Carbon emissions trading price forecasts by multi-perspective fusion

Chong Zhang [a, *], Zhiying Feng [b]

[a] *Business School, Nanjing University, Nanjing, China*
[b] *Business School, The University of Sydney, Sydney, Australia*

ABSTRACT

The precise prediction of carbon emissions trading prices is the foundation for the stable and sustainable development of the carbon financial market. In recent years, influenced by a combi- nation of factors such as the pandemic, trading regulations, and policies, carbon prices have exhibited strong random volatility and clear non-stationary characteristics. Traditional single-perspective prediction methods based on conventional statistical models are increasingly inadequate due to the homogenization of features and are struggling to adapt to China's regional carbon emissions trading market. Therefore, this paper proposes a multi-perspective fusion-based prediction method tailored to the Chinese market. It leverages carbon emissions trading information from key cities as relevant features to predict the price changes in individual cities. Inspired by the development of artificial intelligence, this paper implements various time series models based on deep neural networks. The effectiveness of the multi-perspective approach is validated through multiple metrics. It provides scientific decision-making tools for domestic carbon emissions trading investors, making a significant contribution to strengthening carbon market risk management and promoting the establishment and rational development of a unified carbon market in China.

* Corresponding author: Chong Zhang
E-mail address: zhangc_1024@163.com

1. Introduction and literature review

Carbon emissions trading, as a market mechanism aimed at reducing greenhouse gas emissions, has been widely implemented globally (Carl and Fedor, 2016). For instance, the European Union Emissions Trading System (EU ETS) and China's national carbon market launched in 2021 are indicative of this global trend. With increasing international concern over climate change, carbon prices are generally on the rise in most markets, exerting greater pressure on businesses to reduce emissions. Additionally, to prevent carbon leakage and capacity shifting, some regions like the European Union are considering the implementation of carbon border adjustment mechanisms to ensure fair competition for high-carbon products. Furthermore, in addition to mandatory carbon markets, voluntary carbon markets and other forms are also gradually emerging.

China, as the world's largest emitter of greenhouse gases and a major economic powerhouse, is currently accelerating its transition from a high-carbon to a low-carbon economy (Bi et al., 2019; Liu et al., 2023; Wang et al., 2023). As of December 22nd, the national carbon emissions trading market has achieved a cumulative trading volume exceeding 10 billion yuan. Since the launch of the national carbon market, it has operated for 350 trading days, with a cumulative trading volume of 2.23 billion tons of carbon emissions quotas and a total trading volume of 10.121 billion yuan. While ensuring the smooth operation and healthy development of the trading market, it has played a pivotal role in promoting greenhouse gas emission reductions among enterprises and strengthening awareness of low-carbon development in various sectors of society. In this context, the importance of carbon emissions trading prediction is self-evident. Predictions for the carbon market not only provide critical guidance for China's economic restructuring, the development of green industries, and the innovation of clean technologies but also ensure that China can meet its global emissions reduction commitments, including peaking its carbon emissions by 2030 and achieving carbon neutrality by 2060. This market-oriented mechanism holds profound significance for China's leadership role in global climate governance and its continued economic prosperity.

In the early stages, constrained by the limited development of mathematical tools, scholars predominantly employed qualitative analysis methods for forecasting carbon emissions trading prices (Zhu et al., 2018). However, with the rapid advancement of computer technology, an increasing number of researchers began to adopt quantitative analysis methods for predicting carbon emissions trading prices. Methodologically, the relevant research can be categorized into three main groups: statistical models, artificial intelligence models, and ensemble models. Statistical models primarily encompass Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models (Byun and Cho, 2013), Dynamic Model Averaging (DMA) (Koop and Tole, 2013), Autoregressive Moving Average (ARMA) models (Sanin et al., 2015), Vector Autoregression (VAR) models (Yang, 2023) and similar approaches. Nevertheless, these models often lack consideration of the inherent time series features and may not be suitable for long-term forecasting (Sun and Zhang, 2018). In contrast, artificial intelligence models have gained prominence among scholars due to their robust learning capabilities and the absence of assumptions about data distribution, such as normality or uniformity. Prominent models in this category include Artificial Neural Networks (ANNs) (Atsalakis, 2016), Multilayer Perceptron (MLP) models (Fan et al., 2015), and Least Squares Support Vector Machine (LSSVM) models (Zhu et al., 2017a). However, existing models for carbon emissions trading prediction tend to exhibit structural simplicity, which can result in overfitting issues. Ensemble prediction models partially mitigate such problems and encompass data preprocessing, optimization algorithms, and individual prediction models. Data preprocessing techniques include Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD) (Huan-ying and Xiang-sheng, 2018), and Modified Empirical Mode Decomposition (MEEMD) (Yang et al., 2020). Optimization algorithms comprise Particle Swarm Optimization (PSO) (Zhu et al., 2017b) and Genetic Algorithms (GA) (Wang et al., 2016). Individual prediction models encompass Extreme Gradient Boosting (XGBoost) (Zhu et al., 2018) and Extreme Learning Machine (ELM) (Li et al., 2016), among others. The central philosophy behind ensemble forecasting models is the disaggregation of carbon emissions trading prices

into multiple modalities, enhancing predictive precision through the amalgamation of forecasts for each modality. Nonetheless, an examination of contemporary scholarly literature indicates several deficiencies in pertinent research. Firstly, these methodologies necessitate the cultivation of numerous foundational learners, predicated on intricate optimization sequences. Secondly, a proportion of studies frequently employ a unimodal conceptualization strategy for prognosticating across various modalities. These attributes introduce significant complexities when endeavoring to directly apply ensemble models for the prognostication of profoundly stochastic carbon trading datasets.

The central philosophy behind ensemble forecasting models is the disaggregation of carbon emissions trading prices into multiple modalities, enhancing predictive precision through the amalgamation of forecasts for each modality. Nonetheless, an examination of contemporary scholarly literature indicates several deficiencies in pertinent research. Firstly, these methodologies necessitate the cultivation of numerous foundational learners, predicated on intricate optimization sequences. Secondly, a proportion of studies frequently employ a unimodal conceptualization strategy for prognosticating across various modalities. These attributes introduce significant complexities when endeavoring to directly apply ensemble models for the prognostication of profoundly stochastic carbon trading datasets.

Thus, this paper presents a multi-perspective fusion-based approach for forecasting carbon emissions trading, primarily addressing the pronounced stochasticity and non-stationarity exhibited by carbon emissions trading prices in Chinese cities. On one hand, the method enhances predictive capacity by amalgamating the price changes in carbon trading across multiple cities. On the other hand, inspired by the generalization capabilities inherent in large models (Brown et al., 2020) relying on an extensive set of parameters, the approach employs diverse deep neural network architectures to enhance model fitting capabilities.

The remainder of this paper is organized as follows: Section 2 the research methodology, including deep learning models for carbon trading price prediction and relevant evaluation metrics; Section 3 reports the urban carbon trading dataset used in this study, the basic process of the multi-perspective fusion algorithm, and the experimental results with associated analysis. Finally, our conclusions are reported in Section 4.

2. Research methods

The task of carbon trading price prediction involves time series forecasting. Traditional methods such as ARMA (Hao et al., 2020), SVM (Lu et al., 2020), and ensemble learning (Zhu et al., 2018) rely on high-quality features and stable data distributions, making them less suitable for predicting non-stationary distributions in the context of China's carbon trading prices. With the development of deep neural networks, especially in terms of their remarkable generalization capabilities, this has inspired innovative approaches. In this paper, we introduce cutting-edge work in time series analysis to enhance the predictive ability for the Chinese carbon emission trading market. The specific methodology is as follows:

2.1. Time Series Algorithms

In this subsection, we will provide a detailed overview of some time series models that have been utilized for carbon trading price prediction, as well as introduce novel deep neural network-based time series models that have not yet been applied but are suitable for carbon trading price forecasting tasks.

(1) LSTM (Graves and Graves, 2012):

Within the realm of recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) units represent a significant advancement, particularly in the context of carbon trading price prediction (Zhou et al., 2022). LSTM has been instrumental in addressing a fundamental challenge encountered in traditional RNNs, namely the vanishing

and exploding gradient issues. These issues manifest when the network is trained to identify patterns within extensive sequences, making it arduous for conventional RNNs to retain memory of past inputs within the sequence. In contrast, LSTM, with its specialized architectural design, excels in capturing long-term dependencies inherent in sequential data, which is crucial in the context of predicting carbon trading prices.

The integration of LSTM and Fully Convolutional Network (FCN) stands as a sophisticated deep learning architecture, offering significant potential for enhancing carbon trading price prediction. LSTM's capability to model long-range temporal dependencies within sequential data is synergized with FCN's feature extraction and spatial modeling capabilities. This fusion creates a versatile computational framework well-suited for tackling the intricate task of forecasting carbon trading prices. LSTM, known for encoding complex temporal relationships and effectively handling challenges like the vanishing gradient problem, is highly relevant when dealing with carbon trading data, which often exhibits substantial volatility.

(2) LSTM-FCN (Karim et al., 2017):

Furthermore, FCN, initially designed for image segmentation, gracefully adapts to the one-dimensional nature of time-series data, enabling it to adeptly extract both local and global temporal features from the carbon trading price sequences. This harmonious integration empowers the model to gain a deeper understanding of the intricate patterns within carbon trading price data (Ji et al., 2019), leading to superior predictive performance in the realm of carbon trading price forecasting.

(3) mWDN (Wang et al., 2018):

The Multilevel Wavelet Decomposition Network (mWDN) is an advanced neural network architecture rooted in the principles of multiresolution analysis, particularly wavelet transform theory. mWDN is designed to facilitate the decomposition of complex data into multiple levels of wavelet representations, thus enabling hierarchical feature extraction and analysis. This network incorporates a series of wavelet transformation layers, each responsible for capturing specific frequency and spatial information across different scales.

The resulting multilevel representations are subsequently processed and integrated to facilitate tasks such as signal denoising, feature extraction, or even hierarchical classification. mWDN holds great promise in applications, including carbon trading price predictions, where the inherent multiscale nature of data needs to be effectively leveraged for improved understanding and decision-making. Its efficacy derives from its ability to harness the advantages of wavelet analysis within the context of deep learning paradigms, thereby enhancing the network's capacity for discerning intricate patterns and structures within data across multiple resolutions.

(4) TCN (Bai et al., 2018):

A Temporal Convolutional Network (TCN) is an advanced neural network architecture specialized in modeling sequential data with a focus on exploiting temporal dependencies. TCN leverages a one-dimensional convolutional structure with causal or dilated convolutions, enabling it to efficiently capture long-range dependencies within time-series data. This network architecture is designed to address various temporal modeling tasks, including forecasting, classification, and generative modeling. Notably, TCN's unique attributes, such as its ability to maintain a wide receptive field while preserving data causality, facilitate parallelization for faster training, and adaptability to sequences of varying lengths, make it particularly well-suited for applications like carbon trading price forecasting (Zhang and Wen, 2022). TCNs have consistently demonstrated remarkable performance in a wide range of temporal data analysis tasks, establishing them as a prominent choice in modern deep learning applications for time-series processing.

(5) TST (Zerveas et al., 2021):

The Temporal Set Transformer (TST) is an advanced neural network architecture designed for the analysis and modeling of temporal data sequences, making it particularly well-suited for tasks like carbon trading price predictions. Drawing inspiration from the transformer architecture, originally developed for natural language

processing, TST extends its application to sequential data. This model employs self-attention mechanisms to capture long-range temporal dependencies and interactions within input time-series data, enabling it to discern intricate temporal patterns effectively. TST also utilizes set-based operations to process unordered data, accommodating scenarios with irregular time intervals or variable-length sequences. By combining self-attention and set-based processing, TST offers a versatile and powerful framework for temporal data analysis, demonstrating its effectiveness in various time-series-related applications, including carbon trading price forecasting.

(6) XceptionTime (Ismail Fawaz et al., 2020):

InceptionTime is a sophisticated deep learning architecture tailored for the analysis of time-series data, including its relevance in carbon trading price forecasting. Inspired by the inception module originally introduced in convolutional neural networks, this innovative design comprises multiple parallel convolutional and pooling pathways of varying filter sizes. This enables the network to effectively capture multiscale temporal features within time-series sequences. InceptionTime combines various filter sizes to extract diverse temporal patterns while also incorporating bottleneck layers to reduce computational complexity. With its deep and flexible structure, InceptionTime excels at modeling intricate temporal dependencies, consistently demonstrating exceptional performance in tasks such as time-series classification and forecasting.

(7) XCM (Fauvel et al., 2021):

XCM, an acronym for "Explainable Convolutional Neural Network," is an advanced neural network architecture engineered to enhance the interpretability of deep learning models. In addition to its utility in medical diagnosis and autonomous systems, XCM's transparency-enhancing techniques also make it highly valuable for applications like carbon trading price forecasting. It provides valuable insights into model reasoning, aiding in better understanding and decision-making in this context.

(8) gMLP (Liu et al., 2021):

gMLP, or "gated Multilayer Perceptron," is an innovative neural network architecture that introduces gating mechanisms within the traditional Multilayer Perceptron (MLP) structure. By incorporating learnable gating units similar to those found in recurrent neural networks (RNNs), gMLP enhances the capacity of MLPs to model sequential data. This feature makes gMLP highly applicable in various tasks, including carbon trading price forecasting (Wang et al., 2021), where it can efficiently capture long-range dependencies and contextual information within sequences, improving the accuracy of predictions in deep learning applications.

(9) MiniRocket (Tan et al., 2022):

MiniRocket is a state-of-the-art machine learning technique designed for the analysis of time-series data, including its applicability to tasks such as carbon trading price forecasting. It leverages random convolutional kernels to efficiently perform feature extraction. MiniRocket employs a random selection of convolutional filters, followed by an aggregation operation, enabling it to capture essential temporal patterns and dependencies within time-series sequences. This approach excels in both computational efficiency and predictive accuracy, making MiniRocket a valuable tool for various time-series data analysis tasks, including classification and regression. Its demonstrated effectiveness in handling large-scale and high-dimensional time-series datasets underscores its potential for real-world applications that require robust time-series analysis capabilities.

2.2. Error Metrics

In this section, we will delineate the assessment metrics for several prototypical time series forecasting models.

(1) MSE:

The Mean Squared Error (MSE) loss $\mathcal{L}_{\mathcal{MSE}}$ is a commonly employed criterion in regression tasks and certain optimization problems in the realm of machine learning and statistical modeling. It quantifies the average squared difference between the predicted values and the actual ground truth values. Mathematically, given n predictions

$f(x_i)$ and and corresponding true values y_i, the MSE is defined as:

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2$$

where x_i is the i-th sample and f the predicted function.

The characteristics of the MSE curve are smooth continuity and differentiability. Additionally, when the difference between y_i and $f(x_i)$ exceeds one unit, it increases the error; when the difference is less than one unit, it decreases the error. This behavior is dictated by the properties of squares. In other words, MSE imposes a greater penalty on errors with larger magnitude (> 1) and a smaller penalty on errors with smaller magnitude (< 1). From a training perspective, the model tends to prioritize points with larger errors and assigns greater weight to these locations.

(2) MAE:

The Mean Absolute Error (MAE) is a prevalent metric in the fields of machine learning, statistics, and econometrics when gauging the performance of regression models. Unlike quadratic loss functions, the MAE provides a linear penalty for prediction errors. Specifically, it calculates the average magnitude of errors between the predicted and the actual values, without considering their direction. Mathematically, given n predictions $f(x_i)$ and and corresponding true values y_i, the MAE is defined as:

$$\mathcal{L}_{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)|.$$

The merit of MAE lies in its interpretability, as it directly signifies the average error magnitude in the units of the variable of interest. Furthermore, due to its linear nature, MAE is less sensitive to outliers compared to squared error metrics, making it particularly suitable for applications where large deviations are not intrinsically more penalizing than smaller ones.

(3) MAPE:

The Mean Absolute Percentage Error (MAPE) is a widely-used relative metric for assessing the accuracy of predictive models, especially in forecasting tasks within domains like economics, finance, and operations research. Unlike absolute error metrics, which measure errors in the original units of the data, the MAPE expresses errors as a percentage, making it scale-independent and facilitating comparisons across different data sets and units. Mathematically, given n predictions $f(x_i)$ and and corresponding true values y_i, the MAPE is defined as:

$$\mathcal{L}_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f(x_i)}{y_i}\right|.$$

The outcome is presented in percentage terms, with a smaller MAPE value suggesting a better fit of the predictive model to the actual data.

(4) R^2 Score:

The R^2 score, commonly known as the coefficient of determination, quantifies the proportion of the variance in the dependent variable that is predicted from the independent variables in a regression model. It provides insights into the goodness-of-fit of the model to the observed data and is a metric used to evaluate the performance of regression algorithms. Mathematically, given n predictions $f(x_i)$ and corresponding true values y_i, the R^2 score is defined as:

$$\mathcal{L}_{\mathcal{R}^2} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - f(x_i)\right)}{\sum_{i=1}^{n}(y_i - \bar{y})},$$

where $\bar{y}$ is the mean of the actual values. An R^2 score closer to1 indicates that a larger proportion of variance in the dependent variable is accounted for by the regression model. However, it's important to approach R^2 with caution. A higher R^2 does not necessarily mean the model is suitable for prediction, especially if the model is overfitted. Furthermore, in multiple regression scenarios, R^2 can increase just by adding more variables, regardless of their relevance, so adjusted R^2 is sometimes considered to account for the number of predictors in the model.

3. Carbon Trading Growth Forecast Method

This section primarily focuses on the exposition of our experimental setup and methodological analysis.

3.1. Sample Selection and Augmentation

This study primarily focuses on the carbon emission trading pilots in four Chinese cities: Shenzhen, Shanghai, Hubei, and Guangzhou. Due to the presence of missing data in the original dataset, we adopted a two-pronged approach. Firstly, to ensure timeliness, the sample period was limited to August 1, 2021, through August 30, 2023. Secondly, a combination of mean imputation and zero imputation techniques was employed to ensure data completeness.

Furthermore, to validate the effectiveness of the multi-perspective forecasting method, this paper takes the four cities as subjects and predicts the percentage changes in carbon emission trading for individual cities. Specifically, as illustrated in Figure 1, the left-hand side figure depicts the distribution of percentage changes in the original data, while the right-hand side elucidates the mean and variance of data pertaining to each city.
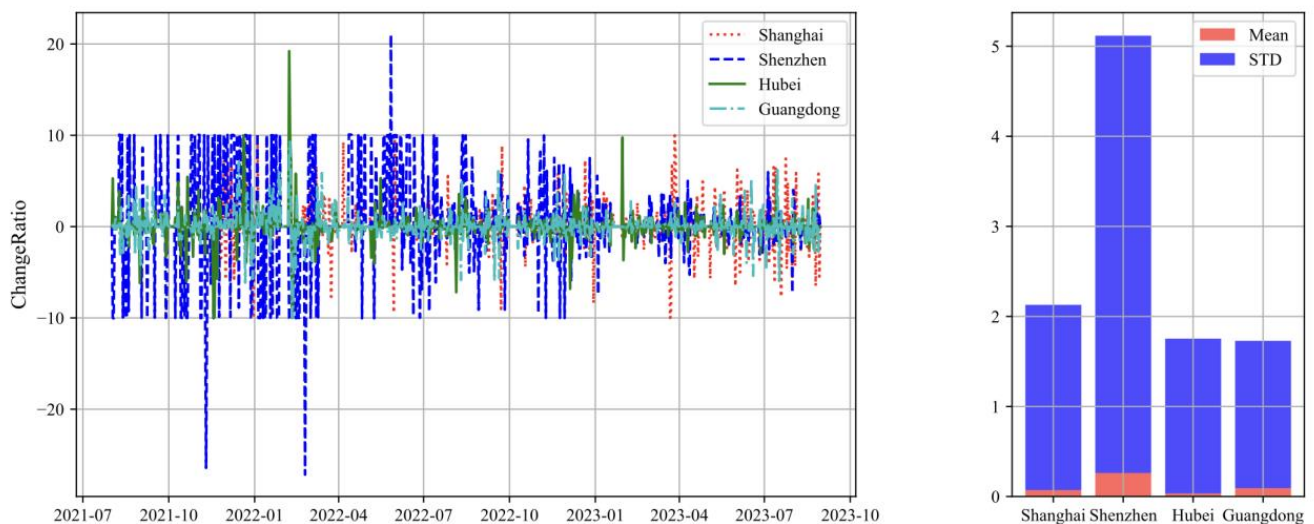

Figure 1. Data distribution.

3.2. Training Process

The algorithm for predicting carbon trading prices based on multi-perspective fusion adheres to the general workflow of time series forecasting with deep learning models. This process includes the following steps:

- Data Collection and Preparation: Gather high-quality carbon trading price increase data that is relevant to the problem. Clean the data by addressing issues like missing values, outliers, and duplicates. Normalize or standardize the data to ensure consistent scaling.
- Feature Aggregation: Combine the increase data from Shanghai, Shenzhen, Hubei, and Guangdong as relevant features. Implement a sliding window approach with a window size of 90 days to forecast the next day's prices.
- Data Segmentation: Use a time-based segmentation method to split the dataset into three subsets: a training set (80%), a validation set (20%), and a testing set.
- Model Selection: Choose deep learning models suitable for time series forecasting, such as LSTM, LSTM-FCN, mWDN, etc. Configure the number of neurons in hidden layers to achieve specific network architectures.
- Hyperparameter Tuning: Train the models by adjusting hyperparameters such as learning rate, batch size, and the number of iterations. Monitor the loss function and other potential evaluation metrics during the training process.
- Performance Evaluation: Assess the model's performance using the validation set and fine-tune model algorithm parameters to prevent overfitting or underfitting.
  This structured approach ensures the development of an effective carbon trading price prediction model that

3.3. Results and Analysis

In this experiment, the baseline models employed are all based on neural network algorithms, as described in Section 2, including LSTM, LSTM-FCN, mWDN, TCN, TST, among others. Furthermore, we ultimately adopted Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared score (R^2 score) as evaluation metrics. Additionally, this experiment utilized the past 90 days' data as samples to predict the percentage increase for the following day, partitioning the dataset with an 80% training set and a 20% testing set. In the subsequent content, we will present the algorithmic performance from both the training and testing perspectives.

As shown in Figure 2, during the training process, we selected Mean Squared Error (MSE) as the evaluation criterion, as shown in the figure. The horizontal axis represents the training iterations, while the vertical axis represents the MSE. We specifically chose Shanghai as representative cases. The red dashed line represents the training curve based on data from a single source, while the blue solid line represents the training curve when integrating data from other cities.

To ensure fairness in the comparison, the algorithm model parameters were kept consistent. The results indicate that among various model performances, the multi-perspective fusion approach consistently outperforms single-source data training.

Furthermore, within the testing dataset, we computed the relevant performance metrics. Specifically, lower values of Mean Squared Error (MES), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are desirable, while higher R^2 scores indicate superior performance. As demonstrated in Table 1, it is evident that the proposed Multi-perspective Fusion method consistently outperforms single-source predictions, with the exception of the mWDN and XCM models. This observation aligns with the training curves based on the Shanghai dataset. In conclusion, we opted for models trained on data from Shanghai City to perform inference on the testing dataset, as illustrated in the Figure 3 and Figure 4. In this context, the red dashed line represents the predictions of the baseline model, while the blue solid line represents the predictions of the multi-perspective model. Taking the data from May to June 2023 as an example, the original data exhibited a downward trend. However, in most of the multi-perspective models, correct predictions were made, whereas the baseline models such as XCM and MiniRocket incorrectly predicted an upward trend. This clearly underscores the superior predictive capability of the multi-perspective

models in capturing trends.
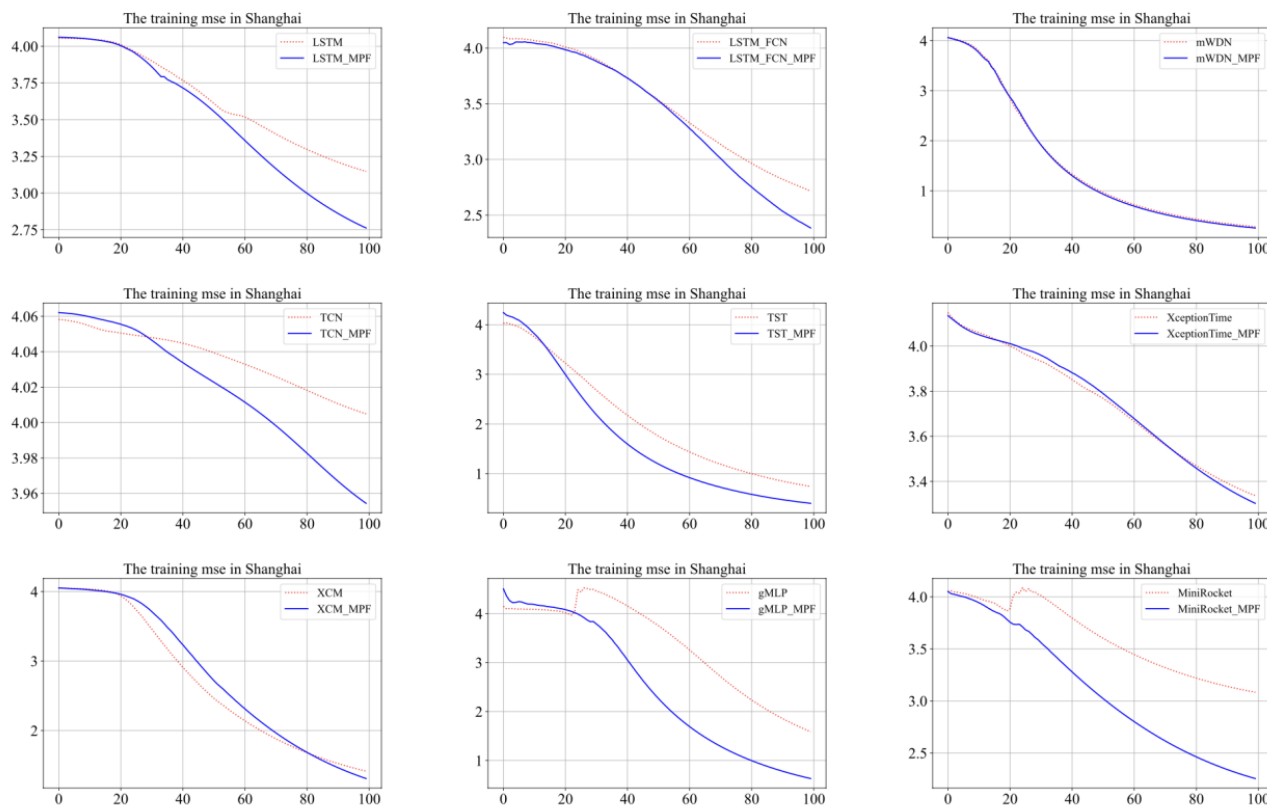
Figure 2. The mean squared error (MSE) loss for different algorithmic models trained on Shanghai's carbon trading percentage increase data.

Table 1. Evaluation Indicators in Shanghai.

Model	Total Params	MSE	MAE	MAPE	R^2 Score
LSTM	565801	6.31	1.66	90794016.0	0.11
LSTM_MPF	568201	6.12	1.71	94586440.0	0.14
LSTM_FCN	341093	7.69	2.03	140405504.0	-0.08
LSTM_FCN_MPF	343781	7.21	1.75	83412760.0	-0.02
mWDN	476773	7.68	2.07	124977992.0	-0.08
mWDN_MPF	477349	10.98	2.61	134746448.0	-0.55
TCN	66676	7.1	1.64	83253008.0	0.0
TCN_MPF	67276	7.1	1.62	77472032.0	0.0
TST	419201	9.31	2.28	140712688.0	-0.31
TST_MPF	419585	7.61	1.95	109542008.0	-0.07
XceptionTime	399090	8.07	2.12	151877472.0	-0.14
XceptionTime_MPF	399570	7.82	2.0	136900032.0	-0.1
XCM	47619	7.45	2.01	145423456.0	-0.05
XCM_MPF	116739	8.01	2.15	138918912.0	-0.13
gMLP	2426101	9.84	2.41	158444368.0	-0.39
gMLP_MPF	2426869	8.43	2.22	124328128.0	-0.19
MiniRocket	29989	10.3	2.6	203984768.0	-0.45
MiniRocket_MPF	29989	7.56	1.99	112000360.0	-0.07
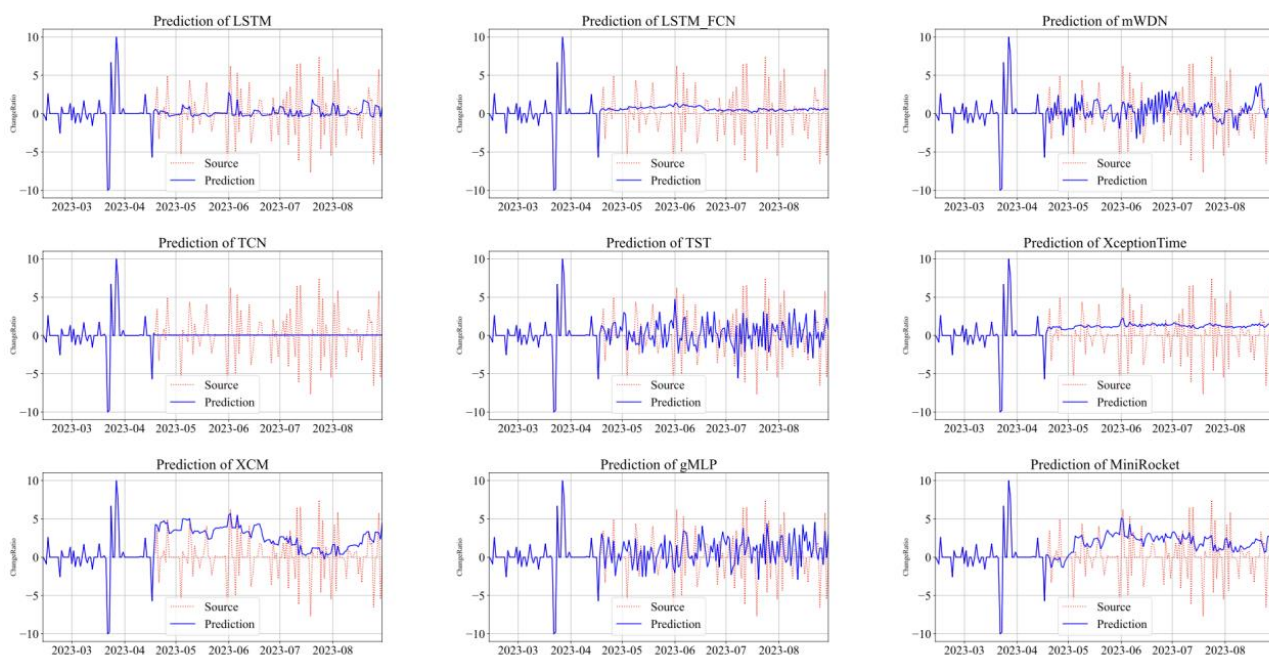
Figure 3. Inference results of the benchmark model trained based on Shanghai data.
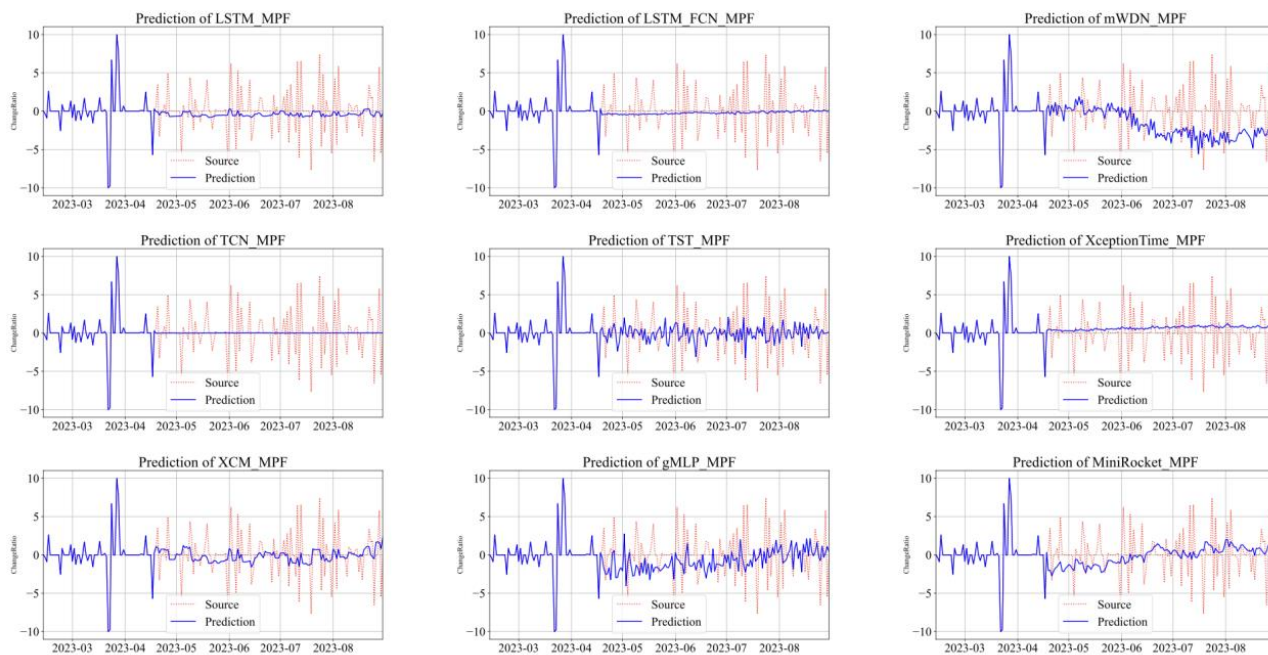
Figure 4. Inference results of the MPF model trained based on Shanghai data.

4. Conclusion

The regional carbon emission trading market in China was established relatively late. In recent years, it has been subject to a combination of factors such as the COVID-19 pandemic, trading regulations, and policies. Consequently, carbon prices have exhibited significant random fluctuations, displaying pronounced nonlinear and

non-stationary characteristics. Existing single-perspective forecasting methods tend to homogenize various distinctive features within the original carbon price time series, only addressing specific aspects of carbon price temporal fluctuations. They often fail to capture the comprehensive and effective information underlying carbon price volatility, making them less suitable for China's regional carbon emission trading market.

Therefore, this paper introduces a multi-perspective fusion-based forecasting approach tailored to the Chinese market. It leverages carbon emission trading information from key cities as relevant features to predict the price changes in individual cities. In the practical implementation, data from four cities, namely Shanghai, Shenzhen, Guangdong, and Hubei, were selected. Comparative analysis was performed using multiple deep forecasting models. Both the training and testing phases validated the effectiveness of our approach. However, there are some limitations in this work, including the high degree of data missingness, substantial randomness, and room for improvement, especially with regard to the R2 score metric. Future research will focus on the following three areas: (1) in-depth exploration of factors influencing trading randomness; (2) construction of deep forecasting models tailored to high randomness scenarios; (3) multi-model fusion, among others.

Funding Statement

Acknowledgment

Conflict of interest

All the authors claim that the manuscript is completely original. The authors also declare no conflict of interest.

References

Atsalakis, G. S. (2016). Using computational intelligence to forecast carbon prices. *Applied Soft Computing*, 43, 107–116. https://doi.org/10.1016/j.asoc.2016.02.029

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

Bi, H., Xiao, H., & Sun, K. (2019). The impact of carbon market and carbon tax on green growth pathway in China: a dynamic cge model approach. *Emerging Markets Finance and Trade*, 55(6), 1312–1325. https://doi.org/10.1080/1540496X.2018.1505609

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901.

Byun, S. J., & Cho, H. (2013). Forecasting carbon futures volatility using GARCH models with energy volatilities. *Energy Economics*, 40, 207–221. https://doi.org/10.1016/j.eneco.2013.06.017

Canakoglu, E., Yahsi, M., & Agrali, S. (2019). Carbon price forecasting models based on big data analytics. *Carbon Management*, 10(2), 175–187.

Carl, J., & Fedor, D. (2016). Tracking global carbon revenues: A survey of carbon taxes versus cap-and-trade in the real world. *Energy Policy*, 96, 50–77. https://doi.org/10.1016/j.enpol.2016.05.023

Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37(1), 28–43. https://doi.org/10.1016/j.ijforecast.2020.02.008

Cui, H-Y., & Dou, X-S. (2018). Carbon price forecasts in Chinese carbon trading market based on emd-ga-bp and emd-pso-lssvm. *Operations Research and Management Science*, 27(7), 133. https://doi.org/10.12005/orms.2018.0166

Fan, X., Li, S., & Tian, L. (2015). Chaotic characteristic identification for carbon price and a multi-layer perceptron network prediction model. *Expert Systems with Applications*, 42(8), 3945–3952. https://doi.org/10.1016/j.eswa.2014.12.047

Fauvel, K., Lin, T., Masson, V., Fromont, E., & Termier, A. (2021). Xcm: An explainable convolutional neural network for multivariate time series classification. *Mathematics*, 9(23), 3137. https://doi.org/10.3390/math9233137

Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., … Petitjean, F. (2020). Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y

Graves, A., Alex. (2012). Long short-term memory. Supervised sequence labelling with recurrent neural networks, pages 37–45.

Hao, Y., Tian, C., & Wu, C. (2020). Modelling of carbon price in two real carbon trading markets. *Journal of Cleaner Production*, 244, 118556. https://doi.org/10.1016/j.jclepro.2019.118556

Ji, L., Zou, Y., He, K., & Zhu, B. (2019). Carbon futures price forecasting based with arima-cnn-lstm model. *Procedia Computer Science*, 162, 33–38. https://doi.org/10.1016/j.procs.2019.11.254

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). Lstm fully convolutional networks for time series classification. *IEEE access*, 6, 1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939

Koop, G., & Tole, L. (2013). Forecasting the European carbon market. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(3), 723–741. https://doi.org/10.1111/j.1467-985X.2012.01060.x

Li, S., Goel, L., & Wang, P. (2016). An ensemble approach for short-term load forecasting by extreme learning machine. *Applied Energy*, 170, 22–29. https://doi.org/10.1016/j.apenergy.2016.02.114

Liu, H., Dai, Z., So, D., & Le, Q. V. (2021). Pay attention to mlps. *Advances in Neural Information Processing Systems*, 34, 9204–9215.

Liu, X., Wang, C., Wu, H., Yang, C., & Albitar, K. (2023). The impact of the new energy demonstration city construction on energy consumption intensity: Exploring the sustainable potential of China's firms. *Energy*, 283(6), 128716. https://doi.org/10.1016/j.energy.2023.128716

Lu, H., Ma, X., Huang, K., & Azimi, M. (2020). Carbon trading volume and price forecasting in China using multiple machine learning models. *Journal of Cleaner Production*, 249, 119386. https://doi.org/10.1016/j.jclepro.2019.119386

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the experimental analysis of behavior*, 76(2), 235–243. https://doi.org/10.1901/jeab.2001.76-235

Sanin, M. E., Violante, F., & Mansanet-Bataller, M. (2015). Understanding volatility dynamics in the eu-ets market. *Energy Policy*, 82, 321–331. https://doi.org/10.1901/jeab.2001.76-235

Sun, W., & Zhang, C. (2018). Analysis and forecasting of the carbon price using multi—resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. *Applied Energy*, 231, 1354–1371. https://doi.org/10.1016/j.apenergy.2018.09.118

Tan, C. W., Dempster, A., Bergmeir, C., & Webb, G. I. (2022). Multirocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, 36(5), 1623–1646. https://doi.org/10.1007/s10618-022-00844-1

Wang, C., Liu, X., Li, H., & Yang, C. (2023). Analyzing the impact of low-carbon city pilot policy on enterprises' labor demand: Evidence from China. *Energy Economics*, 124, 106676. https://doi.org/10.1016/j.eneco.2023.106676

Wang, J., Cui, Q., & Sun, X. (2021). A novel framework for carbon price prediction using comprehensive feature screening, bidirectional gate recurrent unit and Gaussian process regression. *Journal of Cleaner Production*, 314, 128024. https://doi.org/10.1016/j.jclepro.2021.128024

Wang, J., Wang, Z., Li, J., & Wu, J. (2018). Multilevel wavelet decomposition network for interpretable time series analysis. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2437–2446.

Wang, M., Zhu, M., & Tian, L. (2022). A novel framework for carbon price forecasting with uncertainties. *Energy Economics*, 112, 106162. https://doi.org/10.1016/j.eneco.2022.106162

Wang, S., Zhang, N., Wu, L., & Wang, Y. (2016). Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method. *Renewable Energy*, 94, 629–636. https://doi.org/10.1016/j.renene.2016.03.103

Yang, C., Chen, L., & Mo, B. (2023). The spillover effect of international monetary policy on China's financial market. *Quantitative Finance and Economics*, 7(4), 508-537. https://doi.org/10.3934/QFE.2023026

Yang, S., Chen, D., Li, S., & Wang, W. (2020). Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. *Science of the Total Environment*, 716, 137117. https://doi.org/10.1016/j.scitotenv.2020.137117

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A transformer-based framework for multivariate time series representation learning. In Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, 2114–2124.

Zhang, F., & Wen, N. (2022). Carbon price forecasting: a novel deep learning approach. *Environmental Science and Pollution Research*, 29(36), 54782–54795. https://doi.org/10.1007/s11356-022-19713-x

Zhou, F., Huang, Z., & Zhang, C. (2022). Carbon price forecasting based on ceemdan and lstm. *Applied Energy*, 311, 118601. https://doi.org/10.1016/j.apenergy.2022.118601

Zhu, B., Chevallier, J., Zhu, B., & Chevallier, J. (2017a). Carbon price forecasting with a hybrid arima and least squares support vector machines methodology. *Pricing and Forecasting Carbon Markets: Models and Empirical Analyses*, 87–107. https://doi.org/10.1007/978-3-319-57618-3_6

Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., & Wei, Y-M. (2017b). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Applied Energy*, 191, 521–530. https://doi.org/10.1016/j.apenergy.2017.01.076

Zhu, B., Ye, S., Wang, P., He, K., Zhang, T., & Wei, Y-M. (2018). A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. *Energy Economics*, 70, 143–157. https://doi.org/10.1016/j.eneco.2017.12.030