



# Multi-step prediction of carbon emissions based on a secondary decomposition framework coupled with stacking ensemble strategy

Boting Zhang<sup>1</sup> · Liwen Ling<sup>1,2</sup> · Liling Zeng<sup>1</sup> · Huanling Hu<sup>1</sup> · Dabin Zhang<sup>1,2</sup>

Received: 4 November 2022 / Accepted: 15 April 2023 / Published online: 9 May 2023 / Published online: 9 May 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Accurate prediction of carbon emissions is vital to achieving carbon neutrality, which is one of the major goals of the global effort to protect the ecological environment. However, due to the high complexity and volatility of carbon emission time series, it is hard to forecast carbon emissions effectively. This research offers a novel decomposition-ensemble framework for multi-step prediction of short-term carbon emissions. The proposed framework involves three main steps: (i) data decomposition. A secondary decomposition method, which is a combination of empirical wavelet transform (EWT) and variational modal decomposition (VMD), is used to process the original data. (ii) Prediction and selection: ten models are used to forecast the processed data. Then, neighborhood mutual information (NMI) is used to select suitable sub-models from candidate models. (iii) Stacking ensemble: the stacking ensemble learning method is innovatively introduced to integrate the selected sub-models and output the final prediction results. For illustration and verification, the carbon emissions of three representative EU countries are used as our sample data. The empirical results show that the proposed framework is superior to other benchmark models in predictions 1, 15, and 30 steps ahead, with the mean absolute percentage error (MAPE) of the proposed framework being as low as 5.4475% in Italy dataset, 7.3159% in France dataset, and 8.6821% in Germany dataset.

**Keywords** Carbon emission prediction · Multi-step prediction · Stacking ensemble learning · Empirical wavelet transform · Variational mode decomposition · Neighborhood mutual information

## Introduction

### Background and motivation

Global warming is a major concern of the international community. Since the late nineteenth century, the global average surface temperature has increased by 0.4 to 0.8 °C. A warming climate will lead to melting glaciers, rising sea levels, and an increase in the frequency of extreme weather events, which in turn causes serious irreversible damage to

the natural environment (Kong et al. 2022a). Carbon dioxide is a primary greenhouse gas, which is a major contributor to global warming (Qader et al. 2022). Since the 1970s, we have witnessed the astonishing data of average growth rate of carbon emissions per decade as 3%, 1%, 1%, 3%, and 2%, respectively. Global emissions reached 33.3 GtCO<sub>2</sub> in 2020 and 34.9 GtCO<sub>2</sub> in 2021, an increase of 4.8% far exceeding the average growth of the last decade (Liu et al. 2022a).

In order to reduce carbon emissions and address global climate change, the international community has held many talks and signed relevant agreements. Following United Nations Framework Convention on Climate Change and Kyoto Protocol, nearly 200 countries worldwide signed the Paris Agreement in 2016. The long-term goal of the Paris Agreement is to limit global average temperature rise to 2 °C pre-industrial and work towards limiting it to 1.5 °C (Meinshausen et al. 2022). To achieve this goal, a growing number of countries are incorporating carbon neutrality targets into their long-term national strategies, such as China (Shi 2022), the USA (Qin et al. 2022), the UK (Abbasi et al. 2021), and Russia (Safonov et al. 2020).

Responsible Editor: V.V.S.S. Sarma

✉ Dabin Zhang  
zdbff@aliyun.com

<sup>1</sup> College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China

<sup>2</sup> Institute of Rural Revitalization Research, South China Agricultural University, Guangzhou 510642, China

As a global pioneer in the effort against climate change and development of renewable energies, the EU has set ambitious carbon neutral targets for itself. In March 2019, The Resolution on Climate Change of the European Parliament endorsed an overall target of achieving net zero greenhouse gas emissions by 2050. And the EU plans to fully achieve the mid-term climate change target in 2030: a reduction in greenhouse gas emissions of at least 40% from 1990 levels (Salvia et al. 2021). In July 2021, the EU released “Fit for 55 packages” which includes a commitment to reduce greenhouse gas emissions by 55% in 2030 compared to 1990, and 12 aggressive initiatives in energy, industry, transport, and construction.

As shown in Fig. 1, the proportion of EU carbon emissions in the global pool gradually decreased from 2016 to 2019, indicating that EU countries actively strived to reduce carbon emissions effectively after the signing of the Paris Agreement. In 2020, due to the impact of COVID-19, global economic activities and production were severely affected, resulting in a drop in carbon emissions. In 2021, due to further developments occurred among international trade and the global political climate, the proportion of EU carbon emissions rebounded, but was still lower than 8% of global carbon emissions, indicating that the EU’s emission reduction policies have achieved initial success.

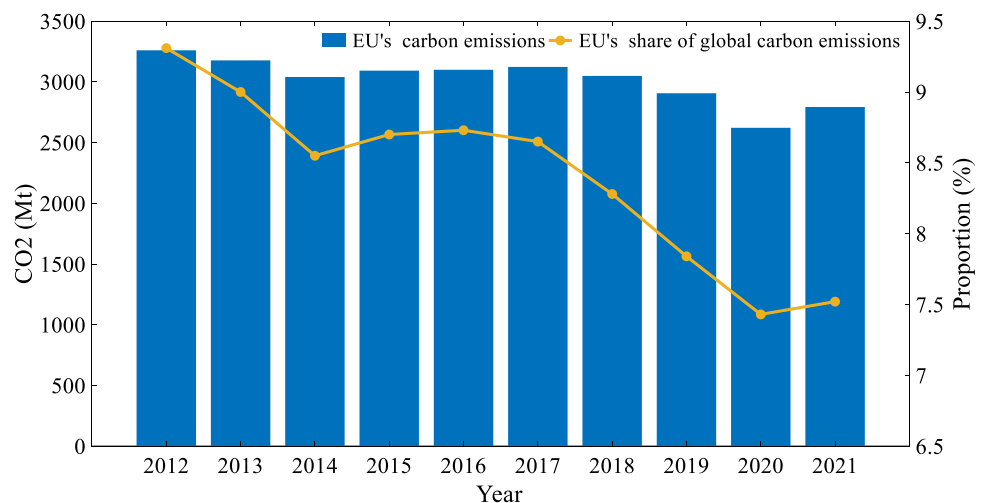
The federal formulation of practical environmental policies plays a key role in achieving carbon neutrality goals, and accurate prediction of carbon emissions can provide a scientific basis for government policy formulation and management decisions (Salvia et al. 2021). Short-term carbon emission prediction can dynamically monitor the progression and trend of carbon emission reduction, which is conducive to flexible adjustment of carbon reduction measures by various functional departments. If the prediction results show high emissions, the government can consider dynamically adjusting its efforts to strengthening carbon taxes

(Cheng et al. 2021) and tightening carbon emission trading scheme to reduce carbon emissions in various sectors of the society. And if the prediction results show low emissions, governments can in turn react appropriately to ease the pressure to reduce emissions, promote economic development, and support the research and development of low-carbon technologies. In addition, short-term carbon emission prediction can provide more meaningful reference drawing data when unexpected events occur. For example, the COVID-19 pandemic has led to an overall reduction in global carbon emissions in 2020. It is difficult for annual prediction to forecast small probability events affecting carbon emission based on historical data, while short-term carbon emission prediction can provide accurate prediction results based on recent data. It can provide scientific data support for policymakers to formulate carbon emission policies according to current socioeconomic conditions, thus better balancing the relationship between economic development and carbon emissions (Zhao et al. 2022).

## Related works

There are currently two main dimensions of carbon emission projection studies: multivariate and univariate prediction methods (Kong et al. 2022b). Multivariate methods forecast carbon emissions by analyzing the influencing factors, such as economic development (Sun and Huang 2022), energy consumption (Nguyen et al. 2021), and population growth (Musah et al. 2021). However, the degree of influence of these factors on carbon emissions has proved hard to determine (Yang and O’Connell 2020); in addition, multivariate methods may lead to generation of cumulative errors and reduction in prediction accuracy. Univariate methods use historical data to make forecast for a certain period of time in the future, relying on only pre-existing solid data (Ziel and Weron 2018). It emphasizes the importance of time factor

**Fig. 1** EU’s carbon emissions and their share of global carbon emissions (<https://ourworldindata.org/>)



in prediction, thus reducing forecast uncertainties caused by the accumulation of multi-factor errors. Therefore, this paper adopts univariate methods for carbon emission prediction.

Table 1 shows the literature review of carbon emission projection. The prediction model, research location, data frequency, and step length are listed. It can be observed that the application scenarios of short-term carbon emissions in the existing literature are mainly focused on China, with fewer studies on other countries and regions. Carbon neutrality is a common goal of the international community, and short-term carbon emission projection can fully portrait the daily dynamics of carbon emissions and provide scientific data support for promoting carbon neutrality. Thus, it is essential to conduct short-term carbon emissions research on other countries and regions. Moreover, most studies focus on only one dataset, their application scopes are too small, and the universality of their models is difficult to verify. Therefore, in this paper, carbon emissions of three representative EU countries are selected as experimental data to verify the validity and universality of the model.

Pre-existing univariate carbon emission prediction can be broadly classified into three main categories: (i) statistical models, (ii) artificial intelligence models, and (iii) hybrid models. Statistical models, such as exponential smoothing models (ETS), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), are more frequently used in carbon emission prediction due to their simple operation and cost-effectiveness. In particular, the ARIMA model, one of the most popular independent statistical methods for predicting greenhouse gas emissions, focuses more on data points with significant autocorrelation without putting more weight on the last observation data point (Yang and O'Connell 2020). However, the statistical model is based on the assumption of approximate linearity (Zeng et al. 2023), which does not capture well the nonlinear and frequent fluctuations of carbon emission time series (Sohail et al. 2022). Artificial intelligence models with powerful nonlinear mapping and adaptive learning capabilities can effectively capture the characteristics of nonlinear series (Lu et al. 2020). For example, back propagation (BP) (Zhao et al. 2018), support vector machine (SVM) (Zhao et al. 2018), extreme learning machine (ELM) (Wang and Wang 2021), long short-term memory (LSTM) (Liu et al. 2022b), *K*-nearest neighbor (KNN) (Lee et al. 2019), random forest (RF) (Smith et al. 2022), and XGB (Sun et al. 2022) have been applied to carbon research forecasting with great results. However, artificial intelligence models suffer from parameter sensitivity and easily fall into overfitting or local optimum (Yu et al. 2022), and their prediction performance require further improvement. Hybrid models use methods such as data processing and optimization algorithms to capture the complex features of nonlinear time series. In short-term carbon emission forecasting, data processing

methods include series decomposition (Sun and Ren 2021) and feature selection (Kong et al. 2022a), and model optimization algorithms include the Whale Optimization Algorithm (WOA) (Sun and Huang 2022) and Genetic Algorithm (GA) (Zhang et al. 2022). The hybrid model can overcome the limitations inherited to single-model solutions, better extract the hidden factors that cannot be captured by traditional methods, and effectively improve the accuracy and stability of prediction. Table 1 shows the literature review of carbon emission projection. The prediction model, research location, data frequency, and research results are listed.

Analyzing the related work, it can be concluded that the decomposition method is applied more in the hybrid models. Yu et al. (2008) proposed the TEI@I complex system research methodology, emphasizing the idea as to decompose before integration. Decompose the complex raw data into relatively stable and regular subsequences, which improves model prediction accuracy (Yue et al. 2022). The decomposition presents promising results in diverse nonlinear time series applications, such as financial (Lv et al. 2022), wind energy (Liu et al. 2021), traffic flow (Tian 2021), and air pollution (Liu et al. 2020). Despite the aforementioned effect, Liu et al. (2014) noted that the first intrinsic mode function (IMF1) obtained using decomposition is highly volatile and irregular, which could affect the overall prediction accuracy. To solve this problem, Yin et al. (2017) used wavelet packet decomposition (WPD) to decompose IMF1 obtained from empirical mode decomposition (EMD) and forecasting success in wind power prediction. Sun et al. (2020) used variational mode decomposition (VMD) to further decompose IMF1 with high complexity obtained by the first decomposition, proving that the secondary decomposition strategy can significantly improve the prediction performance. The signal decomposition methods used in the existing carbon emission literature are mainly EMD, EEMD, and ICEEMDAN, and these methods still possess many problems. For example, EMD has the problem of modal aliasing. Although EEMD overcomes modal aliasing, it still has the problem of residual noise. However, the empirical wavelet transform (EWT) eliminates the modal aliasing by reasonable segmentation and boundary setting (Wang and Wang 2020) and filters out the noise residual by automatically generating an adaptive wavelet (Liu et al. 2018). To some extent, EWT makes up for the shortcomings of the above methods and has a solid theoretical foundation. Therefore, this paper introduces the EWT method, a commonly used decomposition method in engineering (Shi et al. 2021), into the field of carbon emission prediction, taking advantage of its small computational effort and high robustness as the first decomposition method.

Although hybrid models perform effectively in the field of carbon emission prediction, the characteristics of carbon emission data vary widely in different countries and regions,

**Table 1** Summary of various recent carbon emission prediction models

Model	Reference	Location	Frequency	Volume	Results
<b>Statistical models</b>					
Autoregressive fractionally integrated moving average (ARFIMA)	(Belbute and Pereira 2020)	Portugal	Yearly	68	To achieve carbon neutrality by 2050, additional policy efforts would be needed to achieve a reduction of 315 million tonnes of carbon emissions.
GA, fractional GM (1,1) model (GA-FAGM (1,1))	(Hu et al. 2021)	20 countries	Yearly	15	The proposed model is superior to other grey prediction models.
<b>Artificial intelligence models</b>					
Covariance matrix adaptation, evolutionary strategy, neural networks (CMA-ES-NNs)	(Mason et al. 2018)	Ireland	Minutes	2878	Compared with other evolutionary optimization algorithms, the CMA-ES algorithm has better performance.
Generalized regression neural network, grey wolf optimization (GRNN-GWO)	(Heydari et al. 2019)	Iran, Italy, Canada	Yearly	36	This method can accurately predict the multiple time steps of carbon emissions caused by electricity power generation.
Lion swarm optimizer, GA, least squares support vector machine model (LSO-GA-LSSVM)	(Qiao et al. 2020)	12 countries	Yearly	53	The proposed model is superior to single and combination methods.
<b>Hybrid models</b>					
Ensemble empirical mode decomposition, backpropagation neural network, particle swarm optimization (EEMD-PSOBP)	(Sun and Ren 2021)	China	Daily	639	This model has strong global optimization ability, fast convergence speed, and high precision.
EEMD, variational mode decomposition, LSTM (EEMD-VM-D-LSTM)	(Kong et al. 2022b)	China	Daily	912	Compared with 14 control models, this model has the best performance.
Improved complete EEMD with adaptive noise, partial autocorrelation function and ReliefF, improved sparrow search algorithm, ELM (ICCEMDAN-PACF, ReliefF-ISSA-ELM)	(Kong et al. 2022a)	China	Daily	900	The secondary decomposition method can further improve the accuracy of prediction.
Time-varying filter, empirical mode decomposition, objective tangent search algorithm (TVF-EEMD-PSR-MOTSA)	(Liu et al. 2022c)	China, the United States	Daily	639	The proposed CEEFS system can effectively improve the accuracy and stability of prediction.

and no single hybrid model can be considered suitable for all prediction scenarios (Jiang 2021). Combined prediction utilizes several individual models of varying ability to capture nonlinearities and perform differently on various datasets, to obtain better generalization performance (Ling et al. 2019). Allende and Valle (2016) considers that the use of suitable ensemble methods contributes to improve the overall prediction accuracy, and the commonly used integration methods are simple averaging (Graefe et al. 2014) and weighted averaging (Kourentzes et al. 2019). In recent years, the stacking ensemble learning method (Wolpert 1992) has received much attention in the field of machine learning, and introduced the idea of a higher-level meta-learner into combined prediction. Zhao and Cheng (2022) employed the stacking method to ensemble different linear and nonlinear individual stock return prediction models. The experimental results show that the stacking method outperforms Mallows model averaging, simple combination prediction, and others. Cui et al. (2021) used heterogeneous integration of different types of base learners to enhance the generalization ability of the model, and the stacking ensemble learning method can effectively fuse the prediction results of base learners to improve the model performance for earthquake casualty studies. Most of the existing carbon emission literature uses a single model for prediction and seldomly use integrated learning method to improve model performance. Each model has advantages and disadvantages, and using just one model for prediction leads to limitations in the extent of possible improvements to the model's accuracy when processing nonlinear data. The integrated learning approach combines multiple models and utilizes the advantages of each of them, providing better prediction ability and stability than a single model. This paper uses the stacking method to ensemble the results of different prediction models, to enrich the application of the stacking method in the field of carbon emissions.

In general, combined prediction achieves better results than single prediction models (Atiya 2020). The choice of which single models to combine is one of the challenges in the field of prediction. Many studies incorporated single models with more mature applications in this field into the combined model, resulting in a massive computational effort and little effect on prediction performance improvement (Che 2015). Kışınbay (2010) used encompassing test to select sub-models, so as to reduce the correlation of sub-models. It is superior to the benchmark model in the empirical analysis of American macroeconomic datasets. Model trimming is essentially a shrinkage strategy, which generally cuts down candidate single models according to their prediction accuracy. Samuels and Sekkel (2017) proposed a new approach based on model confidence sets, which has better robustness and more space for performance improvement. Based on the perspective of information theory, Cang and Yu (2014) proposed an algorithm for selecting optimal subsets of combined

prediction based on mutual information (MI). When the probability distribution of variables and their joint distribution are known, mutual information can measure the correlation between different random variables (Zhang et al. 2020a). However, probability distributions are difficult to measure in practice (Xiao et al. 2019). Therefore, this paper introduces the neighborhood mutual information (NMI) (Hu et al. 2011) measure to the correlation between two variables, without calculating the probability distribution of dataset.

## Contributions and article organization

Considering the inadequacy of existing studies, this study is aimed at proposing a hybrid prediction model that combines EWT and VMD with prediction models and STACK. The model will preprocess the dataset by secondary decomposition and use NMI to select sub-models for the ensemble, striving to forecast the carbon emissions in a multi-step ahead manner (1, 15, 30 days ahead). The main contributions of this paper can be summarized as follows:

- (1) The stacking ensemble learning method is innovatively introduced into the field of carbon emission prediction. There are few studies on carbon emissions using combined forecasting in existing literature. In this paper, the stacking ensemble learning method is used for combination prediction, and the results show that the stacking ensemble learning method is better than traditional linear method.
- (2) A new hybrid model, which involves secondary decomposition, sub-model selection, and stacking ensemble learning, is proposed to conduct carbon emission multi-step prediction. The ablation experiments show that all three components of this hybrid model are effective and can improve the overall prediction accuracy.
- (3) We apply the proposed decomposition-ensemble framework to carbon emission datasets of three countries: Italy, France, and Germany. Numerical experiment indicates that the proposed model has higher level prediction accuracy compared with other models in multi-step prediction. And the experimental results of multiple datasets also confirm the effectiveness and generality of the framework.

The rest of this paper is structured as follows. The “**Methodologies**” section introduces the basic methods used in this paper and describes the modeling process of the proposed model in detail. The “**Empirical analysis**” section introduces the experimental data sources and details and conducts comparative experiments from three perspectives. The “**Conclusions**” section summarizes this study and provides an outlook for future research.

## Methodologies

In this section, the basic methods taken in the proposed model are described. In the “[Data preprocessing methods](#)” section, we briefly introduce the two decomposition methods and the method for quantifying sequence complexity. Then, the “[Neighborhood mutual information \(NMI\)](#)” section describes the stacking ensemble learning method. Next, the “[Stacking ensemble learning](#)” section introduces the method NMI to determine sequence similarity. Finally, the “[Construction of the proposed model](#)” section describes the proposed model framework in detail.

### Data preprocessing methods

#### Empirical wavelet transform (EWT)

The EWT decomposition method is a novel adaptive signal processing method proposed by Gilles (2013). The amplitude modulation-frequency modulation (AM-FM) components of the Fourier spectrum are extracted by adaptive segmentation of the signal spectrum, construction of wavelet functions with compact support characteristics in the segmentation interval, and construction of appropriate orthogonal wavelet filters. It can not only solve the problem of mode aliasing existing in empirical mode decomposition (EMD) but also has good noise robustness.

Suppose that the Fourier support interval  $[0, \pi]$  is partitioned into  $N$  consecutive components, and the boundaries of its partitioned segments are denoted as  $\Lambda_n = [\omega_{n-1}, \omega_n]$ , and then,  $\cup_{n=1}^N \Lambda_n = [0, \pi]$ .  $\omega_n$  is the boundary of each segment and defines a transition region  $T_n$  with  $\omega_n$  as the center point and  $2\lambda_n$  as the width. The empirical wavelet is a bandpass filter on interval  $\Lambda_n$ . Drawing on the construction method of Meyer wavelet, for any  $n > 0$ , the empirical scaling function  $\hat{\varphi}_n(\omega)$  and the empirical wavelet function  $\hat{\psi}_n(\omega)$  are obtained.

$$\hat{\varphi}_n(\omega) = \begin{cases} 1 & |\omega| \leq (1 - \gamma)\omega_n \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\gamma\omega_n} (|\omega| - (1 - \gamma)\omega_n) \right) \right] & (1 - \gamma)\omega_n < |\omega| < (1 + \gamma)\omega_n \\ 0 & \text{else} \end{cases} \quad (1)$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & (1 + \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_{n+1} \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\gamma\omega_{n+1}} (|\omega| - ((1 - \gamma)\omega_{n+1})) \right) \right] & (1 - \gamma)\omega_{n+1} < |\omega| \leq (1 + \gamma)\omega_{n+1} \\ \sin \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\gamma\omega_n} (|\omega| - (1 - \gamma)\omega_n) \right) \right] & (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0 & \text{else} \end{cases} \quad (2)$$

where  $\forall x \in [0, 1], 0 < \gamma < \min_n \left( \frac{\omega_{n-1} - \omega_n}{\omega_{n-1} + \omega_n} \right), \lambda_n = \gamma\omega_n$ , and  $\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3)$ .

According to the traditional wavelet transform method, the empirical wavelet transform is reconstructed, and the detail function is obtained from the inner product of the empirical wavelet function and the signal, and the approximate coefficient is obtained from the inner product of the empirical scale function and the signal. The original time series can be reconstructed as

$$f(t) = W_f^\varepsilon(0, t) * \varphi_1(t) + \sum_{n=1}^N W_f^\varepsilon(n, t) * \psi_n(t) \\ = F^{-1} \left[ \hat{W}_f^\varepsilon(0, w) * \hat{\varphi}_1(w) + \sum_{n=1}^N \hat{W}_f^\varepsilon(n, w) * \hat{\psi}_n(w) \right] \quad (3)$$

where  $*$  denotes the convolution symbol;  $\hat{W}_f^\varepsilon(0, w), \hat{W}_f^\varepsilon(n, w)$  denote Fourier transform of  $W_f^\varepsilon(0, t), W_f^\varepsilon(n, t)$  respectively. The empirical mode function  $f_k(t)$  is

$$f_0(t) = w_f^\varepsilon(0, t) * \varphi_1(t) \quad (4)$$

$$f_k(t) = w_f^\varepsilon(k, t) * \psi_k(t) \quad (5)$$

#### Variational mode decomposition (VMD)

VMD decomposition is a fully intrinsic, adaptive non-recursive decomposition technique proposed by Dragomiretskiy et al. (2014). It uses an alternating direction multiplier algorithm to continuously iteratively calculate each modal function and its central frequency to solve the problem of signal noise and avoid modal confusion to some extent. The specific decomposition steps are as follows:

- (1) For each mode  $u_k$ , the one-sided spectrum is obtained by calculating the corresponding resolved signal through the Hilbert transform. Then, an exponential term is added to adjust the respective center frequency, and the spectrum of each mode function is modulated to the baseband. Applying Gaussian smoothing to the demodulated signal to estimate the corresponding bandwidth can be regarded as a constrained variational problem.

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \quad (6)$$

where  $t$  is the time,  $\delta(t)$  is the unit impact function,  $u_k$  is the decomposition mode,  $w_k$  is the center frequency corresponding to the mode, and the constraint condition is that the sum of all modes of  $\sum_k u_k = f(t)$  is equal to the original signal  $f(t)$ .

(2) Introduction of quadratic penalty factors  $\alpha$  and Lagrange multipliers  $\lambda$  to transform the variational problem into an unconstrained optimization problem.

$$L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda, f(t) - \sum_{k=1}^K u_k(t) \right\rangle \tag{7}$$

(3) The alternating direction multiplier method (ADMM) is used to solve the above equation, and the original signal is decomposed into  $K$  narrowband modal components to obtain the optimal solution of the constrained variational model. The solution expressions of modal components  $u_k$  and  $w_k$  are as follows:

$$u_k^{n+1}(w) = \frac{f(w) - \sum_{i \neq k} u_i(w) + \frac{\lambda(w)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{8}$$

$$w_k^{n+1} = \frac{\int_0^\infty w |u_k^{n+1}(w)|^2 dw}{\int_0^\infty |u_k^{n+1}(w)|^2 dw} \tag{9}$$

where  $f(w)$ ,  $\lambda(w)$ ,  $u_i(w)$ , and  $u_k^{n+1}(w)$  are the Fourier transforms of  $f(t)$ ,  $\lambda(t)$ ,  $u_i(t)$ , and  $u_k^{n+1}(t)$ , respectively.

**Sample entropy (SE)**

Sample entropy is a method proposed by Richman and Moorman to measure the complexity of time series. This method does not depend on the length of data and has better robustness and consistency than approximate entropy. The higher the complexity of time series, the higher the entropy value, and vice versa, the smaller the entropy value. The algorithm is as follows:

(1) Calculate the absolute value distance of corresponding elements between time series vectors of length  $N$ , and define its maximum value as the distance between vectors.

$$d[X_m(i), X_m(j)] = \max |x(i+k) - x(j+k)| \tag{10}$$

where  $k$  is an integer and  $k \in [1, m-1]$ ,  $1 \leq i, j \leq N-m+1$ ,  $i \neq j$ .

For given threshold  $r$ , calculate its ratio  $B_i^m(r)$  and the ratio average  $B^m(r)$ .

$$B_i^m(r) = \frac{1}{N-m-1} \text{num}\{d[X_m(i), X_m(j)] < r\} \tag{11}$$

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{12}$$

(2) Sequence sample entropy can be expressed as:

$$SE(m, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \left( \frac{B^{m+1}(r)}{B^m(r)} \right) \right\} \tag{13}$$

**Neighborhood mutual information (NMI)**

Due to the difficulty of probability density calculation, the commonly used Shannon information entropy is hard to quantify the correlation between features. Hu et al. extended the concept of neighborhood mutual information on this basis. The greater the neighborhood mutual information value, the greater the correlation between the two variables.

Let a nonempty finite set  $X = \{x_1, x_2, \dots, x_n\}$ ,  $f \subseteq F$  on the space of real numbers inscribed by the characteristic set  $F$  be any characteristic subset. For any object  $x_i$ ,  $\delta \geq 0$  on  $x$  defines its  $\delta$  neighborhood on the characteristic subset  $f$  as

$$\delta_f(x_i) = \{x | x \in X, \Delta^f(x, x_i) \leq \delta\} \tag{14}$$

where  $\Delta(x, x_i)$  represents the distance between sample  $x$  and sample  $x_i$ , and the distance function usually uses the Euclidean distance.  $\delta$  is the neighborhood parameter, which determines the size of the neighborhood.

Given  $S \subseteq F$  is a subset of the feature set  $F$ , define the average neighborhood entropy of the sample set  $x$  on the feature subset  $S$  as

$$NH_\delta(S) = -\frac{1}{n} \sum_{i=1}^n \log \frac{\|\delta_S(x_i)\|}{n} \tag{15}$$

in which  $\|\delta_S(x_i)\|$  is the base of the set  $\delta_S(x_i)$ . For any  $x_i$ , there are  $\delta_S(x_i) \subseteq X$ ,  $\frac{\|\delta_S(x_i)\|}{n} \leq 1$ . Therefore,  $0 \leq NH_\delta(S) \leq \log n$ . This shows that the size of  $\delta$  determines the neighborhood entropy of the sample set.

Given  $S$  and  $R$  are two subsets of the feature set  $F$ , and define the joint neighborhood entropy of the sample set  $x$  on the feature subset  $S \cup R$  as

$$NH_\delta(S, R) = -\frac{1}{n} \sum_{i=1}^n \log \frac{\|\delta_{S \cup R}(x_i)\|}{n} \tag{16}$$

where  $\delta_{S \cup R}(x_i)$  is the neighborhood of sample  $x_i$  on  $S \cup R$ .

Define the neighborhood mutual information of the sample set  $x$  on the feature subset  $S \cup R$  as

$$\text{NMI}_{\delta}(R, S) = -\frac{1}{n} \sum_{i=1}^n \log \frac{\|\delta_S(x_i)\| \cdot \|\delta_R(x_i)\|}{n \|\delta_{S \cup R}(x_i)\|} \quad (17)$$

## Stacking ensemble learning

The stacking ensemble learning method introduces the idea of an advanced meta-learner. By combining different prediction models in the same model, an asymptotically optimal learning system is constructed and obtains higher accuracy by reducing the bias of the generalization. The stacking ensemble learning method is generally divided into two levels. Level 0 is the base learners for training different models, and level 1 is the meta-learner for learning the prediction results of level 0 models. The flow diagram of stacking ensemble learning method is shown in Fig. 2.

The base learners, also known as weak models, train different weak models on the dataset, and the prediction results are used as the input of the meta-learner. The base learners generated by different induction algorithms will contain various objective functions, hyperparameters, and covariates, which can increase the diversity of the base learners. The more multiformity the base learners are, the more generalization and flexibility the model has, and it can adapt to data with different characteristics (Mendes-Moreira et al. 2012).

In this paper, BP, SVM, ELM, random vector functional-link (RVFL) (Zhang et al. 2019), LSTM, RF, KNN, XGB, ARIMA, and THETA (Spiliotis et al. 2020) are selected as base learners. Among them, BP, SVM, ELM, RVFL, LSTM, RF, KNN, and XGB are machine learning models, which have achieved good application results in the field of carbon emissions. ARIMA and THETA are linear models, which are better at capturing linear features of sequences. The characteristics of the above models are completely different, which ensures the diversity of base learners for better prediction results.

Instead of using simple averaging for combination prediction, the meta-learner trains a generalized model for aggregation. This allows every single model to be weighted differently, and some single models with better prediction performance receive greater weight, which improves the

prediction quality. In addition, the complex aggregation helps the model maintain overall stability (Ribeiro and Dos Santos Coelho 2020). Some single models cannot identify sharp changes in the series, and some single models forecast well when the series fluctuates drastically. The aggregation of the two models can get a better prediction result even when the series shows non-stationary fluctuations.

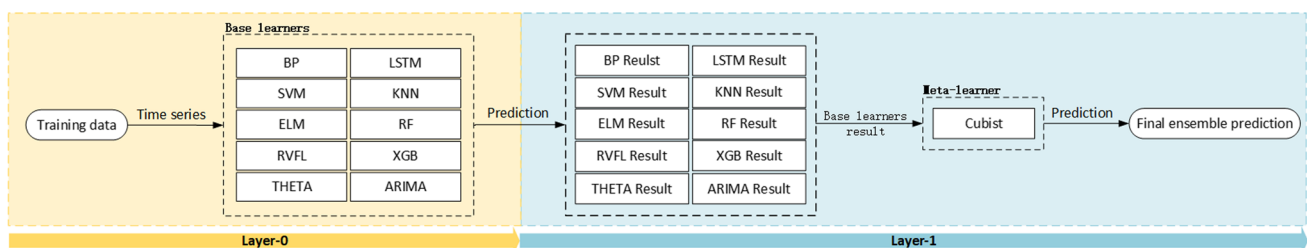
In this paper, the Cubist model is used as a meta-learner. The Cubist model is a regression model based on the proposed M5 model tree. The Cubist model creates a tree structure where each path through the tree is collapsed into a rule. Unlike the M5 model tree, the Cubist model creates a series of “if after after” rules that can overlap. That is, a sample can be assigned to multiple rules, and all predictions are averaged to produce the final value (Yang et al. 2017). The Cubist model adds boosting with training committees, which is similar to developing a series of trees with adjusted weights to make the weights more balanced (Zhou et al. 2019). The Cubist model has been applied to wind speed (Da Silva et al. 2021), electric power (Moon et al. 2022), and engineering (Zhou et al. 2019), which has achieved good prediction results.

## Construction of the proposed model

As shown in Fig. 3, the proposed model in this study includes three modules: data decomposition, prediction and selection, and stacking ensemble. The specific modeling steps are as follows:

**Step 1: Data decomposition.** Firstly, the original carbon emission time sequence is decomposed by the EWT decomposition method, autonomously resulting in a certain amount of IMF components. Next, the complexity of IMF is calculated using SE, to find the IMF with higher complexity than the original time series. Lastly, the selected IMF is decomposed for a second time by VMD.

**Step 2: Prediction and selection.** Firstly, ten models are used to multi-step forecast all IMF separately, and the results of the same model for each IMF are summed to obtain the prediction results. Secondly, sub-model selection is performed in the validation set section. The model



**Fig. 2** Stacking ensemble learning method framework



with the highest accuracy is selected as the preferred model and included in the set of selected sub-models. Then, calculate the NMI for the remaining 9 models and the preferred model, and find the average of all NMIs as reference. At last, models below the average NMI are also added to the set of selected sub-models.

Step 3: Stacking ensemble. The selected sub-models are used as base learners, and their prediction results are input to the meta-learner built by Cubist. The final carbon emission prediction result is obtained by the stacking ensemble learning method.

## Empirical analysis

### Data description

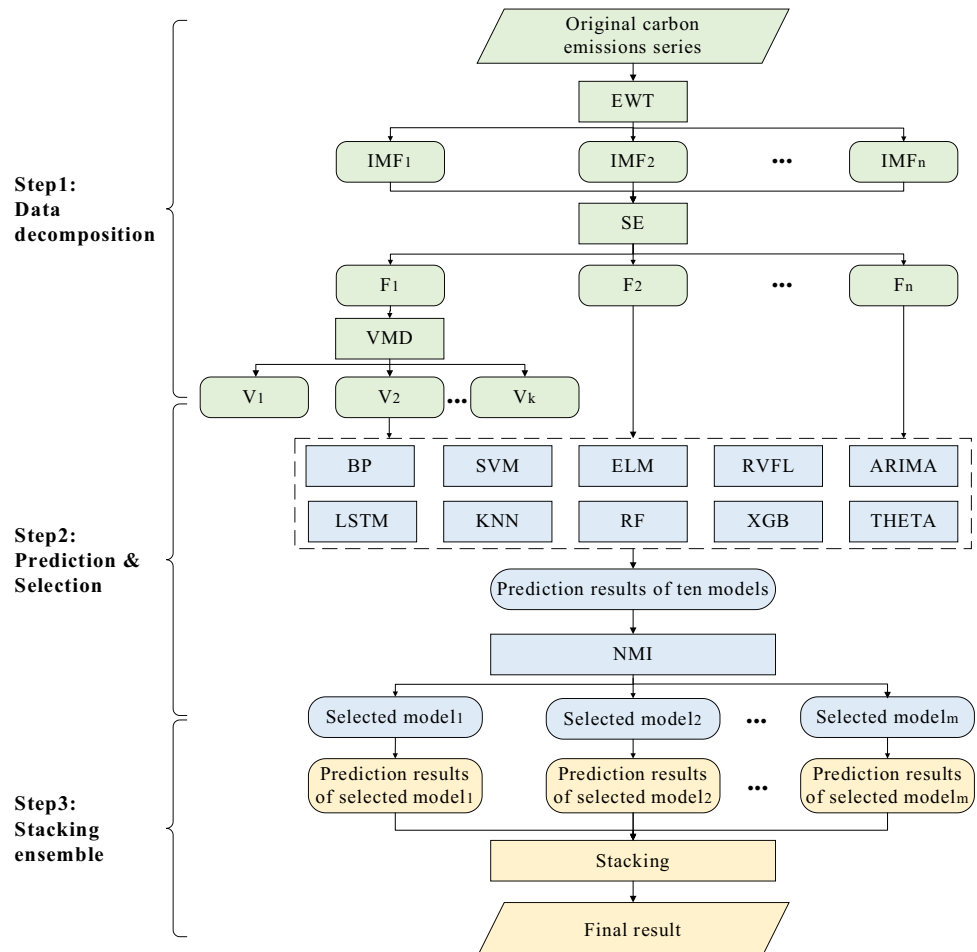
The EU has 27 member states, and the five countries with the highest carbon emissions in the last decade are Germany, Italy, France, Poland, and Spain. Figure 4 shows the carbon emission trends of these five countries from 2012 to 2021, which clearly shows that Germany is the highest carbon

emitter in the EU. While Germany’s carbon emissions are trending downward, they are still twice as high as those of second-placed Italy. Therefore, it is necessary to research Germany’s carbon emissions.

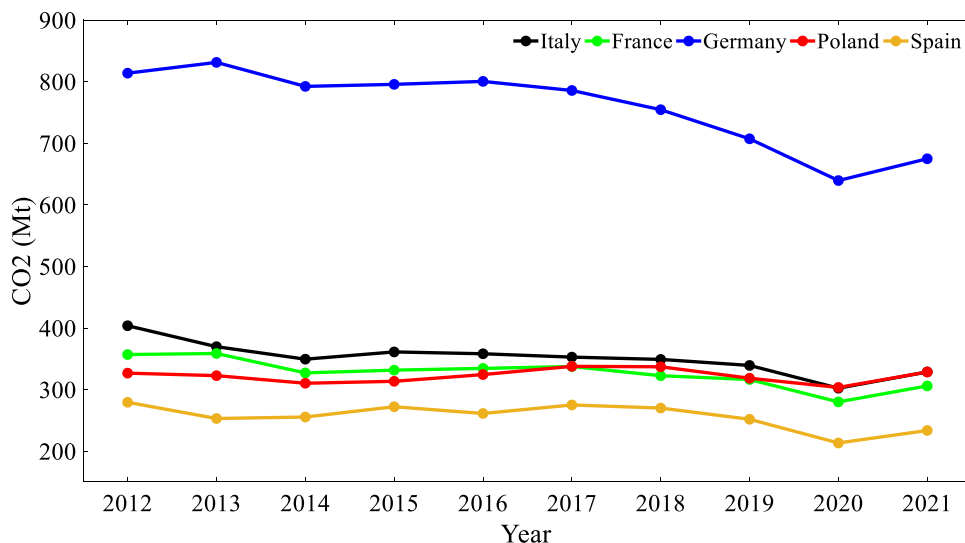
Figure 5 shows the per capita carbon emissions of the five countries in the last decade. It can be seen that the per capita carbon emissions of Germany and Poland are higher than the EU average, and the per capita carbon emissions of Italy and Spain have similar trends. Significantly, France’s per capita carbon emissions over the past 2 years are lower than the world’s per capita carbon emissions, which indicates that France has made faster progress in carbon neutrality and is worth learning from by other countries. Thus, we conclude that the data for Italy and France are more representative.

In addition, these three countries have different progress in policy-making on carbon neutrality, which we consider are sufficient to represent the EU governments’ focus on carbon neutrality. Germany adopted the Climate Protection Act in 2019, which sets the goal of achieving a carbon neutrality target by 2050 in legal form. Furthermore, a new Climate Protection Act was passed in Germany in 2021, aiming to advance the date of achieving carbon neutrality to 2045. France developed the French

Fig. 3 The process of proposed model



**Fig. 4** Carbon emissions of the top five countries in the EU over the past decade (<https://ourworldindata.org/>)



National Low-carbon Strategy in 2015, intending to become carbon neutral by 2050 and reduce the carbon footprint of national consumption. And Italy has not yet issued a relevant policy document (Zhao et al. 2022). Therefore, in this paper, Germany, Italy, and France are chosen as the subjects of the study.

This article selects January 1, 2019, to May 31, 2022, carbon emission daily detection data to carry on the empirical analysis (data from <https://carbonmonitor.org.cn/downloads/>), with 1247 data, as shown in Fig. 6. The data from January 1, 2019, to September 22, 2021, are selected as the training set to train the model. The data from September 23, 2021, to January 26, 2022, with 250 data, are selected as the validation set to perform a preliminary evaluation of the model. The data from January 27, 2022, to May 31, 2022, with 249 data, are selected as the test set to obtain the final prediction results. The ratio of the training set, validation set, and test set is about 60%, 20%, and 20%. According to the central limit theorem (Bianucci

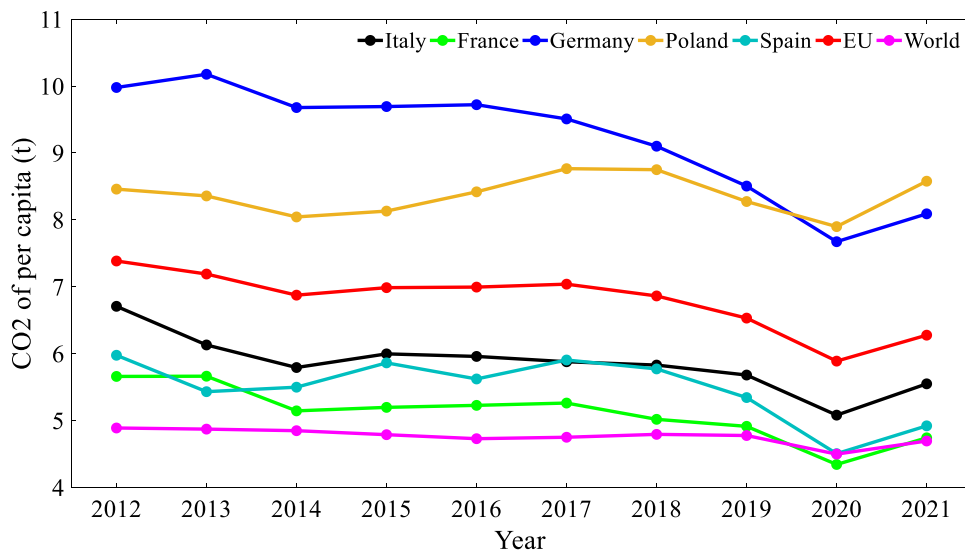
2021), this experiment assumes that the carbon emission time series can be reasonably extended, and the overall distribution follows normally distribution when the sample size is large enough. The descriptive statistics of carbon emissions of Italy, France, and Germany are shown in Table 2.

**Experimental settings**

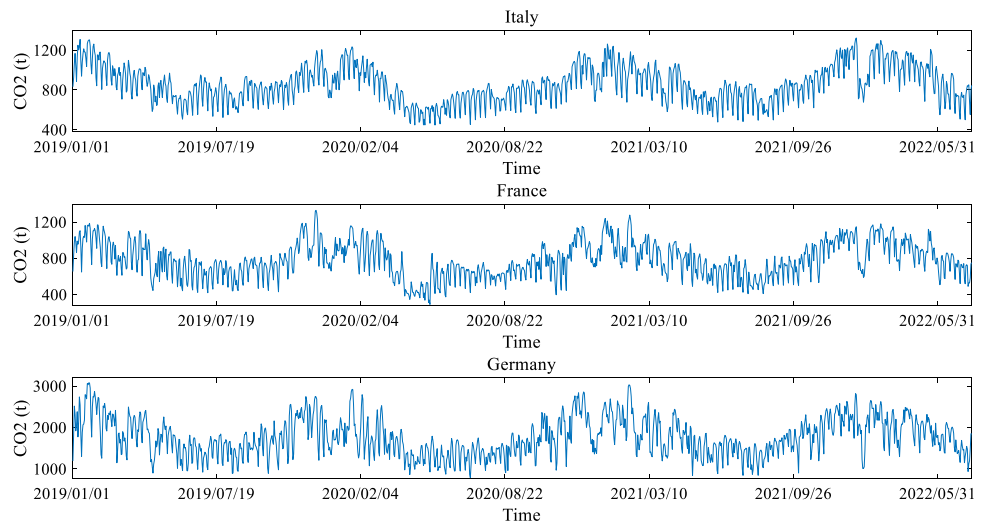
**Evaluation indicators**

To better evaluate the prediction results of the model, root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as evaluation criteria. RMSE reflects the statistical characteristics of the error between the predicted value and the observed value, that is, the dispersion of the sample. MAPE does not deal with the absolute value of deviation to avoid the situation of positive and negative

**Fig. 5** Per capita carbon emissions of the EU’s top five countries over the past decade (<https://ourworldindata.org/>)



**Fig. 6** Time series data of carbon daily emissions in three countries



value cancellation, to reflect the real prediction effect. The smaller RMSE and MAPE are, the closer the prediction result is to the actual value, and the higher the prediction accuracy of the model is. The math is as follow:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \tag{18}$$

$$MAPE = 100 \times \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{19}$$

where  $N$  is the number of observations,  $y_t$  is the actual value at time  $t$ , and  $\hat{y}_t$  is the predicted value at time  $t$ .

**Parameter setting**

The setting of model parameters affects the prediction performance of the model, and appropriate parameter setting can improve the prediction accuracy. In this paper, the core parameters are determined by referring to previous studies

**Table 2** Basic descriptive statistics of three carbon emission time series

	Italy	France	Germany
Maximum	1323.1640	1330.2918	3077.3849
Minimum	447.2859	291.3712	789.6739
Median	840.7706	765.1621	1707.6856
Mean	855.0156	781.1145	1767.0738
Standard deviation	183.9744	194.9637	459.8242
Skewness	0.2214	0.1465	0.3353
Kurtosis	2.4468	2.4339	2.5466
Statistic for Jarque-Bera	26.2148	21.2524	34.1279
$P$ value for Jarque-Bera	0.0010	0.0010	0.0010

and trial-and-error tests, as shown in Tables 3 and 4, in particular, ARIMA and THETA forecast by the R language package “predict” and CUBIST forecast by the R language package “Cubist,” using the default parameters.

**Experimental particulars**

**Data decomposition**

The original carbon emission time series is highly volatile and irregular, making it hard to achieve great results in direct prediction. Therefore, in this paper, the original series is primary decomposed by EWT and divided into a series of IMFs with more regularity and stability, which can facilitate the analysis of the fluctuation characteristics of the original series.

Among them, IMF1 is a noise series, which is greatly affected by various uncertain factors and has complex fluctuations without obvious regularity. The fluctuation trend of IMF2 is consistent with the original time series, effectively capturing the inflection point of rapid change of carbon emissions in the short term. For example, Germany’s carbon emissions decreased by 296 ton from December 11, 2020, to December 12, 2020, and also showed a sharp downward trend from point 711 to point 712 in Fig. 9b. France’s carbon emissions increased by 139 ton from December 2, 2019, to December 3, 2019, and also showed a sharp upward trend from point 336 to point 337 in Fig. 8b. Italy’s carbon emissions decreased by 53 ton from December 24, 2019, to December 25, 2019, also falling sharply from points 358 to 359 in Fig. 7b. It can be seen that the carbon emissions in winter vary considerably in each country. The energy consumption of these three countries is dominated by fossil fuels, the burning of which produces large amounts of carbon dioxide. Winter temperature changes; energy

**Table 3** Parameter values of decomposition methods

Method	Parameters	Value
EWT	Global trend	Poly
	Degree	10
	Reg	None
	Type detect	Empirical law
VMD	Intrinsic mode function	Determined by SE
	Tolerance of convergence criterion	1.00E-07

consumption changes (Shirizadeh and Quirion 2022), resulting in a large change in the overall carbon emissions.

Compared with IMF3, the vibration amplitude of IMF4 increased, and the regularity of fluctuations was enhanced. Different time-scale features are extracted as different frequency component series, which effectively reduces the difficulty of modeling work. IMF5 fluctuates gently with an overall U-shaped trend, reflecting the long-term changes in carbon emissions under the influence of temperature. The value decreases in cold weather and increases in hot weather. Germany's IMF5 ranges from 1249 to 2297, about twice as high as France and Italy. Germany is the biggest emitter of carbon dioxide in the EU, with overall carbon emissions roughly twice as high as the other two countries. This is due to the fact that Germany, as a large industrial country, requires higher energy consumption to fuel its economic growth (Jia et al. 2022), which leads to higher carbon emissions. And Germany is strongly opposed to the development of nuclear power and is gradually shutting down nuclear power plants. However, no clean energy alternative to nuclear power has yet been found, leading to increased generation from coal-fired power plants and increased carbon emissions.

Although the IMF trend after EWT decomposition is enhanced, some sub-series are still highly unstable, which will affect the overall prediction accuracy. For this reason, this paper uses SE to quantify the complexity of IMF and selects the selected IMF with secondary decomposition. As shown in Table 5, after the primary decomposition of each dataset, there is an IMF with SE higher than the original sequence, which means the complexity is higher than the original sequence. If this IMF is forecasted directly, the decomposition effect will be weakened, and the prediction accuracy will be lowered. To solve this problem, this IMF is decomposed twice to reduce the sequence complexity and improve the model prediction accuracy.

The VMD is used to secondarily decompose the IMF which is selected. Choosing a reasonable decomposition modulus number  $K$  is the key to effective decomposition of VMD. In this paper, SE is used to optimize the value of decomposition  $K$  (Zhang et al. 2020b). Firstly, decomposition is attempted using VMD with different  $K$  values.

And then, calculate the decomposition to get the SE value of each IMF. In the next step, the SE average value of IMF is obtained when comparing different  $K$  values. When the value of  $K$  is small, the SE value is large, indicating that the IMF complexity after decomposition is high, and the time series is not fully decomposed. With the increase of  $K$  value, the SE value gradually decreases, and the decomposition effect is significantly improved. Meanwhile, in order to avoid excessive decomposition, this paper chooses the inflection point where SE tends to be stable as the  $K$  value of VMD, the number of decomposition. In Germany dataset, when the value of  $K$  is higher than 8, the value of SE tends to be stable, so the value of  $K$  is chosen as 8. Analogously, Italy and France both chose  $K$  values of 7. The result of the secondary decomposition of three dataset are shown in Figs. 7c, 8c, and 9c.

### Sub-model selection based on NMI

Using all models as base learners does not necessarily lead to better prediction results. There may be redundancy among models, which increases unnecessary computation. Furthermore, different data characteristics are different, and one model cannot be suitable for any dataset. Combining all models which includes an underperforming single model will lead to a poor final prediction effect. Since putting diverse and distinct single models in the base learner can improve the effectiveness of stacking, this paper adopts NMI

**Table 4** Parameter values of prediction models

Model	Parameters	Value
BP	Iterations number	1000
	Learning rate	0.1
	Training requirements precision	0.001
SVM	Iterations number	1000
	Kernel function	RBF
	Gamma	0.03125
ELM	Iterations number	1000
	Kernel function	RBF
RVFL	Iterations number	1000
	Kernel function	RBF
LSTM	Iterations number	500
	Initial learn rate	0.005
	Number of neurons	128
KNN	$n\_neighbors$	8
	Weights	Distance
RF	$n\_estimators$	50
	Random state	50
XGB	Learning rate	0.1
	Max depth	4
	Booster	gbtree

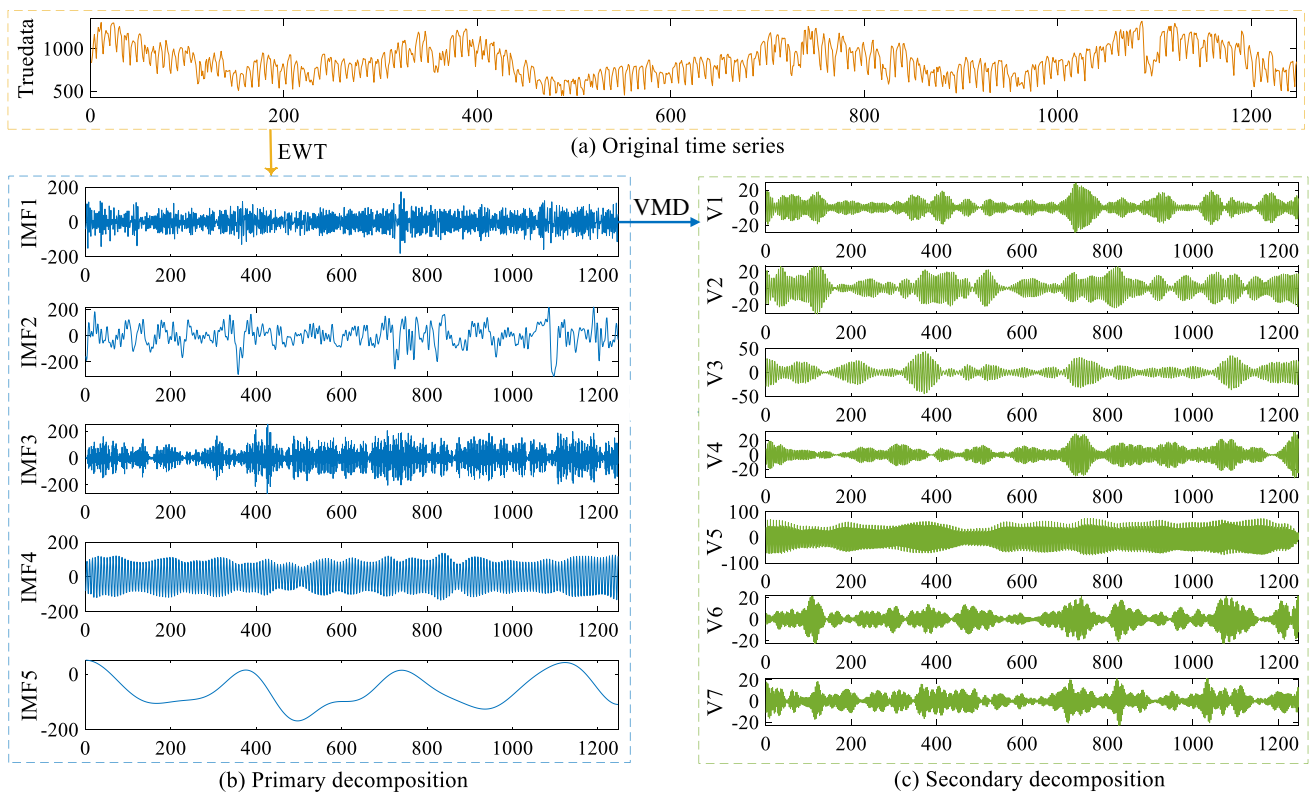


Fig. 7 Secondary decomposition results of Italy

to measure the correlation between different single models. The process of selecting sub-model is performed in the validation set. Firstly, the single model with the best outcome is selected according to the prediction accuracy index. Then, the NMI of this single model and other models is calculated. Finally, the model whose NMI value is higher than the average NMI of the optimal model and all other models is selected as the base learner.

Figure 10 shows the models selected at various steps for three datasets, where orange stands for the single model with the best prediction performance, yellow stands for the sub-model selected based on NMI, and blue stands for unselected models. It can be seen that the selected models of different datasets are not completely the same. For example, at the advance one step, only RVFL in Italy’s selected model is consistent with the other two datasets, which is because the

different models can capture different dataset features more effectively, indicating the necessity of sub-model selection. In addition, as shown in Table 6, the number of sub-model with different step selection is also inconsistent, but none of them exceed 7, which is in line with the number of sub-model commonly used in combined prediction (Liu et al. 2022c).

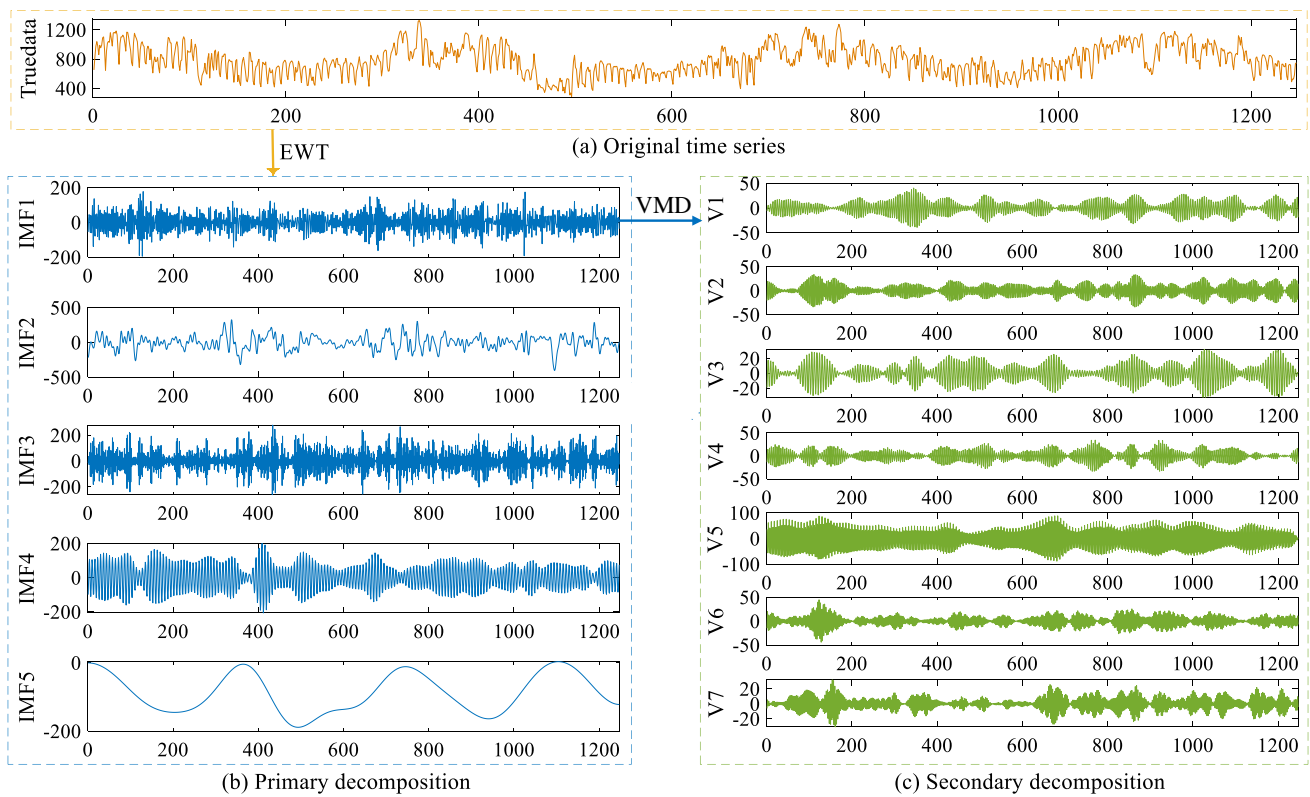
**Comparison experiments**

In order to fully verify the effectiveness and feasibility of the proposed model, three types of comparative experiments are designed in this paper. In each subsection in 3.5, we compared our proposed model with itself, making several alterations commonly used in the field and representative of the references mentioned in the introduction section, to realistically test the effectiveness of our proposed model.

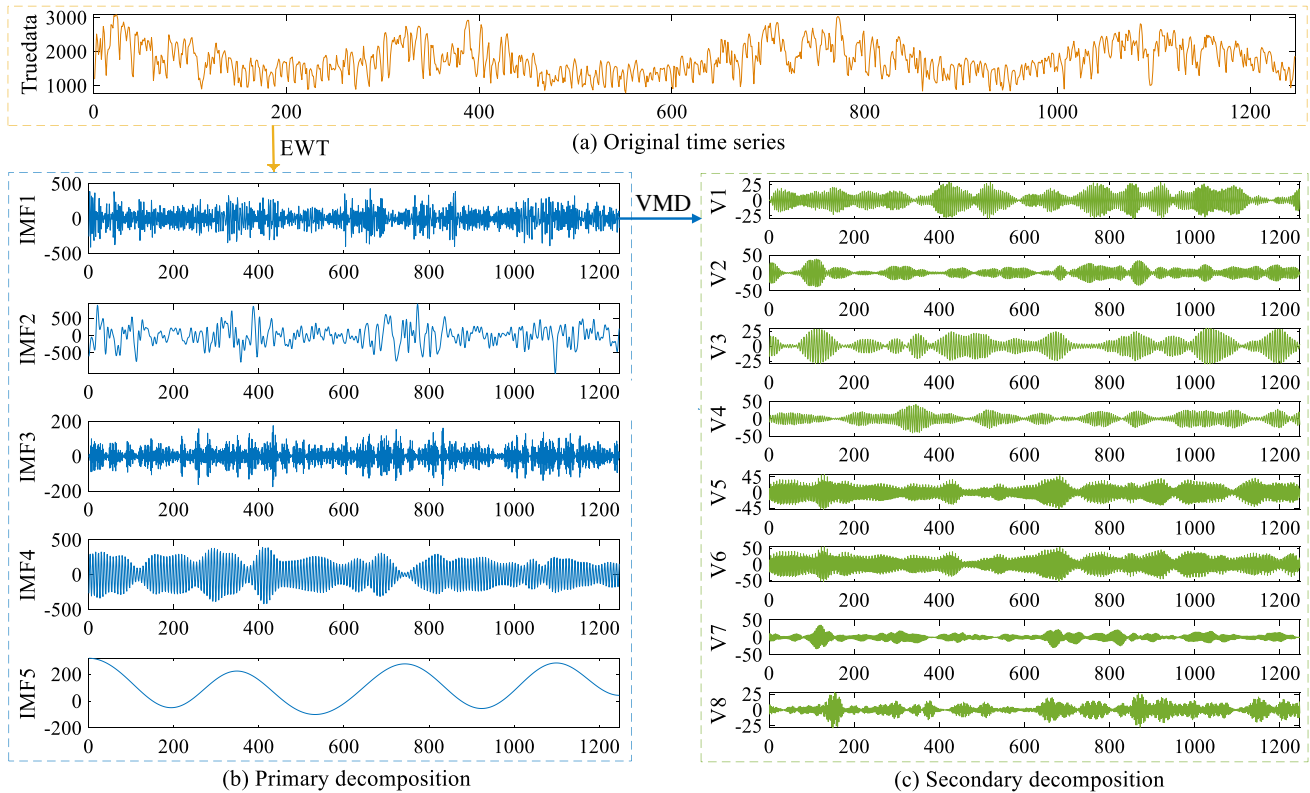
Table 5 SE result for IMF obtained by EWT decomposition

	Original	IMF1	IMF2	IMF3	IMF4	IMF5
Italy	1.3301	<b>1.373</b>	0.8177	0.6437	0.2755	0.0445
France	1.4174	<b>1.4568</b>	0.6858	0.7034	0.3665	0.0466
Germany	1.4692	<b>1.5564</b>	0.7266	0.7215	0.3385	0.0442

Note: The bolded numbers represent the complexity of the corresponding IMF is higher than the original sequence



**Fig. 8** Secondary decomposition results of France



**Fig. 9** Secondary decomposition results of Germany

Model	BP	0.78	0.73	0.76	0.72	0.80	0.75	0.72	0.74	0.75
	SVM	0.79	0.65	0.78	0.73	0.73	0.72	0.72	0.76	0.85
	ELM	0.80			0.72			0.70	0.88	
	RVFL		0.74	0.77		0.79	0.82			0.86
	LSTM	0.81	0.76	0.67	0.71	0.84	0.73	0.71	0.74	0.74
	KNN	0.76	0.70	0.71	0.76	0.78	0.67	0.78	0.76	0.76
	RF	0.77	0.69	0.71	0.77	0.79	0.66	0.75	0.86	0.70
	XGB	0.76	0.70	0.71	0.74	0.74	0.65	0.73	0.77	0.72
	THETA	0.82	0.86	0.87	0.68	0.78	0.78	0.65	0.71	0.69
	ARIMA	0.80	0.81	0.83	0.63	0.79	0.88	0.59	0.82	0.85
		ItalyH1	ItalyH15	ItalyH30	FranceH1	FranceH15	FranceH30	GermanyH1	GermanyH15	GermanyH30
		Dataset								

Fig. 10 NMI between selected model and candidate models

We construct a naming system for ease of reference, as shown in Table 7. Each model’s name consists of three components: decomposition level, sub-model selection method, and combination method. Our proposed model can be denoted as “secondary-NMI-stacking,” which refers to the hybrid model using a secondary decomposition method, sub-model selection method based on NMI, and stacking ensemble learning method. Furthermore, in the following figures and tables, the black bold text indicates the optimal error value, and H1, H15, and H30 show 1, 15, and 30 steps ahead of prediction.

**Comparison with different decomposition levels**

In order to verify the effectiveness of the secondary decomposition method in improving the prediction performance, the proposed model is compared with the same model but without secondary decomposition.

Considering the situation where no NMI sub-model selection method and stacking ensemble learning method

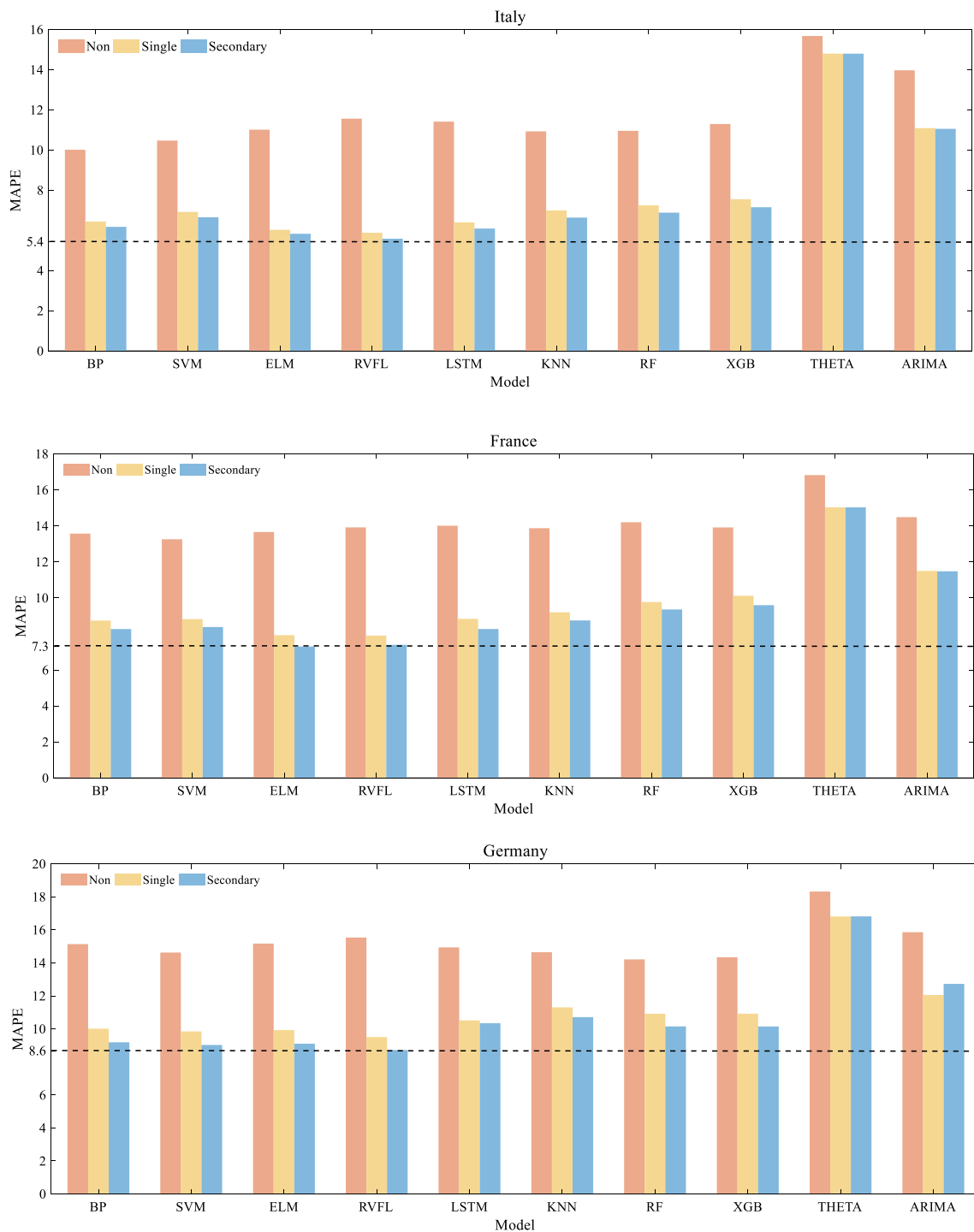
Table 7 Introduction of benchmark models

Method	Symbol	Details
Decomposition	Non	Non-decomposition
	Single	Single decomposition
	Secondary	Secondary decomposition
Sub-model selection	Full	No sub-model selection
	TRI	Model trimming
	NMI	Neighborhood mutual information
Combination	AVE	Simple average
	ER	Error reciprocal based on RMSE
	Stacking	Stacking ensemble learning

are used, the single models with different decomposition levels are compared with the proposed model in this paper. Figure 11 shows the comparison of the prediction performance of the 10 single models and the proposed model in three datasets. It can be seen that compared with the non-decomposition model, the effect of the 1-step decomposition model was significantly improved. The prediction accuracy was improved by more than 50% on average for 1 step ahead, more than 25% for 15 steps ahead, and more than 30% for 30 steps ahead. The shorter the prediction step, the better the improvement effect, and the prediction error was greatly reduced. Furthermore, the IMF with high complexity after the first decomposition was decomposed for secondary. In the experiments of 1-step, 15-step, and 30-step experiments, the prediction error RMSE of RVFL was reduced by 15.43%, 5.5%, and 3.33%, respectively, and the prediction accuracy MAPE of 10 single models is improved by 7.48%, 4.29%, and 2.07% on average. The results showed that the secondary decomposition has a good effect in advanced multi-step prediction and is an effective method to improve the prediction accuracy and perform data preprocessing. In addition, this paper proposed model with lower prediction accuracy than that of all the single models with different decomposition levels. A black dotted line in Fig. 11 represents the prediction accuracy by MAPE of the proposed model, and it can be seen that the black line is lower than the cylinder of the 10 single models. For example, in Italy dataset, the average improvement in prediction accuracy of the proposed model was 44.54% relative to the 10 single models, indicating that the hybrid model proposed in this paper is better than the single model without preprocessing and ensemble learning.

Table 6 Sub-model selected for multi-step prediction of three countries

	H1	H15	H30
Italy	RVFL, XGB, KNN, RF, BP	ELM, SVM, RF, KNN, XGB, BP, RVFL	ELM, LSTM, RF, KNN, XGB
France	RVFL, ARIMA, THETA, LSTM	ELM, SVM, XGB, KNN	ELM, XGB, RF, KNN, SVM, LSTM
Germany	RVFL, ARIMA, THETA, ELM	RVFL, THETA, LSTM, BP, KNN, SVM, XGB	ELM, THETA, XGB, LSTM, BP, KNN



**Fig. 11** Prediction results of models using multi-layer decomposition of three countries

Considering the situation where the NMI sub-model selection method and stacking ensemble learning method are used to integrate single models with different decomposition levels, to verify the effectiveness of the secondary decomposition in the proposed hybrid model, Fig. 12 shows the prediction results of the non-decomposed

model, the single decomposition model, and the secondary decomposition model using the same integration method. It can be seen that the integration effect of the non-decomposed model is the worst, the single decomposition is the next best, and the secondary decomposition integration is the best. In the 1-step prediction, the secondary



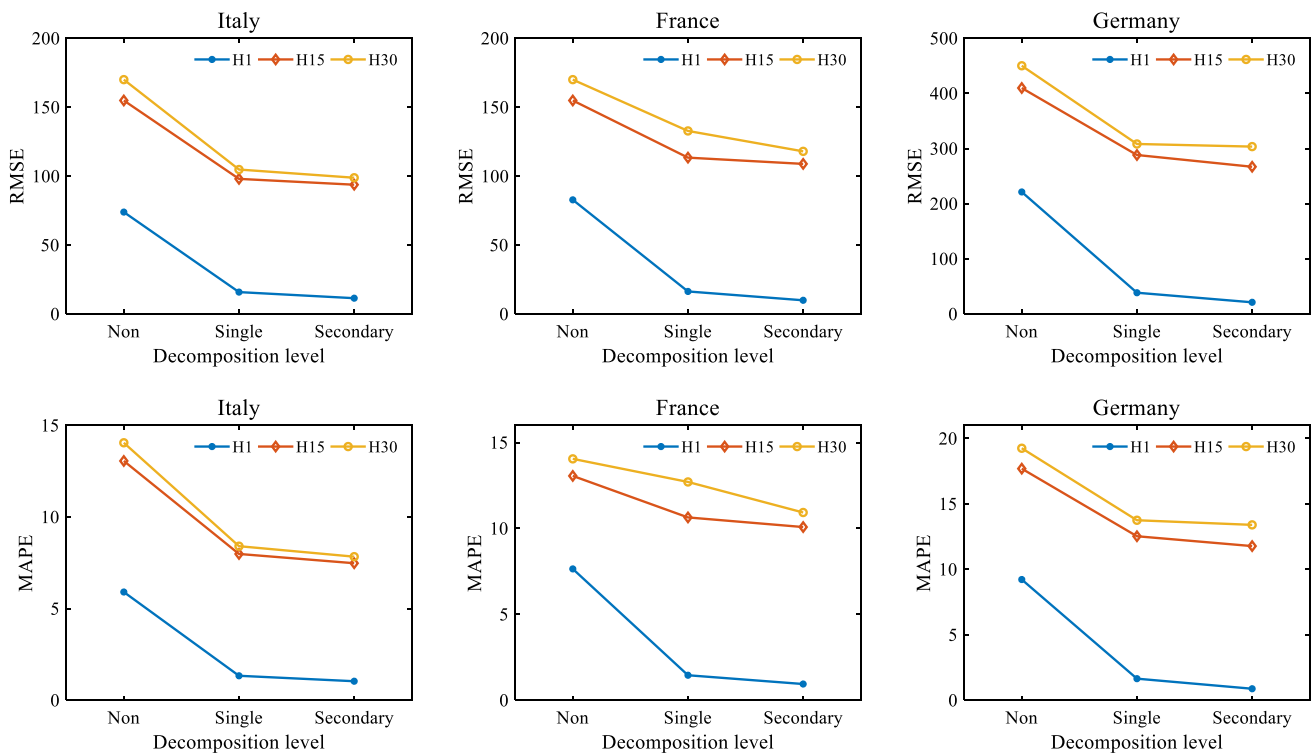


Fig. 12 Performance comparison of different decompose levels of three countries

decomposition of Germany dataset has the best integration effect, with 90.43% and 45.06% improvement in RMSE compared with the non-decomposed and single decomposition, respectively. In the 15-step prediction, the MAPE of the secondary decomposition integration of Italy dataset improved from 16.0238 to 7.4751, which was 53.35% better than that of the non-decomposed integration, indicating that the performance of the secondary decomposition is still outstanding in the multi-step prediction. In the 30-step prediction, our proposed model improved the prediction error RMSE by 20.88% on average in France dataset, and the MAPE decreased from 14.0491 to 10.9263. The prediction error is larger than that of 1 step ahead and 15 steps ahead, mainly because there are more uncertain factors in multi-step ahead prediction. The longer the prediction step, the higher the prediction difficulty, and the prediction error will increase relatively. However, the performance of the model proposed in this paper still maintains good prediction performance at 30 steps in advance, which further demonstrates the effectiveness of the proposed model.

To sum up, the above experimental results illustrate three conclusions. First of all, the secondary decomposition method is more effective than non-decomposed method and single decomposition method, which is due to the fact that the secondary decomposition method can decompose the time series high-frequency data more completely and effectively and reduce the difficulty of prediction. Secondly,

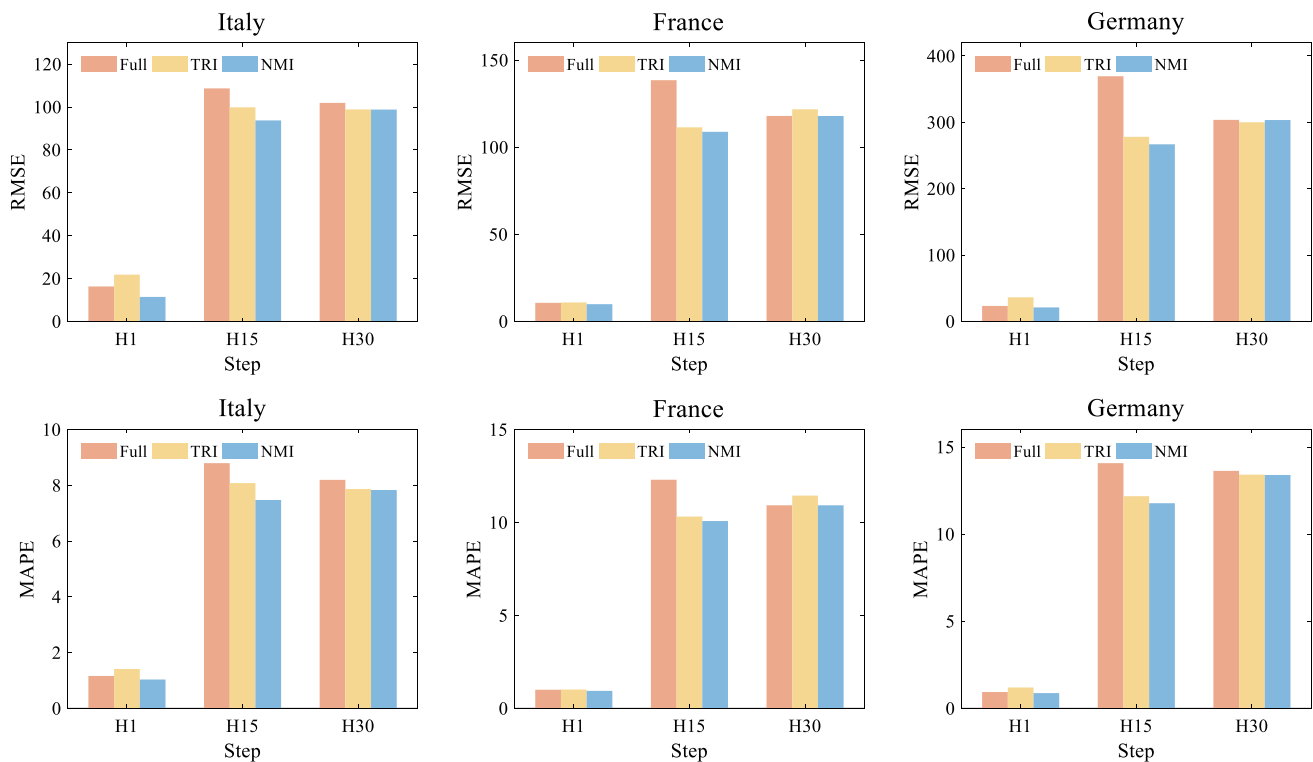
EWT is used for single decomposition and VMD is used for secondary decomposition, which can capture the valid data characteristics of nonlinear and non-stationarity of carbon emissions. Finally, preprocessing the data using the secondary decomposition method is very necessary in the proposed hybrid model, which enhances the overall prediction performance.

**Comparison with different sub-model selection methods**

To further ensure the effectiveness of sub-model selection in the proposed hybrid framework, we used other sub-model selection methods and the NMI method for comparison under the premise that the decomposition method and the integration method remain unchanged.

First, the proposed model is compared with and without the sub-model selection method, that is, the prediction results of the 10 single models after secondary decomposition were directly integrated by the stacking ensemble learning method. Next, comparing of two different sub-model selection methods using trimming and NMI. Model trimming is a commonly used method for sub-model selection, which censors candidate single models according to their prediction accuracy. In this paper, we choose to remove the single models with the smallest and largest prediction errors, which means that eight single models are integrated.

As can be seen from Fig. 13, using the NMI sub-model selection method has the smallest error at all three steps

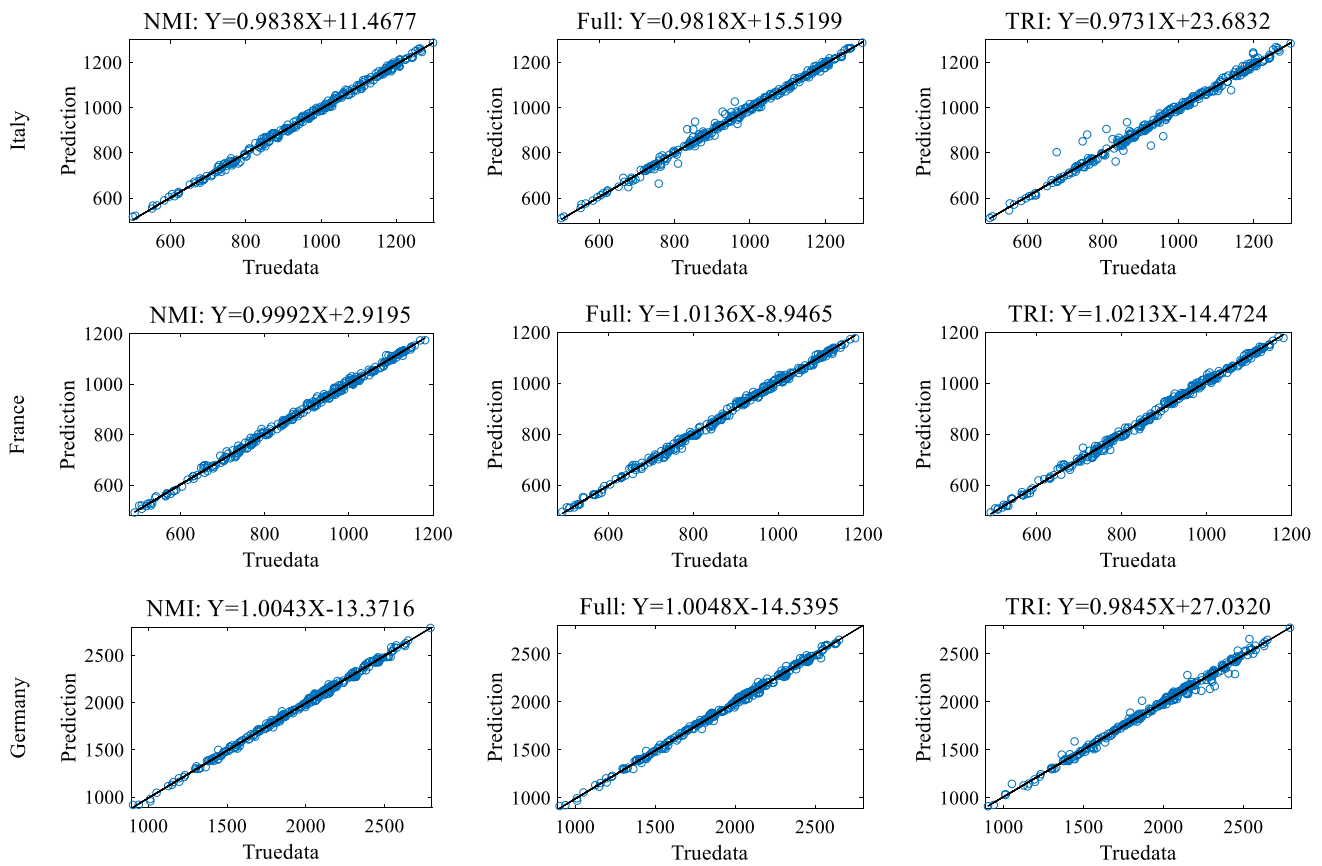


**Fig. 13** Performance comparison of different sub-model selection methods of three countries

and achieves the best prediction in all three datasets. In Germany dataset, the proposed model performs well in both 1-step ahead and 15-step ahead predictions. However, in 30-step ahead prediction, the NMI sub-model selection method has a higher RMSE than trimming by 1.11% and a lower MAPE than trimming by 0.13%. This is due to the increased difficulty of advance multi-step prediction, and the individual extreme values in the model prediction values selected by the NMI sub-model selection method. RMSE will amplify the prediction error, resulting in poor effect of RMSE. And MAPE represents the relative size of the deviation between predicted and true values, indicating that the NMI sub-model selection method has improved in the overall prediction effect. In France dataset, the average improvement of prediction accuracy RMSE in 1-step and 15-step ahead was 9.96%, which also achieved a good improvement effect. But in 30-step ahead prediction, the prediction accuracy of the model using the NMI sub-model selection method was the same as that of the model without the sub-model selection method. This is because when using the stacking method, we choose the Cubist model, which is a decision tree model, and the model itself will prune the data in complexity. In other words, the selected models obtained by the NMI sub-model selection method at this point are the same as the result of pruning 10 single models into the Cubist model.

But its prediction accuracy is much higher than that of the single model without integration, and the necessity of sub-model selection cannot be denied.

The model prediction results can be visualized from Fig. 14 scatter plot. The straight line  $y = wx + b$  ( $w$  represents the line gradient and  $b$  represents the intercept) represents the line fitted from the predicted value and which line has  $w$  closer to one indicates that the model has a better fitting effect. In other words, when the fitted straight line is closer to the diagonal, the higher the prediction accuracy of the model. It can be seen that the model without sub-model selection and trimming-based selection have an unsatisfactory prediction, and the deviation from the real value is large. The prediction performance of the model selected by the NMI sub-model is better, and the fitting degree is higher compared with the real value. In Italy dataset, compared with the model without sub-model selection, the proposed model prediction accuracy of RMSE was increased by 29.89%, and MAPE was increased by 10.95%, which fully demonstrates the effectiveness of sub-model selection. And compared with the trimming sub-model selection method, the proposed model has better prediction accuracy, with an average improvement of 37.17%. In 1-step ahead prediction, the NMI sub-model selection method performed well on both France and Germany datasets. In France dataset, the



**Fig. 14** Scatter plot of predicted and true values 1-step ahead of three countries

proposed model prediction error RMSE decreased from 10.9002 to 9.9292, and MAPE improved by 6.55%, which also achieved a good improvement effect. According to Germany data, the prediction accuracy of using NMI sub-model selection method was significantly higher than that using trimming, with RMSE decreasing from 36.3815 to 21.1808 and MAPE decreasing from 1.1984 to 0.8762. In summary, it is important to use the sub-model selection algorithm before integrating, and the NMI sub-model selection method adopted in this paper is superior to the trimming method.

From the above experiments, it can be shown that it is necessary to use the sub-model selection method for integration input selection. Because there will be redundancy among the candidate single models, the effect of combining all candidate models may not be better than that of single models. The sub-model selection method can remove the redundant models and select the optimal sub-model. In addition, the prediction performance of the NMI sub-model selection method is higher than trimming because of the specificity of the integration method. Using selection models with differences as a base learner can effectively improve the prediction accuracy of the stacking ensemble learning method. The NMI sub-model selection

method can select the sub-models with a high difference and strong diversity among the candidate models, which is more suitable for the stacking ensemble learning method.

**Comparison with different ensemble methods**

To evaluate the performance of stacking ensemble learning method used in this paper, the commonly used simple average method and error reciprocal method are used as control methods for the experiments, while the secondary decomposition method and NMI sub-model selection method remained unchanged. The simple average method is the simplest weight determination method, that is, to give the same weight to each prediction model. The error reciprocal method gives more weight to models with smaller errors. In this paper, RMSE is used as the error measure index. Assuming that the errors of two single models in the validation set are RMSE1 and RMSE2, the weights of model 1 is  $\frac{RMSE1}{RMSE1+RMSE2}$ , and model 2 is  $\frac{RMSE2}{RMSE1+RMSE2}$ .

Table 8 shows the MAPE and RMSE indicators of the error between the actual value and the multi-step predicted value. It can be seen that comparing with simple average and error reciprocal methods, the stacking

**Table 8** The error results of multi-step prediction for three datasets

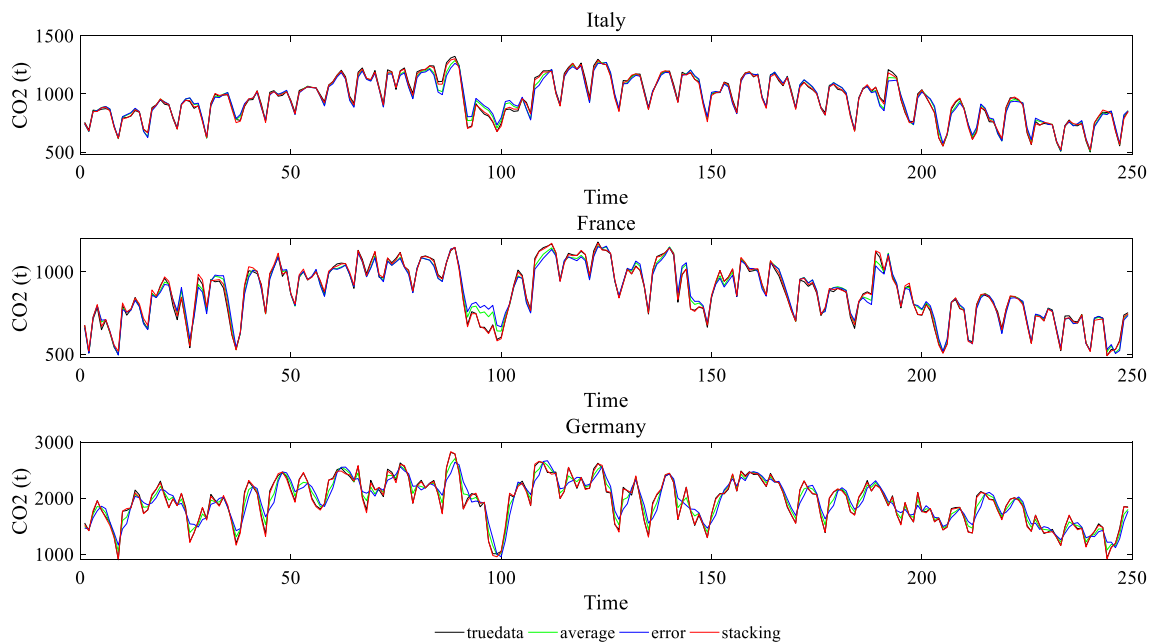
Dataset	Model	1-step		15-step		30-step	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Italy	Non-NMI-stacking	73.91	5.90	166.47	16.02	182.86	17.05
	Single-NMI-stacking	15.87	1.33	97.99	7.98	104.77	8.41
	Secondary-full-stacking	16.34	1.16	108.69	8.97	101.93	8.20
	Secondary-TRI-stacking	21.85	1.41	99.86	8.08	98.89	7.87
	Secondary-NMI-AVE	25.08	2.05	95.37	7.69	104.97	8.48
	Secondary-NMI-ER	34.09	2.78	96.10	7.74	105.53	8.53
	Proposed model	<b>11.45</b>	<b>1.04</b>	<b>93.76</b>	<b>7.48</b>	<b>98.82</b>	<b>7.83</b>
France	Non-NMI-stacking	82.79	7.64	154.85	13.06	169.97	14.05
	Single-NMI-stacking	16.34	1.45	113.39	10.64	132.76	12.70
	Secondary-full-stacking	10.71	1.00	138.48	12.30	117.96	10.93
	Secondary-TRI-stacking	10.90	1.01	111.43	10.32	121.79	11.45
	Secondary-NMI-AVE	24.29	2.17	113.56	10.60	125.39	11.78
	Secondary-NMI-ER	34.06	2.97	113.82	10.63	126.04	11.85
	Proposed model	<b>9.93</b>	<b>0.94</b>	<b>108.86</b>	<b>10.08</b>	<b>117.96</b>	<b>10.93</b>
Germany	Non-NMI-stacking	221.25	9.22	409.77	17.69	450.09	19.25
	Single-NMI-stacking	38.55	1.65	288.26	12.53	308.26	13.75
	Secondary-full-stacking	23.41	0.94	372.04	15.30	303.73	13.63
	Secondary-TRI-stacking	36.38	1.20	278.20	12.18	<b>300.16</b>	13.41
	Secondary-NMI-AVE	90.08	3.95	278.77	12.40	300.43	13.45
	Secondary-NMI-ER	160.84	7.03	284.07	12.79	300.20	13.47
	Proposed model	<b>21.18</b>	<b>0.88</b>	<b>266.94</b>	<b>11.77</b>	303.51	<b>13.40</b>

ensemble learning method has lower prediction error in three countries. In Italy dataset, at 15-step ahead prediction, the stacking ensemble learning method improved the prediction accuracy by 2.612% on average. And at 30-step ahead prediction, the simple average method prediction error RMSE was 104.9658, and MAPE was 8.479, and the stacking method prediction accuracy RMSE was improved by 5.86%, and MAPE was improved by 7.62%, with better prediction performance. In France dataset, the average prediction error RMSE for the stacking ensemble learning method multiple steps ahead is 68.01, and MAPE is 5.45. Relative to simple average method reduces the prediction error RMSE by 9.49% and MAPE by 8.73%. Relative to error reciprocal method, the prediction error was reduced by 13.41% for RMSE and 12.76% for MAPE. In Germany dataset, the average prediction error MAPE of the stacking ensemble learning method is less than 28.58%, and the RMSE is less than 29.86% for the other two methods.

To allow readers a better visualization of the effect of the model predictions, we selected prediction results for 1-step ahead of the three countries as indicated in a line graph. From Fig. 15, the use of the stacking ensemble learning method still has a certain impact on the prediction results. In Italy dataset, compared with the error reciprocal method, RMSE decreased from 34.085 to 11.4523, MAPE decreased from 2.7781 to 1.0355, and the prediction accuracy increased by 64.56% on average. In France

dataset, the ensemble method used in this paper has superior performance in advance multi-step prediction and a better fitting effect. Compared with the simple average method and error reciprocal method, RMSE increased by 59.12% and 70.85%, and MAPE increased by 56.66% and 68.33%, indicating that the stacking ensemble learning method is very effective in 1-step prediction. In Germany dataset, the stacking ensemble learning method performed more prominently in the 1-step ahead prediction, with a more significant improvement in prediction accuracy. In the RMSE error index, the simple average method was 90.0833, the inverse error method was 169.8419, and the stacking ensemble learning method was 21.1808, with an average improvement of 81.66% in prediction accuracy.

In general, the combined effect of the stacking ensemble learning method is significantly better than that of simple average method and error reciprocal method. This is because the stacking ensemble learning method can make full use of different learning strategies of the input model, enhance the fitting ability of a single model, and combine them for accurate prediction. Unlike the other two combination methods, the combined function of stacking ensemble learning method is a complex model, rather than a simple function such as average or error reciprocal, which is more suitable for non-stationary, noisy and volatile data. In addition, it can be found through experiments that the performance of the error reciprocal method in carbon



**Fig. 15** Prediction result of 1-step ahead of three countries

emission data is not stable, especially on Germany dataset, and the prediction accuracy of 30-step prediction is poor. The simple averaging method has poor performance in 1-step prediction, and the prediction error is much higher than that of stacking method. In comparison, the stacking ensemble learning method has stronger generalization ability and has stable and excellent performance in all three datasets.

## Results and discussion

The prediction results of the three comparative experiments can be found as follows.

- (1) Consistent with the experimental results of Kong et al. (2022a), the data decomposition method can effectively improve the prediction accuracy of carbon emissions. By using the signal decomposition method to preprocess the original time series data, the data can be decomposed into subsequences with stronger fluctuation regularity and lower complexity, and the accuracy of model prediction can be improved.
- (2) The secondary decomposition of the subsequence with the highest sample entropy value can make the sequence decomposition more complete, and the prediction model can better grasp the sequence characteristics and reduce the modeling difficulty. This is the same as the experimental conclusion of Li et al. (2021). The EWT-VMD decomposition method adopted in this paper has obtained better prediction results in the datasets of three

- countries, indicating that the decomposition method can decompose complex carbon emission data effectively.
- (3) The RVFL and ELM models perform outstandingly in the field of carbon emission prediction, among which RVFL performs better in single-step prediction and ELM performs well in multi-step prediction. In addition, the ARIMA model, while not a single model with the lowest predictive accuracy, can capture data points that AI models ignore and perform well in combined prediction (Yang and O'Connell 2020).
- (4) Sub-model selection method is crucial for combined forecasting. For different datasets, not all submodels can play an active role in combined prediction due to different data characteristics (Fu and Zhang 2022). Sub-model selection can select appropriate models from the candidate models of that dataset and effectively improve forecasting accuracy. The NMI sub-model selection method is superior to the trimming-based selection method. For example, in the Italy single-step prediction, the prediction accuracy using the NMI sub-model selection method is higher than that of the unused model, with RMSE and MAPR improving by 26.74% and 29.89%, respectively.
- (5) The stacking ensemble learning method is an effective combined prediction method for carbon emission. The meta-learner of the stacking ensemble learning method uses complex functions, which is better than the traditional simple average method. This is consistent with the experimental results of Zhao and Cheng (2022). For example, in the France single-step prediction,

the prediction accuracy of the model using the stacking ensemble learning method is significantly higher than that of the simple average method, with MAPE increased by 56.66% and RMSE increased by 59.11%.

In addition, it can be found that the carbon emissions of these three countries have seasonal characteristics, with a general trend of high in winter and low in summer, and a U-shaped curve throughout the year. The cold winter weather increases the demand for heating, leading to the corresponding increase in energy generation, which produces large amounts of carbon dioxide, resulting in the annual peak of carbon emissions occurring in winter. Then, as temperatures rise, society's demand for heating decreases and so does the supply of energy. And the vegetation gradually flourishes and the carbon sequestration capacity increases, which makes the carbon emissions in the summer lower.

The three countries exhibit different volatility characteristics due to different carbon neutral policies and economic environments. Germany, whose carbon emissions fluctuate wildly, is the biggest carbon emitter in the EU, accounting for roughly one-fifth of the EU's greenhouse gas emissions. This is caused by Germany's high degree of industrialization; the industrial sector's share of total emissions is significantly higher than the EU average. On the other hand, Germany's electricity structure is dominated by coal and natural gas, supplemented by renewable energy sources. Renewable energy sources mainly refer to wind and solar energy. During undesirable weather, weather, gas, and coal-fired power plants have to compensate for reduced wind and solar capacity, which doubles carbon emissions. After phasing out nuclear energy, Germany needs to minimize the use of fossil energy and continuously explore renewable energy sources (Yang 2022).

France's carbon emissions fluctuate more moderately, with per capita emissions below the world average over the past 2 years. France's energy structure is one of the most decarbonized countries in the world. The low carbon emissions are mainly due to nuclear energy, which is the main source of electricity generation in France (Lebrouhi et al. 2022). Besides, in 2019, France codified into law its goal of achieving net zero emissions by 2050, and the following year updated its energy transition framework. Proactive government action has enabled the carbon neutrality process to move forward effectively. However, the energy consumption of the whole French economy is still dominated by fossil fuels, and the goal of carbon neutrality should be achieved through the development of wind energy resources and other clean sources, such as solar energy.

Italy's carbon emissions fluctuate similarly to France's, with smaller extreme differences and higher per capita emissions. The distribution of energy-intensive industries in Italy is higher than the EU average, and the share of fossil energy

represented by coal, oil, and natural gas in the energy structure is over 70%. Although natural gas, which has a lower carbon footprint, has become the primary choice for energy consumption, it is mainly dependent on imports and is seriously affected by international policy situations. In recent years, the Italy government has been vigorously developing their hydrogen energy industry. Hydrogen energy has various advantages, such as large storage capacity, low pollution, and high efficiency. It is an ideal new energy source for Italy to achieve carbon neutrality (Pastore et al. 2022).

## Conclusions

The EU plans to be carbon neutral by 2050. Reliable short-term carbon emission prediction can depict the imminent change in carbon emissions and provide timely and accurate data for policymakers to use as reference. In this paper, we propose a hybrid model based on secondary decomposition and stacking ensemble learning method to improve the prediction accuracy of carbon emissions. The model is mainly divided into three parts: data decomposition, model prediction and selection, and stacking ensemble. The generalization ability and robustness of the hybrid model are demonstrated by ablation experiments in Italy, France, and Germany. According to the prediction results, the following conclusions can be drawn:

- (1) The hybrid model proposed in this paper is suitable for multi-step prediction of carbon emissions. In the experiments of 1 step, 15 steps, and 30 steps in advance, the prediction effect is basically better than that of the control model and achieves higher prediction accuracy.
- (2) The secondary decomposition method is superior to single decomposition and non-decomposition method, and the EWT-VMD decomposition used in this paper can effectively reduce the complexity of carbon emission data.
- (3) The sub-model selection method can select the single model more suitable for combination from candidate models. In this paper, the sub-model selection method based on NMI can enhance the prediction accuracy of the hybrid model successfully.
- (4) Combined prediction can contain more comprehensive predictive information and has higher stability. In this paper, the stacking ensemble learning method is used to combine the advantages of single models, so that the overall prediction performance is significantly improved.

Based on the forecast results of three representative EU countries, some suggestions are put forward for policymakers in EU countries. First, there is a need to shift the energy

structure away from fossil fuels towards renewable energy sources, such as solar, wind, tidal, nuclear, and hydrogen. There is also a need to increase investment in renewable energy and ensure the development of clean industries. Second, these countries should improve their carbon fixation capacities and effectively promote the development of carbon capture, utilization, and sequestration technology to improve the resourceful use of carbon dioxide emitted from fossil fuels. At the same time, vegetation coverage should be increased, and plant photosynthesis made full use of to enhance carbon fixation capacity. Finally, the carbon neutrality targets should be written into law, and strict legal measures will be taken to ensure the timely completion of carbon neutrality targets. The formulation of laws and regulations cannot be divorced from the national economic situation, which is because most countries have not yet decoupled economic development from carbon emissions, and cannot propose emission reduction strategies at the expense of economic development.

However, there are still some limitations in this study. For example, the change of hyperparameter and sample range can lead to changes in machine learning model prediction accuracy, while the change of candidate models could potentially affect the stability and accuracy of NMI sub-model selection. In addition, the limited base models selected for the combined prediction in this paper and the application of many models in the field of carbon emissions remain to be explored. These aspects will continue to be investigated in future work.

**Acknowledgements** Thanks are due to Dabin Zhang, Liwen Ling, Huanling Hu, and Liling Zeng for valuable discussion.

**Author contribution** Boting Zhang: methodology, software, and writing original draft. Liwen Ling: writing—review and editing. Liling Zeng: visualization and formal analysis. Huanling Hu: literature and revised the draft. Dabin Zhang: supervision and finalizes proof. All authors have read and approved the manuscript.

**Funding** This work was supported by the National Natural Science Foundation of China (nos. 71971089 and 72001083), the Natural Science Foundation of Guangdong Province (no. 2022A1515011612), and the Special Project in Key Areas of Universities of Guangdong Province (no. 2020ZDZX3009).

**Data availability** The data used in this paper is publicly available at the CEADs' repository, <https://carbonmonitor.org.cn/downloads/>.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

- Abbasi KR, Hussain K, Radulescu M, Ozturk I (2021) Does natural resources depletion and economic growth achieve the carbon neutrality target of the UK? A way forward towards sustainable development. *Resour Policy* 74:102341. <https://doi.org/10.1016/j.resourpol.2021.102341>
- Allende H, Valle C (2016) *Ensemble Methods for Time Series Forecasting*. Springer International Publishing, Cham, pp 217–232. [https://doi.org/10.1007/978-3-319-48317-7\\_13](https://doi.org/10.1007/978-3-319-48317-7_13)
- Atiya AF (2020) Why does forecast combination work so well? *Int J Forecast* 36:197–200. <https://doi.org/10.1016/j.ijforecast.2019.03.010>
- Belbute JM, Pereira AM (2020) Reference forecasts for CO<sub>2</sub> emissions from fossil-fuel combustion and cement production in Portugal. *Energy Policy* 144:111642. <https://doi.org/10.1016/j.enpol.2020.111642>
- Bianucci M (2021) Operators central limit theorem. *Chaos Solit Fractals* 148:110961. <https://doi.org/10.1016/j.chaos.2021.110961>
- Cang S, Yu H (2014) A combination selection algorithm on forecasting. *Eur J Oper Res* 234:127–139. <https://doi.org/10.1016/j.ejor.2013.08.045>
- Che J (2015) Optimal sub-models selection algorithm for combination forecasting model. *Neurocomputing* 151:364–375. <https://doi.org/10.1016/j.neucom.2014.09.028>
- Cheng Y, Sinha A, Ghosh V, Sengupta T, Luo H (2021) Carbon tax and energy innovation at crossroads of carbon neutrality: designing a sustainable decarbonization policy. *J Environ Manage* 294:112957. <https://doi.org/10.1016/j.jenvman.2021.112957>
- Cui S, Yin Y, Wang D, Li Z, Wang Y (2021) A stacking-based ensemble learning method for earthquake casualty prediction. *Appl Soft Comput* 101:107038. <https://doi.org/10.1016/j.asoc.2020.107038>
- Da Silva RG, Ribeiro MHD, Moreno SR, Mariani VC, Coelho LDS (2021) A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting. *Energy* 216:119174. <https://doi.org/10.1016/j.energy.2020.119174>
- Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. *IEEE T Signal Proces* 62:531–544. <https://doi.org/10.1109/TSP.2013.2288675>
- Fu T, Zhang S (2022) Wind speed forecast based on combined theory, multi-objective optimisation, and sub-model selection. *Soft Comput* 26:13615–13638. <https://doi.org/10.1007/s00500-022-07334-y>
- Gilles J (2013) Empirical wavelet transform. *IEEE T Signal Process* 61:3999–4010. <https://doi.org/10.1109/TSP.2013.2265222>
- Graefe A, Armstrong JS, Jones RJ, Cuzán AG (2014) Combining forecasts: an application to elections. *Int J Forecast* 30:43–54. <https://doi.org/10.1016/j.ijforecast.2013.02.005>
- Heydari A, Garcia DA, Keynia F, Bisegna F, Santoli LD (2019) Renewable energies generation and carbon dioxide emission forecasting in microgrids and national grids using GRNN-GWO methodology. *Energy Procedia* 159:154–159. <https://doi.org/10.1016/j.egypro.2018.12.044>
- Hu Q, Zhang L, Zhang D, Pan W, An S, Pedrycz W (2011) Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Syst Appl* 38:10737–10750. <https://doi.org/10.1016/j.eswa.2011.01.023>
- Hu Y, Jiang P, Tsai J, Yu C (2021) An optimized fractional grey prediction model for carbon dioxide emissions forecasting. *Int J Env Res Pub He* 18:587. <https://doi.org/10.3390/ijerph18020587>
- Jia L, Chang T, Wang M (2022) Revisit economic growth and CO<sub>2</sub> emission nexus in G7 countries: mixed frequency VAR model. *Environ Sci Pollut R*. <https://doi.org/10.1007/s11356-022-24080-8>
- Jiang PLZNX (2021) A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods

- for short-term wind speed forecasting. *Energy* 217:119361. <https://doi.org/10.1016/j.energy.2020.119361>
- Kışınbay T (2010) The use of encompassing tests for forecast combinations. *J Forecast* 29:715–727. <https://doi.org/10.1002/for.1170>
- Kong F, Song J, Yang Z (2022a) A daily carbon emission prediction model combining two-stage feature selection and optimized extreme learning machine. *Environ Sci Pollut R*. <https://doi.org/10.1007/s11356-022-21277-9>
- Kong F, Song J, Yang Z (2022b) A novel short-term carbon emission prediction model based on secondary decomposition method and long short-term memory network. *Environ Sci Pollut R* 29:64983–64998. <https://doi.org/10.1007/s11356-022-20393-w>
- Kourentzes N, Barrow D, Petropoulos F (2019) Another look at forecast selection and combination: evidence from forecast pooling. *Int J Prod Econ* 209:226–235. <https://doi.org/10.1016/j.ijpe.2018.05.019>
- Lebrouhi BE, Schall E, Lamrani B, Chaibi Y, Kousksou T (2022) Energy transition in France. *Sustainability* 14:5818. <https://doi.org/10.3390/su14105818>
- Lee TR, Wood WT, Phrampus BJ (2019) A machine learning (kNN) approach to predicting global seafloor total organic carbon. *Global Biogeochem Cy* 33:37–46. <https://doi.org/10.1029/2018GB005992>
- Li H, Jin F, Sun S, Li Y (2021) A new secondary decomposition ensemble learning approach for carbon price forecasting. *Knowl Based Syst* 214:106686. <https://doi.org/10.1016/j.knsys.2020.106686>
- Ling L, Zhang D, Mugeraw AW, Chen S, Xia Q (2019) A forecast combination framework with multi-time scale for livestock products' price forecasting. *Math Probl Eng* 2019:1–11. <https://doi.org/10.1155/2019/8096206>
- Liu B, Han Z, Li J, Yan B (2022a) Comprehensive evaluation of municipal solid waste power generation and carbon emission potential in Tianjin based on grey relation analysis and long short term memory. *Process Saf Environ* 168:918–927. <https://doi.org/10.1016/j.psep.2022.10.065>
- Liu H, Wu H, Li Y (2018) Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. *Energy Convers Manage* 161:266–283. <https://doi.org/10.1016/j.enconman.2018.02.006>
- Liu H, Yin S, Chen C, Duan Z (2020) Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. *J Clean Prod* 277:124023. <https://doi.org/10.1016/j.jclepro.2020.124023>
- Liu Z, Deng Z, Davis SJ, Giron C, Ciais P (2022b) Monitoring global carbon emissions in 2021. *Nat Rev Earth Environ* 3:217–219. <https://doi.org/10.1038/s43017-022-00285-w>
- Liu Z, Hara R, Kita H (2021) Hybrid forecasting system based on data area division and deep learning neural network for short-term wind speed forecasting. *Energy Convers Manage* 238:114136. <https://doi.org/10.1016/j.enconman.2021.114136>
- Liu Z, Jiang P, Wang J, Zhang L (2022c) Ensemble system for short term carbon dioxide emissions forecasting based on multi-objective tangent search algorithm. *J Environ Manage* 302:113951. <https://doi.org/10.1016/j.jenvman.2021.113951>
- Liu Z, Sun W, Zeng J (2014) A new short-term load forecasting method of power system based on EEMD and SS-PSO. *Neural Comput Applic* 24:973–983. <https://doi.org/10.1007/s00521-012-1323-5>
- Lu H, Ma X, Huang K, Azimi M (2020) Carbon trading volume and price forecasting in China using multiple machine learning models. *J Clean Prod* 249:119386. <https://doi.org/10.1016/j.jclepro.2019.119386>
- Lv P, Shu Y, Xu J, Wu Q (2022) Modal decomposition-based hybrid model for stock index prediction. *Expert Syst Appl* 202:117252. <https://doi.org/10.1016/j.eswa.2022.117252>
- Mason K, Duggan J, Howley E (2018) Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks. *Energy* 155:705–720. <https://doi.org/10.1016/j.energy.2018.04.192>
- Meinshausen M et al (2022) Realization of Paris Agreement pledges may limit warming just below 2 °C. *Nature* 604:304–309. <https://doi.org/10.1038/s41586-022-04553-z>
- Mendes-Moreira J, Soares C, Jorge AM, Sousa JFD (2012) Ensemble approaches for regression. *Acm Comput Surv* 45:1–40. <https://doi.org/10.1145/2379776.2379786>
- Moon J, Park S, Rho S, Hwang E (2022) Interpretable short-term electrical load forecasting scheme using Cubist. *Comput Intell Neurosci* 2022:1–20. <https://doi.org/10.1155/2022/6892995>
- Musah M, Kong Y, Vo XV (2021) Predictors of carbon emissions: an empirical evidence from NAFTA countries. *Environ Sci Pollut Res Int* 28:11205–11223. <https://doi.org/10.1007/s11356-020-11197-x>
- Nguyen DK, Huynh T, Nasir MA (2021) Carbon emissions determinants and forecasting: evidence from G6 countries. *J Environ Manage* 285:111988. <https://doi.org/10.1016/j.jenvman.2021.111988>
- Pastore LM, Lo Basso G, Sforzini M, de Santoli L (2022) Technical, economic and environmental issues related to electrolyzers capacity targets according to the Italian Hydrogen Strategy: a critical analysis. *Renew Sustain Energy Rev* 166:112685. <https://doi.org/10.1016/j.rser.2022.112685>
- Qader MR, Khan S, Kamal M, Usman M, Haseeb M (2022) Forecasting carbon emissions due to electricity power generation in Bahrain. *Environ Sci Pollut R* 29:17346–17357. <https://doi.org/10.1007/s11356-021-16960-2>
- Qiao W, Lu H, Zhou G, Azimi M, Yang Q, Tian W (2020) A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J Clean Prod* 244:118612. <https://doi.org/10.1016/j.jclepro.2019.118612>
- Qin M, Su C, Zhong Y, Song Y, Lobont O (2022) Sustainable finance and renewable energy: promoters of carbon neutrality in the United States. *J Environ Manage* 324:116390. <https://doi.org/10.1016/j.jenvman.2022.116390>
- Ribeiro MHD, Dos Santos CL (2020) Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl Soft Comput* 86:105837. <https://doi.org/10.1016/j.asoc.2019.105837>
- Safonov G, Potashnikov V, Lugovoy O, Safonov M, Dorina A, Bolotov A (2020) The low carbon development options for Russia. *Clim Change* 162:1929–1945. <https://doi.org/10.1007/s10584-020-02780-9>
- Salvia M et al (2021) Will climate mitigation ambitions lead to carbon neutrality? An analysis of the local-level plans of 327 cities in the EU. *Renew Sustain Energy Rev* 135:110253. <https://doi.org/10.1016/j.rser.2020.110253>
- Samuels JD, Sekkel RM (2017) Model confidence sets and forecast combination. *Int J Forecast* 33:48–60. <https://doi.org/10.1016/j.ijforecast.2016.07.004>
- Shi G, Qin C, Tao J, Liu C (2021) A VMD-EWT-LSTM-based multi-step prediction approach for shield tunneling machine cutterhead torque. *Knowl Based Syst* 228:107213. <https://doi.org/10.1016/j.knsys.2021.107213>
- Shi M (2022) Forecast of China's carbon emissions under the background of carbon neutrality. *Environ Sci Pollut R* 29:43019–43033. <https://doi.org/10.1007/s11356-021-18162-2>
- Shirizadeh B, Quirion P (2022) The importance of renewable gas in achieving carbon-neutrality: insights from an energy system optimization model. *Energy* 255:124503. <https://doi.org/10.1016/j.energy.2022.124503>
- Smith JE, Domke GM, Woodall CW (2022) Predicting downed woody material carbon stocks in forests of the conterminous United States. *Sci Total Environ* 803:150061. <https://doi.org/10.1016/j.scitotenv.2021.150061>
- Sohail A, Du J, Abbasi BN, Ahmed Z (2022) The nonlinearity and nonlinear convergence of CO2 emissions: evidence from top 20



- highest emitting countries. *Environ Sci Pollut R* 29:59466–59482. <https://doi.org/10.1007/s11356-022-19470-x>
- Spiliotis E, Assimakopoulos V, Makridakis S (2020) Generalizing the theta method for automatic forecasting. *Eur J Oper Res* 284:550–558. <https://doi.org/10.1016/j.ejor.2020.01.007>
- Sun W, Huang C (2020) A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *J Clean Prod* 243:118671. <https://doi.org/10.1016/j.jclepro.2019.118671>
- Sun W, Huang C (2022) Predictions of carbon emission intensity based on factor analysis and an improved extreme learning machine from the perspective of carbon emission efficiency. *J Clean Prod* 338:130414. <https://doi.org/10.1016/j.jclepro.2022.130414>
- Sun W, Ren C (2021) Short-term prediction of carbon emissions based on the EEMD-PSOBP model. *Environ Sci Pollut R* 28:56580–56594. <https://doi.org/10.1007/s11356-021-14591-1>
- Sun Y, Lü L, Cai Y, Lee P (2022) Prediction of black carbon in marine engines and correlation analysis of model characteristics based on multiple machine learning algorithms. *Environ Sci Pollut R* 29:78509–78525. <https://doi.org/10.1007/s11356-022-20496-4>
- Tian Z (2021) Approach for short-term traffic flow prediction based on empirical mode decomposition and combination model fusion. *Ieee T Intell Transp* 22:5566–5576. <https://doi.org/10.1109/TITS.2020.2987909>
- Wang B, Wang J (2020) Deep multi-hybrid forecasting system with random EWT extraction and variational learning rate algorithm for crude oil futures. *Expert Syst Appl* 161:113686. <https://doi.org/10.1016/j.eswa.2020.113686>
- Wang W, Wang J (2021) Determinants investigation and peak prediction of CO<sub>2</sub> emissions in China's transport sector utilizing bio-inspired extreme learning machine. *Environ Sci Pollut R* 28:55535–55553. <https://doi.org/10.1007/s11356-021-14852-z>
- Wolpert DH (1992) Stacked generalization. *Neural Netw* 5:241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Xiao L, Wang C, Dong Y, Wang J (2019) A novel sub-models selection algorithm based on max-relevance and min-redundancy neighborhood mutual information. *Inform Sci* 486:310–339. <https://doi.org/10.1016/j.ins.2019.01.075>
- Yang H, O'Connell J (2020) Short-term carbon emissions forecast for aviation industry in Shanghai. *J Clean Prod* 275:122734. <https://doi.org/10.1016/j.jclepro.2020.122734>
- Yang L, Liu S, Tsoka S, Papageorgiou LG (2017) A regression tree approach using mathematical programming. *Expert Syst Appl* 78:347–357. <https://doi.org/10.1016/j.eswa.2017.02.013>
- Yang P (2022) Urban expansion of Energiewende in Germany: a systematic bibliometric analysis and literature study. *Energy Sustain Soc* 12. <https://doi.org/10.1186/s13705-022-00373-1>
- Yin H et al (2017) An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. *Energ Conver Manage* 150:108–121. <https://doi.org/10.1016/j.enconman.2017.08.014>
- Yu L, Liang S, Chen R, Lai KK (2022) Predicting monthly biofuel production using a hybrid ensemble forecasting methodology. *Int J Forecast* 38:3–20. <https://doi.org/10.1016/j.ijforecast.2019.08.014>
- Yu L, Wang S, Lai KK (2008) Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ* 30:2623–2635. <https://doi.org/10.1016/j.eneco.2008.05.003>
- Yue W, Liu Q, Ruan Y, Qian F, Meng H (2022) A prediction approach with mode decomposition-recombination technique for short-term load forecasting. *Sustain Cities Soc* 85:104034. <https://doi.org/10.1016/j.scs.2022.104034>
- Zeng L, Ling L, Zhang D, Jiang W (2023) Optimal forecast combination based on PSO-CS approach for daily agricultural future prices forecasting. *Appl Soft Comput* 132:109833. <https://doi.org/10.1016/j.asoc.2022.109833>
- Zhang D, Chen S, Liwen L, Xia Q (2020a) Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons. *IEEE Access* 8:28197–28209. <https://doi.org/10.1109/ACCESS.2020.2971591>
- Zhang S, Huo Z, Zhai C (2022) Building carbon emission scenario prediction using STIRPAT and GA-BP neural network model. *Sustainability* 14:9369. <https://doi.org/10.3390/su14159369>
- Zhang Y, Pan G, Chen B, Han J, Zhao Y, Zhang C (2020b) Short-term wind speed prediction model based on GA-ANN improved by VMD. *Renew Energy* 156:1373–1388. <https://doi.org/10.1016/j.renene.2019.12.047>
- Zhang Y, Wu J, Cai Z, Du B, Yu PS (2019) An unsupervised parameter learning model for RVFL neural network. *Neural Netw* 112:85–97. <https://doi.org/10.1016/j.neunet.2019.01.007>
- Zhao AB, Cheng T (2022) Stock return prediction: stacking a variety of models. *J Empir Financ* 67:288–317. <https://doi.org/10.1016/j.jempfin.2022.04.001>
- Zhao X, Han M, Ding L, Calin AC (2018) Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA. *Environ Sci Pollut R* 25:2899–2910. <https://doi.org/10.1007/s11356-017-0642-6>
- Zhao Y et al (2022) Have those countries declaring “zero carbon” or “carbon neutral” climate goals achieved carbon emissions-economic growth decoupling? *J Clean Prod* 363:132450. <https://doi.org/10.1016/j.jclepro.2022.132450>
- Zhou J, Li E, Wei H, Li C, Qiao Q, Armaghani DJ (2019) Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Appl Sci* 9:1621. <https://doi.org/10.3390/app9081621>
- Ziel F, Weron R (2018) Day-ahead electricity price forecasting with high-dimensional structures: univariate vs. multivariate modeling frameworks. *Energy Econ* 70:396–420. <https://doi.org/10.1016/j.eneco.2017.12.016>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.