INTRODUCTION

Due to the “reform and opening-up” policy, economic growth of China has greatly promoted. However, with rapid economic growth, China’s carbon emissions (CE) have threatened the environment. This should be caused by huge energy consumption since Granger causal relationship has been found between economic growth and energy consumption by Wang et al. Overall, the huge energy consumption supports China's economic growth and, in the meanwhile, greatly damages the environment through carbon emissions.

Due to the relationship between CE and energy consumption, China has become both the largest energy consumer and the largest carbon emitter in the world. Many factors affect CE, such as residential building sector, gas consumption, and carbon trading volume and price. Previous studies have discussed their relationship with CE, and some of them developed methods to predict them. Under the pressure of all countries in the world and the importance of sustainable development, the Chinese government promised to reach the carbon emissions peak in 2030 (United Nations Climate Change Summit 2009). Therefore, Chinese energy consumption and CE have received attention all over the world. The future changing trend of China in CE which can directly impact world energy structure and environmental change is a concern.
Accordingly, many researchers are devoted to develop calculation methods to predict CE. We can divide their methods into two categories: time series prediction and regression.

For time series prediction, researchers use CE data as time series and use CE from previous years to forecast CE in the next years. Because many factors which cannot be fully counted can affect CE, this kind of method considers that the historical data of CE contain all the information affecting CE. Therefore, the input and output of this kind of method are only CE data. Gray prediction model (GM) and autoregressive integrated moving average (ARIMA) are used by Hsiao-Tien Pao and Chung-Ming Tsai to forecast CO2 emissions, energy, and gross domestic product (GDP) in Brazil in 2011. ARIMA was also used to predict greenhouse gas (GHG) emission in India by Sen et al. GM and ARIMA are suitable for stationary linear time series, but cannot perform well in nonlinear situations. Then, in 2017, Pao et al. proposed nonlinear gray Bernoulli model (NGBM) to predict same factors in China. They compared their method with ARIMA and GM and found that NGBM performed best among these methods. In addition, different forms of the GM’s discrete equation and the continuous equation determination could not equal accurately. To overcome this problem, discrete gray model (DGM) was developed. GM (1,1) and DGM (1,1) were used to predict CE, renewable energy, and economic growth in Vietnam by Ho et al. Wu et al. developed rolling multivariable gray forecasting model to predict CE in Brazil, Russia, India, China, and South Africa (BRIC nations). In addition, Wenqing Wu et al. also proposed an improved nonhomogeneous gray model to predict CE in BRIC nations. Other methods like extended environmental Kuznets curve (EKC) was also applied into CE prediction.

In summary, traditional methods such as GM and ARIMA are the most common methods to predict CE as time series. However, two problems are existed in this type of method. (1) Novel and powerful methods need to be applied to improve the precision. (2) Applying CE historical data to predict CE results in poor interpretation.

For regression methods, researchers use CE-related variables such as GDP and energy consumption to predict CE. Therefore, the input of this type of method is related variables and its output is CE. Zhou et al. considered residential sector, commercial sector, industrial sector, transport sector, and power sector to regress CE in China. Industry, transport, and buildings sectors are used to regress CE by Gambhir et al. System dynamic model (SDM) was employed to estimate the energy consumption and CE of China in 2008-2020 by Liu et al. GDP and different types of energy were used to regress CE by SDM. Wei Sun and Mohan Liu discussed the impact of three major industries on CE by least-square support-vector machine (LSSVM). Zhu et al. also used LSSVM to predict carbon price in 2017. However, too many support vectors of LSSVM results in poor sparsity and poor accuracy than SVM. Zheng-Xin Wang and De-Jun Ye predicted CE by GDP using nonlinear gray multivariable model. Huang et al. integrated sixteen potential influencing factors by principal component analysis (PCA) and used long short-term memory (LSTM) method to predict CE in China. LSTM has been proved to be very suitable for long series prediction.

In summary, economic factors and energy-related factors are the most common variables to be used to regress CE, which makes the result easily to be understood. However, there are two drawbacks of this type of method. (1) The model built by this method can only explain the relationship between variables and CE, if we want to predict CE in the next few years, we should know the related variables in the next few years firstly. This leads to two types of error: error of related variables and error of model. These two types of error will cause much larger errors than linear superposition. (2) Since we need to know the related variables in the next few years firstly, this kind of method is not real prediction, but a kind of estimation.

To solve the drawback of these two kinds of methods, we developed KLS which is a fusion of Kalman Filter (KF), LSTM, and support vector machine (SVM) to borrow ideas of both time series prediction and regression. Firstly, LSTM is used to predict CE as time series, which can solve the problem of two types of error and fake prediction. Then, SVM is used to regress CE by 14 variables, which can solve the problem of poor interpretation. Finally, KF can be a bridge between time series prediction and regression to integrate results from LSTM and SVM. The final prediction of KF makes the accuracy better than both LSTM and SVM.

2 | METHODS

2.1 | Work frame

The work frame of our method is as Figure 1.

As we can see in the Figure 1, the workflow of KLS can be divided into three steps.

The first step is to predict CE as time series by LSTM. We can obtain the prediction result and error covariance $Q_{\text{LSTM}}$ of LSTM by this step (details in Section 2.2). The second step is to predict CE by other variables using SVM. 14 related variables are prepared at first. Since some of them may not strongly related to CE and multiple collineairities may occur between some variables, ridge regression (RR) is used to selected significantly related variables (details in Section 2.3.1). $P$ variables ($P < 14$) are selected to do regression by SVM. To achieve the aim of prediction, the values of variables in last year would be used to predict CE in the next year. In the end, prediction result and error covariance $R_{\text{SVM}}$ of SVM could be obtained by this step (details in Section 2.3). Finally, KF is used to integrate results from LSTM and SVM. Kalman gain can be obtained.
Time series prediction by LSTM

1. **Build LSTM model**
   - $CE_i = LSTM(\{CE_{i-5}, CE_{i-4}, \ldots, CE_{i-1}\})$

2. **Calculate error covariance** $Q_{LSTM}$

Integrate results by KF

1. **Calculate Kalman gain**
   - $K_i = f(Q_{LSTM}, R_{SVM})$

2. **Update prediction**
   - $CE_{exp} = CE_{LSTM} + K_i(CE_{SVM} - CE_{LSTM})$

3. **Update error covariance** $R_{SVM}$

**FIGURE 1** Workflow of KLS

From $Q_{LSTM}$ and $R_{SVM}$, then can be used to update the final prediction. At last, the whole error covariance of system could be updated. This covariance is critical for calculating Kalman gain in the next step of prediction (details in Section 2.4).

### 2.2 Predict CE by LSTM

Because many classical linear methods are difficult to adapt to multi-input prediction problems, LSTM,\(^{29}\) as a representative algorithm of deep learning in time series prediction, has gradually replaced GM, ARIMA, etc LSTM solved the vanishing gradient problem for RNNs by creating “forget gate.” Now LSTM is a power tool to do time series prediction which is widely used in the field of stock,\(^{30}\) energy,\(^{31}\) and economy.\(^{32}\)

#### 2.2.1 Preprocessing of time series

Firstly, we should encode CE as a time series so that LSTM can use it to do prediction. Formula (1) shows the encoding way:

$$
\begin{bmatrix}
data_1, & \ldots, & data_m \\
data_2, & \ldots, & data_{m+1} \\
\vdots, & \ddots, & \vdots \\
data_n-m, & \ldots, & data_{n-1}
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
data_{m+1} \\
data_{m+2} \\
\vdots \\
data_{n}
\end{bmatrix}
\quad(1)
$$

In this paper, the previous 5 years data are used to predict the data of 6th year, so $m$ should be 5 in formula (1). The previous 5 years data are shown in left side, and target data of 6th year are shown in right side. Since we want to use previous 5 years data to predict the data of 6th year, $m$ should be 5. In addition, $n$ should be the length of whole CE.

#### 2.2.2 Building model by LSTM

Forget gate makes LSTM quite different from normal recurrent neural network (RNN), and it is the key point why LSTM is suitable to predict time series.\(^{33}\) Forget gate can help LSTM “forget” some historical information so that LSTM can focus on the new updated data which contain more useful information. We can understand this by this way, for example, if we
want to predict CE in 2020, the usefulness of CE in 1965 must be less than that in 2019.

LSTM can process data by following formulas:

\[ i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \]  

(2)

Firstly, we should input data. \( i_t \) is the input gate. \( h_{t-1} \) denotes the output of previous step. \( x_t \) is our input data. \( b_i \) and \( W_i \) denote the bias and weight of input gate respectively. \( \sigma() \) is sigmoid function.

\[ f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \]  

(3)

Then, we should put data into forget gate. \( f_t \) is the forget gate and \( W_f \) is the weight of forget gate. \( h_{t-1} \) denotes the output of previous step. \( b_f \) denotes bias of forget gate.

\[ \overline{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \]  

(4)

Then, we can update our previous state. \( \overline{C}_t \) denotes the state of previous step and \( W_c \) denotes its weight and \( b_c \) denotes its bias.

\[ C_t = f_t \cdot C_{t-1} + i_t \cdot \overline{C}_t \]  

(5)

Then, we can obtain our new state by the previous state. Since we can obtain \( i_t, f_t \), and \( \overline{C}_t \) by (2), (3) and (4), we can put them into (5) and obtain the cell state of current moment.

\[ o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \]  

(6)

Finally, we can put the cell state of current moment into output gate. \( o_t \) denotes output gate, and \( W_o \) and \( b_o \) are its weight and bias respectively.

\[ h_t = o_t \cdot \tanh(C_t) \]  

(7)

The last step is to generate output for the next iteration. The structure of LSTM is shown as Table 1.

2.2.3 Calculate error covariance of LSTM

Since we could not know the real error covariance of our prediction result, we could use error covariance of training results to approximate.

The error covariance of training results would be obtained by formula (8).

\[ Q_{LSTM} = \text{std} \left( \frac{\text{LSTM}(\text{train}) - \text{CE}_{\text{train}}}{\text{CE}_{\text{train}}} \right) \]  

(8)

\text{LSTM(train)} \text{ means the training results. It can be obtained by putting training data into LSTM model. \text{CE}_{\text{train}} \text{ denotes the real CE value in the training dataset. std()} \text{ denotes standard deviation function.}

By formula (8), the estimation of error covariance of LSTM \( Q_{LSTM} \) could be obtained.

If the prediction result of LSTM obeys normal distribution, \( Q_{LSTM} \) is the standard deviation of this distribution. In the other way, the prediction result obeys \( N(\text{result}, Q_{LSTM}) \).

2.3 Predict CE by SVM

Since we obtained data from 1965 to 2014, only 50 samples are obtained. SVM is suitable for building model from small samples, so it is chosen to regress CE-related variables to CE.

2.3.1 Select variables by RR

The first step is to select CE-related variables by RR. RR is a typical way to select most significantly relevant variables to do regression since it can effectively avoid multiple collinearity.\textsuperscript{34,35} Although RR has biasedness due to using partial information, the coefficients are more reliable.\textsuperscript{36}

Because we prepared 14 potential variables which may relate to CE and the Pearson correlation coefficients (PCC) between them and CE are all very high with low \( P \)-value, to avoid multiple collinearity and reduce redundant information, RR is used to screen variables.

The biggest difference between RR and normal linear regression is loss function. The loss function of normal linear regression is as following:

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\theta}(x^i) - y^i)^2 \]  

(9)

However, the loss function of RR is as following:

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\theta}(x^i) - y^i)^2 + \lambda \sum_{j=1}^{n} \theta_j^2 \]  

(10)

As we can see, regularization parameter \( \lambda \) is introduced in RR, which makes the coefficients of unimportant variables very small. In this way, we can screen the most significantly relevant variables from 14 variables to do regression.

2.3.2 Construct feature

The most important process of building model is to extract feature. As we mentioned in Section 2.3.1, we prepared 14 variables which are potentially related to CE. RR was used to screen the most significantly CE-related variables. Therefore,
SVM would only use these variables to regress CE. Suppose we obtain \( P \) variables which are significantly related to CE, we take these \( P \) variables as the features of CE. To avoid doing fake prediction, the value of \( P \) variables in the previous year is taken as the feature of CE in the next year. In this way, if we obtain the value of \( P \) variables in this year, we can predict the CE of next year.

The encoding method in formula (11) is as same as the way in formula (1).

### 2.3.3 Building model by SVM

The input of SVM is the feature constructed in the previous section and the output is corresponding CE. Since constructing SVM model needs two parameters, Grid search method with 10-cross validation has been used to select best parameters. Finally, we set the regularization parameter as 1.4, kernel width parameter as 2.9. In addition, radical kernel function is selected. Finally, the model of SVM could be denoted as:

\[
\begin{bmatrix}
V_{1t}, & \ldots, & VP_{t} \\
V_{1t+1}, & \ldots, & VP_{t+1} \\
\vdots & & \vdots \\
V_{1t+m}, & \ldots, & VP_{t+m}
\end{bmatrix} \Rightarrow \begin{bmatrix}
CE_{t+1} \\
CE_{t+2} \\
\vdots \\
CE_{t+m+1}
\end{bmatrix}
\] (11)

The encoding method in formula (11) is as same as the way in formula (1).

2.3.4 Calculate error covariance of SVM

This step shares the same situation as LSTM’s. Error covariance of training results is used to approximate the error covariance of prediction.

\[
R_{SVM} = \text{std} \left( \frac{\text{SVM}(\text{train}) - CE_{\text{train}}}{CE_{\text{train}}} \right)
\] (13)

### 2.4 Integrate results by KF

Kalman Filter can estimate the state of a dynamic system by combining many uncertainties information based on statistical method. It can obtain more accurate estimation by combining the state of previous moment with the measured values.

In our work frame, LSTM can obtain the prediction from the value of historical moments and SVM can obtain the prediction from other measured values. Therefore, the two kinds of important information that KF needs can be obtained, so KF can obtain a more accurate estimation.

2.4.1 State equation and measurement equation of KF

There are two main equations in KF: state equation and measurement equation. In this paper, the model of LSTM could be the state equation and measurement equation is the model of SVM.

Therefore, the two equations should be:

\[
x_k = A x_{k-1} + B u_{k-1} + w_{k-1} \\
= \text{LSTM}(x_{k-1}, x_{k-2}, \ldots, x_{k-5}) + w_{k-1}
\] (14)

\[
z_k = H x_k + v_k = \text{SVM}([V_{1t-1}, V_{2t-1}, \ldots, VP_{t-1}]) + v_k
\] (15)

\( w \) is the uncertainty of the model of LSTM (state equation), and it obeys \( N(0, Q_{\text{LSTM}}) \) and \( v \) is the uncertainty of the model of SVM (measurement equation) which obeys \( N(0, R_{\text{SVM}}) \).
2.4.2 | Obtain final prediction by iteration of KF

There are five steps for KF doing iteration.

Step 1. Input the CE data of previous years and obtain $Q_{\text{LSTM}}$:

$$x_k = \text{LSTM}(x_{k-1}, x_{k-2}, x_{k-3}, x_{k-4}, x_{k-5}) + w_{k-1} \quad (16)$$

Step 2. Obtain prior covariance $P'_k$ by setting initial $P_0$.

$$P'_k = AP_{k-1}A^T + Q \quad (17)$$

Step 3. Kalman gain $K_k$ is obtained by inputting $R_{\text{SVM}}$ and $P'_k$.

$$K_k = P'_kH^T (HP'_kH^T + R)^{-1} \quad (18)$$

Step 4. Obtain the final prediction $\hat{x}_k$ of KF.

$$\hat{x}_k = \hat{x}'_k + K_k (z_k - H\hat{x}'_k) \quad (19)$$

Step 5. Obtain current covariance $P_k$ by $P'_k$ in the step 2.

$$P_k = (I - K_kH)P'_k \quad (20)$$

By these five steps, KF can filter the prediction noise of LSTM and SVM effectively.

Then, after setting up $P_0 = 0.05$, the final prediction of CE could be obtained.

The work frame of KF is as Figure 2.

The noise variance of LSTM and SVM could be used to obtain Kalman gain. Then, the final prediction of CE could be obtained based on Kalman gain and the prediction results of LSTM and SVM. Finally, the covariance could be updated by Q, R and previous covariance. The final covariance would be used in the next round of prediction.

3 | RESULTS

3.1 | Data description

We obtained energy consumption, coal consumption, oil consumption, natural gas consumption, nuclear consumption, Hydroelectricity consumption, other renewable resource consumption, population, primary industry, secondary industry, tertiary industry, GDP, and industrial value-added of each year from “China Statistical Yearbook.” The CE data are obtained from World Bank. Since World Bank only released CE data of China from 1965 to 2014, all the data we used in this paper are from this period.

3.2 | Variable selection

RR coefficients and PCCs between 14 variables and CE are shown in Table 2.

As we can see in Table 2, PCC cannot help us screen variables. However, 10 of 14 variables are screened by RR ($P$-value lower than .05). Because PCC can only test the relevance between variables and CE, although these variables all have strong relationship with CE, these variables have multiple collinearity. In other words, the information some variables offered have already been given by other variables, which makes some variables not that useful anymore.

Since the units of these 14 variables are different, the value of RR coefficients is not important. But the plus and minus sign of RR coefficients are crucial for us to know the contribution of each variables to CE. As our common sense, oil, hydroelectricity, nuclear energy, other renewable resource, and tertiary industry can help us reduce CE, so their RR coefficients are negative. However, the RR coefficient of natural gas is positive, which means consumption of natural gas can increase CE. It is not hard to understand since burning natural gas still causes carbon dioxide production. The reason it is cleaner than oil is...
that natural gas combustion does not produce other pollutants, and the cost of processing and transporting natural gas is lower.

Overall, since the RR $P$-value of energy consumption, coal consumption, natural gas consumption, nuclear consumption, other renewable resource consumption, population, primary industry, secondary industry, GDP, and industrial value-added are lower than 0.05, and these 10 variables are selected to predict CE by SVM.

### 3.3 | Power of KF

In this section, we used data from 1965 to 2005 to train LSTM and SVM model and, then, test the prediction result from 2006 to 2014.

We compared prediction results of LSTM, SVM, and KLS to show the power of KF.

#### TABLE 2
Ridge regression (RR) coefficients and Pearson correlation coefficients (PCCs) between 14 variables and carbon emissions

| Variable                      | RR coefficients | RR $P$-value | PCC | PCC $P$-value |
|-------------------------------|-----------------|--------------|-----|---------------|
| Year                          | 3.752e + 02     | .3449        | 0.90| 2.21e−19      |
| Energy consumption            | 8.962e + 01     | <2e−16       | 0.99| 6.68e−62      |
| Coal consumption              | 1.862e + 02     | <2e−16       | 0.99| 2.19e−52      |
| Oil consumption               | −3.036e + 01    | .6136        | 0.99| 3.86e−42      |
| Natural gas consumption       | 1.248e + 03     | 2.95e−10     | 0.95| 1.54e−25      |
| Hydroelectricity consumption  | −1.022e + 02    | .5360        | 0.98| 7.01e−37      |
| Nuclear consumption           | −2.350e + 03    | .0200        | 0.96| 6.89e−30      |
| Other renewable               | −1.082e + 03    | .0317        | 0.79| 4.50e−12      |
| Population                    | 1.037e + 00     | 3.28e−05     | 0.84| 2.81e−14      |
| Primary industry              | 1.086e + 00     | 0.0377       | 0.98| 4.80e−38      |
| Secondary industry            | 4.817e−01       | <2e−16       | 0.97| 8.83e−32      |
| Tertiary industry             | −9.534e−02      | .1113        | 0.95| 5.58e−27      |
| GDP                           | 9.872e−02       | 1.13e−07     | 0.97| 3.26e−30      |
| Industrial value-added        | 8.704e−02       | 3.06e−07     | 0.97| 3.48e−30      |

*Note: The Bold values are the values significantly related to carbon emissions.*

#### FIGURE 3
Prediction results of long short term memory, support vector machine and KLS

#### TABLE 3
Error indicators for the three methods

| Method  | MSE   | MAE   | MAXE  |
|---------|-------|-------|-------|
| KLS     | 0.0039| 0.061 | −0.089|
| LSTM    | 0.0070| 0.071 | −0.17 |
| SVM     | 0.0081| 0.081 | −0.11 |

*Note: The Bold values are the best model in the test.*

The blue line denotes prediction result of KLS. The red line denotes true value of CE. The black line with points is the prediction result of SVM, and the black dotted line is the prediction result of LSTM.

As we can see in Figure 3, the result of KLS is the combination of results of LSTM and SVM. Although LSTM performed best from 2009 to 2011, KLS performed best in the aspect of the whole period. In order to further compare these three methods, we use the mean square error (MSE), the average absolute
error (MAE), and the maximum error (MAXE) to evaluate these three methods. The results are shown in Table 3.

We can clearly see in Table 3, KLS performed best in all three error indicators.

This experiment shows the power of KF. KF can fix the error of each method based on variances and obtain a more accuracy prediction result.

3.4 Power of KLS

In this section, we compared our method KLS with some common used methods. GM(1,1), nonlinear gray multivariate model (NGMM), ARIMA, and back-propagation artificial neural network (BP-ANN) are used to compare with KLS.

The blue line denotes prediction result of KLS. The red line denotes true value of CE. The black line denotes prediction result of NGMM.26 The line with points is the prediction result of GM (1,1). The prediction result of BP-ANN is shown as dotted line, and ARIMA is shown as dotted line with triangles. The prediction result of Kernel-based nonlinear multivariate gray model (KGM)38 is shown as dotted line with squares.

As we can see in Figure 4, KLS still performed best among these methods. The accuracy of ARIMA is the worst. Table 3 shows the error of these methods.

Table 4 shows the performance of each method. KLS has the smallest error, followed by NGMM, then BP-ANN. ARIMA is the worst.

This experiment shows KLS is an effective method for forecasting CE, and it is better than the previous methods.

3.5 Prediction of Chinese CE in 2030 by KLS

Since we have proved the effectiveness of KLS in the Section 3.4, we used KLS to predict CE of China from 2015 to 2030 in this section.

There are two steps of this prediction. The first step is to estimate CE of China from 2015 to 2019. Because 10 variables selected in the Section 3.2 from 2014 to 2018 can be obtained from “China Statistical Yearbook,” we can input these variables into KLS model directly to estimate the CE of China from 2015 to 2019. The second step is to predict CE of China from 2020 to 2030. Because all the variables from 2019 to 2029 cannot be obtained, we should predict the trend of them firstly and then put them into the model to predict.

Therefore, for the first step, we trained the model by the data from 1965 to 2014 and estimated CE from 2015 to 2019. Because 10 variables, energy consumption, coal consumption, natural gas consumption, nuclear consumption, other renewable resource consumption, population, primary industry, secondary industry, GDP, and industrial value-added of each year, were obtained from “China Statistical Yearbook” and historical CE from 1965 to 2014 were known; therefore, it is easy to achieve the estimation.

Figure 5 shows the estimation results by KLS. The results showed that CE leveled off from 2014 to 2016, and then, it gradually increased from 2017.

The next step is to predict CE of China from 2020 to 2030. Since we cannot obtain 10 variables from 2019 to 2029, we
summarized their trends from 2010 to 2018 and predicted their trends in three situations: optimistic situation, normal situation, and pessimistic situation.

As shown in Figure 6, the annual growth rate of all variables does not change much, except for ‘other renewable resource’. Based on the boxplot, first quartile, median, and third quartile were used to describe the trend of these variables from 2019 to 2029.

Table 5 shows the predicted growth trend of 10 variables from 2019 to 2029. However, the growth of some variables promotes the increase in CE, while the growth of nuclear consumption and other renewable resource consumption help to reduce...
CE. Therefore, for variables other than nuclear consumption and other renewable resource consumption, the optimistic growth rate should be less than pessimistic growth rate, that is, the values of pessimistic situation and optimistic situation should be reversed.

Finally, based on the predicted growth rate, the value of these 10 variables from 2019 to 2029 could be predicted. Then, KLS was used to predict the CE of China from 2020 to 2030.

As shown in Figure 7, we predicted CE of China from 2015 to 2030 in three situations. The red line is the prediction results in normal situation. The black line is the results in pessimistic situation, and green line denotes the optimistic situation. The prediction results showed the CE peak would be reached in about 2024 and then CE would gradually decrease, but before that, CE in China would keep increasing.

4 | CONCLUSIONS

China has become the world's highest carbon emitter, and its impact on the environment has received much attention. Forecasting China's CE can show the energy demand and environmental impact in advance, which is conducive to the improvement of low-carbon development policies.

Many researchers have developed methods to forecast CE. Due to the complex nonlinear relationship between CE and other variables, the accuracy of these methods is unsatisfactory. In addition, most of these methods do not have the ability to predict CE. They can only predict CE by obtaining corresponding variables in advance. Most scholars first roughly estimate the corresponding variables and then bring them into the regression model to predict CE. This results in the secondary transmission of errors, which directly affects the accuracy of the results.

To solve these problems, KLS is proposed in this paper. KLS combines the advantages of time series prediction and regression and perfectly integrates the two kinds of methods in the way of statistical filtering. LSTM is used to predict CE as time series and SVM is used to predict CE by 10 relevant variables. Finally, KF fixes the results of LSTM and SVM, and gives a more accurate results. 10 relevant variables are selected by RR, which effectively avoid the problem of multicollinearity. In addition, to achieve the real prediction by regression, 10 variables of this year are used to regress the CE of next year. In this way, we can predict the CE in 2015 by the ten variables of 2014.

Then, two experiments were completed to show the superiority of KLS. In these experiments, we trained KLS model by CE from 1965 to 2005 and tested the model by the data from 2006 to 2014. The first experiment compared KLS with LSTM and SVM. The result showed that using KF to fuse these two methods can achieve better accuracy than using only one of them. The other experiment compared KLS with GM (1,1), ARIMA, BP-ANN, and NGMM. The result showed that KLS performed best among these method. Overall, these experiments proved that KLS is very suitable for forecasting CE.

After verifying the effectiveness of KLS, we used it to predict CE of China in 2030. Since we had the CE data from 1965 to 2014 and 10 variables’ value from 2014 to 2018, we predicted the CE from 2015 to 2019 in the first step. In the next step, we firstly estimated the 10 variables from 2019 to 2029 based on the historical trend of these 10 variables. We estimated these 10 variables in three situations: pessimistic situation, normal situation, and optimistic situation. Then, we put these variables into KLS model to predict CE from 2020 to 2030. Finally, we obtained the CE prediction results in three situations. The results show that CE in China will keep increasing and then reach the peak in 2024 and then gradually decrease.

ACKNOWLEDGMENT

This work was supported by the National Social Science Foundation of China (No: 16CJY074).
REFERENCES

1. Wang S, Fang C, Wang Y, Huang Y, Ma H. Quantifying the relationship between urban development intensity and carbon dioxide emissions using a panel data analysis. *Ecol Ind.* 2015;49:121-131.

2. Wang S, Li Q, Fang C, Zhou C. The relationship between economic growth, energy consumption, and CO2 emissions: empirical evidence from China. *Sci Total Environ.* 2016;542:360-371.

3. Nejat P, Jomehzadeh F, Taheri MM, Gohari M, Majid MZA. A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). *Renew Sustain Energy Rev.* 2015;43:843-862.

4. Ma M, Ma X, Cai W, Cai W. Carbon-dioxide mitigation in the residential building sector: a household scale-based assessment. *Energy Convers Manage.* 2019;198:111915.

5. Lu H, Behbahani SS, Azimi M, Matthews J, Han S, Iseley T. Trenchless construction technologies for oil and gas pipelines: State-of-the-Art Review. *J Construct Eng Manage.* 2019;127(6):476-484.

6. Lu H, Ma A, Azimi M. US natural gas consumption prediction using an improved kernel-based nonlinear extension of the Arps decline model. *Energy.* 2020;194:116905.

7. Lu H, Ma X, Huang K, Azimi M. Carbon trading volume and price forecasting in China using multiple machine learning models. *J Clean Prod.* 2020;249:119386.

8. Lu CC, Chen X, Hsieh CL, Chou KW. Dynamic energy efficiency of slack-based measure in high-income economies. *Energy Sci Eng.* 2019;7:943-961.

9. Esbati S, Amooie MA, Sadeghzadeh M, Ahmadi MH, Pourfayaz F, Ming T. Investigating the effect of using PCM in building materials for energy saving: case study of Sharif Energy Research Institute. *Energy Sci Eng.* 2019;1:1-14. https://doi.org/10.1002/esc3.328

10. Li Y. Prediction of energy consumption: variable regression or time series? A case in China. *Energy Sci Eng.* 2019;7:2510-2518.

11. Martínez-Rodríguez D, Colmenar JM, Hidalgo JJ, Villanueva Micó RJ, Salcedo-Sanz S. Particle swarm grammatical evolution for energy demand estimation. *Energy Sci Eng.* 2019;1:1-12.

12. Pao H-T, Tsai C-M. Modeling and forecasting the CO2 emissions, energy consumption, and economic growth in Brazil. *Energy.* 2011;36:2450-2458.

13. Sen P, Roy M, Pal P. Application of ARIMA for forecasting energy consumption and GHG emission: a case study of an Indian pig iron manufacturing organization. *Energy.* 2016;116:1031-1038.

14. Pao H-T, Fu H-C, Tseng C-L. Forecasting of CO2 emissions, energy consumption and economic growth in China using an improved grey model. *Energy.* 2012;40:400-409.

15. Xie N-M, Liu S-F. Discrete grey forecasting model and its optimization. *Appl Math Model.* 2009;33:1173-1186.

16. Ho H-XT. Forecasting of CO2 emissions, renewable energy consumption and economic growth in Vietnam using grey models. In: *2018 4th International Conference on Green Technology and Sustainable Development (GTSD).* IEEE, 2018:452-455.

17. Wu L, Liu S, Liu D, Fang Z, Xu H. Modelling and forecasting CO2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. *Energy.* 2015;79:489-495.

18. Wu W, Ma X, Zhang Y, Li W, Wang Y. A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries. *Sci Total Environ.* 2020;707:135447.

19. Pérez-Suárez R, López-Menéndez AJ. Growing green? Forecasting CO2 emissions with environmental Kuznets curves and logistic growth models. *Environ Sci Policy.* 2015;54:428-437.

20. Pan WL, Pan W, Liu S, Tsai SB, Hu C, Tsu H. China’s provincial energy-related carbon emissions-economy nexus: A two-stage framework based on decoupling analysis and panel vector autoregression. *Energy Sci Eng.* 2019;7:1201-1213.

21. Zhou N, Fridley D, Khanna NZ, Ke J, McNeil M, Levine M. China’s energy and emissions outlook to 2050: perspectives from bottom-up energy end-use model. *Energy Pol.* 2013;53:51-62.

22. Gambari A, Schulz N, Napp T, et al. A hybrid modelling approach to develop scenarios for China’s carbon dioxide emissions to 2050. *Energy Pol.* 2013;59:614-632.

23. Liu X, Mao G, Ren J, Li RYM, Guo J, Zhang L. How might China achieve its 2020 emissions target? A scenario analysis of energy consumption and CO2 emissions using the system dynamics model. *J Clean Prod.* 2015;103:401-410.

24. Sun W, Liu M. Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China. *J Clean Prod.* 2016;122:144-153.

25. Zhu B, Han D, Wang P, Wu Z, Zhang T, Wei Y-M. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Appl Energy.* 2017;191:521-530.

26. Wang Z-X, Ye D-J. Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. *J Clean Prod.* 2017;142:600-612.

27. Huang Y, Shen L, Liu H. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. *J Clean Prod.* 2019;209:415-423.

28. Duan Y, Lv Y, Wang F-Y. Travel time prediction with LSTM neural network. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).* IEEE; 2016:1053-1058.

29. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks.* 2005;18:602-610.

30. Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: A case study of China stock market, Big Data (Big Data). In: *2015 IEEE International Conference on.* IEEE; 2015:2823-2824.

31. Gensler A, Henze J, Sick B, Raabe N. Deep Learning for solar power forecasting—an approach using AutoEncoder and LSTM Neural Networks, Systems, Man, and Cybernetics (SMC). In: *2016 IEEE International Conference on.* IEEE; 2016:2858-2865.

32. Ameyaw B, Yao L. Analyzing the impact of GDP on CO2 emissions and forecasting Africa's total CO2 emissions with non-assumption driven bidirectional long short-term memory. *Sustainability.* 2018;10:3110.

33. Zhao T, Cheng L, Zang T, Hu Y. Peptide-major histocompatibility complex class I binding prediction based on deep learning with novel feature. *Front Genet.* 2019;10:1-8.

34. Guilkey DK, Murphy JL. Directed ridge regression techniques in cases of multicollinearity. *J Am Stat Assoc.* 1975;70:769-775.

35. Kasarda JD, Shih W-FP. Optimal bias in ridge regression approaches to multicollinearity. *Social Method Res.* 1977;5:461-470.
36. Zhao T, Hu Y, Zang T, Cheng L. Identifying Alzheimer’s disease-related proteins by LRRGD. *BMC Bioinform.* 2019;20:570.

37. Zhang F, Zhang M, Hansen JA. Coupling ensemble Kalman filter with four-dimensional variational data assimilation [J]. *Advances in Atmospheric Sciences.* 2009;26(1):1-8.

38. Ma X, Liu Z-B. The kernel-based nonlinear multivariate grey model. *Appl Math Model.* 2018;56:217-238.

**How to cite this article:** Li Y. Forecasting Chinese carbon emissions based on a novel time series prediction method. *Energy Sci Eng.* 2020;8:2274–2285. [https://doi.org/10.1002/ese3.662](https://doi.org/10.1002/ese3.662)