecasting method based on random forest and WRF model. *Engineering Letters, 28*(2), 312–321. https://www.engineeringletters.com/issues_v28/issue_2/EL_28_2_07.pdf.

Jin, B., Cruz, L., & Gonçalves, N. (2020). Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. *IEEE Access, 8*, 123649–123661. https://doi.org/10.1109/ACCESS.2020.3005687

Jin, B., & Vai, M. I. (2015). An adaptive ultrasonic backscattered signal processing technique for accurate object localization based on the instantaneous energy density level. *Journal of Medical Imaging and Health Informatics*, 5(5), 1059–1064. Doi: 10.1166/jmihi.2015.1493.

Jin, X., Yang, N., Wang, X., et al. (2019). Integrated predictor based on decomposition mechanism for PM2.5 long-term prediction. *Applied Sciences, 9*(21), 4533. https://doi.org/10.3390/app9214533

Ketsarapong, S., Punyangarm, V., Phusavat, K., et al. (2012). An experience-based system supporting inventory planning: A fuzzy approach. *Expert Systems with Applications, 39*(8), 6994–7003. https://doi.org/10.1016/j.eswa.2012.01.048

Kumharn, W., Sudhibrabha, S., Hanprasert, K., et al. (2022). Improved hourly and long-term PM2.5 prediction modeling based on MODIS in Bangkok. *Remote Sensing Applications: Society and Environment*, 28, 100864. Doi: 10.1016/j.rsase.2022.100864.

Lee, H. J. (2019). Benefits of high resolution PM2.5 prediction using satellite MAIAC AOD and land use regression for exposure assessment: California examples. *Environmental Science & Technology*, 53(21), 12774–12783. Doi: 10.1021/acs.est.9b03799.

Liang, X., Li, S., Zhang, S., Huang, H., & Chen, S. (2016). PM2.5 data reliability, consistency, and air quality assessment in five Chinese cities. *Journal of Geophysical Research: Atmospheres, 121*(17), 220–236. https://doi.org/10.1002/2016JD024877

Liu, H., & Chen, C. (2020). Prediction of outdoor PM2.5 concentrations based on a three-stage hybrid neural network model. *Atmospheric Pollution Research, 11*(3), 469–481. https://doi.org/10.1016/j.apr.2019.11.019

Liu, J., Paisley, J., Schwartz, J., et al. (2019). Bayesian nonparametric ensemble for PM2.5 prediction and uncertainty characterization. *Environmental Epidemiology, 3*, 206. https://doi.org/10.1097/01.EE9.0000608104.75601.7a

Liu, W., Guo, G., Chen, F., et al. (2019). Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by PSO algorithm. *Atmospheric. Pollution Research, 10*(5), 1482–1491. https://doi.org/10.1016/j.apr.2019.04.005

Li, J., Li, X., & Wang, K. (2019). Atmospheric PM2.5 concentration prediction based on time series and interactive multiple model approach. *Advances in Meteorology, 2019*, 1–11. https://doi.org/10.1155/2019/1279565

Li, S., Xie, G., Ren, J., et al. (2020). Urban PM2.5 concentration prediction via attention-based CNN-LSTM. *Applied Sciences, 10*(6), 1953. https://doi.org/10.3390/app10061953

Lohat, S., Jain, S., & Kumar, R. (2023). AROA: Adam remora optimization algorithm and deep Q network for energy harvesting in Fog-IoV network. *Applied Soft Computing, 136*, Article 110072. https://doi.org/10.1016/j.asoc.2023.110072

Lu, X., Zhou, W., Qi, C., et al. (2021). Prediction into the future: A novel intelligent approach for PM2.5 forecasting in the ambient air of open-pit mining. *Atmospheric Pollution Research, 12*(6), Article 101084. https://doi.org/10.1016/j.apr.2021.101084

Ma, J., Yu, Z., Qu, Y., Xu, J., & Cao, Y. (2020). Application of the XGBoost machine learning method in PM2.5 prediction: A case study of Shanghai. *Aerosol and Air Quality Research, 20*(1), 128–138. https://doi.org/10.4209/aaqr.2019.08.0408

Ma, J., Ding, Y., Cheng, J. C. P., et al. (2019). Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. *Journal of Cleaner Production, 244*, Article 118955. https://doi.org/10.1016/j.jclepro.2019.118955

Ma, J., Ding, Y., Gan, V. J. L., Lin, C., & Wan, Z. (2019). Spatiotemporal prediction of PM2.5 concentrations at different time granularities using IDW-BLSTM. *IEEE Access, 7*, 107897–107907. https://doi.org/10.1109/ACCESS.2019.2932445

Nguyen, M. H., Nguyen, P. L., Nguyen, K., et al. (2021). PM2.5 prediction using genetic algorithm-based feature selection and encoder-decoder model. *IEEE Access, 9*, 57338–57350. https://doi.org/10.1109/ACCESS.2021.3072280

Pak, U., Ma, J., Ryu, U., et al. (2020). Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Science of The Total Environment, 699*, Article 133561. https://doi.org/10.1016/j.scitotenv.2019.07.367

Pozna, C., Precup, R. E., Tar, J. K., et al. (2010). New results in modelling derived from Bayesian filtering. *Knowledge-Based Systems, 23*(2), 182–194. https://doi.org/10.1016/j.knosys.2009.11.015

Precup, R. E., Duca, G., Travin, S., & Zinicovscaia, I. (2022). Processing, neural network-based modeling of biomonitoring studies data and validation on republic of moldova data. *Proceedings of the Romanian Academy, Series A: Mathematics, Physics, Technical Sciences, Information Science*, 23(4), 403–410. https://academiaromana.ro/sectii2002/proceedings/doc2022-IP/ip2022_i2_1682-Precup.pdf.

Pu, Q., & Yoo, E. H. (2021). Ground PM2.5 prediction using imputed MAIAC AOD with uncertainty quantification. *Environmental Pollution, 274*, Article 116574. https://doi.org/10.1016/j.envpol.2021.116574

Reyers, B., Moore, M. L., Haider, L., Schlüter, M. (2022). The contributions of resilience to reshaping sustainable development. *Nature Sustainability*, 5(8), 657–664. https://doi.org /s41893-022-00889-6.

Seng, D., Zhang, Q., Zhang, X., Chen, G., & Chen, X. (2021). Spatiotemporal prediction of air quality based on LSTM neural network. *Alexandria Engineering Journal, 60*(2), 2021–2032. https://doi.org/10.1016/j.aej.2020.12.009

Shi, L., Zhang, H., Xu, X., Han, M., & Zuo, P. (2022). A balanced social LSTM for PM2.5 concentration prediction based on local spatiotemporal correlation. *Chemosphere, 291*, Article 133124. https://doi.org/10.1016/j.chemosphere.2021.133124

Sun, W., & Xu, Z. (2021). A novel hourly PM2.5 concentration prediction model based on feature selection, training set screening, and mode decomposition-reorganization. *Sustainable Cities and Society, 75*, Article 103348. https://doi.org/10.1016/j.scs.2021.103348

Sun, Y., Jiang, Q., Wang, Z., et al. (2014). Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013. *Journal of Geophysical Research: Atmospheres, 119*(7), 4380–4398. https://doi.org/10.1002/2014JD021641

Tan, J., Liu, H., Li, Y., Yin, S., & Yu, C. (2022). A new ensemble spatio-temporal PM2.5 prediction method based on graph attention recursive networks and reinforcement learning. *Chaos, Solitons & Fractals*, 162, 112405. Doi: 10.1016/j.chaos.2022.112405.

Teng, M., Li, S., Song, G., et al. (2022). Including the feature of appropriate adjacent sites improves the PM2.5 concentration prediction with long short-term memory neural network model. *Sustainable Cities and Society, 76*, Article 103427. https://doi.org/10.1016/j.scs.2021.103427

Verma, A., Meenpal, T., & Acharya, B. (2022). Computational cost reduction of convolution neural networks by insignificant filter removal. *Romanian Journal of Information Science & Technology, 25*(2), 150–165. http://www.romjist.ro/full-texts/paper713.pdf.

Wang, Y., Wang, H., & Zhang, S. (2020a). Quantifying prediction and intervention measures for PM2.5 by a PDE model. *Journal of Cleaner Production, 268*, Article 122131. https://doi.org/10.1016/j.jclepro.2020.122131

Wang, Y., Wang, H., & Zhang, S. (2020b). Prediction of daily PM2.5 concentration in China using data-driven ordinary differential equations. *Applied Mathematics and Computation, 375*, Article 125088. https://doi.org/10.1016/j.amc.2020.125088

Wong, P. Y., Su, H. J., Lung, S. C. C., & Wu, C. (2023). An ensemble mixed spatial model in estimating long-term and diurnal variations of PM2.5 in Taiwan. *Science of The Total Environment*, 2023, 161336. https://doi.org/10.1016/j.scitotenv.2022.161336.

Wu, J.-B., Xu, J., Pagowski, M., et al. (2015). Modeling study of a severe aerosol pollution event in December 2013 over Shanghai China: An application of chemical data assimilation. *Particuology, 20*, 41–51. https://doi.org/10.1016/j.partic.2014.10.008

Xiao, Q., Geng, G., Cheng, J., et al. (2021). Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models. *Atmospheric Environment, 244*, Article 117921. https://doi.org/10.1016/j.atmosenv.2020.117921

Xiao, Y., Zhao, J., Liu, H., et al. (2020). Dynamic prediction of PM2.5 diffusion in urban residential areas in severely cold regions based on an improved urban canopy model. *Sustainable Cities and Society, 62*, Article 102352. https://doi.org/10.1016/j.scs.2020.102352

Xiao, Q., Geng, G., Liang, F., et al. (2020). Changes in spatial patterns of PM2.5 pollution in China 2000–2018: Impact of clean air policies. *Environment International, 141*, Article 105776. https://doi.org/10.1016/j.envint.2020.105776

Xia, C., Wang, Z., Zheng, C., et al. (2019). A new coupled disease-awareness spreading model with mass media on multiplex networks. *Information Sciences, 471*, 185–200. https://doi.org/10.1016/j.ins.2018.08.050

Xu, Y., Yang, W., & Wang, J. (2017). Air quality early-warning system for cities in China. *Atmospheric Environment, 148*, 239–257. https://doi.org/10.1016/j.atmosenv.2016.10.046

Xu, M., & Wang, Y. X. (2015). Quantifying PM2.5 concentrations from multi-weather sensors using hidden Markov models. *IEEE Sensors Journal, 16*(1), 22–23. https://doi.org/10.1109/JSEN.2015.2485665

Yang, X., Zhang, T., Zhang, Y., et al. (2021). Global burden of COPD attributable to ambient PM2.5 in 204 countries and territories, 1990 to 2019: A systematic analysis for the Global Burden of Disease Study 2019. *Science of the Total Environment, 796*, Article 148819. https://doi.org/10.1016/j.scitotenv.2021.148819

Yang, H., Liu, Z., & Li, G. (2022). A new hybrid optimization prediction model for PM2.5 concentration considering other air pollutants and meteorological conditions. *Chemosphere, 307*, Article 135798. https://doi.org/10.1016/j.chemosphere.2022.135798

Yeo, I., Choi, Y., Lops, Y., & Sayeed, A. (2021). Efficient PM2.5 forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Computing and Applications, 33*(22), 15073–15089. https://doi.org/10.1007/s00521-021-06082-8

Yu, Y., Li, H., Sun, S., & Li, Y. (2022). PM2.5 concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy. *Sustainable Cities and Society, 85*, Article 104049. https://doi.org/10.1016/j.scs.2022.104049

Yu, T., Wang, Y., Huang, J., et al. (2022). Study on the regional prediction model of PM2.5 concentrations based on multi-source observations. *Atmospheric. Pollution Research, 13*(4), Article 101363. https://doi.org/10.1016/j.apr.2022.101363

Zhang, L., Lin, J., Qiu, R., et al. (2018). Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model. *Ecological Indicators*, 2018, 95, 702–710. Doi: 10.1016/j.ecolind.2018.08.032.

Zhang, X., & Gan, H. (2023). STF-Net: An improved depth network based on spatio-temporal data fusion for PM2.5 concentration prediction. *Future Generation Computer Systems, 144*, 37–49. https://doi.org/10.1016/j.future.2023.02.023

Zhang, Y., & Yan, Q. (2023). A spatiotemporal model for PM2.5 prediction based on the K-Core idea and label distribution. *Meteorological Applications, 30*(1), e2115.

Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory-Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere, 220*, 486–492. https://doi.org/10.1016/j.chemosphere.2018.12.128

Zhao, K., Hu, J., Shao, H., & Hu, J. (2023). Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy. *Reliability Engineering & System Safety, 262*, Article 109246. https://doi.org/10.1016/j.ress.2023. 109246

Zhao, K., Jia, F., & Shao, H. (2023). A novel conditional weighting transfer Wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains. *Knowledge-Based Systems, 262*, Article 110203. https://doi.org/10.1016/j.knosys.2022.110203

Zheng, Q., Tian, X., Yu, Z., et al. (2023a). Application of wavelet-packet transform driven deep learning method in PM2.5 concentration prediction: A case study of Qingdao, China. *Sustainable Cities and Society, 92*, Article 104486. https://doi.org/10.1016/j.scs.2023.104486

Zheng, Q., Tian, X., Yu, Z., et al. (2023b). DL-PR: Generalized automatic modulation classification method based on deep learning with priori regularization. *Engineering Applications of Artificial Intelligence, 122*, Article 106082. https://doi.org/10.1016/j.engappai.2023. 106082

Zheng, Q., Zhao, P., Wang, H., Elhanashi, A., & Saponara, S. (2022). Fine-grained modulation classification using multi-scale radio transformer with dual-channel representation. *IEEE Communications Letters, 26*(6), 1298–1302. https://doi.org/10.1109/LCOMM.2022.3145647

Zheng, Q., Zhao, P., Li, Y., Wang, H., & Yang, Y. (2021). Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification. *Neural Computing and Applications, 33*(13), 7723–7745. https://doi.org/10.1007/s00521-020-05514-1

Zheng, Q., Zhao, P., Zhang, D., & Wang, H. J. (2021). MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. *International Journal of Intelligent Systems, 36*(12), 7204–7238. https://doi.org/10.1002/int.22586

Zheng, Q., Yang, M., Tian, X., et al. (2020). A full stage data augmentation method in deep convolutional neural network for natural image classification. *Discrete Dynamics in Nature and Society, 4706576*. https://doi.org/10.1155/2020/4706576

Zheng, Q., Yang, M., Yang, J., et al. (2018). Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process. *IEEE Access, 6*, 15844–15869. https://doi.org/10.1109/ACCESS.2018.2810849

Zhou, S., & Lin, R. (2019). Spatial-temporal heterogeneity of air pollution: The relationship between built environment and on-road PM2.5 at micro scale. *Transportation Research Part D: Transport and Environment, 76*, 305–322. https://doi.org/10.1016/j.trd. 2019.09.004

Zhu, J., Deng, F., Zhao, J., & Zheng, H. (2021). Attention-based parallel networks (APNet) for PM2.5 spatiotemporal prediction. *Science of The Total Environment, 769*, Article 145082. https://doi.org/10.1016/j.scitotenv.2021.145082

Zhu, M., & Xie, J. (2023). Investigation of nearby monitoring station for hourly PM2.5 forecasting using parallel multi-input 1D-CNN-biLSTM. *Expert Systems with Applications, 211*, Article 118707. https://doi.org/10.1016/j.eswa.2022.118707

Zoran, M. A., Savastru, R. S., Savastru, D. M., & Tautan, M. N. (2020). Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. *Science of the Total Environment, 738*, Article 139825. https://doi.org/10.1016/j.scitotenv.2020.139825