ENERGY CONSUMPTION PREDICTION OF CEMENT CALCINATION PROCESS BASED ON DUAL-CHANNEL TEMPORAL CONVOLUTION NEURAL

by

Zijian WANG and Hongtao KAO*

College of Materials Science and Engineering, Nanjing Tech University, Nanjing, China

Original scientific paper
https://doi.org/10.2298/TSCI240318168W

The cement industry has consistently consumed large amounts of coal and electricity resources. Optimizing energy scheduling and production process control can typically save energy and improve production efficiency. Therefore, the prediction of energy consumption holds great significance in the cement industry and other energy-intensive sectors. However, predicting energy costs is challenging due to multiple production factors, variable coupling, and time lags. In this research, we proposed the use of a dual-channel temporal convolution neural network to forecast coal and electricity consumption in the cement calcination process for the upcoming production hour. Additionally, we employ the Spearman correlation coefficient method to select variables for the calcination system, aiming to reduce feature data dimensions and improve model training efficiency. To address parameter redundancy and mitigate the risk of overfitting, we devise a dual-channel structure. For comparison, we utilized various models including recurrent neural network, gate recurrent unit, long short-term memory, convolutional neural network, and back-propagation in prediction experiments using actual cement calcination process energy consumption data. The results indicated that with a kernel size of 13, dilation rates of [2^1, 2^2, …, 2^6] and a filter size of 36, the temporal convolution neural model achieves an accuracy of 97.65%. Relative to other models, the temporal convolution neural model achieved a reduction of at least 40% and 24% in the mean squared error for coal and electricity consumption forecasts, respectively, meeting the expected requirements.

Key word: *energy consumption prediction, temporal convolution neural, cement clinker production, spearman coefficient*

Introduction

Over a long-term, the cement manufacturing industry, as a high energy consumption industry, should prioritize both high-quality production and energy conservation with emission reduction. This includes optimizing coal and electricity consumption reduction. Predicting energy consumption in advance based on characteristic variables of the cement calcination process not only provides a reference for energy allocation to reduce consumption but also informs engineers on plant production improvements. This is vital for achieving energy efficiency, emission reduction, and effective energy scheduling [1].

*Corresponding author, e-mail: kaoht@163.com

Neural network models are widely used in the fields of machine learning and artificial intelligence. Due to the diverse needs of various application scenarios, different kinds of models have emerged. Representative models include RNN, LSTM, GRU, BPNN, CNN, among others. The BPNN is widely used in electric load forecasting [2] and various industrial equipment, such as centrifugal pump [3], motors [4], and bearings [5]. Its structure can be flexibly adjusted to model complex nonlinear relationships. The RNN excels in sequential prediction tasks, applicable to natural language processing [6], stock price forecasting [7], and speech recognition [8]. Its design retains past information within its internal state, an essential feature for processing sequence data. The LSTM, a variant of RNN, addresses the issues of vanishing or exploding gradients with its unique structure [9], making it suitable for industrial energy consumption forecasting [10], disaster monitoring [11], and traffic prediction [12]. The GRU, a streamlined version of RNN, offers a simpler structure and more efficient computations than LSTM, holding promising applications in fault diagnosis [13] and automatic control [14]. The CNN, one of the most popular neural network models, is primarily used in visual recognition [15], including autonomous driving [16], and the medical field [17], specializing in capturing local features through convolutional computations.

The cement calcination process contains various physicochemical reactions as well as process variables [18]. Physicochemical reactions such as carbonate decomposition, fuel combustion, and solid-phase raw material reactions occur. The neural network structure proposed in this study can efficiently learn the intrinsic relationships between these physicochemical processes and energy consumption [19], thus enhancing computational efficiency [20]. Process variables like raw material feed rate, secondary air temperature, and outlet temperature of the decomposition furnace are essential. Spearman's correlation coefficient method is used to eliminate redundant and low-correlation features, thereby improving model robustness.

In the domain of thermal engineering, ANN have demonstrated significant predictive capabilities. Kocyigit and Bulgurcu [21] enhanced the accuracy of predicting the overall heat transfer coefficient in coaxial double-pipe heat exchangers, by integrating curve fitting with ANN techniques. This advancement illustrates that ANN can be effectively utilized in optimizing and simulating thermal efficiency, offering innovative approaches and methodologies for industrial applications.

In cement manufacturing, the temperature of the rotary kiln firing zone significantly influences clinker quality. Shi *et al*. [22] predicted the energy consumption of the cement clinker firing process by establishing a sliding window-based CNN dual-channel (SWDC-CNN) prediction model. The results indicate that the accuracy of energy consumption prediction using this structure is 94.75%. Meanwhile, the TCN model employed in this study achieves an accuracy of 97.65%, suggesting that dilated convolution might be more effective than the sliding window technique. It is important to note that inappropriate window sizes may hinder predictions. For further comparison, the CNN model is used in this study as a benchmark against other models.

Zheng *et al*. [23] used an improved RNN structure with an attention mechanism to model the rotary kiln and achieved prediction accuracy of 91.6% for the residence time in the calcination process, which indicates desirable results. Since RNN models are better suited for processing short sequence data, this study considers it as a comparative model. Liu *et al*. [24] developed a rotary kiln temperature prediction model based on the CNN-BILSTM-OC model by combining CNN and LSTM. This model possesses memory functions and feature extraction capabilities, which are important for time series prediction. The remarkable accuracy of 99.8% highlights the importance of feature extraction in this context. Similarly, the TCN model

enhances these predictions through its dilated convolutions, which also play a significant role in feature extraction.

Liu *et al*. [25] integrated the spatial attention mechanism with LSTM for predicting the electricity consumption of cement raw mill systems. The model achieved an R^2 prediction accuracy of 0.908, facilitating optimal energy consumption realization over time. The LSTM has limitations in capturing essential information when dealing with highly coupled data. Therefore, in this study, LSTM is used as a comparative model. Ali *et al*. [26] used BPNN to forecast power for waste heat recovery systems in cement plants, achieving an impressive accuracy of 99.9%. This serves as an alternative to thermodynamic modeling , thereby avoiding extensive computations.

The TCN, a relatively new and effective neural network architecture, combines the advantages of RNN and CNN to offer high flexibility and has been widely employed in various fields. For example, it achieves an error rate that is 5.27% lower than the LSTM model in energy prediction [27]. It also has a predictive accuracy of 94.1% in motion recognition [28]. Additionally, it has been applied to industrial data forecasting [29], and lifespan prediction [30]. The TCN structure excels in preventing gradient explosion, and its extensive receptive field effectively captures long-term dependencies. These characteristics make it highly suitable for the cement manufacturing industry [31].

Whether it's RNN, LSTM, or sliding window techniques, there might be inherent structural issues such as exploding gradients, vanishing gradients, and insufficient memory history, particularly in the context of energy consumption in cement manufacturing. To address these deficiencies, the TCN structure can solve the problem of short memory history by increasing dilated convolutions, and prevent exploding and vanishing gradients through residual connections. Therefore, TCN demonstrates significant potential for application in cement manufacturing, suggesting that it can provide robust technological support for optimizing and predicting the manufacturing process, thereby enhancing production efficiency and stability.

Analysis of cement clinker calcination system and selection of characteristic variables

As illustrated in fig. 1, once the preparation of cement raw material is completed, it is conveyed to the kiln tail suspended preheater and decomposition furnace.

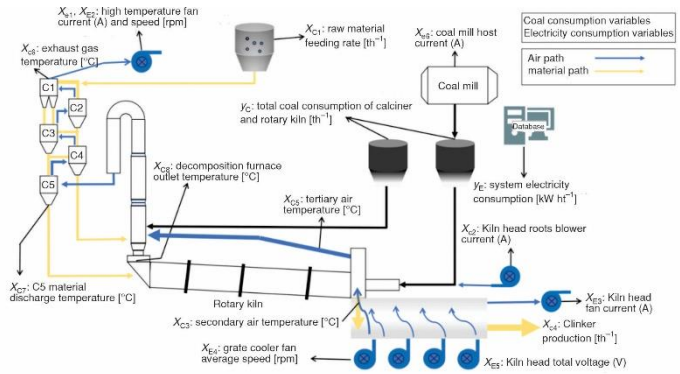

Figure 1. Flow chart of cement clinker calcination system

The high temperature material, after preheating and decomposition, enters the rotary kiln for calcination into clinker. After cooling by the grate cooler, the clinker is then delivered

to the clinker storage. The fuel coal, dried and grounded by the coal mill, is stored in the form of pulverized coal in the pulverized coal silo and then supplied to the decomposition furnace and rotary kiln. Coal consumption is mainly concentrated during the calcination process of the calciner and kiln, while electricity consumption is mainly concentrated in the kiln's main motor, high-temperature fan, coal mill, exhaust fan, and other components.

The variation in energy costs in the cement clinker calcination system is influenced by numerous factor variables, which must be selected based on experience and theory. Data labels and ranges, including maximum, minimum, average values, and standard deviations, are shown in tab. 1. It can be concluded that the relative dispersion of coal consumption is higher than that of electricity consumption, indicating that coal consumption prediction is more challenging.

Table 1. Training data details

Input Variables	Maximum	Minimum	Average	Standard deviation
High temperature fan current, X_{E1} [A]	138.00	62.00	105.63	11.76
High temperature fan speed, X_{E2} [rpm]	915.00	741.00	864.01	35.41
Kiln head fan current, X_{E3} [A]	2176.00	1120.00	1852.60	213.01
Grate cooler fan average speed, X_{E4} [rpm]	1040.00	592.00	804.73	94.71
Kiln head total voltage, X_{E5} [V]	4176.00	1632.00	3227.69	504.07
Coal mill hosts current, X_{E6} [A]	7.49	0.15	5.35	2.20
Raw material feeding rate, X_{C1} [th^{-1}]	569.00	257.00	358.70	26.04
Kiln head roots blower current, X_{C2} [A]	30.80	14.90	22.66	2.54
Secondary air temperature, X_{C3} [°C]	1249.00	1232.50	1240.96	3.28
Clinker production, X_{C4} [th^{-1}]	263.00	167.00	235.53	16.49
Tertiary air temperature, X_{C5} [°C]	1200.00	869.00	987.48	45.56
Exhaust gas temperature, X_{C6} [°C]	374.00	257.00	298.26	19.76
C5 material discharge temperature, X_{C7} [°C]	1091.00	825.00	889.72	38.15
Decomposition furnace outlet temperature, X_{C8} [°C]	129.10	73.39	99.60	4.11
System electricity consumption, y_E [kWht^{-1}]	7222.00	4158.00	6260.72	605.02
Total coal consumption of calciner and rotary kiln, y_C [th^{-1}]	35.00	22.50	29.00	2.97

In this section, the Spearman rank correlation coefficient method is employed to validate and select the original feature variables. This method, a non-parametric statistical approach, calculates the correlation between two variables based on their ranks, reflecting the monotonic relationship between them. The formula is [32]:

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \tag{1}$$

where ρ is the Spearman's value, n – the number of samples, and d_i – the difference between the i^{th} data pair rank. The range of Spearman's values range from –1 to 1, closer to 1 (or –1) indicate a higher degree of positive (or negative) correlation.

After obtaining production data from the cement plant and removing null values and outliers, the correlation coefficients between the feature variables and coal consumption, as well as power consumption, are calculated using eq. (1). The heatmap generated from these coefficients is shown in fig. 2. A feature can be chosen as an input feature to the model when the coefficient exceeds 0.2, indicating a certain level of correlation between the two variables.

From fig. 2(a), it can be concluded that the fluctuation in electricity consumption most closely resembles that of the total kiln head voltage X_{E5}, which includes the voltages of both the rotary kiln and the grate cooler. This is notable since the rotary kiln and grate cooler are the major consumers of electricity in actual production. Figure 2(b) shows that the secondary air temperature X_{C3} is closely associated with the fluctuation in coal consumption, with a Spearman rank correlation coefficient of 0.9, indicating a very strong correlation. The secondary air is the air after the grate cooler cools the clinker. When the consumption of coal increases, the temperature of the clinker from the rotary kiln increases, and consequently, the temperature of the secondary air also increases. Upon comparison between the charts in figs. 2(a) and 2(b), it is observed that the correlation between coal consumption and its features is generally lower than that between power consumption and its features. This is attributed to the complexity of coal consumption in the rotary kiln and decomposition furnace, where various variables interact, exhibiting a stronger degree of coupling compared to power consumption. The Spearman rank correlation coefficient is advantageous for capturing fluctuation patterns, which makes it particularly effective at identifying nonlinear relationships and mitigating the impact of noise on model performance.

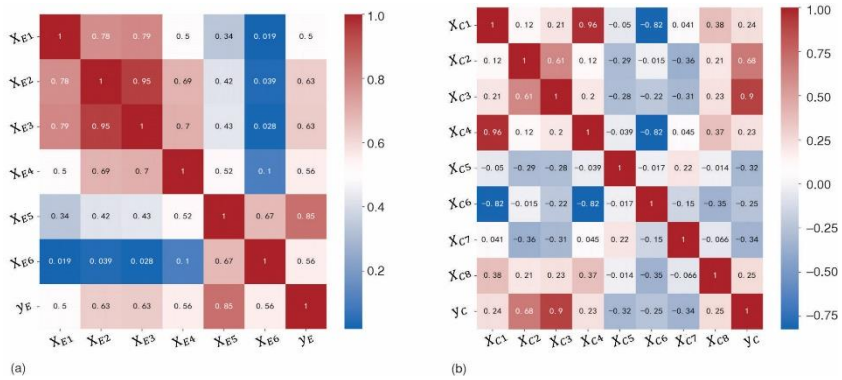

Figure 2. Spearman correlation analysis between process variables; (a) electricity variables analysis and (b) coal variables analysis

Construction of the TCN forecasting model

Previous section outlines significant challenges in predicting energy consumption of cement production. These challenges include complex physicochemical reactions, redundant process parameters, temporality, and non-linearity. The dilated convolution computation and residual block structure of the TCN model can effectively address these challenges. The model operates with the following inputs:

$$X = \{x_1, x_2, x_3, ..., x_t\} \varepsilon R_{t\times 14},\ x_t = (x_{E1}, x_{E2}, ..., x_{C1}, x_{C2}, ...) \tag{2}$$

where x_t is the feature values matrix at a certain time t. The prediction matrix of the model is then:

$$\tilde{Y}_t = \{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, ..., \tilde{y}_t\} \varepsilon R_{t\times 2},\ \tilde{y}_t = (\tilde{y}_C, \tilde{y}_E) \tag{3}$$

where $\tilde{y}_t$ is the output values matrix at a certain moment t. That is, there is a function F such that:

$$\tilde{Y}_t = F(x_1, x_2, x_3, ..., x_t) \tag{4}$$

The task of the neural network model is to search for underlying patterns between X_t and the true value Y_t o that the loss between the predicted value $\tilde{Y}_t$ and Y_t is minimized to optimize the prediction of the model.

In this section, TCN is employed to construct a model for predicting the clinker burning energy consumption in a cement plant in Guangxi, China. As illustrated in fig.3, the first part of the model involves normalizing the data, scaling the values of input variables to the range (0~1) to avoid the impact of data with different scales on the predictive performance of the model.

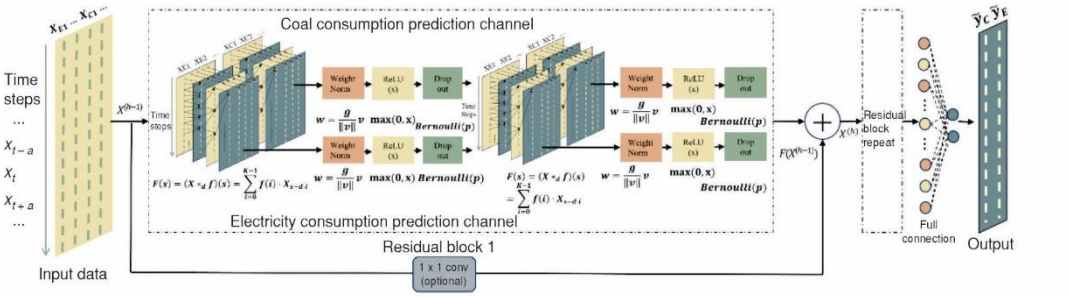

Figure 3. Dual-channel TCN model structure

The second part involves feeding the data into a dual-channel residual block for training. One channel is responsible for predicting coal consumption, while the other is dedicated to predicting electrical consumption. The weights of these two channels are independent of each other. The third part involves feeding the data into the fully connected layer for predicting coal consumption and electricity consumption. The MSE between the predicted results and the actual production record values is calculated, and the weights are adjusted using the Adam algorithm.

Residual connection

The presence of residual connections helps address the issue of model degradation as the depth increases by introducing skip connections that allow data to bypass one or more layers, effectively alleviating the problems of vanishing and exploding gradients. As demonstrated in fig. 4, where the outcome of the h^{th} residual module $F(X^{(h-1)})$ is summed with the input $X^{(h-1)}$ to obtain the new input $X^{(h)}$ [32]:

$$X^{(h)} = \delta\left[F\left(X^{(h-1)}\right) + X^{(h-1)}\right] \tag{5}$$

where δ is the activation function.

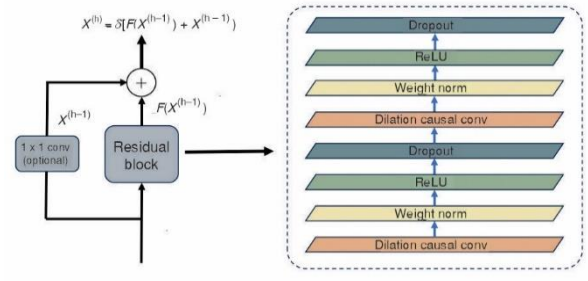

Figure 4. The structure of Residual block

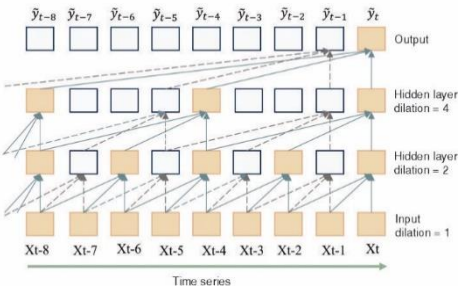

Figure 5. The structure of dilation convolution

The residual module is an important part of the TCN network, composed of stacked layers including dilated convolutional layers, Weight Norm layers, ReLU layers, and Dropout layers. As illustrated in fig. 5, dilated convolution increases the receptive field by inserting gaps in the convolutional kernel, where the gap size is the dilation rate. The dilation rate is represented as $d = [d_1,d_2,d_3, \ldots] = [2^0,2^1,2^2, \ldots]$, and the formula for dilated convolution on the sequence element, s, is as [30, 32]:

$$F(s) = (X * f)(s) = \sum_{i=0}^{K-1} f(i) X_{s-d \cdot i} \tag{6}$$

where X is the input sequence, filter $f = (f_1,f_2, \ldots, f_K)$, K means filter size, d means dilation factor, $s - d \cdot i$ denotes the past direction and $*$ denotes the convolution operator.

As dilation factor grows, the range of the receptive field expands, and the model is able to get a longer history of input data. The size of the receptive field r was calculated as:

$$r = 1 + nl(K-1) \tag{7}$$

where nl is the number of layers, and the appropriate dilation rate and convolution kernel size are selected depending on the duration of the timeseries and the actual need during the modeling process.

The Weight Norm layer is employed in the residual module for normalizing the weights of the convolution filter, expressed as:

$$w = \frac{g}{\|v\|} v \tag{8}$$

where w is the weight magnitude, v – the multidimensional vector, and g – the scalar. The activation layer utilizes the ReLU function, a non-linear activation function with the expression [22]:

$$\mathrm{Re}\,LU\left(x^{(l)}\right) = \max\left(0, x^{(l)}\right) \tag{9}$$

Finally, the dropout layer is employed to reduce overfitting by randomly setting the outputs of neurons to zero (dropout) during training, calculated as:

$$x^{(l)} = f\left(w_i^{(l)} r^{(l-1)} x^{(l-1)} + b_i^{(l)}\right) \tag{10}$$

where $(l-1)^{\text{th}} r^{(l-1)} \sim Bernoulli(p)$ represents a binary vector generated from a Bernoulli distribution with dropout rate, p

Parameter adjustment algorithm

In neural network computations, various optimization algorithms including gradient descent, momentum, Adagrad, root mean square propagation (RMSprop), Nesterov Accelerated Gradient, and Adam are utilized. These algorithms play a crucial role during training as they determine how the model learns and improves from its training data. In subsequent experiments, the Adam algorithm has proven to be effective. Therefore, in this article, the Adam algorithm is chosen to automatically adjust the model weights. For the weights of the model w, the formula for calculating the gradient at each iteration step t is given by [19]:

$$g_t = \nabla\left(w_t\right) \tag{11}$$

Then, first-order moment estimates m_t and the second-order moment estimate v_t of the gradient are calculated using [19]:

$$m_t = \beta_1 m_t - 1 + (1-\beta_1) g_t,\ v_t = \beta_2 v_t - 1 + (1-\beta_2) g_t^2 \tag{12}$$

Correction for bias in the first-order moment estimate and second-order moment estimate is performed as [19]:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t},\ \hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{13}$$

Finally, the new model weights were obtained by the weight update formula [19]:

$$w_{t+1} = w_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{14}$$

where α is the learning rate, usually $\varepsilon = 10^{-8}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ [19].

Prediction and outcome analysis

In this section, the data was preprocessed and neural networks were trained, adjusting various important parameters of the model. The objective was to set the simultaneous prediction of coal consumption and electricity consumption as the predicted values for the next hour. A dual-channel TCN model was compared with five classical neural network models to verify the superiority of this approach.

Data pretreatment

The data collected for this study are the production data from March to June of a cement factory in Guangxi, China. The sensors recorded the data per hour, resulting in a total of 2760 pieces of production data. These data need to be normalized due to the different magnitudes of the data variables. Normalization was done using equation [32]:

$$q' = \frac{q - q_{\min}}{q_{\max} - q_{\min}} \tag{15}$$

where q is the original value of the variable, q' – the normalized value of the variable, $q_{\max}$ – the maximum value of the input variables, and $q_{\min}$ – the minimum value of the input variables.

Evaluating the performance of the prediction model is primarily done by calculating the error between forecasted values and true values. In this study, the selected calculation error functions include MSE, RMSE, MAPE, MAE, with [30]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2,\ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2} \tag{16}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\tilde{y}_i - y_i}{y_i}\right|,\ MAE = \frac{1}{n}\sum_{i=1}^{n}|\tilde{y}_i - y_i| \tag{17}$$

wehere $\tilde{y}_i$ is the predicted value, y_i – the true value, and n – the sample size.

Parameter and structural adjustment experiments of TCN

According to the loss curves of the TCN model under different gradient descent strategies as shown in fig. 6, it can be seen that the model's convergence speed is relatively slow

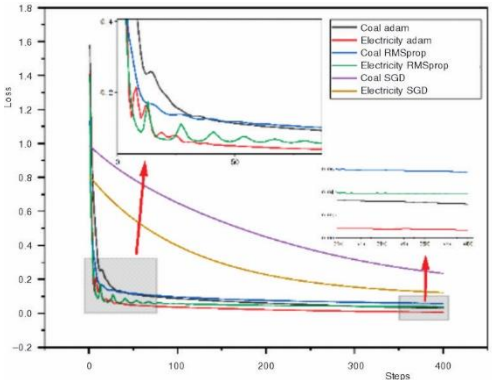

Figure 6. The TCN training loss with different strategy

under the SGD strategy. However, under the Adam and RMSprop strategies, the descent basically stops after 350 iterations, indicating that the model can converge faster under these strategies. In the prediction of electricity consumption required for the cement clinker calcination process, the fluctuation of the loss curve of RMSprop algorithm is more obvious compared to the Adam algorithm, which may hinder the model's ability to find a better solution. Therefore, the Adam algorithm is more suitable for the TCN model. In the prediction of coal consumption, the descent speed of RMSprop algorithm is faster than Adam at the beginning. However, after multiple iterations, the TCN model using the Adam algorithm achieves a lower loss value, indicating that the Adam algorithm can adapt better to the update of TCN model parameters.

From the process of model establishment in section *Construction of the TCN forecasting model*, it is evident that the convolutional kernel size, dilation rate, filters and training batch size are important parameters affecting the model's predictive performance. As shown in tab. 2, we selected different parameters for model training, and different codes represent different parameter configurations. We tested the model's predictive MSE that calculated according to eq. (16) by varying a single variable, and the final results are shown in figs. 7 and 8. The model's prediction error for different convolutional kernels first decreases and then increases, it is evident that the overall error of the model is lowest when the convolutional kernel is set to 13.

Table 2. Different parameters of TCN model

Parameters\code	1	2	3	4	5
Kernel size	11	12	13	14	15
Dilation	[2, 4, 8, 16]	[2, 4, 8, 16, 32]	$[2^1\text{-}2^6]$	$[2^1\text{-}2^7]$	$[2^1\text{-}2^8]$
Filter size	18	36	54	72	90
Batch rate	0.2	0.4	0.6	0.8	1.0

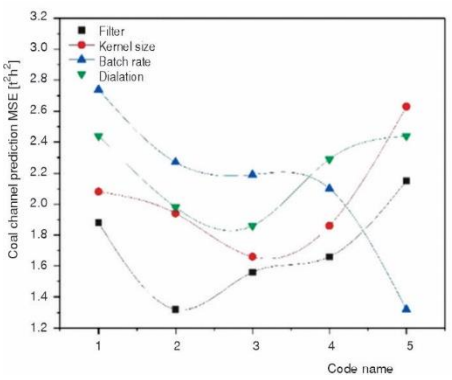

Figure 7. The MSE under various conditions of coal channel

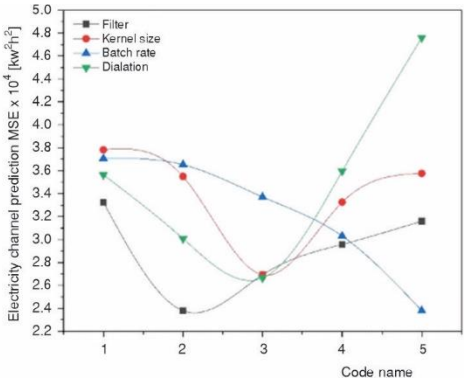

Figure 8. The MSE under various conditions of electricity channel

The dilation rate determines the model's receptive field. As shown in the graph, the overall error of the model is minimized when the dilation rate is set to [2^1-2^6]. The filter size directly affects the number of parameters in the convolutional layer. As shown in the graph, the overall model error decreases and then increases with an increase in filter size, with the minimum error achieved when the filter size is 36.

The batch size is an important hyperparameter during the training process. We set the batch size to different proportions of the training set. As shown in the graph, the overall model error increases with an increase in batch size, with the minimum error achieved when the batch size is 1. Therefore, the final chosen configuration includes a convolutional kernel size of 13, dilation rate of [2^1-2^6], filter size of 36, and a batch size of 1.

Comparing the predictions of the TCN model with other mode

In this section, the practical prediction accuracy of BPNN, LSTM, GRU, RNN, CNN, and TCN is assessed, alongside simulating the application scenarios of different models in real production processes. Prediction experiments are conducted using a test set separate from the training data. The prediction error of coal consumption by different models, calculated according to eqs. (16) and (17), is shown in tab. 3, and the comparison of real and predicted values is presented in fig. 9. It can be clearly observed that the CNN model exhibits a significant bias in predicting coal consumption, indicating overfitting, and leading to a significant prediction error on the test set.

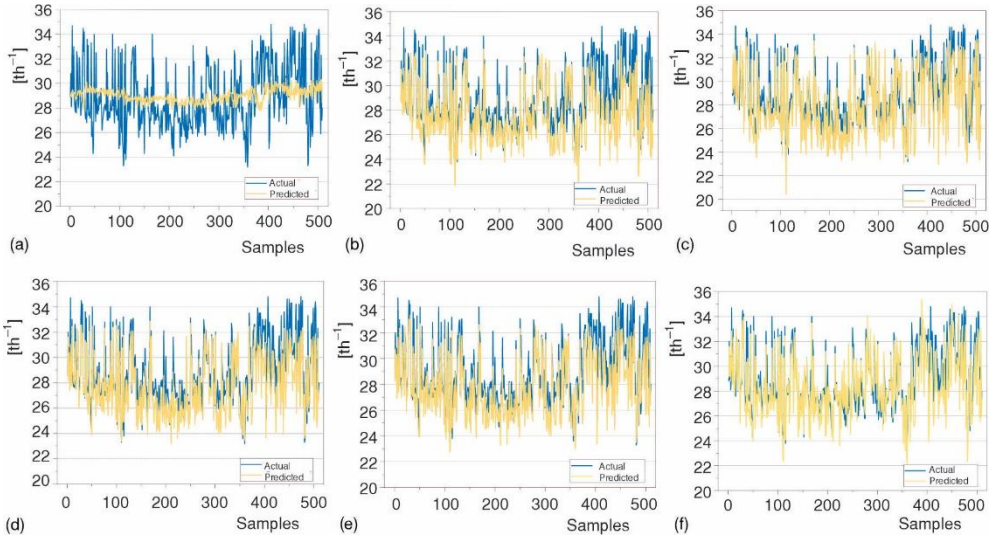

Figure 9. Comparison of coal consumption prediction results of different models; (a) test outcome of CNN, (b) test outcome of LSTM, (c) test outcome of GRU, (d) test outcome of RNN, (e) test outcome of BPNN, and (f) test outcome of TCN

In contrast, the LSTM, GRU, RNN, and BPNN models show more satisfactory prediction results, consistent with the actual coal consumption values within a certain error allowance. As shown in fig. 9(f), TCN performs better in the latter part of the samples, while the other models exhibit more noticeable errors. This is attributed to the lack of generalization ability when the models encounter unseen data. This indicates that the TCN model exhibit superior performance in capturing sequence features of coal consumption curves.

Table 3. Test errors of coal consumption

Model	Parameter	MSE	RMSE	MAPE	MAE
CNN	$k_1 = 5$, $k_2 = 10$	6.16	2.48	7%	2.05
LSTM	Hidden size = 13	3.10	1.76	5%	1.41
GRU	Hidden size = 13	2.48	1.58	4%	1.25
RNN	Hidden size = 13	2.47	1.57	4%	1.27
BPNN	Hidden size = 13	2.21	1.49	4%	1.23
TCN	Dilation = [1, 2, 4, 8, 16, 32, 64], kernel size = 13	1.32	1.15	3%	0.92

While BPNN, GRU, RNN, and LSTM models show significant results in coal consumption prediction, however, as indicated in tab. 4 and fig.10, they perform less satisfactorily in electricity consumption prediction with dual-channel structure.

Table 4. Test errors of electricity consumption

Model	Parameter	MSE	RMSE	MAPE	MAE
BPNN	Hidden size = 13	45453.8	213.20	2.5%	172.38
GRU	Hidden size = 13	38587.9	196.40	2.3%	154.69
RNN	Hidden size = 13	32250.0	179.58	2.0%	136.45
CNN	$k_1 = 5$, $k_2 = 10$	31646.6	177.90	2.0%	137.59
LSTM	Hidden size = 13	31140.2	176.47	2.0%	134.45
TCN	Dilation = [1, 2, 4, 8, 16, 32, 64], Kernel size = 13	23792.6	154.25	1.7%	114.12

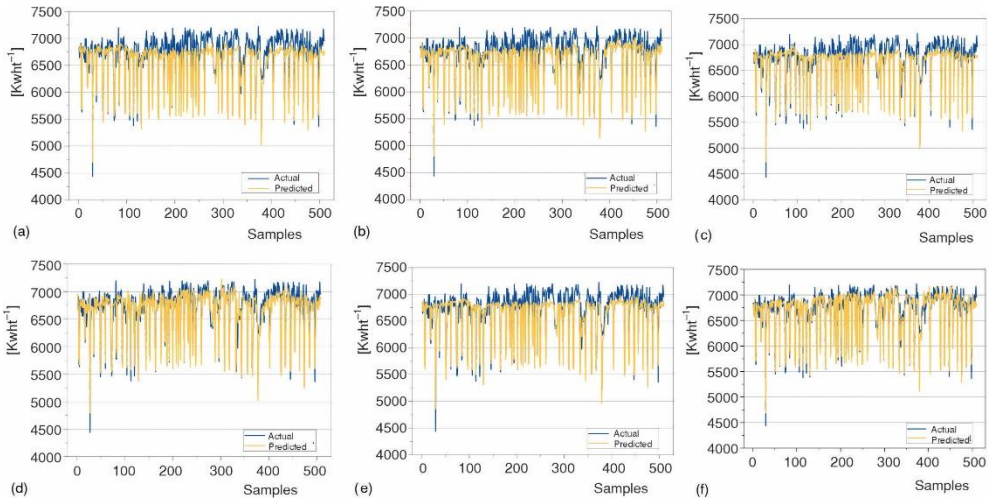

Figure 10. Comparison of electricity consumption prediction results of different models; (a) test outcome of BPNN, (b) test outcome of GRU, (c) test outcome of RNN, (d) yest outcome of CNN, (e) test outcome of LSTM, and (f) test outcome of TCN

Conversely, the CNN model performs poorly in coal consumption prediction, but shows satisfactory accuracy in electricity consumption prediction. This suggests a tendency for models to favor the prediction results of a specific single channel in dual-channel synchronous prediction tasks.

As shown in figs. 10(a)-10(c), and 10(e), during the electricity consumption prediction process, BPNN, GRU, RNN, and LSTM perform poorly when the electricity consumption exceeds 6500 kWh/t exhibiting significant underfitting. As illustrated in figs. 10(d) and 10(f), the TCN model with longer memory outperforms the slightly better performing CNN model in predicting accuracy between electricity consumption values of 6500 kWh/t and 7000 kWh/t.

Unlike the previous models, the TCN model demonstrates excellent performance in dual-channel synchronous prediction tasks, bridging this limitation. Its dilated convolutional structure can capture long-term dependencies in time series, while the use of residual block structures enhances the accuracy of the model, leading to a well-balanced performance in dual-channel forecasting.

As shown in tabs. 3 and 4, this paper employs MSE, RMSE, MAE, and MAPE as the testing error evaluation metrics. For the TCN model, the coal consumption prediction errors are 1.32, 1.15 , 3%, and 0.92, compared to traditional prediction models, the MSE of TCN has decreased by a minimum of 40% and a maximum of 79%, while the electricity consumption prediction errors are 23792.6, 154.25, 1.7%, and 114.12, compared to traditional prediction models, the MSE of TCN has decreased by a minimum of 24% and a maximum of 48%. This confirms the TCN model's capability to extract time-series features of the power and coal consumption data in the cement clinker manufacturing process.

Conclusions

- A synchronous forecasting model for coal and electricity consumption using TCN is proposed. We found that the TCN achieves the best prediction performance in this study when the kernel size is 13, the dilation layers are [2, ..., 16], and the number of filters is 36. The model was trained using actual production data, saved for application in the cement calcination system. Additionally, the model autonomously preprocesses production data, predict energy consumed for the next hour, and provide feedback to on-site engineers or the factory's automatic control system. This model has the potential to offer effective and scientifically sound recommendations for cement production control and resource scheduling.
- In this study, the Spearman correlation coefficient method was used to select the engineering variables for the prediction process. Among these, the variable with the highest correlation to coal consumption was the secondary air temperature, with a correlation coefficient of 0.9, while the variable most correlated with power consumption was the total voltage at the kiln head, with a correlation coefficient of 0.85. This approach aimed to enhance prediction precision by reducing model parameter redundancy and improving the nonlinear fit of the TCN to the nonlinearities between the energy consumption of the cement clinker calcination system and its influencing factors.
- The proposed model adopts a dilated convolutional structure design, facilitating a longer retain of time series data and enabling it to consider a broader data spectrum during predictions and thus enhancing accuracy. In predicting coal and power consumption, the dual-channel TCN structure achieves MAPE of just 3% and 1.7%, respectively, excellently fulfilling the precision requirements for energy consumption forecasts.
- The TCN is a novel and effective model designed specifically for time-series data, making it highly suitable for use in the cement manufacturing industry. It can be utilized to predict other key indicators within the industry, such as the exit temperature of the decomposer and the discharge temperature of the five-stage cyclone. Additionally, there is potential for TCN to be combined with other types of neural network models to develop a new model with a

broader application range and enhanced performance. We will continue our research based on these ideas.

Nomenclature

d_i^2 – difference between the i^{th} data pair rank
F – function
g – scalar
g_t – gradient at each iteration step
m_t – first-order moment estimate
n – number of samples
p – dropout rate
q – normalized value of the variable
r – receptive field
v – multidimensional vector
v_t – second-order moment estimate
w – weight magnitude
X – input matrix
x_t – feature values matrix at a certain time
$x^{(l)}$ – output of l layer level
Y_t – true value matrix
$\tilde{Y}_t$ – prediction matrix of the model
$\tilde{y}_t$ – output values matrix at a certain time

Greek syimbols

α – learning rate
$\varepsilon,\beta_1,\beta_2$ – constant value
δ – activation function
ρ – spearman correlation value

Subscripts and superscripts

C – coal
d – dilation factor
E – electricity
h – residual block sequence number
i – index
K – filter size
max – maximum
min – minimum
t – time step

Acronyms

RNN – recurrent neural network
GRU – gate recurrent unit
LSTM – long short-term memory
CNN – convolutional neural network
BP – back-propagation
TCN – temporal convolution network
MSE – mean squared error
ANN – artificial neural network
RMSE – root mean square error
MAPE – mean absolute percentage error
MAE – mean absolute error
SGD – stochastic gradient descent
RMSprop – root mean square propagation

References

[1] Wang, J. Q., *et al*., LSTM Based Long-Term Energy Consumption Prediction with Periodicity, *Energy, 197* (2020), 117197
[2] Hu, Y., *et al*., Short Term Electric Load Forecasting Model and Its Verification for Process Industrial Enterprises Based on Hybrid GA-PSO-BPNN Algorithm – A Case Study of Papermaking Process, *Energy, 170* (2019), Mar., pp. 1215-1227
[3] Wu, Y., *et al*., Application of GA-BPNN on Estimating the Flow Rate of a Centrifugal Pump, *Engineering Applications of Artificial Intelligence, 119* (2023), 105738
[4] Zhang, P., *et al*., Application of BPNN Optimized by Chaotic Adaptive Gravity Search and Particle Swarm Optimization Algorithms for Fault Diagnosis of Electrical Machine Drive System, *Electrical Engineering, 104* (2022), 2, pp. 819-831
[5] Xiao, M., *et al*., Fault Diagnosis of Rolling Bearing Based on Back Propagation Neural Network Optimized by Cuckoo Search Algorithm, *Multimedia Tools and Applications, 81* (2022), 2, pp. 1567-1587
[6] Dhyani, M., Kumar, R., An Intelligent Chatbot Using Deep Learning with Bidirectional RNN and Attention Model, *Materials Today: Proceedings, 34* (2021), Part 3, pp. 817-824
[7] Saud, A. S., Shakya, S., Analysis of Look Back Period for Stock Price Prediction with RNN Variants: A Case Study on Banking Sector of NEPSE, *Procedia Computer Science, 167* (2020), 2, pp. 788-798
[8] Yadav, S. P., *et al*., Survey on Machine Learning in Speech Emotion Recognition and Vision Systems Using a Recurrent Neural Network (RNN), *Archives of Computational Methods in Engineering, 29* (2022), 3, pp. 1753-1770
[9] Sherstinsky, A., Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, *Physica D: Nonlinear Phenomena, 404* (2020), 132306
[10] Zagrebina, S. A., *et al*., Electrical Energy Consumption Prediction is based on the Recurrent Neural Network, *Procedia Computer Science, 150* (2019), Lan., pp. 340-346

[11] Dey, P., *et al*., Hybrid CNN-LSTM and IoT-Based Coal Mine Hazards Monitoring and Prediction System, *Process Safety and Environmental Protection, 152* (2021), Aug., pp. 249-263
[12] Zhao, Z., *et al*., LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast, *Iet Intelligent Transport Systems, 11* (2017), 2, pp. 68-75
[13] Yuan, J., Tian, Y., An Intelligent Fault Diagnosis Method Using GRU Neural Network towards Sequential Data in Dynamic Processes, *Processes, 7* (2019), 3, p. 152
[14] Sun, J., *et al*., GRU-Based Model-Free Adaptive Control for Industrial Processes, *Neural Computing and Applications, 35* (2023), 24, pp. 17701-17715
[15] Zhang, K., *et al*., Computer Vision Detection of Foreign Objects in Coal Processing Using Attention CNN, *Engineering Applications of Artificial Intelligence, 102* (2021), 104242
[16] Sonata, I., *et al*., Autonomous Car Using CNN Deep Learning Algorithm, *Journal of Physics. Conference Series, 1869* (2021), 1, p. 12071
[17] Arena, P., *et al*., Image Processing for Medical Diagnosis Using CNN, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, *Detectors and Associated Equipment, 497* (2003), 1, pp. 174-178
[18] John, J. P., Parametric Studies of Cement Production Processes, *Journal of Energy, 2020* (2020), Feb., pp. 1-17
[19] Shi, X., *et al*., A Synchronous Prediction Model Based on Multi-Channel CNN with Moving Window for Coal and Electricity Consumption in Cement Calcination Process, *Sensors, 21* (2021), 13, p. 4284
[20] Lara-Benítez, P., *et al*., Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting, *Applied Sciences, 10* (2020), 7, p. 2322
[21] Kocyigit, N., Bulgurcu, H., Modeling of Overall Heat Transfer Coefficient of a Concentric Double Pipe Heat Exchanger with Limited Experimental Data by Using Curve Fitting and ANN Combination, *Thermal Science, 23* (2019), 6A, pp. 3579-3590
[22] Shi, X., *et al*., Sliding Window and Dual-Channel CNN (SWDC-CNN): A Novel Method for Synchronous Prediction of Coal and Electricity Consumption in Cement Calcination Process, *Applied Soft Computing, 129* (2022), 109520
[23] Zheng, J., *et al*., Hybrid Model of a Cement Rotary Kiln Using an Improved Attention-Based Recurrent Neural Network, *Isa Transactions, 129* (2022), Part B, pp. 631-643
[24] Liu, Z., *et al*., Research on Lime Rotary Kiln Temperature Prediction by Multi-Model Fusion Neural Network Based on Dynamic Time Delay Analysis, *Thermal Science, 28* (2023), 3B, p. 2703-2715
[25] Liu, G., *et al*., SA-LSTMs: A New Advance Prediction Method of Energy Consumption in Cement Raw Materials Grinding System, *Energy (Oxford), 241* (2022), 122768
[26] Ali, A., *et al*., Power Prediction of Waste Heat Recovery System for a Cement Plant Using Back Propagation Neural Network and Its Thermodynamic Modeling, *International Journal of Energy Research, 45* (2021), 6, pp. 9162-9178
[27] Shaikh, A. K., *et al*., A New Approach to Seasonal Energy Consumption Forecasting Using Temporal Convolutional Networks, *Results in Engineering, 19* (2023), 101296
[28] Wang, H., *et al*., Exploring Hybrid Spatio-Temporal Convolutional Networks for Human Action Recognition, *Multimedia Tools and Applications, 76* (2017), 13, pp. 15065-15081
[29] Kumar Sharma, D., *et al*., Data Driven Predictive Maintenance Applications for Industrial Systems with Temporal Convolutional Networks, *Computers & Industrial Engineering, 169* (2022), 108213
[30] Qiu, H., *et al*., A Piecewise Method for Bearing Remaining Useful Life Estimation Using Temporal Convolutional Networks, *Journal of Manufacturing Systems, 68* (2023), June, pp. 227-241
[31] Hao, X., *et al*., A Multi-Indicator Prediction Method for NOx Emission Concentration and Ammonia Escape Value for Cement Calciner System, *Journal of Computational Science, 76* (2024), 102212
[32] Cheng, W., *et al*., High-Efficiency Chaotic Time Series Prediction Based on Time Convolution Neural Network, *Chaos, Solitons & Fractals, 152* (2021), 111304

Paper submitted: March 18, 2024
Paper revised: April 28, 2024
Paper accepted: May 1, 2024