Article

Dynamic Multi-Factor Correlation Analysis for Prediction of Provincial Carbon Emissions in China's Bohai Rim Region

Yanfen Qi [1,*] **, Xiurui Zhang** [2] **, Jiaan Zhang** [3] **and Yu Sun** [1]

[1] School of Public Management, Tianjin University of Commerce, Tianjin 300134, China; sunyuyao@163.com
[2] School of Electronic and Computer Engineering, Peking University, Beijing 100871, China; 2200013090@stu.pku.edu.cn
[3] School of Electrical Engineering, Hebei University of Technology, Tianjin 300401, China; 2011086@hebut.edu.cn
[*] Correspondence: 920132573@tjcu.edu.cn

Abstract: This study presents a dynamic multi-factor correlation analysis method designed to predict provincial carbon dioxide emissions (CDE) within China's Bohai Rim region, including Tianjin, Hebei, Shandong, and Liaoning. By employing the sliding window technique, dynamic correlation curves are computed between various influencing factors and CDE at different time intervals, thereby facilitating the identification of key feature attributes. A novel metric, the Consistency Index of Influencing Factors (CIIF), is introduced to evaluate the consistency of these factors across regions. Furthermore, the Accurate Predictive Capability Indicator (APCI) is defined to measure the impact of different feature categories on the prediction accuracy. The findings reveal that models relying on a single influencing factor exhibit limited accuracy, whereas combining multiple factors with diverse correlation features significantly improves the prediction accuracy. This study introduces a refined analytical framework and a comprehensive indicator system for CDE prediction. It enhances the understanding of the complex factors that influence CDE and provides a scientific rationale for implementing effective emission reduction strategies.

Keywords: carbon emissions; prediction; multi-factor analysis; dynamic correlation; accuracy enhancement

1. Introduction

With the intensifying global warming problem, CDE management has become a pivotal issue in the global battle against climate change and the advancement of sustainable development. Accurate CDE predictions are crucial for formulating effective emission reduction policies, optimizing energy structures, adjusting industrial frameworks, and fostering technological innovation. Consequently, CDE prediction is at the forefront of environmental science research and is an indispensable component of global environmental governance and sustainable development planning.

In models based on neural networks, the selection of key influencing factors significantly impacts both the accuracy and reliability of the prediction results. Therefore, a comprehensive analysis of the correlation between CDE and relevant influencing factors, as well as an exploration of their underlying mechanisms, is essential for improving the accuracy of the CDE predictions. In terms of CDE management, elucidating the intricate relationships between CDE and its influencing factors is also of great significance for the formulation and implementation of carbon management policies.

This study selects the Bohai Rim region in China as the research subject, which primarily comprises four provinces and cities: Tianjin, Hebei, Shandong, and Liaoning. The main reasons for choosing this region are as follows:

(1) Economic Significance: the Bohai Rim region is one of China's key economic areas, characterized by developed industries and a significant economic output, which exerts a notable influence on the country's carbon emissions.

(2) Industrial Structural Similarity: These four provinces and cities share certain similarities in their industrial structures, especially in heavy industry and chemical industry. This study aims to better understand the overall carbon emission characteristics of the Bohai Rim region.

(3) Data Availability: Given the feasibility of the research and the completeness of the data, we utilized panel data spanning from 1999 to 2021 for this region, relying on reference data sources from Guan, Y. et al. [1] and websites [2,3]. This ensures that sufficient data are available to support the analysis throughout the research process.

To investigate the intricacies of provincial CDE predictions in China's Bohai Rim region, influenced by a multitude of complex factors, we have adopted a dynamic multi-factor correlation analysis method. This approach aims to improve the prediction accuracy and address regional variations. Additionally, it meets the scientific requirements for informing carbon reduction strategies. With this method, we seek to contribute to the existing literature on CDE prediction and management, ultimately facilitating practical applications in environmental science and policy formulation. To achieve this objective, the following questions will be addressed:

(1) What are the mechanisms by which different categories of influencing factors affect CDE?

(2) How can the impact of influencing factors on CDE be dynamically described?

(3) How can the differences in the effects of the mechanisms affecting CDE across regions be quantified?

(4) How can key influencing factors be selected to improve prediction accuracy?

The remainder of this paper is organized as follows: Section 2 reviews the influence mechanism of different categories of factors on CDE and their representation. Section 3 describes the dynamic correlation features between CDE and its influencing factors. Section 4 analyzes the consistency of influencing factors according to different features. Section 5 compares the prediction accuracy of models using influencing factors with different features. Section 6 presents the discussion, and Section 7 concludes the paper.

2. Literature Review

The current prediction of CDE often incorporates various categories of factors, including economic development, urbanization, and technological advancements. As highlighted in [4–6], exploring the mechanisms of the influencing factors on CDE and conducting in-depth identification of key influencing factors significantly impact the accuracy of CDE predictions.

The relationship between economic development and CDE is evident in the correlation observed between economic growth and CDE across various regions and stages of development, further manifesting in regional disparities and individual differences in CDE levels. Multiple factors, such as economic growth, industrial structure, energy composition, and policy formulation, influence this correlation. Abid, M. [7] investigated the positive correlation between economic development and CDE. Liao, H. et al. [8] validated the nonlinear relationship between economic growth and CDE. The Environmental Kuznets Curve (EKC) hypothesis suggests that as levels of economic development increase, environmental pollution initially rises and then declines, following an inverted U-shaped trajectory. However, Mikayilov, J.I. et al. [9] analyzed the long-term impact of economic growth in Azerbaijan using various cointegration methods while proposing that the EKC hypothesis may not be applicable in that specific context. Furthermore, the relationship between economic development and CDE demonstrates significant variations across different regions and individuals. An analysis by Li, W. et al. [10] showed that countries form distinct higher-order clusters in terms of their relationships with economic development and CDE, reflecting differences in stages of economic development, energy structures,

policy frameworks, and other factors among nations. Both industrial structure and energy composition have a substantial impact on CDE; a higher proportion of heavy industry and high-energy-consuming sectors often results in increased CDE. Nie, Y. et al. [11] examined the disparities in industrial structure among the eastern, central, and western regions of China, leading to divergent trends in CDE.

Urbanization significantly influences CDE by driving infrastructure development, population aggregation, and the expansion of economic activities. On the one hand, the increase in energy consumption and changes in the industrial structure resulting from urbanization often lead to an increase in CDE [12,13]. On the other hand, urbanization also fosters technological innovation and enhances energy efficiency, which can help mitigate the growth of CDE to some extent [14,15]. Zhang, Y. et al. [16] conducted research using Beijing as a case study to examine the effects of policy interventions during the urbanization process on CDE, highlighting the critical role of strategic urban planning and effective governance practices in reducing CDE. Wang, S. et al. [17] examined the various effects of urbanization on CDE using panel data analysis. Furthermore, Abdallh, A.A. et al. [18] highlighted the mediating role of energy consumption in shaping the relationship between urbanization and CDE, emphasizing that improving energy efficiency is essential for curbing the growth of CDE. Meanwhile, Musah, M. et al. [19] revealed the importance of optimizing the industrial structure and pursuing low-carbon transformations for reducing CDE. Notably, significant differences exist in how urbanization relates to CDE across various regions and countries. Li, J. et al. [20] confirmed that factors such as economic structure, energy efficiency, and policy environment have different impacts on CDE across regions, while Wang, Y. et al. [21] identified a nonlinear relationship between levels of urbanization and CDE among different countries.

Technological innovation serves as an effective strategy to address the increase in CDE by reducing energy consumption, enhancing energy efficiency, and effectively driving economic growth activities. However, the impact of technological innovation on CDE exhibits regional disparities. For instance, in China, effects vary across eastern, central, and western regions [22]; similarly, in Malaysia, although specific regions are not explicitly distinguished, it can be inferred that differing levels of technological innovation and environmental conditions may result in variations in emission reduction outcomes [23]. Consequently, the relationship between technological innovation and CDE is complex and multidimensional, particularly in developing economies. Research conducted by Cheng, S. et al. [24] indicated that technological innovations in renewable energy positively affect CDE intensity in low quantile regions while exerting a negative impact in high quantile areas; conversely, the effect of fossil fuel-related innovation is the opposite. Zhang, M. et al. [25] performed studies utilizing provincial panel data from China to investigate how technological innovation influences CDE. Their findings revealed that such innovations indirectly contribute to reductions in CDE through improvements in energy efficiency and exhibit spatial spillover effects on neighboring provinces. Ali, W. et al. [23] identified a bidirectional causal relationship between energy consumption and economic growth, as well as between economic growth and technological innovation in the short term. Furthermore, Erdogan, S. [26] emphasized that CDE is a cumulative variable influenced by historical values, thus requiring consideration of dynamic effects. Eslamipoor, R. et al. [27] proposed a green supply chain model under a carbon cap and highlighted the critical role of policymakers and the importance of setting allowable emission limits.

In summary, economic development and industrial structure are the primary determinants of CDE, particularly in developing countries [28,29]. Meanwhile, the advancement of urbanization, accompanied by increased energy consumption and transport-related activities, has a significant impact on CDE [30]. Furthermore, technological advancements have become a crucial factor influencing CDE by optimizing industrial and energy structures [31]. The intricate interactions between these factors, along with other relevant elements, collectively determine the trajectory and magnitude of CDE trends [32].

Currently, the primary approaches for describing the relationship between CDE and its influencing factors predominantly encompass statistical analysis, decomposition analysis, grey relational analysis, and artificial intelligence analysis. Statistical analysis commonly employs techniques such as correlation, regression, factor, and principal component analysis. Raihan, A. et al. [33] utilized Dynamic Ordinary Least Squares (DOLS) and Canonical Correlation Regression (CCR) to analyze the dynamic implications of factors such as economic growth and energy consumption on CDE, revealing the causal relationships among these factors. Wang, Z. et al. [34] identified the principal influencing factors and their interactions using factor analysis (FA) and subsequently employed a Bayesian Neural Network (BNN) to capture the nonlinear relationship between the inputs and CDE. Chang, L. et al. [35] exploited the capability of Projection Pursuit Regression (PPR) in handling high-dimensional data and extracted the most critical information for predicting CDE from large-scale datasets. Yang, H. et al. [36] transformed the complex CDE time series into manageable components through decomposition and reconstruction, and subsequently employed a deep learning model to capture the patterns of each component and make predictions. Ding, Y.K. et al. [37] quantified the influence of technology-upgrading policies on CDE through factor analysis and expressed the flows of CDE between different regions and industries based on Graph Representation Learning (GRL), thereby predicting CDE. Chen, Y.X. et al. [38] captured the spatiotemporal correlations of CDE by means of a hybrid deep learning model integrating a Gated Recurrent Unit (GRU) and Graph Convolutional Network (GCN), thereby accomplishing CDE prediction.

Based on the above analysis, the potential limitations of the current research are as follows:

(1) Regarding the studied regions and datasets: Some research focuses on carbon emission studies in specific areas, such as in Egypt or certain urban agglomerations in China. These studies often rely on limited datasets, potentially failing to adequately capture the variations and similarities across diverse regions.

(2) Regarding the depth and breadth of influencing factor analysis: some studies may primarily concentrate on the impact of a few key influencing factors (e.g., economic growth and energy use) on carbon emissions, resulting in a somewhat oversimplified comprehension of the matter.

(3) With respect to models and methodologies: although these methods may excel in addressing simple or linear relationships, they may face limitations when dealing with complex, nonlinear carbon emission data.

(4) In terms of the practicality and specificity of policy recommendations: the policy suggestions proposed in some research may indeed be relatively high-level or general, lacking specific implementation plans tailored to particular regions or situations.

The primary objective of this study is to comprehensively analyze the dynamic evolution relationship between various influencing factors and CDE, specifically manifested in the following aspects:

(1) Dynamism: distinct from static correlation analysis, this study employs a sliding window technique to capture the time-varying relationships between carbon emissions and various influencing factors.

(2) Multi-factor Analysis and New Indicators: This study simultaneously considers multiple influencing factors, thereby providing a more comprehensive understanding of the drivers of carbon emissions. Furthermore, we introduce two novel indicators: the Consistency Index of Influencing Factors (CIIF) and the Accurate Predictive Capability Indicator (APCI), which together form a robust framework for evaluating and comparing prediction models.

(3) Comprehensive Analysis: by analyzing the consistency and comparing the prediction accuracy of influencing factors across different feature categories, our method offers deeper insights into the complex mechanisms of carbon emissions.

3. Dynamic Characterization of Correlation Features between CDE and Influencing Factors

The most prominent issue encountered in the panel data of the Bohai Rim region is the presence of missing values. To address this, the following data preprocessing measures were adopted in this study:

(1) Data Cleaning: Variables with excessive missing values and outliers were removed. To ensure the scientific validity of the research findings, interpolation methods were employed to correct missing data when the amount for a variable was less than 10%. Otherwise, the variable was excluded from the analysis, which inevitably led to certain differences in the datasets of various provinces.

(2) Data Standardization: To eliminate the influence of different variable dimensions, all variables were standardized. In this study, the Min–Max standardization method was primarily employed.

Additionally, the partitioning of the dataset into training and testing sets will be introduced in Section 5.

By employing the Pearson correlation coefficient, as outlined in Equation (1), we define the correlation between various influencing factors and CDE within a sliding window, as illustrated in Equation (2).

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu X)(Y - \mu Y)]}{\sigma_X \sigma_Y} \tag{1}$$

In Equation (1), X and Y represent the random variables, respectively. ρ_{XY} denotes the Pearson correlation coefficient between these two random variables. $\text{cov}(.,.)$ represents the covariance, σ indicates the standard deviation, μ signifies the expected value of a specified random variable, and $E[.]$ denotes the mathematical expectation function.

$$\rho_{XY\_Wi} = \frac{\text{cov}(X_i, Y_i)}{\sigma_{X_i} \sigma_{Y_i}}, i \in [1, n - w + 1] \tag{2}$$

In Equation (2), ρ_{XY_Wi} represents the Pearson correlation coefficient of the random variable within the i-th sliding window. X_i or Y_i denote the values of the random variable within the i-th sliding window, respectively. n indicates the number of values taken by the random variable, and w indicates the width of the window.

The process of the dynamic multi-factor correlation analysis is summarized as follows:

(1) Obtain the length n of the time series data for both CDE and its influencing factors. Set the width w of the sliding window and define the moving step size as n_{Step}. Calculate the number of sliding windows, denoted as n_w. In this case, set $n_w = (n - w)/n_{Step} + 1$.

(2) Let Y represent the CDE time series data. Identify and denote the number of influencing factors as N_i. Initialize the variable $k = 1$ to represent the sequence number of the influencing factors.

(3) Represent the time series of the k-th influencing factor with X. Initialize the variable $i = 1$, which corresponds to a specific sliding window.

(4) Calculate the starting position, denoted as n_{i_Begin} and the ending position, denoted as n_{i_End}, for the i-th sliding window. Here, $n_{i\_Begin} = 1 + (i - 1) \cdot n_{Step}$, and $n_{i\_End} = n_{i\_Begin} + w$.

(5) Apply Equation (2) to calculate the correlation coefficient ρ_{XY_Wi} for the variables X and Y within the interval $[n_{i\_Begin}, n_{i\_End})$.

(6) Increase the value of i by 1. If $i \leq n_w$, go to step (4).

(7) At this step, the correlation curve between the k-th influencing factor and the CDE, which is represented by the time series $[\rho_{XY\_Wi} | i = 1, 2 \ldots n_w]$, has been obtained.

(8) Increase the value of k by 1. If $k \leq N_i$, proceed to step (3).

(9) Finish the process.

Using Tianjin as a case study, the correlation between various influencing factors and CDE was calculated based on Equation (1). The representative results of this calculation are presented in Table 1.

Table 1. Overall correlation values between CDE and different influencing factors in Tianjin.

Influencing Factor	Unit	Comprehensive Correlation Value
Merchandise Sales	billion yuan	0.9704
Primary Plastics Output	ten thousand tons	0.976
Transaction Volume	billion yuan	0.8159
Electricity Generation	billion kWh	0.9387
Number of Markets	-	0.5644
Total Current Assets (Industry Enterprises)	billion yuan	0.9523
Total Wastewater Discharge	ten thousand tons	0.952
Second Industry Value-added Share in GDP	%	−0.4437
Artificial Gas Users	ten thousand people	0.5736
Urban Population	ten thousand people	0.9401
Number of Industry State-owned Enterprises	-	−0.952
Rural Population Share	%	−0.9394
Tertiary Industry Value-added Share in GDP	%	0.566
Primary Industry Value-added Share in GDP	%	−0.9223
Oil Production	ten thousand tons	−0.7574
Green Coverage in Built-up Areas	%	−0.7766
Yarn Production	ten thousand tons	−0.9562
Unemployment Insurance Recipients	ten thousand people	−0.4642
Crude Salt Production	ten thousand tons	−0.883
Cleaning and Maintenance Area	ten thousand sqm	0.8204

Using Equation (2), the dynamic correlation curves for each influencing factor presented in Table 1 were computed individually, as illustrated in Figure 1. The window width employed in these calculations was set to 10 units, with a moving step of 1.

Through a comparative analysis of Table 1 and Figure 1, the following conclusions can be drawn:

(1) Variables that exhibit a high overall correlation in Table 1 may demonstrate relatively low correlation within specific sliding windows illustrated in Figure 1. These variables may exhibit various fluctuation features, such as initially high then low, initially low then high, low in the middle with high values on both sides, or high in the middle with low values on both sides.

(2) Conversely, variables exhibiting low overall correlation in Table 1 may show relatively high correlation within certain sliding windows presented in Figure 1, also displaying diverse features similar to those mentioned above.

Thus, by utilizing Equation (2), the sliding window descriptive method offers a more comprehensive understanding of the intricate correlation characteristics between CDE and its various influencing factors. Within this analytical framework, the correlation between independent and dependent variables is categorized as follows: a correlation coefficient exceeding 0.6 is classified as High ("H"), a coefficient ranging from 0 to 0.6 is considered Low ("L"), a coefficient falling between −0.6 and 0 is labeled as Shallow ("S"), and a coefficient below -0.6 is denoted as Deep ("D").
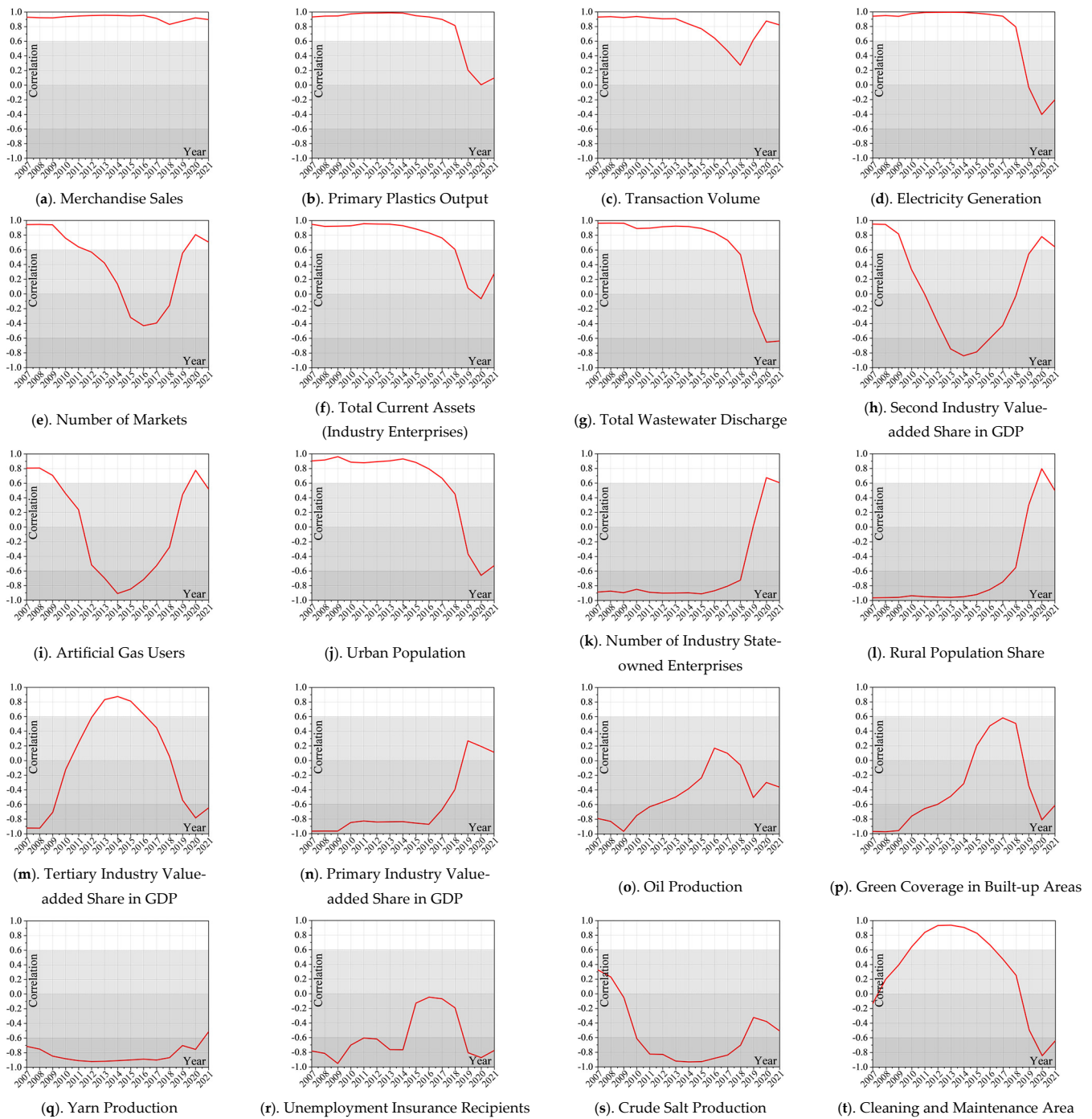
Figure 1. Dynamic correlation curves between CDE and different influencing factors in Tianjin.

Based on these classifications, the correlation features illustrated in Figure 1 can be further represented using the format presented in Table 2. This representation aligns with the "inverted U-shaped relationship between environmental pollution and per capita income" described in reference [9]. In Table 2, the correlation characteristic between "Tertiary Industry Value-added Share in GDP" and CDE in Tianjin is classified as DHD, indicating an inverted U-shaped feature that can be seen in Figure 1m. When Figure 1 and Table 2 are considered together, it becomes evident that the relationship between CDE and its influencing factors also encompasses shapes such as "V" (Figure 1i), "L" (Figure 1s), and other more complex forms. These features are generally summarized as "complexity" and "diversity," and they are manifested through the aforementioned "HLSD" combinations.

This provides a reference for the dynamic description and in-depth analysis of correlation mining, modeling, and visualization.

Table 2. Correlation features between CDE and different influencing factors in Tianjin.

Influencing Factors	Correlation Curves in Figure 1	Correlation Features
Merchandise Sales	Figure 1a	HH
Primary Plastics Output	Figure 1b	HL
Transaction Volume	Figure 1c	HLH
Electricity Generation	Figure 1d	HS
Number of Markets	Figure 1e	HSH
Total Current Assets (Industry Enterprises)	Figure 1f	HSL
Total Wastewater Discharge	Figure 1g	HD
Second Industry Value-added Share in GDP	Figure 1h	HDH
Artificial Gas Users	Figure 1i	HDHL
Urban Population	Figure 1j	HDS
Number of Industry State-owned Enterprises	Figure 1k	DH
Rural Population Share	Figure 1l	DHL
Tertiary Industry Value-added Share in GDP	Figure 1m	DHD
Primary Industry Value-added Share in GDP	Figure 1n	DL
Oil Production	Figure 1o	DLS
Green Coverage in Built-up Areas	Figure 1p	DLD
Yarn Production	Figure 1q	DS
Unemployment Insurance Recipients	Figure 1r	DSD
Crude Salt Production	Figure 1s	LDS
Cleaning and Maintenance Area	Figure 1t	SHD

In practical applications of the dynamic multi-factor correlation analysis method, the step size of the sliding window primarily influences the density of data points along the resulting curve. The width of the sliding window is pivotal in determining the accuracy of the data analysis, which in turn affects its robustness and reproducibility. Here, taking Tianjin as an example, Table 3 presents the correlation features between CDE and different influencing factors across various sliding window widths. The considered sliding window widths include 6, 8, 9, 10, 11, 12, 14, and 16, respectively. Notably, a sliding window width of 10 was utilized for the analysis presented in this article.

As shown in Table 3, when the sliding window width is narrow, the correlation analysis captures features with more frequent fluctuations. This is exemplified by the impact of "Crude Salt Production" on CDE. Specifically, when the sliding window width is set to 6, the correlation features are characterized as "SHDLSL". Consequently, an excessively narrow sliding window width may result in an abundance of features, potentially diminishing the practical significance of the results. Conversely, when the sliding window width is broad, some important correlation features may be overlooked. For instance, considering the influencing factor "Urban Population", when the sliding window width is 10, its impact on CDE is "HDS". However, with a width of 16, the result changes to "HH". It is evident that a wider window may lead to an incomplete capture of certain intermediate process features and distort the view of recent trends, particularly given the limited length of the panel data.

Therefore, when setting the parameters of the sliding window, it is recommended to consider two primary factors. The first is the length of the panel data's time series. When the data length is limited, the sliding window step size should be set to 1 to maximize the use of available data. If the data length is substantial, this value can be appropriately increased to balance computational efficiency and feature capture. The second important factor is the policy-making cycle. In China, major plans are typically formulated every 5 years, accompanied by corresponding policy adjustments. Consequently, in this article, the sliding window width is set at 10, which encompasses two consecutive 5-year plans and the policy adjustments that occur within those periods. When applying this method to

different datasets, this factor should also be taken into account to ensure that the sliding window parameters align with the relevant policy cycles.

Table 3. Correlation features of CDE and different influencing factors in Tianjin under different sliding window widths.

Influencing Factor	Sliding Window Width							
	6	8	9	10	11	12	14	16
Merchandise Sales	HH	HH	HH	HH	HH	HH	HH	HH
Primary Plastics Output	HSH	HL	HSL	HL	HL	HL	HH	HH
Transaction Volume	HLHL	HL	HLH	HLH	HLH	HLH	HL	HL
Electricity Generation	HL	HSL	HS	HS	HS	HS	HH	HH
Number of Markets	HDHS	HSL	HSH	HSH	HSH	HSL	HS	HS
Total Current Assets (Industry Enterprises)	HSL	HSL	HSL	HSL	HS	HL	HH	HH
Total Wastewater Discharge	HDLS	HDS	HD	HD	HD	HS	HH	HH
Second Industry Value-added Share in GDP	HDHS	HDL	HDH	HDH	HDH	HDL	LDS	LDS
Artificial Gas Users	HDHS	HDL	HDL	HDHL	HDH	HDL	HDS	LD
Urban Population	HDS	HDS	HDS	HDS	HD	HS	HH	HH
Number of Industry State-owned Enterprises	DSDHS	DHL	DHL	DH	DH	DL	DD	DD
Rural Population Share	DHS	DHL	DHL	DHL	DH	DL	DD	DD
Tertiary Industry Value-added Share in GDP	DHDL	DHS	DHD	DHD	DHD	DHS	SHL	LHL
Primary Industry Value-added Share in GDP	DHS	DL	DL	DL	DL	DL	DD	DD
Oil Production	SDHDS	DHS	DLS	DLS	DLS	DLS	DL	DS
Green Coverage in Built-up Areas	DHDH	DHS	DHS	DLD	DLD	DLS	DL	DL
Yarn Production	SDL	SDS	DS	DS	DD	DD	DD	DD
Unemployment Insurance Recipients	SDLDL	DS	DSD	DSD	DLD	DLD	DL	DL
Crude Salt Production	SHDLSL	SLDS	LDS	LDS	LDS	LDS	SD	DD
Cleaning and Maintenance Area	SHDL	SHDS	SHD	SHD	LHD	LHS	HL	HH

4. Consistency Analysis of Influencing Factors According to Different Features

The influencing factors of CDE exhibit diverse correlation features across different provinces and cities. In this paper, four provinces and cities in the Bohai Rim region of China, including Tianjin, Hebei, Shandong, and Liaoning, have been selected, and the features of some influencing factors mentioned in previous references are compared, as illustrated in Table 4.

Table 4. Correlation features of influencing factors across different provinces and cities.

Influencing Factors	Correlation Features in Provinces and Cities			
	Tianjin	Hebei	Shandong	Liaoning
Economic Growth (GDP)	HS	HD	HH	HSL
Agricultural Growth	HS	HL	HH	HL
Industrial Growth	HDS	HD	HH	HLH
Service Sector Growth	HD	HD	HH	HL
Population Size	HS	HD	HH	HD
Urbanization Rate	HDS	HD	HH	HLH
Investment in Pollution Control	HDS	HD	HH	HLH
Government Public Expenditure	LHD	HS	HH	HLH
Foreign Direct Investment	HS	HSLS	HSH	LHLH

As illustrated in Table 4, even within the Bohai Rim region, which shares certain similarities in comprehensive structures, such as economy and energy, the correlation features of identical influencing factors can exhibit significant variations across different areas. This observation indicates that generalizing the impact of factors with identical nomenclature across diverse regions and time periods using a single correlation feature is not feasible. For instance, economic growth, represented by GDP, has an influence on CDE. At the end of the last century in the Bohai Rim region, this influence displayed a consistent

H-type correlation feature. However, Shandong consistently maintained a high level of correlation, while Liaoning's correlation gradually weakened. In contrast, Tianjin and Hebei's correlations transitioned from positive to negative over time. Regarding industrial growth's impact on CDE, Tianjin and Hebei exhibited similar correlation characteristics, influenced by regional integration and development initiatives within the Beijing–Tianjin–Hebei region. Both regions demonstrated a gradual transition from positive to negative correlations. In contrast, Shandong, distinguished by its prominent heavy industrial capabilities nationwide, and Liaoning, a pivotal center of heavy industry in Northeast China, consistently exerted a positive influence on CDE through industrial expansion, manifesting HH and HLH correlation features, respectively. The disparate correlation features of the identical influencing factors across various provinces and cities inherently affect the identification of primary CDE drivers, highlighting the importance of this aspect for accurate predictions.

In this study, after accounting for data completeness, the four aforementioned provinces and cities encompass nearly 400 common influencing factors, which are categorized into approximately 50 distinct correlation feature categories. To further quantify the consistency of these influencing factors with varying features among provinces and cities, the Consistency Index of Influencing Factors (CIIF), as illustrated in Equation (3), is defined.

$$CIIF_{i\_AB} = \frac{card(C_{i\_A} \cap C_{i\_B})}{card(C_{i\_B})}, i \in [1, n_C] \tag{3}$$

In Equation (3), $CIIF_{i\_AB}$ represents the CIIF value for the *i*-th correlation feature from province *B* relative to province *A*, where n_C denotes the total number of correlation feature categories and *i* is the index of the correlation feature. The variables C_{i_A} and C_{i_B} represent the sets of influencing factors corresponding to the *i*-th correlation feature for provinces A and B, respectively, and *card*(.) is the counting function for set elements.

The *CIIF* calculation process involves the following steps:

(1) Identify the number (n_C) of shared correlation features between two provinces, such as *A* and *B*, and initialize the variable *i* to 1, where *i* corresponds to a specific feature category;

(2) Calculate the number of influencing factors associated with correlation feature *i* for *A* and *B*, respectively, thus obtaining the values of $card(C_{i\_A})$ and $card(C_{i\_B})$;

(3) Assess the intersection of influencing factors within feature category *i* for both *A* and *B*. The count of these factors is the value of $card(C_{i\_A} \cap C_{i\_B})$, then *CIIF* is calculated using Equation (3);

(4) Increase *i* by 1. If $i \leq n_C$, go to step (2); otherwise, finish the process.

Based on Equation (3) and the above steps, the CIIF was calculated separately for the commonly shared correlation features among the four provinces and cities. This analysis specifically targeted representative features such as DH, DHD, DHS, DL, DLS, DS, HD, HDH, HDL, HDS, HH, HL, HLH, HLHS, HS, HSH, HSL, LDH, LDL, LDS, LHD, LHDS, LHS, LSH, LSHS, SDSD, SHD, and SHSL. The results of these calculations are illustrated in Figure 2.

In each graph of Figure 2, the curve values represent the magnitude of the CIIF based on each correlation feature within a specific province or city. This indicator does not exhibit a clear numerical distribution. From the CIIF curves of the correlation features in each province and city, it is evident that the LSHS feature demonstrates higher values in Tianjin, Liaoning, and Hebei, indicating that the influencing factors associated with this feature are similar across these regions. Conversely, the CIIF values corresponding to the DHS and DL features are zero in all provinces and cities, suggesting that the influencing factors related to these correlation features are entirely distinct.
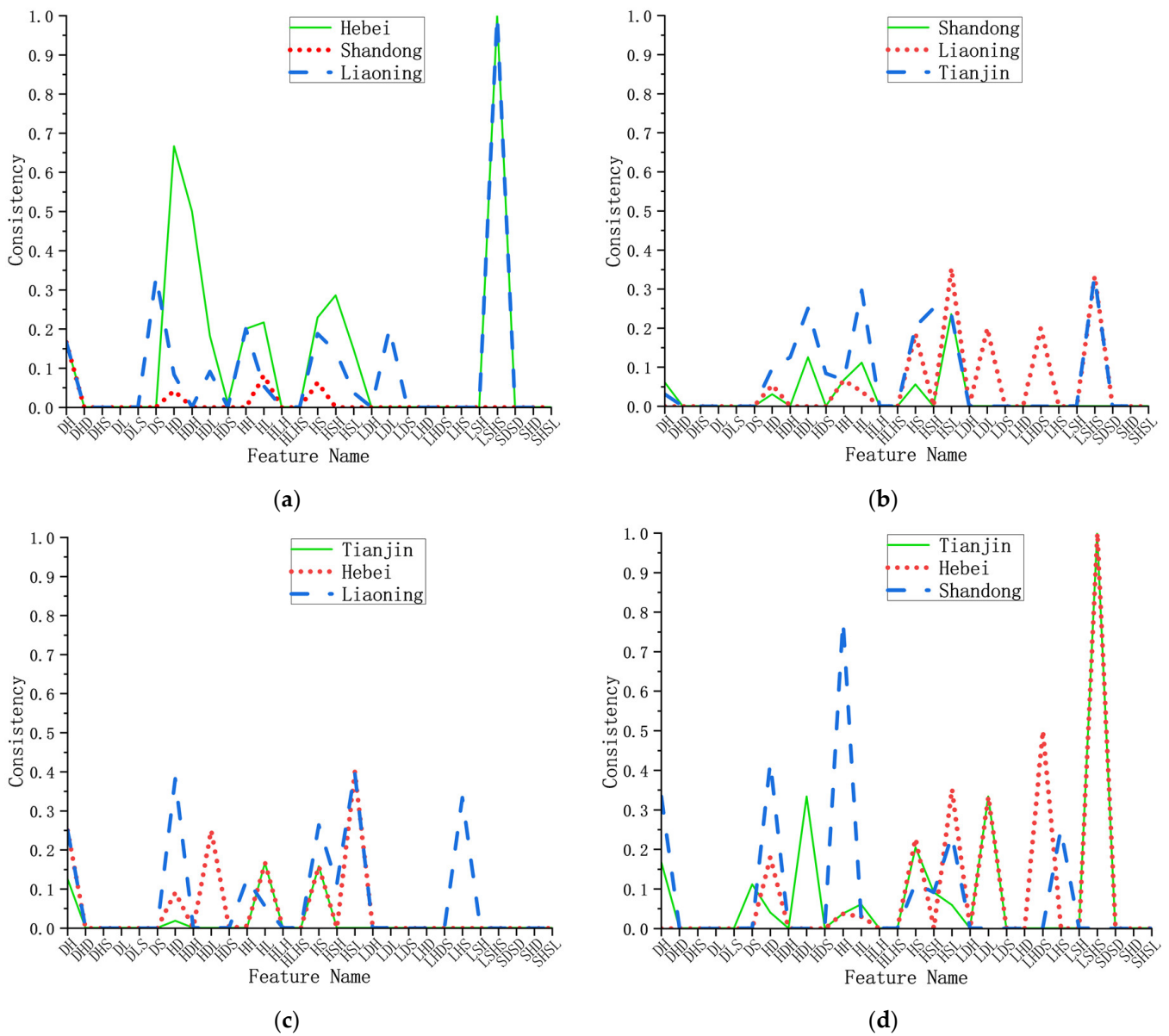
Figure 2. Consistency Index of Influencing Factors (CIIF) of different correlation features across four provinces and cities. (**a**). Consistency of various features of Tianjin with the other three regions. (**b**). Consistency of various features of Hebei with the other three regions. (**c**). Consistency of various features of Shandong with the other three regions. (**d**). Consistency of various features of Liaoning with the other three regions.

In Figure 2a, the CIIF for features such as HD, HDH, and HDL from Hebei relative to Tianjin is non-zero, which consequently indicates that these features will also exhibit non-zero CIIF values for Tianjin in relation to Hebei in Figure 2b. However, the corresponding values do not necessarily equal one another.

From this perspective, the CIIF reflects, to some extent, the transfer capability of various influencing factors on CDE, as manifested through correlation features. It is evident that the relationships between different influencing factors and CDE are complex and exhibit significant variations among provinces and cities, despite potential deviations in the size of effective datasets among them.

5. Comparison of CDE Prediction Accuracy Based on Influencing Factors with Different Features

In this study, we assume that the accuracy of neural network prediction depends on the correlation between the selected factors and the predicted value, which is highlighted in Jebli, I. et al. [39].

Given the diversity of correlated features, predicting CDE inherently presents complex challenges, particularly concerning the identification of key influencing factors. In this section, we evaluate the accuracy of CDE predictions based on influencing factors with distinct features. Taking the aforementioned four provinces and cities as case studies, we conduct a comparative analysis of the prediction accuracy for CDE, employing correlation feature categories of influencing factors as the primary analytical unit. The comparisons include both single and multiple categories of influencing factors.

For the prediction testing, a Long Short-Term Memory (LSTM) neural network was utilized with the following baseline parameters: sequence length = 3, hidden size = 5, learning rate = 0.003, batch size = 4, and number of epochs = 300. Considering the variations in the correlation between influencing factors and CDE across different years, the years 2006, 2011, 2016, and 2021 were selected as test years. This means that the accuracy calculation for each set of influencing factors includes four separate training and testing cycles.

To assess the capacity of influencing factors belonging to distinct correlation feature categories and combinations of multiple categories in predicting CDE accurately, the Accurate Predictive Capability Indicator (APCI) is defined as presented in Equation (4).

$$APCI = \frac{\Sigma_{i=1}^{n_C} TP_i}{N} \times 100\% \tag{4}$$

In Equation (4), TP_i represents the number of samples correctly predicted for the i-th category, while N denotes the total number of samples.

After the CDE prediction, the APCI can be calculated using statistical methods based on a specific prediction accuracy level setting.

In this study, predictions were primarily categorized into two scenarios: first, predictions based on influencing factors exhibiting a single category of correlation features; second, predictions derived from a combination of influencing factors across multiple categories of correlation features. In the former scenario, predictions were made by selecting either one factor or multiple factors (in this case, three) from a single category of correlation features. In the latter scenario, predictions were executed by selecting one factor from each of three entirely distinct categories of correlation features.

5.1. Prediction Based on Influencing Factors with One Correlation Feature Category

Table 5 presents the results of the prediction deviations for CDE using a single influencing factor. For comparison, only categories common to all provinces and cities are included. The maximum prediction deviation and the corresponding year for each category among the aforementioned four provinces and cities are listed. The last column of the table provides the overall deviation for each province and city, representing the maximum value observed within them.

Table 6 presents the APCI values corresponding to the calculation results displayed in Table 5, including both the individual and comprehensive results for the four provinces and cities. The APCI values offer a quantitative assessment of the predictive capability of each influencing factor and their combinations, facilitating a comparison of their effectiveness in accurately predicting CDE.

Table 5. Prediction deviations for a single influencing factor.

Correlation Features	Maximum Deviation Across Provinces and Cities								Overall Maximum Deviation
	Tianjin	Year	Hebei	Year	Shandong	Year	Liaoning	Year	
HH	0.399	2016	0.24	2011	0.167	2016	0.093	2011	0.399
HL	0.422	2016	0.33	2016	0.181	2021	0.182	2016	0.422
HLH	0.345	2016	0.605	2011	0.329	2016	0.175	2016	0.605
HLHS	0.588	2011	0.81	2011	0.242	2016	1.729	2006	1.729
HS	0.246	2006	0.325	2016	0.266	2016	0.34	2011	0.34
HSH	0.421	2021	0.773	2011	0.238	2011	0.387	2011	0.773
HSL	0.315	2006	0.838	2011	0.555	2016	0.266	2016	0.838
HDH	0.48	2006	0.394	2016	0.945	2021	0.247	2016	0.945
HDL	0.505	2016	0.329	2016	0.399	2016	0.109	2021	0.505
HD	0.243	2006	0.195	2006	0.176	2016	0.148	2021	0.243
HDS	0.217	2006	0.323	2016	0.539	2016	0.037	2011	0.539
LHS	0.436	2006	0.592	2011	0.776	2006	0.19	2006	0.776
LHD	0.452	2011	0.516	2011	0.552	2016	0.265	2011	0.552
LHDS	0.341	2006	0.486	2006	2.22	2006	0.431	2011	2.22
LSH	0.647	2006	1.943	2006	0.188	2011	1.163	2006	1.943
LSHS	1.334	2011	1.202	2011	0.584	2011	0.829	2006	1.334
LDH	0.299	2011	1.383	2011	0.429	2011	0.362	2006	1.383
LDL	0.504	2006	1.541	2021	1.417	2006	0.744	2021	1.541
LDS	0.643	2011	1.221	2006	0.166	2016	2.668	2006	2.668
SHSL	0.346	2011	0.979	2011	0.617	2021	1.191	2016	1.191
SHD	0.579	2006	0.534	2006	1.005	2006	0.356	2011	1.005
SDSD	0.409	2016	0.649	2021	0.53	2006	0.391	2006	0.649
DH	0.274	2016	0.246	2016	0.183	2011	0.148	2016	0.274
DHS	0.438	2016	1.324	2006	1.481	2006	0.247	2011	1.481
DHD	2.501	2006	0.763	2006	1.453	2021	0.665	2011	2.501
DL	0.36	2016	0.436	2011	0.375	2016	0.547	2021	0.547
DLS	0.416	2016	0.436	2016	0.4	2016	0.638	2011	0.638
DS	0.419	2016	0.402	2016	0.519	2016	0.11	2006	0.519

Table 6. The Accurate Predictive Capability Indicator (APCI) for predictions based on a single influencing factor.

Prediction Accuracy Threshold	Tianjin	Hebei	Shandong	Liaoning	Comprehensive Value
80%	0%	2%	7.33%	10.56%	5.68%
85%	0%	0%	0%	5%	1.65%
90%	0%	0%	0%	2.22%	0.73%

The prediction deviations for CDE based on a single influencing factor, as illustrated in Table 5, are generally substantial. Factors belonging to different categories exhibit varying degrees of accuracy, and even factors within the same category demonstrate significant performance disparities across different provinces and cities. Specifically, the smallest deviations for Tianjin and Liaoning are observed in the HDS correlation feature category, with values of 0.217 and 0.037, respectively. For Hebei and Shandong, the minimum deviations are found in the HD and LDS categories, with values of 0.195 and 0.166, respectively. Considering the overall maximum deviation across the four provinces and cities, the HD correlation feature category exhibits the smallest deviation at 0.243.

Table 6 presents the APCI values for each province and city, corresponding to varying prediction accuracies. Specifically, at a prediction accuracy threshold of 80%, Tianjin demonstrates an APCI of 0. When the threshold is increased to 85%, both Hebei and Shandong also exhibit an APCI of 0. In contrast, at a prediction accuracy level of 90%, Liaoning achieves an APCI of 2.22%.

It is evident that relying on a single influencing factor is insufficient for achieving universally high-accuracy predictions of CDE. This limitation primarily arises from the

complex interplay of multiple factors affecting CDE, rendering it impossible for any single factor to fully capture their intricate and dynamic variation characteristics. Furthermore, prediction models that depend solely on individual influencing factors tend to exhibit heightened sensitivity to hyperparameters, necessitating stringent conditions to achieve satisfactory prediction accuracy. Consequently, this paper will not further explore the optimization of prediction models.

Table 7 presents the CDE prediction results based on three influencing factors that belong to the same category. These factors have been selected to investigate the potential of combining multiple influencing factors from a single category to enhance the prediction accuracy. The table outlines the prediction results for various combinations of these factors along with their corresponding deviation values. Additionally, Table 8 displays the APCI values associated with the calculation results presented in Table 7.

Table 7. Prediction deviations of three factors derived from single feature category.

Correlation Features	Maximum Deviation across Regions								Overall Maximum Deviation
	Tianjin	Year	Hebei	Year	Shandong	Year	Liaoning	Year	
HH	0.371	2016	0.155	2021	0.266	2016	0.27	2011	0.371
HL	0.193	2016	0.285	2021	0.204	2016	0.403	2021	0.403
HLH	0.354	2016	0.309	2011	0.757	2016	0.116	2021	0.757
HLHS	0.245	2016	0.728	2011	0.478	2011	0.373	2016	0.728
HS	0.262	2016	0.093	2006	0.32	2016	0.4	2021	0.4
HSH	0.105	2016	0.129	2016	0.372	2016	0.433	2011	0.433
HSL	0.233	2016	0.319	2021	0.362	2021	0.241	2006	0.362
HDH	0.273	2016	0.401	2011	0.692	2021	0.182	2016	0.692
HDL	0.197	2021	0.281	2011	0.343	2021	0.651	2021	0.651
HD	0.244	2016	0.135	2011	0.244	2016	0.148	2021	0.244
HDS	0.375	2016	0.188	2016	0.189	2016	0.17	2006	0.375
LHS	0.347	2006	0.24	2021	0.558	2016	0.358	2021	0.558
LHD	0.402	2006	0.329	2006	0.216	2006	0.181	2011	0.402
LHDS	0.391	2006	0.852	2011	0.641	2021	0.293	2021	0.852
LDL	0.226	2021	0.874	2016	0.239	2016	0.486	2006	0.874
LDS	0.467	2011	1.926	2006	0.77	2011	0.599	2011	1.926
SHSL	0.431	2016	0.428	2006	0.386	2016	0.17	2021	0.431
SHD	0.543	2011	0.729	2006	0.391	2006	0.116	2011	0.729
SDSD	0.272	2011	0.408	2011	0.266	2011	0.378	2011	0.408
DH	0.198	2016	0.248	2016	0.182	2016	0.075	2011	0.248
DHS	0.527	2016	0.175	2021	0.162	2016	0.175	2006	0.527
DHD	0.228	2016	0.695	2011	0.261	2016	0.345	2021	0.695
DL	0.193	2016	0.34	2016	0.813	2016	0.604	2021	0.813
DLS	0.314	2016	0.463	2021	0.183	2016	0.368	2021	0.463
DS	0.431	2016	0.211	2011	0.249	2016	0.148	2021	0.431

Table 8. The Accurate Predictive Capability Indicator (APCI) for predictions of three influencing factors derived from the single feature category.

Prediction Accuracy Threshold	Tianjin	Hebei	Shandong	Liaoning	Comprehensive Value
80%	15.15%	25.93%	16.67%	32.35%	22.88%
85%	3.03%	11.11%	0%	17.65%	8.47%
90%	0%	3.7%	0%	2.94%	1.69%

The analysis of Table 7 indicates significant prediction deviation across different provinces and cities, exhibiting similar variations to those observed in Table 5. For Tianjin and Hebei, the minimum deviations are noted in the HSH and HS categories, with values of 0.105 and 0.093, respectively. For Shandong and Liaoning, the lowest deviations are recorded in the DHS and DH categories, with values of 0.162 and 0.075, respectively. When

assessing the overall maximum deviation across the four provinces and cities, the HD correlation feature category consistently demonstrates the smallest deviation at 0.244.

The comparison between Tables 5 and 7 indicates that, despite a reduction in the minimum prediction deviations for each province and city, the correlation feature categories associated with these deviations have changed. Nevertheless, the overall maximum deviation and its corresponding correlation feature category remain unchanged.

As presented in Table 8, when the prediction accuracy threshold is established at 80%, Shandong Province exhibits an APCI of 16.67%, indicating a significant increase. For Tianjin, the APCI rises to 3.03% with a prediction accuracy of 85%. When the prediction accuracy criterion is set at 90%, Hebei and Liaoning provinces demonstrate respective APCI enhancements of 3.7% and 2.94%. Furthermore, the overall APCI across the four provinces and cities also shows improvement.

The comparison of results between Tables 5 and 7, as well as Tables 6 and 8, indicates that utilizing three influencing factors from a single correlation feature category for CDE prediction leads to a reduction in prediction deviations across the four provinces and cities. This suggests that by concentrating on one category of correlation features and incorporating multiple influencing factors, there is potential to enhance the prediction accuracy to a certain degree. However, it also underscores the limitation that a combination of factors within the same correlation feature category cannot fully capture the diversity of CDE variations.

5.2. Combined Prediction Based on Multiple Correlation Feature Categories

Table 9 presents the statistical deviations associated with CDE predictions, derived from a model that incorporates three influencing factors, each originating from distinct categories of correlation features. Given the large number of potential combinations arising from these diverse categories, only the top ten combinations exhibiting the smallest deviations across the provinces and cities are displayed, ensuring consistent representation of these combinations across all provinces and cities. Subsequently, Table 10 illustrates the APCI values corresponding to the computational results presented in Table 9.

Table 9. Prediction deviations of combined influencing factors derived from three distinct feature categories.

Combined Features	Top 10 Minimum Deviations								Overall Maximum Deviation
	Tianjin	Year	Hebei	Year	Shandong	Year	Liaoning	Year	
HDHL, DLS, DS	0.032	2011	0.113	2006	0.147	2016	0.136	2021	0.147
LDHL, LDHL, HLH	0.039	2011	0.284	2011	0.429	2016	0.396	2011	0.429
HDHL, DHS, DS	0.039	2016	0.214	2021	0.137	2016	0.321	2021	0.321
HSH, HSH, HSH	0.047	2006	1.015	2011	0.372	2016	0.433	2011	1.015
HDHL, HDHL, HLH	0.048	2021	0.153	2006	0.981	2016	0.481	2021	0.981
HDHL, HLH, DS	0.051	2006	0.239	2006	0.237	2016	0.081	2011	0.239
SDL, DHS, DHS	0.053	2006	0.587	2006	0.209	2016	0.150	2016	0.587
SHDS, HSH, DL	0.054	2006	0.412	2011	0.164	2006	0.399	2021	0.412
HSH, LDS, DS	0.055	2016	0.776	2021	0.208	2016	0.191	2011	0.776
HSHL, HDHL, DS	0.059	2016	0.301	2006	0.258	2016	0.313	2021	0.313
DHD, LSHD, DLS	0.242	2016	0.052	2006	0.181	2021	0.210	2021	0.242
HS, LHD, DS	0.342	2006	0.054	2016	0.312	2016	0.206	2011	0.342
HSL, LSHD, LSHD	0.210	2016	0.060	2016	0.191	2011	0.170	2006	0.21
DH, HSL, SHS	0.204	2021	0.064	2021	0.196	2016	0.105	2011	0.204
HSL, LSHD, DLS	0.306	2006	0.069	2021	0.184	2006	0.372	2016	0.372
HH, DHL, LSHS	0.152	2021	0.070	2011	0.144	2016	0.585	2021	0.585
HD, HD, HSL	0.281	2016	0.071	2021	0.208	2016	0.166	2016	0.281
HH, LHD, DLS	0.255	2016	0.072	2006	0.323	2011	0.249	2016	0.323
HDS, DHL, DS	0.329	2006	0.072	2016	0.187	2016	0.044	2021	0.329
HS, LHD, DLS	0.249	2016	0.077	2006	0.312	2016	0.376	2016	0.376

Table 9. *Cont.*

Combined Features	Top 10 Minimum Deviations								Overall Maximum Deviation
	Tianjin	Year	Hebei	Year	Shandong	Year	Liaoning	Year	
LH, HSL, LHSL	0.166	2021	0.355	2016	0.052	2011	0.250	2011	0.355
LSH, LHS, LHSL	0.397	2006	0.186	2006	0.035	2016	0.162	2021	0.397
HS, HL, LHSL	0.318	2006	0.286	2006	0.069	2006	0.396	2021	0.396
HSL, HDL, LHD	0.517	2006	0.479	2006	0.077	2021	0.089	2006	0.517
HSL, LHSL, SD	0.340	2006	0.583	2021	0.098	2021	0.330	2011	0.583
HS, HSL, HDL	0.318	2006	0.278	2006	0.095	2011	0.344	2021	0.344
SH, LHSL, SD	0.170	2021	0.247	2021	0.057	2011	0.297	2011	0.297
SH, LSH, HLHS	0.236	2016	0.398	2021	0.236	2011	0.198	2021	0.398
HS, HDL, LHS	0.228	2016	0.362	2006	0.112	2016	0.471	2021	0.471
HSL, LDS, LHSL	0.233	2006	1.286	2021	0.078	2016	0.632	2016	1.286
HL, LHLH, SHD	0.199	2021	0.137	2006	0.165	2016	0.009	2016	0.199
HD, LSH, HLHD	0.179	2016	0.247	2006	0.179	2016	0.025	2011	0.247
HLH, DHLH, DD	0.275	2006	0.175	2006	0.212	2016	0.027	2011	0.275
DH, HDS, LD	0.285	2016	0.234	2016	0.202	2016	0.027	2011	0.285
HL, LD, DHLH	0.297	2016	0.159	2021	0.185	2016	0.028	2021	0.297
DLD, SHL, DHL	0.141	2006	0.122	2016	0.211	2016	0.031	2021	0.211
DSD, LHS, HLHD	0.421	2006	0.444	2021	0.166	2016	0.032	2021	0.444
HH, LSH, LD	0.142	2016	0.253	2006	0.166	2006	0.034	2011	0.253
HL, SHLH, SHLH	0.173	2016	0.202	2006	0.371	2016	0.034	2006	0.371
HSL, HSL, DHLH	0.232	2016	0.185	2006	0.243	2016	0.035	2006	0.243

Table 10. The Accurate Predictive Capability Indicator (APCI) for predictions of combined influencing factors from three distinct feature categories.

Prediction Accuracy Threshold	Tianjin	Hebei	Shandong	Liaoning	Comprehensive Value
80%	21.34%	19.8%	39.40%	37.3%	28.95%
85%	7.7%	6.33%	9.9%	20.8%	5.81%
90%	1.86%	0.97%	1.32%	7.06%	3.42%

The data presented in Table 9 demonstrate that, by incorporating combinations of influencing factors from diverse correlation feature categories, the accuracy of CDE predictions has been significantly enhanced across various provinces and cities, compared to the results shown in Tables 5 and 7. Notably, certain feature combinations achieve high prediction accuracy specifically for Tianjin; however, the overall maximum deviation may still be considerable. This observation is also applicable to combinations that demonstrate effectiveness in other provinces and cities, reinforcing the notion that key influencing factors for CDE predictions differ among various administrative regions. Therefore, a comprehensive analysis tailored to the specific correlation features of influencing factors within each province and city is essential.

Table 10 indicates that at a 90% prediction accuracy threshold, all four provinces and cities exhibit non-zero APCI values, suggesting unique combinations of correlation feature categories that can achieve notably high accuracy in CDE predictions. The relatively modest APCI values reported, both individually and collectively, can be attributed to the extensive range of potential category combinations inherent within each provincial or city context.

In summary, the integration of multiple influencing factors from several distinct correlation feature categories significantly enhances the accuracy of CDE predictions. This suggests that combining various factor categories provides a more comprehensive understanding of the multifaceted drivers behind CDE. Consequently, this approach mitigates the model's sensitivity to hyperparameters, resulting in improved prediction accuracy. However, it is essential to note that not all combinations achieve optimal prediction accuracy; therefore, the analysis of effective combinations must be specifically tailored to different provinces and cities, along with varying categories of influencing factors.

6. Discussion

This study employed a comprehensive dynamic multi-factor correlation analysis methodology to investigate CDE in China's Bohai Rim region. We computed and analyzed the dynamic correlation curves between CDE and various influencing factors, thereby elucidating their intricate interrelationships and dynamic characteristics. To quantify these features, we implemented the CIIF and the APCI indices. Among these indicators, the CIIF serves as a valuable tool for policymakers to assess the applicability of successful emission reduction strategies across various regions. Nevertheless, despite its advantages, the proposed method does present certain limitations. A constructive discussion is provided to address these shortcomings and improve future research methodologies.

(1) Policy Implications and Regional Variations: The determinants influencing the variations in relevant characteristics among provinces and cities are intricate and multifaceted, encompassing factors that may have been elaborated upon in the literature review or remain unaddressed in this paper. Despite the geographical proximity of Tianjin, Hebei, Shandong, and Liaoning within the Bohai Rim region, the correlation features identified in this analysis exhibit significant disparities, as illustrated in Table 4 and Figure 2. In Table 4, with the exception of "Government Public Expenditure" and "Foreign Direct Investment", the dynamic correlations between the other factors and CDE initiate with a high "H" across all provinces. This phenomenon is partially attributable to early regional influences; however, as reforms deepen and economic globalization progresses, diversifications in correlation features subsequently become inevitable. Taking the factor "Economic Growth (GDP)" as an example, the reasons for the differences in its correlation with CDE among different provinces are as follows: the negative correlation observed in Hebei and Tianjin is mainly attributed to the optimization and improvement of industrial structures, policy guidance, and controls on CDE. Additionally, coordinated governance in the Beijing–Tianjin–Hebei area has also played a significant role. In Shandong, the strong correlation is primarily due to its heavy industrial structure, limited variety in its energy consumption mix, and expanding economic scale. The weak correlation in Liaoning Province is mainly due to its heavy industrial structure, reliance on coal-based energy, rapid economic development, and unresolved issues related to pollution emissions. It is therefore recommended that future research efforts undertake a more comprehensive survey encompassing a greater number of provinces and cities nationwide. In doing so, more suitable indicators should be established, referencing the CIIF, to accurately reflect both the synergies and heterogeneities among CDE drivers. Furthermore, the use of clustering methodologies should be considered a viable strategy, and the development of prediction models tailored to specific regions and industrial sectors could significantly enhance their practical applicability.

(2) Multi-Factor Consideration in Policy Formulation: Our findings indicate that prediction models based on multiple influencing factors with diverse correlation features outperform those based on a single factor. This highlights the complexity of CDE dynamics and suggests that policymakers should consider the combined effects of multiple factors when devising emission reduction strategies. The CIIF presented in this study provides a quantitative metric to identify factors that exhibit consistent influence across regions, facilitating the formulation of targeted policies. When devising emission reduction policies, governments and policymakers can leverage the CIIF to prioritize coordinated actions on consistently influential driving factors, thus improving the efficacy of policy execution. Furthermore, insights from the CIIF regarding regional variations in driving factors can optimize resource allocation, enabling a more precise and effective distribution of resources for emission reduction measures. Specifically, Table 4 presents the influencing factors and their correlation characteristics derived from panel data. Tianjin and Hebei both exhibit a downward trend in these characteristics, indicating a need to strengthen the implementation of existing CDE-related policies. Conversely, Shandong demonstrates an upward trend, suggesting that enhancement of corresponding CDE control measures is necessary. In Liaoning, industrial growth, urbanization rate, and investment in pollution control are

strongly correlated with CDE, thereby indicating the need for targeted CDE management strategies that account for the varying stages of industrialization and urbanization.

(3) Implications for Neural Network-Based Prediction Models: This paper establishes the APCI index by evaluating the prediction accuracy of various influencing factors and their combinations through an LSTM neural network. The selection of the LSTM architecture is due to its remarkable memory capabilities, proficiency in managing long-term dependencies within sequential data, and adaptability to dynamic changes in influencing factors. Simultaneously, the limitations of LSTM neural networks have also been recognized: firstly, the model's complexity, characterized by a relatively intricate structure containing a substantial number of parameters, leading to a time-consuming training process; secondly, the risk of overfitting, which occurs when the training data volume is insufficient; and thirdly, parameter sensitivity, as the model's performance is notably sensitive to the choice of hyperparameters. To mitigate these limitations, corresponding measures were adopted during the research. For instance, cross-validation was employed to prevent overfitting, and grid search was utilized to optimize the hyperparameters. To further enhance the prediction performance, future research could integrate this model with other neural network architectures, such as Gated Recurrent Units (GRUs), Convolutional Neural Networks (CNNs), Feedforward Neural Networks (FFNNs), and Transformers, thereby improving both the prediction accuracy and stability. Although different neural network models can significantly influence CDE predictions and the APCI, whether these impacts lead to transformative conclusions depends on several factors, such as model characteristics, task complexity, data quality and attributes, as well as the degree of parameter tuning and optimization. In practical applications, it is advisable to empirically compare the performance of diverse models through cross-validation while developing tailored APCIs for each specific model.

(4) Insights into Factor Interplay and Prediction Reliability: This study elucidates the intricate interplay of factors influencing CDE and their associated processes. It provides valuable insights into the accuracy and reliability of CDE predictions across various application scenarios, which have profound implications for environmental protection, policy formulation, and resource management. Given the multidimensional nature of these influencing factors, the correlation features between them and CDE are classified in practical applications. Next, an APCI index is constructed to analyze the predictive ability of different categories of influencing factors on CDE. Moreover, given the complexity and multidimensionality of the CDE impact mechanism, selecting an appropriate combination of influencing factors is crucial. Ultimately, by acknowledging regional variations in key influencing factors, strategies for mutual learning and transfer learning across prediction models in different provinces and cities can be formulated using the CIIF, enhancing both the learning efficiency and prediction accuracy.

(5) Temporal Trend Analysis and Future Enhancements: The current dynamic correlation analysis model established between influencing factors and CDE primarily reflects instances in which variable values traverse multiple threshold regions; nevertheless, it does not incorporate rigorous temporal trend information. Future research can implement several enhancements. Firstly, additional dynamic feature indicators should be integrated alongside time series analysis techniques to effectively capture long-term trends, short-term fluctuations, and random variations within the data sequences. This approach will provide a more comprehensive understanding of the temporal dynamics inherent in the data. Secondly, by refining the data resolution over time, for example, by changing the data collection frequency from annual to quarterly or monthly intervals, this will enable a more detailed observation of how correlations evolve over time, thereby allowing for a more precise analysis of temporal variations.

Collectively, these improvements will further enhance both the precision of the dynamic correlation analysis and the overall predictive capabilities of the model, particularly in capturing long-term trends and short-term fluctuations in the data.

7. Conclusions

Based on panel data from 1999 to 2021, this study conducted an in-depth analysis of the dynamic correlations between provincial CDE and various influencing factors in the Bohai Rim region of China, including Tianjin, Hebei, Shandong, and Liaoning provinces and cities. The main findings and their implications are summarized as follows:

(1) Limitations of Single Influencing Factor: The study reveals that CDE prediction models relying solely on individual influencing factors often fail to achieve high accuracy. This underscores the complexity of CDE dynamics, where multiple factors interact to drive CDE. Therefore, single-factor models cannot fully capture the complex changes in CDE.

(2) Advantages and Limitations of Combining Similar Factors: While combining multiple factors with similar correlation features can slightly improve the prediction accuracy, this approach remains limited in fully capturing the breadth of CDE variations. In contrast, integrating factors with diverse correlation features offers a more comprehensive solution. This highlights the need for a more diverse set of influencing factors to improve the predictive performance.

(3) Benefits of Integrating Multiple Feature Categories: Integrating multiple types of influencing factors with different correlation features significantly enhances the accuracy of CDE predictions. This approach more comprehensively captures the multifaceted driving mechanisms behind CDE, thus improving the robustness and reliability of the predictions.

Implications for Future Research and Policy:

(1) Methodological Innovation: The dynamic multi-factor correlation analysis method introduced in this study provides a novel perspective and tool for advancing future research on CDE prediction. It enhances the understanding of the intricate mechanisms underlying CDE dynamics and offers a pathway for refining predictive models.

(2) Data Integration and Advanced Techniques: As data accessibility and technological advancements continue to evolve, future research should aim to integrate a broader spectrum of dynamic feature indicators and sophisticated time series analysis methodologies. This will further refine the accuracy and reliability of CDE predictions, enabling more informed decision making.

(3) Policy Implications: The research findings emphasize the importance of considering the cumulative impacts of multiple factors when developing emission reduction strategies. Policymakers should utilize the Consistency Index of Influencing Factors (CIIF) as a quantitative tool to pinpoint factors that consistently impact CDE across regions, thereby enabling the development of targeted and coordinated emission reduction strategies.

In summary, this study not only uncovers the complexity of factors influencing CDE but also provides robust support for optimizing CDE prediction models, scientifically formulating emission reduction policies, refining energy structures, adjusting industrial frameworks, and nurturing technological innovations. The insights gained from this research hold significant potential for guiding future environmental policy and sustainable development initiatives.

Author Contributions: Conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing—original draft, writing—review and editing, Y.Q.; data curation, methodology, software, validation, writing—original draft, writing—review and editing, X.Z.; conceptualization, data curation, formal analysis, investigation, resources, software, writing—review and editing, J.Z.; conceptualization, formal analysis, funding acquisition, project administration, resources, supervision, writing—review and editing, Y.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guan, Y.; Shan, Y.; Huang, Q.; Chen, H.; Wang, D.; Hubacek, K. Assessment to China's Recent Emission Pattern Shifts. *Earth's Future* **2021**, *9*, e2021EF002241. [CrossRef]
2. China Regional Economy Data. Available online: https://www.epsnet.com.cn (accessed on 17 May 2024).
3. China's Provincial Total Apparent Carbon Dioxide Emission Data. Available online: https://www.ceads.net.cn (accessed on 17 May 2024).
4. Shuai, C.; Shen, L.; Jiao, L.; Wu, Y.; Tan, Y. Identifying key impact factors on carbon emission: Evidences from panel and time-series data of 125 countries from 1990 to 2011. *Appl. Energy* **2017**, *187*, 310–325. [CrossRef]
5. Yang, Y.; Zhao, T.; Wang, Y.; Shi, Z. Research on impacts of population-related factors on carbon emissions in Beijing from 1984 to 2012. *Environ. Impact Assess. Rev.* **2015**, *55*, 45–53. [CrossRef]
6. Shao, S.; Liu, J.; Geng, Y.; Miao, Z.; Yang, Y. Uncovering driving factors of carbon emissions from China's mining sector. *Appl. Energy* **2016**, *166*, 220–238. [CrossRef]
7. Abid, M. The close relationship between informal economic growth and carbon emissions in Tunisia since 1980: The (ir)relevance of structural breaks. *Sustain. Cities Soc.* **2015**, *15*, 11–21. [CrossRef]
8. Liao, H.; Cao, H. How does carbon dioxide emission change with the economic development? Statistical experiences from 132 countries. *Glob. Environ. Chang.* **2013**, *23*, 1073–1082. [CrossRef]
9. Mikayilov, J.I.; Galeotti, M.; Hasanov, F.J. The impact of economic growth on CO2 emissions in Azerbaijan. *J. Clean. Prod.* **2018**, *197*, 1558–1572. [CrossRef]
10. Li, W.; Yang, G.; Li, X.; Sun, T.; Wang, J. Cluster analysis of the relationship between carbon dioxide emissions and economic growth. *J. Clean. Prod.* **2019**, *225*, 459–471. [CrossRef]
11. Nie, Y.; Li, Q.; Wang, E.; Zhang, T. Study of the nonlinear relations between economic growth and carbon dioxide emissions in the Eastern, Central and Western regions of China. *J. Clean. Prod.* **2019**, *219*, 713–722. [CrossRef]
12. Wang, W.; Liu, L.; Liao, H.; Wei, Y. Impacts of urbanization on carbon emissions: An empirical analysis from OECD countries. *Energy Policy* **2021**, *151*, 112171. [CrossRef]
13. Wang, Y.; Li, L.; Kubota, J.; Han, R.; Zhu, X.; Lu, G. Does urbanization lead to more carbon emission? Evidence from a panel of BRICS countries. *Appl. Energy* **2016**, *168*, 375–380. [CrossRef]
14. Al-mulali, U.; Che Sab, C.N.B.; Fereidouni, H.G. Exploring the bi-directional long run relationship between urbanization, energy consumption, and carbon dioxide emission. *Energy* **2012**, *46*, 156–167. [CrossRef]
15. Zhang, N.; Yu, K.; Chen, Z. How does urbanization affect carbon dioxide emissions? A cross-country panel data analysis. *Energy Policy* **2017**, *107*, 678–687. [CrossRef]
16. Zhang, Y.; Yi, W.; Li, B. The Impact of Urbanization on Carbon Emission: Empirical Evidence in Beijing. *Energy Procedia* **2015**, *75*, 2963–2968. [CrossRef]
17. Wang, S.; Zeng, J.; Huang, Y.; Shi, C.; Zhan, P. The effects of urbanization on CO2 emissions in the Pearl River Delta: A comprehensive assessment and panel data analysis. *Appl. Energy* **2018**, *228*, 1693–1706. [CrossRef]
18. Abdallh, A.A.; Abugamos, H. A semi-parametric panel data analysis on the urbanisation-carbon emissions nexus for the MENA countries. *Renew. Sustain. Energy Rev.* **2017**, *78*, 1350–1356. [CrossRef]
19. Musah, M.; Kong, Y.; Mensah, I.A.; Antwi, S.K.; Donkor, M. The connection between urbanization and carbon emissions: A panel evidence from West Africa. *Environ. Dev. Sustain.* **2021**, *23*, 11525–11552. [CrossRef]
20. Li, J.; Huang, X.; Kwan, M.; Yang, H.; Chuai, X. The effect of urbanization on carbon dioxide emissions efficiency in the Yangtze River Delta, China. *J. Clean. Prod.* **2018**, *188*, 38–48. [CrossRef]
21. Wang, Y.; Chen, L.; Kubota, J. The relationship between urbanization, energy use and carbon emissions: Evidence from a panel of Association of Southeast Asian Nations (ASEAN) countries. *J. Clean. Prod.* **2016**, *112*, 1368–1374. [CrossRef]
22. Liu, Y.; Tang, L.; Liu, G. Carbon dioxide emissions reduction through technological innovation: Empirical evidence from Chinese provinces. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9543. [CrossRef]
23. Ali, W.; Abdullah, A.; Azam, M. The dynamic linkage between technological innovation and carbon dioxide emissions in Malaysia: An autoregressive distributed lagged bound approach. *Int. J. Energy Econ. Policy* **2016**, *6*, 389–400.
24. Cheng, S.; Meng, L.; Xing, L. Energy technological innovation and carbon emissions mitigation: Evidence from China. *Kybernetes* **2022**, *51*, 982–1008. [CrossRef]
25. Zhang, M.; Li, B.; Yin, S. Is technological innovation effective for energy saving and carbon emissions reduction? Evidence from China. *IEEE Access* **2020**, *8*, 83524–83537. [CrossRef]
26. Erdogan, S. Dynamic nexus between technological innovation and building sector carbon emissions in the BRICS countries. *J. Environ. Manag.* **2021**, *293*, 112780. [CrossRef]
27. Eslamipoor, R.; Sepehriyar, A. Promoting green supply chain under carbon tax, carbon cap and carbon trading policies. *Bus. Strategy Environ.* **2024**, *33*, 4901–4912. [CrossRef]
28. Wu, Y.; Shen, L.; Zhang, Y.; Shuai, C.; Yan, H.; Lou, Y.; Ye, G. A new panel for analyzing the impact factors on carbon emission: A regional perspective in China. *Ecol. Indic.* **2019**, *97*, 260–268. [CrossRef]
29. Khan, M.T.; Imran, M. Unveiling the Carbon Footprint of Europe and Central Asia: Insights into the Impact of Key Factors on CO2 Emissions. *Arch. Soc. Sci. J. Collab. Mem.* **2023**, *1*, 52–66. [CrossRef]

30. Shuai, C.; Chen, X.; Wu, Y.; Tan, Y.; Zhang, Y.; Shen, L. Identifying the key impact factors of carbon emission in China: Results from a largely expanded pool of potential impact factors. *J. Clean. Prod.* **2018**, *175*, 612–623. [CrossRef]

31. Shi, Q.; Chen, J.; Shen, L. Driving factors of the changes in the carbon emissions in the Chinese construction industry. *J. Clean. Prod.* **2017**, *166*, 615–627. [CrossRef]

32. Dong, F.; Hua, Y.; Yu, B. Peak carbon emissions in China: Status, key factors and countermeasures-A literature review. *Sustainability* **2018**, *10*, 2895. [CrossRef]

33. Raihan, A.; Ibrahim, S.; Muhtasim, D.A. Dynamic impacts of economic growth, energy use, tourism, and agricultural productivity on carbon dioxide emissions in Egypt. *World Dev. Sustain.* **2023**, *2*, 100059. [CrossRef]

34. Wang, Z.; Li, Y.P.; Huang, G.H.; Gong, J.W.; Li, Y.F.; Zhang, Q. A factorial-analysis-based Bayesian neural network method for quantifying China's CO_2 emissions under dual-carbon target. *Sci. Total Environ.* **2024**, *920*, 170698. [CrossRef]

35. Chang, L.; Mohsin, M.; Hasnaoui, A.; Taghizadeh-Hesary, F. Exploring carbon dioxide emissions forecasting in China: A policy-oriented perspective using projection pursuit regression and machine learning models. *Technol. Forecast. Soc. Chang.* **2023**, *197*, 122872. [CrossRef]

36. Yang, H.; Wang, M.Z.; Li, G.H. A multi-factor forecasting model for carbon emissions based on decomposition and swarm intelligence optimization. *Measurement* **2023**, *222*, 113554. [CrossRef]

37. Ding, Y.K.; Li, Y.P.; Zheng, H.; Mei, M.Y.; Liu, N. A graph-factor-based random forest model for assessing and predicting carbon emission patterns—Pearl River Delta urban agglomeration. *J. Clean. Prod.* **2024**, *469*, 143220. [CrossRef]

38. Chen, Y.X.; Xie, Y.X.; Dang, X.; Huang, B.; Wu, C.; Jiao, D.L. Spatiotemporal prediction of carbon emissions using a hybrid deep learning model considering temporal and spatial correlations. *Environ. Model. Softw.* **2024**, *172*, 105937. [CrossRef]

39. Jebli, I.; Belouadha, F.Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by pearson correlation using machine learning. *Energy* **2021**, *224*, 120109. [CrossRef]