

Article

A Data-Driven Approach for the Ultra-Supercritical Boiler Combustion Optimization Considering Ambient Temperature Variation: A Case Study in China

Zhi Wang ¹, Guojia Yao ¹, Wenyuan Xue ¹, Shengxian Cao ^{1,*}, Shiming Xu ^{2,3} and Xianyong Peng ^{2,3,*}

¹ School of Automation Engineering, Northeast Electric Power University, Jilin 132012, China; 1201900035@neepu.edu.cn (Z.W.)

² School of Low-Carbon Energy and Power Engineering, China University of Mining and Technology, No. 1, Daxue Road, Xuzhou 221116, China

³ Jiangsu Smart Energy Technology and Equipment Engineering Research Center, No. 1, Jinshan Dong Road, Xuzhou 221116, China

* Correspondence: csxlb_jl@163.com (S.C.); xianyong.peng@cumt.edu.cn (X.P.)

Abstract: To reduce coal consumption, nitrogen oxide (NO_x), and carbon emissions for coal-fired units, combustion optimization has become not only a hot issue for scientists but also a practical engineering for engineers. A data-driven multiple linear regression (MLR) model is proposed to solve the time-consuming problems of boiler online combustion optimization systems. Firstly, A whole year's worth of the historical operating data preprocessing procedure of a coal-fired boiler in a power station including data resampling, data cleaning, steady-state selection, and cluster analysis is performed. In order to meet the applicable conditions of the linear model, the historical operating data are divided into different sub-datasets (combination mode of coal mills, main steam flow, ambient temperature, lower heating value of coal). Secondly, the multi-objective optimization strategy of economical, carbon, and NO_x emissions indexes is employed to select operating optimum data packets, and a new dataset is established that is better than the average value of the optimization target in each sub-dataset. On this basis, a stepwise regression algorithm (SRA) is used to select the specific manipulated variables (MVs) that are significant to the multiple optimization targets from 47 candidate MVs in each sub-dataset (different partitions have different types of MVs), and an MLR prediction model is developed. In order to further realize combustion optimization control, the MVs are optimized by employing the MLR model. According to the deviation between the optimal value and the real-time value of the MVs, a boiler combustion closed-loop control system is developed, which is connected with the DCS using the sum of the deviation signal and the corresponding original one. Then, a boiler combustion application test was carried out under some working conditions to verify the feasibility and effectiveness of the approach. The update time of the system signals running on industrial computers is less than 1 s and suitable for online applications. Finally, a full-scale test of the combustion optimization online control system (OCS) is executed. The results show that the boiler thermal efficiency increased by 0.39% based on standard coal, the NO_x emissions reduced by 2.85% and the decarbonization effect is significant.

Keywords: combustion optimization; closed-loop control; working condition internals; multiple linear regression; stepwise regression algorithm; data mining; carbon emissions



Citation: Wang, Z.; Yao, G.; Xue, W.; Cao, S.; Xu, S.; Peng, X. A Data-Driven Approach for the Ultra-Supercritical Boiler Combustion Optimization Considering Ambient Temperature Variation: A Case Study in China. *Processes* **2023**, *11*, 2889. <https://doi.org/10.3390/pr11102889>

Academic Editors: Ziming Yan, Rui Wang, Chuan He, Tao Chen and Zhengmao Li

Received: 17 September 2023

Revised: 28 September 2023

Accepted: 28 September 2023

Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A large amount of coal consumption in the process of electricity production has led to global warming and air pollution worldwide [1]. The “Paris Agreement”, signed by 178 contracting parties around the world, is a unified arrangement for the global response to climate change after 2020. Its long-term goal is to control the global average temperature rise within 2 °C compared with the pre-industrial period and strive to limit the temperature

rise within 1.5 °C. The agreement is legally binding on the parties. In this context, China, as one of the signatories, announced a new carbon peak target and carbon neutral vision at the 2020 UN General Assembly. Carbon dioxide emissions strive to reach a peak before 2030 and strive to achieve carbon neutrality by 2060 [2]. In the field of energy consumption, coal-powered electricity is the industry that emits the most carbon dioxide [3]. Among the top ten coal-powered electricity countries in the world, as shown in Table 1, five countries have committed to carbon neutrality. As of October 2022, countries that commit to carbon neutrality have reached 127 around the world. Throughout the 20th century, developed countries in Europe and America took nearly 80 years of structural adjustment from carbon peak to carbon neutrality. China has nearly half of the world's coal-fired power generation and promises that the time from carbon peak to carbon neutrality will be much shorter than developed countries [4]. The main carbon reduction measures currently include shutting down coal-fired power generation, adopting renewable energy [5–8], improving the efficiency of coal-fired power generation [9–12], CCS [13,14], etc.

Table 1. The top ten countries with global coal-powered electricity share committed to a carbon-neutral target time.

Commitment or Not	Year	Coal Consumption Ranking	Countries	Global Coal-Powered Electricity Share in 2022
✓	2060	1	China	52.3%
×	-	2	India	13.6%
×	-	3	United States	8.9%
✓	2050	4	Japan	3.0%
✓	2050	5	South Korea	2.0%
×	-	6	Indonesia	2.0%
✓	2050	7	South Africa	1.9%
×	-	8	Russia	1.9%
✓	2050	9	Germany	1.8%
×	-	10	Australia	1.3%

About 66% of China's electricity production comes from coal, and this dependence on coal is more serious than the European Union (20%) or the United States (30%) [15]. In order to fulfill the above commitments, China has issued a carbon reduction action plan, the "14th Five-Year Plan" [16], which incorporates carbon peak and carbon neutrality into the overall layout of ecological civilization construction. In 2020, Shandong Province shut down 52 coal-fired units below 300 MW that did not meet environmental protection standards [17]. Huadian Group, one of the five largest power generation groups, plans to shut down 3GW of thermal power installed capacity by 2025. State Power Investment Corporation promised to take the lead in achieving the carbon peak target in 2023.

Yue et al. [18] took China as an example to develop a modeling framework that integrates industrial efficiency and power plant unit information. By improving industrial efficiency, we eliminated and shut down the most polluting units to achieve the goal of reducing emissions. The shutdown of thermal power units will have an immediate effect on reducing carbon emissions, but by 2040, global electricity will still be provided by 31% coal power [19]. Carrying out combustion optimization research on existing coal-fired boilers is an effective measure to continuously reduce carbon emissions. At present, the methods for coal-fired boiler combustion optimization include boiler equipment modification [20–22], temperature field monitoring and optimization control [23,24], computational fluid dynamics (CFDs) [25,26], and the combustion optimization control parameter model [27–30]. Among them, the boiler equipment modification requires a large number of on-site tests to make adjustments, occupying valuable power generation resources; the CFD modeling process also takes several weeks or even longer due to time and calculation requirements, and this is impractical for most power plants [9]. Compared with the other three methods, the establishment of the combustion optimization control parameter model has the advantages of less capital investment and shorter calculation time.

This method generally adopts a data-driven method for modeling and establishes a data mapping or mathematical expression of operating parameters and performance parameters. Since boiler combustion is a complex and non-stationary dynamic process, with the development of computer technology and data science, people's ability to deal with non-linear problems has gradually improved, and data-driven methods, including machine learning, provide new means for model construction. Zhou et al. [27] used unburned carbon content to characterize boiler efficiency, used an artificial neural network (ANN) model to establish the predictive relationship between input parameters and boiler efficiency and NOx emissions, and used the genetic algorithm (GA) to find the optimal solution, but the training data are too small, resulting in low simulation accuracy. Wu et al. [28] used the support vector regression (SVR) algorithm to establish a multi-objective optimization model for coal-fired power plant boilers for the multi-objective optimization problem in the literature [27] combined with the cytogenetic algorithm to calculate the Pareto optimum of boiler efficiency and NOx emissions. Support vector machines (SVMs) have better non-linear mapping capabilities than ANNs, but it is difficult for SVMs to obtain the optimal solution when there are more input variables [29], and the training process using large datasets will become difficult to calculate [30].

Due to the high complexity of the above-mentioned neural network and machine learning optimization intelligent control algorithms based on biological evolution theory, the optimization process often takes a few seconds or even tens of seconds, which is difficult to apply to real-time online optimization and the adjustment of coal-fired power plants [31]. In response to the above problems, this paper is guided by the online application of combustion optimization, and a data-driven multiple linear regression (MLR) model is proposed. The MLR model is a method developed in the field of multiple statistical analysis. Due to its simple and clear mathematical background, it is widely used in power plant database modeling to improve linear regression models [32]. The technical roadmap of this paper is shown in Figure 1. We extract the historical operating data (nearly 600,000 records) of the unit for the past year from the DCS historical database of the target unit. The historical data contain rich and valuable unit status information [33], including unit equipment characteristics, operator operating habits, and experience. The steady-state operation data of the unit are screened through the sliding window method, and the steady-state data are used to establish an approximate linear partition of the main steam flow (MSFlow), ambient temperature (AT), the lower heating value of coal (LHVC), and the combination mode of coal mills (COM-Mill). The MLR model is used in each partition to establish the mapping relationship between the controlled parameters and the boiler efficiency. In order to reduce the multicollinearity existing between the control parameters, the stepwise regression method (SRM) is used in the modeling process to select the regression variables from the candidate variables. It should be noted that there are differences in the combustion control laws of different partitions, so the regression parameters included in the regression models of different partitions will also be different. The optimization process is used to calculate the optimized value of the corresponding manipulated variable. According to the partial regression coefficient in the model, the optimization response time is less than 1 s, which is suitable for online optimization control. Finally, connect the optimized signal to the unit's DCS logic, participate in online real-time optimization control, and compare the boiler efficiency changes before and after optimization.

The key contributions of this paper are as follows:

- We quantitatively analyzed the impact of AT changes on boiler combustion control and used the fuzzy C-means (FCM) clustering algorithm to use the AT as the basis for working conditions to improve the accuracy of the linear model;
- Subdivide the complex non-linear operating conditions of the boiler into simple operating conditions that can be linearly processed. Aiming at the industrial production data of coal-fired boilers in the Weifang Power Plant, an MLR model is used to establish the mapping relationship between manipulated variables and boiler efficiency, and the model is suitable for online application;

- Using partial differential derivation to calculate the optimal control deviation of model variables and participating in the boiler online real-time closed-loop control by embedding the DCS configuration logic is a practical attempt under carbon neutrality.

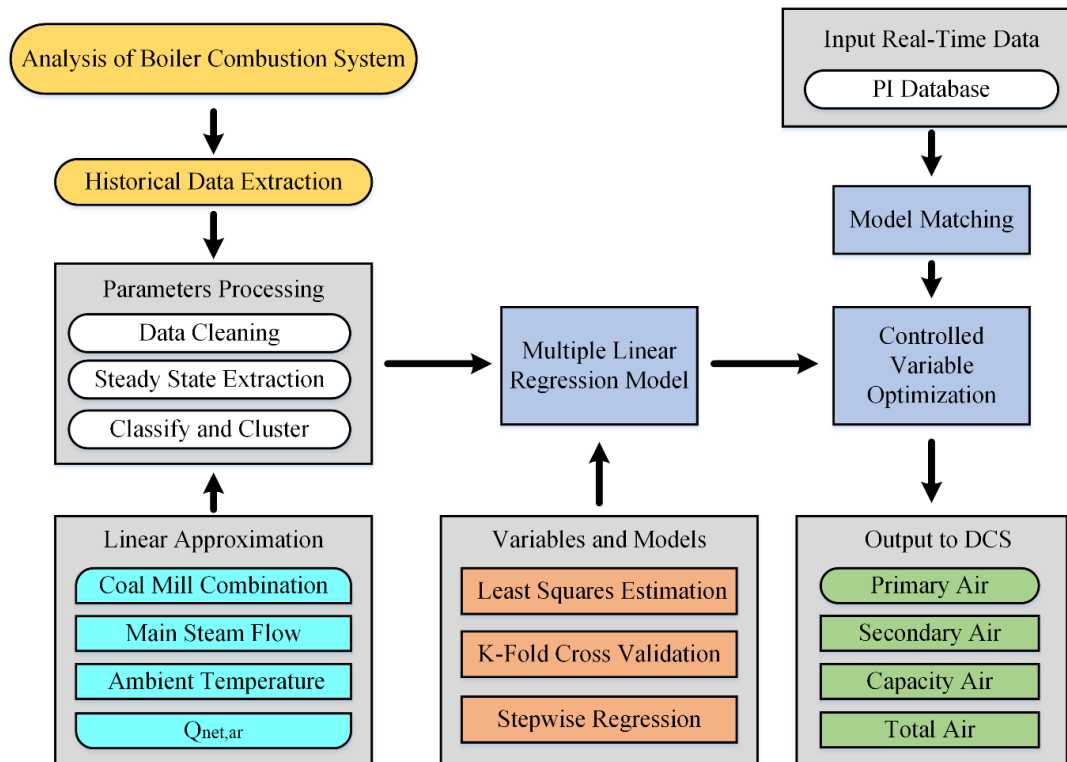


Figure 1. Flowchart of the optimization method used in the present work.

The rest of the paper is organized as follows. In Section 2, we describe typical boilers and variable selection. We preprocessed the historical data of the Weifang Power Plant in Section 3. In Section 4, we introduced the optimization methods for the MLR model and manipulated variables. In Section 5, we conducted model training and test experiments using historical boiler operating data, introduced the structure and composition of the combustion optimization OCS, deployed a full-scale combustion optimization online closed-loop control test, and analyzed the test results. Finally, in Section 6, we draw conclusions.

2. Analysis of the Boiler Combustion System

2.1. Description of the Boiler Combustion System

The Weifang Power Plant has a total installed capacity of 2000 MW and four coal-fired units. The studied boiler is number 3: a 670 MW tangentially fired, which is designed based on the supercritical boiler technology from ALSTOM combined with the lean coal-fired experience of Shanghai Boiler Works in China. The boiler is a supercritical parameter variable pressure operation once-through furnace, equipped with four low NO_x coaxial combustion system (LNCFS) burners, which are arranged in an equal air distribution method. The burner adopts a double inlet and double outlet coal mill, cold primary air, positive pressure direct blowing pulverizing system design, and six coal mills transport pulverized coal, respectively, for six-layer coarse coal nozzles and six-layer fine coal nozzles.

As shown in Figure 2, each combustor assembly is composed of thirty-four layers of damper baffles; the bottom of the burner in the vertical direction is one layer of an Underfire Air (UFA) nozzle (AAA), and upwards are six layers of coarse coal nozzles (AR\BR\CR\DR\ER\FR), six layers of fine coal nozzles (AL\BL\CL\DL\EL\FL), and two layers of CCOFA nozzles (CCOFA-A\B). There is a SOFA burner on the top, including six layers of SOFA nozzles (SOFA-A\B\C\D\E\F). There is end air (AA) between UFA

and AR, and two layers of auxiliary air nozzles (AB\ABB, BC\BCC, CD\CDD, DE/DEE, EF/EFF) are arranged between every adjacent two layers of coarse coal nozzles. Finally, there is a secondary air nozzle FF arranged in the middle of the coarse and fine coal nozzles, and BCL is arranged between BL and CL.

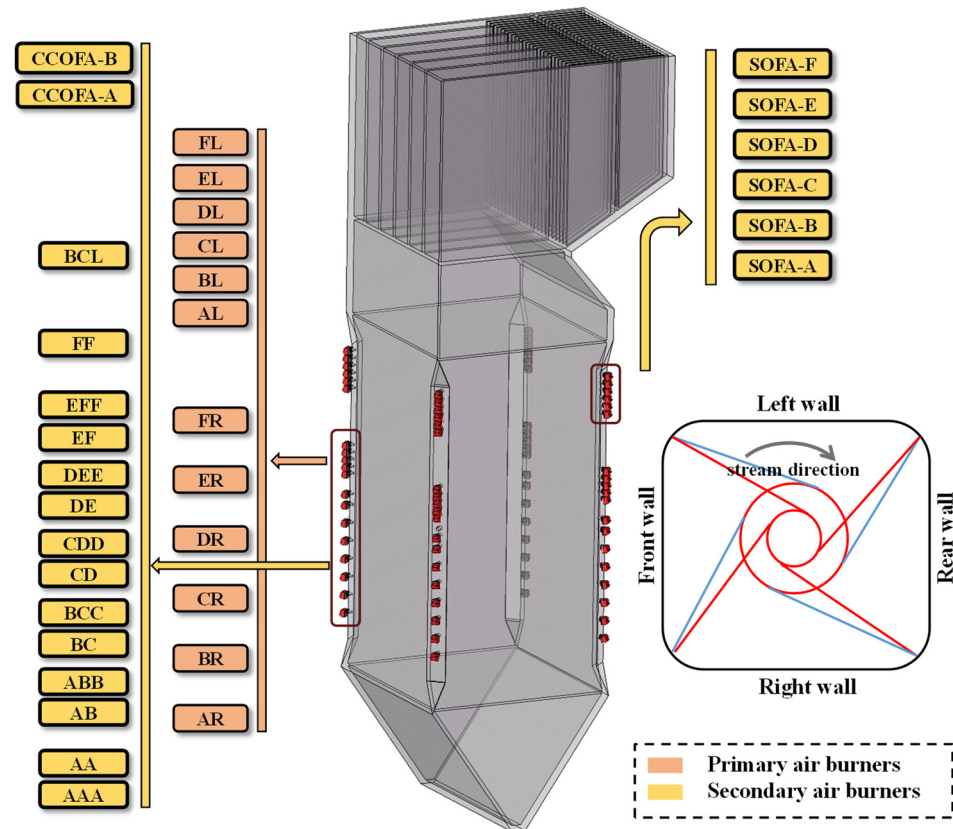


Figure 2. Schematic diagram of the Weifang Power Plant unit boiler furnace.

2.2. Variable Analysis and Selection

2.2.1. The Influence of AT on the Total Air Flow of the Boiler

Zhou et al. [34] conducted a unified investigation on 66 coal samples in China and calculated the amount of air required for each coal sample to release 100 MJ of heat. The statistical results show that the average air volume required is approximately equal to 26,265 L. The deviation between each required air volume and the average air volume is less than 5% for all coal samples, except one coal sample. The deviation of the air volume required for 14 coal samples is between ± 2.5 and $\pm 5\%$, and the deviation of the air volume required for the remaining 51 coal samples and the average air volume is less than $\pm 2.5\%$. The conclusion is that if the theoretical air volume is calculated according to the unit mass of fuel, the difference between different coal types is very large; if the theoretical air volume is calculated according to the unit heat, the difference is not much. A certain load requires a certain amount of heat; that is, a certain amount of air. Therefore, the optimal air volume of the boiler should be an amount that only varies with load, not coal. So, a novel concept was presented, the air-to-carbon ratio, which is defined as the mass ratio of air volume required and carbon in coal. Theoretically, the air-to-carbon ratio is more accurate and reasonable than the air-to-coal ratio in the boiler's air supply control system. Therefore, precise air volume control will effectively improve boiler efficiency. However, the accuracy of the boiler air supply control is affected by the AT, which changes with seasons and time. To the ideal gas state equation is shown as Equation (1):

$$pV_{\varphi} = nRT \quad (1)$$

where P represents the pressure (Pa), V_ϕ represents the air volume (m^3), n represents the amount of air matter (mol), T represents the air thermodynamic temperature (K), and R represents the ideal air constant. Equations (1)–(3) can be obtained:

$$pM = \rho nRT \quad (2)$$

$$\frac{\rho_{T_{\max}}}{\rho_{T_{\min}}} = \frac{T_{\min}}{T_{\max}} \quad (3)$$

where M represents the air quantity, ρ represents the air density, $T_{\max}$ and $T_{\min}$, respectively, represent the highest and lowest values of AT year-round in the Weifang Power Plant, and $\rho_{T_{\max}}$ and $\rho_{T_{\min}}$, respectively, represent the gas density corresponding to the highest and lowest temperature values throughout the year.

Let us take the target unit as an example to quantitatively analyze the influence of AT on the air flow control of the boiler. Figure 3 represents the temperature curves at the inlet of the air preheater of the target boiler of the Weifang Power Plant for one whole year and three days, respectively. The figure shows that the annual temperature interval is $-6 \sim 38^\circ\text{C}$, and the whole day temperature span is close to 15°C . According to Equation (3), the air density varies close to 13% throughout the year and close to 5% throughout the day. Under the existing control strategy, a certain load matches the certain total air flow, but the mass flow of the total air changes with the density at different ATs; as a result, under a certain load, the mass of the total air varies with AT, which affects the optimal air supply volume and decrease the boiler efficiency. Therefore, the total air supply control strategy should be improved by considering the impact of the AT.

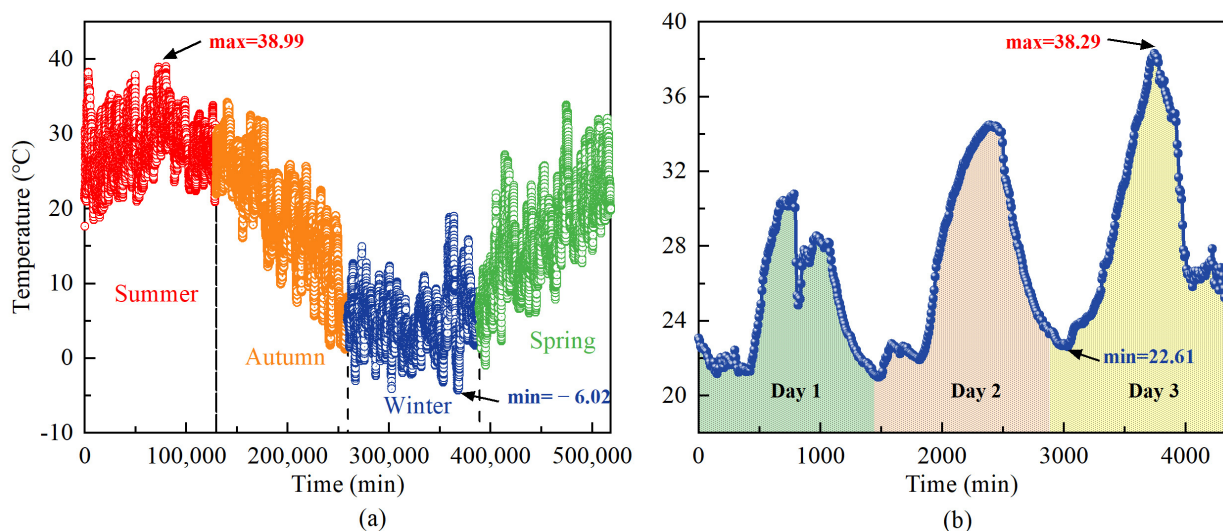


Figure 3. Air temperature profile at the inlet of the air preheater. (a) Annual variation curve, (b) random extraction three-day variation curve.

2.2.2. The Influence of COM-Mill on Damper Control

The six coal mills of the unit convey pulverized coal, respectively, for the six-layer coarse pulverized coal nozzle and the six-layer fine pulverized coal nozzle. For example, mill A is connected to the AR and AL layer nozzles, and the COM-Mill directly affects the opening control of the air damper. Table 2 shows the running number of coal mills corresponding to different load intervals. It can be seen in Table 2 that there is a difference in the running number of coal mills in the load overlap interval. Therefore, the utilization of load and COM-Mill to subdivide the combustion conditions of the boiler contributes to improving the accuracy of the model.

Table 2. The operation mode of the coal mill under different loads of boilers.

Operation Mode of Coal Mill	Boiler Load MCR
Operation of six coal mills	80–100%
Operation of five coal mills	60–100%
Operation of four coal mills	45–80%
Operation of three coal mills	35–60%
Operation of two coal mills	
(Mixed combustion of coal and fuel oil in 10–50% BMCR)	10–40%
Oil gun operation	0–30%

2.2.3. Variable Selection

According to the above variable analysis and boiler control principle, Table 3 lists the combustion MVs and data range. The combustion MVs include the opening of the air dampers and the total air, and the MVs are numbered x_1 – x_{47} in sequence. The performance variables affected by the MVs in Table 3 include the boiler efficiency and NOx emissions listed in Table 4 because unit 3 usually delivers the low-pressure steam to partly meet the needs of municipal heating from November of each year to March of the next year. MSFlow is selected to describe the unit load, and boiler efficiency is represented as the ratio of MSFlow to total coal quantity in this paper. The steam coal ratio is calculated by Equation (4), which represents the boiler efficiency under the classification of LHVC; the sensors of NOx emissions are located at the SCR inlet.

$$R_{sc} = \text{Main steam flow} / \text{Total coal flow} \quad (4)$$

Table 3. Manipulated variables.

Variable Description	Regression Variable	Symbol	Operation Range (%)
Primary air	x_1 – x_{12}	AR\BR\CR\DR\ER\FR AL\BL\CL\DL\EL\FL	[0, 100]
Secondary air	x_{13} – x_{26}	AAA\AA\AB\ABB\BC\BCC\CD CDD\DE\DEE\EF\EFF\FF\BCL CCOFA-A\CCOFA-B	[0, 100]
Over-fire air	x_{27} – x_{34}	SOFA-A\SOFA-B\SOFA-C SOFA-D\SOFA-E\SOFA-F	[0, 100]
Coal mill capacity air flow	x_{35} – x_{46}	MA1\MA2\MB1\MB2\MC1\MC2 MD1\MD2\ME1\ME2\MF1\MF2	[0, 100]
Total air	x_{47}	TA	[0, 100]

Table 4. Performance variables.

Parameter Description	Symbol	Operation Range
NOx emissions (mg/m ³)	-	[0, 1000]
Boiler efficiency (%)	η	[3, 15]

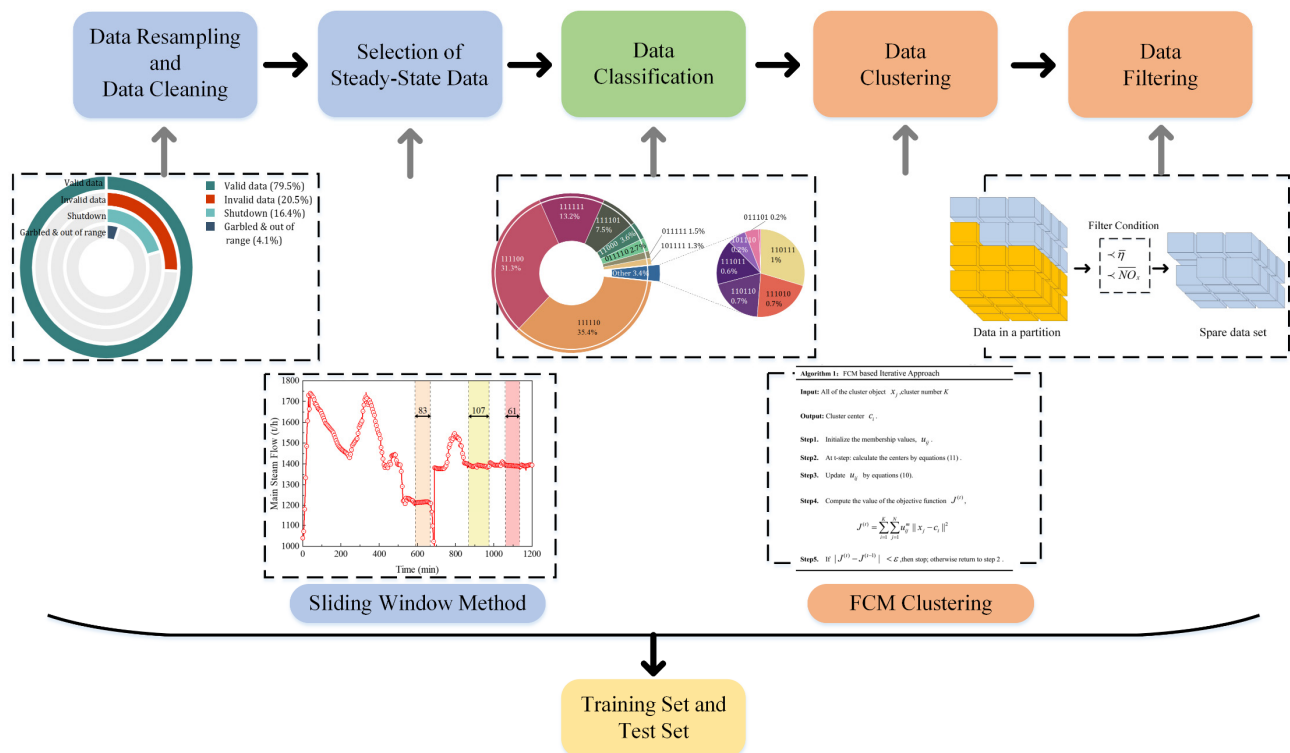
The performance variables of boiler combustion in Table 4 are affected by MVs and are also closely related to boiler operating conditions. In order to approximate the applicable conditions of the linear model, this paper takes the MSFlow, AT, LHVC, and COM-Mill as the basis for subdividing working conditions, as shown in Table 5. When the MSFlow optimization interval is 50–100%, the AT interval is selected between the lowest and highest temperature at the inlet of the forced draft fan all year. The LHVC comes from the test results of coal entering the furnace by the fuel operation department of the power plant. The COM-Mill ranges from three to six.

Table 5. Condition variables.

Parameter Description	Symbol	Operation Range
Main steam flow (t/h)	MSFlow	[1000, 2000]
Ambient temperature (°C)	AT	[−8, 38]
Lower heating value of coal (KJ/kg)	$Q_{\text{net,ar}}$	[17642, 26991]
Combination mode of coal mills	-	3–6

3. Data Processing

In this paper, a linear model is presented to approximate the non-linear process. The linear model cannot be used directly in the data processing of the original sampled data, so data preprocessing is utilized to support modeling. Firstly, resample and clean the original data. Secondly, select steady-state data. Then, the data are classified according to the COM-Mill, and FCM is used to cluster the conditional variables after classification. Finally, the sampling values that are better than the average boiler efficiency $\bar{\eta}$, the average Nox emissions $\overline{NO}_x$ in the partition are selected in each partition, and then the model spare dataset is formed. The procedure of data preprocessing is shown in Figure 4.

**Figure 4.** Data preprocessing process.

3.1. Data Cleaning

Researchers can independently select the sampling interval when extracting data from the DCS historical database. In order to reduce the amount of data, Smrekar et al. [35] resampled data for 0.5 min, 1 min, and 1.5 min, respectively. The results show that the 1 min resampling interval not only follows the dynamic process of the measured signals but also filters small fluctuations. The historical operating data of unit 3 from May 2021 to June 2022 was extracted by utilization of a 1 min sampling interval method in this paper, with a total of 604,800 sampling points.

Due to the complex environment of the industrial production site, it is inevitable that sensor failures and information transmission errors will occur, resulting in incorrect data, such as garbled codes (data with *), in the original sampled data. Erroneous data, outage data, and data beyond the optimized range are deleted through data cleaning. Figure 5

shows the statistical results of data cleaning, in which valid data accounted for 79.5% and invalid data accounted for 20.5%.

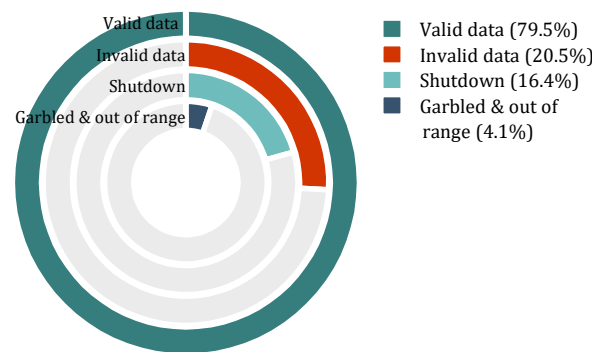


Figure 5. Statistical results of data cleaning.

3.2. Selection of Steady-State Data

Based on the data-driven MLR model, the steady-state operation data of the boiler is used to establish a combustion optimization model. When load is frequently adjusted, the dynamic characteristics of the boiler will change accordingly and the thermal variables that represent mainly the operation state of the unit also fluctuate, leading to the historical operating data at the same period cannot be a true reflection of input-output relationship of the system. For data-driven boiler combustion optimization, it is necessary to obtain historical data under steady-state conditions [36,37]. This paper uses the sliding window method [38,39] to filter the steady-state data in the historical data.

The AT changes relatively slowly, and the coal quality batch changes periodically, so the MSFlow is selected as the criterion for the steady-state data of the unit. The steady-state criterion of the sliding window method is shown in Equation (5):

$$\sigma_{x_j} = \sqrt{\frac{1}{N-1} \sum_{\tau=t}^{t+N-1} (x_j^{\tau} - \bar{x}_j)^2} < \sigma_{x_j}^{set} \quad (5)$$

where t represents the start time, x_j^{τ} represents the value of the j -th characteristic variable at time t , N represents the length of the sliding window, in minutes; $\bar{x}_j$ represents the mean value of the j -th feature variable when the length of the sliding window is N , σ_{x_j} represents the steady-state test result of the j -th characteristic variable, $\sigma_{x_j}^{set}$ represents the preset steady-state critical threshold of the j -th characteristic variable.

We scanned all the valid data of the MSFlow and recorded the data interval in accordance with Equation (5). Figure 6 shows an example of steady state detection—a historical data segment of 1200 min, the steady-state critical threshold is 20, and the length of the sliding window is $N = 60$, which is 1 h. Figure 6 shows that three steady-state intervals have been found, with time lengths of 83, 107, and 61 min respectively.

3.3. Cluster Analysis

3.3.1. Data Classification

The literature [39] clustered the load and coal quality coefficient in the clustering process and then classified according to the COM-Mill in the clustering interval. This classification method can be improved logically because when the number of running coal mills is small, such as two mills running, the load range covered in the historical data is less. There may be only one load cluster under this COM-Mill that cannot be clustered anymore. This paper proposes data classification according to the COM-Mill first, and then clusters based on MSFlow, AT, and LHVC. This method can ensure that the division of working conditions is more reasonable. Figure 7 shows the proportion of all COM-Mills in the steady-state effective data.

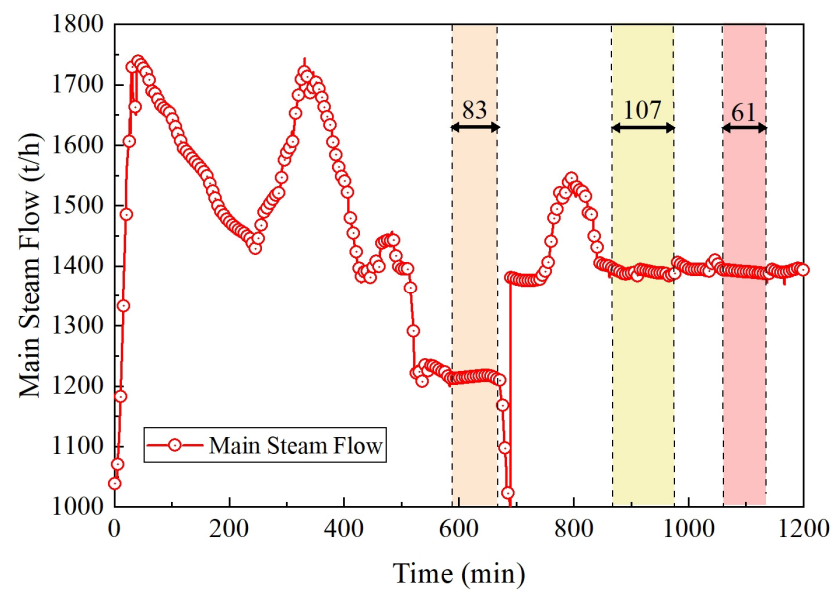


Figure 6. Steady state detection example.

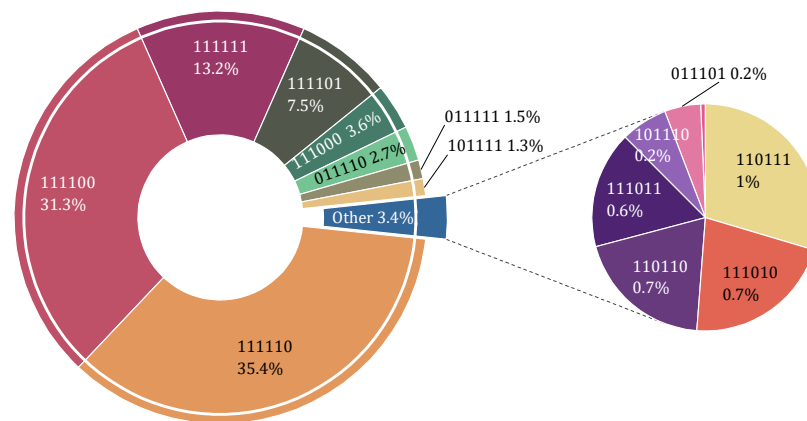


Figure 7. Proportion of combination mode of coal mills.

3.3.2. Clustering Method

At present, the antecedent identification methods widely used in thermal process TS modeling are FCM clustering, GK clustering, GG clustering, etc. The above clustering methods provide the possibility for local linearization modeling of non-linear thermal processes [40]. In this paper, the FCM algorithm is used to cluster analysis of MSFlow, AT, and LHVC, and the minimization of the objective function of the FCM algorithm is shown in Equation (6):

$$\min J(u_{ij}, c_i) = \sum_{i=1}^K \sum_{j=1}^N u_{ij}^m \|x_j - c_i\|^2 \quad (6)$$

$$\text{s.t. } \sum_{i=1}^K u_{ij} = 1, j = 1, 2, \dots, N; i = 1, 2, \dots, K$$

where x_j represents the cluster object, c_i represents the cluster center, u_{ij} represents the object x_j belonging to the weight of category i , and m is the weight index. For this type of equation-constrained optimization problem, the Lagrangian multiplier λ_j is introduced, and the corresponding Lagrangian function is Equation (7):

$$L(u_{ij}, c_i, \lambda_j) = \sum_{i=1}^K \sum_{j=1}^N u_{ij}^m \|x_j - c_i\|^2 - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^K u_{ij} - 1 \right) \quad (7)$$

Calculate the partial derivatives of c_i and u_{ij} through Equation (7), and then calculate the cluster center point c_i and the degree of membership u_{ij} :

$$u_{ij} = \frac{1}{\sum_{l=1}^K \left(\frac{\|x_j - c_i\|}{\|x_j - c_l\|} \right)^{\frac{2}{m-1}}} \quad (8)$$

$$c_i = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m} \quad (9)$$

The input of the FCM algorithm includes the number of clusters K . This paper calculates the number of clusters of MSFlow, AT, and LHVC according to Equations (10)–(12). According to the control characteristics of the boiler and the recommendations of the operation engineer, the initial interval is set to 50, 4, and 1000. After the effective data are classified by the combination mode of coal mills, the maximum and minimum values of the MSFlow and AT and the low-level coal calorific values in each combination are extracted. Divide the difference by the initial interval and take an integer to obtain it, as shown in Equations (10)–(12):

$$[K_{MSFlow}] = \frac{MSFlow_{\max} - MSFlow_{\min}}{50} \quad (10)$$

$$[K_{AT}] = \frac{T_{\max} - T_{\min}}{4} \quad (11)$$

$$[K_{Q_{net,ar}}] = \frac{Q_{net,ar_{\max}} - Q_{net,ar_{\min}}}{1000} \quad (12)$$

where $[\times]$ denotes that the largest integer is no more than $\times$ and $[K_{MSFlow}]$, $[K_{AT}]$ and $[K_{Q_{net,ar}}]$ are the numbers of clusters of MSFlow, AT, and LHVC.

It is worth noting that because the historical data under different coal mill combinations are different, the MSFlow rate, AT, and LHVC range are different, so the number of clusters under different coal mill combinations is also different.

Algorithm 1 illustrates the FCM-based iterative approach, and the default number of iterations t is 100 times.

Algorithm 1: FCM-based Iterative Approach

Input: All of the cluster object x_j , cluster number K

Output: Cluster center c_i

Step1. Initialize the membership values, u_{ij}

Step2. At t -step: calculate the centers by Equation (8).

Step3. Update u_{ij} by Equation (9).

Step4. Compute the value of the objective function $J^{(t)}$,

$$J^{(t)} = \sum_{i=1}^K \sum_{j=1}^N u_{ij}^m \|x_j - c_i\|^2$$

Step5. If $|J^{(t)} - J^{(t-1)}| < \varepsilon$, then stop; otherwise return to step 2.

3.4. Data Filtering

A data-driven multi-objective combustion optimization method is studied in this paper. The optimization objectives are boiler efficiency and NO_x emissions. The partitioned data are filtered according to the mean value of boiler efficiency $\bar{\eta}$ and NO_x emissions $\overline{NO_x}$ in the interval to establish a model spare dataset. Figure 8 shows a schematic diagram of data filtering in a certain partition.

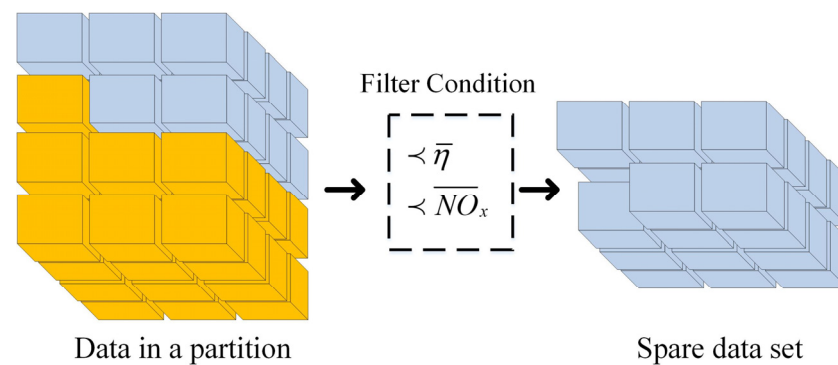


Figure 8. Schematic diagram of data filtering.

3.5. A Case Study

Taking the COM-Mill 111110 as an example (that is, the A-E mill runs and the F mill stops), this paper deploys a cluster analysis on 142,619 steady-state historical sampling points. The relevant data interval is analyzed under this combination and shown in Table 6.

Table 6. Condition variables of 111110 coal mill combinations.

Variable Description	Symbol	Operation Range
Main steam flow (t/h)	MSFlow	[1091.99, 1991.28]
Ambient temperature (°C)	AT	[−4.38, 38.09]
Lower heating value of coal (KJ/kg)	Q _{net,ar}	[19,699, 24,758]
Coal mill combination	-	111,110

According to the condition variable distribution interval of the COM-Mill (111110), the cluster number of MSFlow, AT, and LHVC is calculated according to Equations (10)–(12). As shown in Table 7, the FCM algorithm is used to calculate the clustering results of each clustering object. Among them, the weight index is $m = 2$, the maximum number of iterations is $J = 100$, and the iteration termination condition is $\varepsilon = 0.00001$.

Table 7. Condition variable cluster result of 111110 coal mill combinations.

Condition Variables	Parameter Description	Cluster Result				
Main steam flow	[K _{MSFlow}]			17		
	Cluster center	1259.81	1319.76	1373.22	...	1959.59
	Cluster interval	[1203.8, 1289.7]	[1289.82, 1346.49]	[1346.5, 1387.08]	...	[1906.13, 1991.28]
	Number of samples	2831	5241	8714	...	1090
Ambient temperature	[K _{AT}]			10		
	Cluster center	0.04	3.63	6.83	...	32.59
	Cluster interval	[−4.38, 1.83]	[1.84, 5.23]	[5.24, 8.87]	...	[30.75, 38.09]
	Number of samples	11,767	16,391	18,277	...	8786
Low calorific value	[K _{Q_{net,ar}}]			5		
	Cluster center	19,989	20,805	22,330	23,243	24,273
	Cluster interval	[19,699, 20,375]	[20,452, 21,557]	[21,600, 22,758]	[22,795, 23,736]	[23,763, 24,758]
	Number of samples	7652	22,262	34,785	46,612	31,308

4. MLR Model

4.1. Model Description

The regression analysis method is to study the correlation and structural state of the relationship between variables by establishing a statistical model and an effective tool for model prediction [41]. This paper uses an MLR model to establish the mapping between boiler efficiency and manipulated variables, as shown in Equation (13):

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (13)$$

where η represents boiler efficiency, which can be obtained by Equation (4), x_1-x_p represents the 47 manipulated variables listed in Table 3, β_0 represents the regression constant, and $\beta_1-\beta_p$ represents the regression coefficients.

4.2. Least Squares Estimation of Regression Parameters

Regression parameter estimation methods include least squares estimation, regression value, residual error, and maximum likelihood estimation. The number of boiler sampling points n extracted in this paper is much larger than the number of regression variables p , so this paper uses least squares estimation to solve the regression parameters. The regression parameter fitting model is shown in Equation (14):

$$\eta_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} \quad (i = 1, 2, \cdots, n) \quad (14)$$

Then, the corresponding Jacobian matrix $\mathbf{J}$ is shown in Equation (15):

$$\mathbf{J} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix} \quad (15)$$

Sign:

$$\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad (16)$$

Then, the least squares estimator of the regression parameter $\boldsymbol{\beta}$ can be calculated by Equation (17):

$$\boldsymbol{\beta} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \boldsymbol{\eta} \quad (17)$$

4.3. Optimal Model Selection

4.3.1. SRM to Determine the Regression Variable

When selecting variables in this paper, the combustor damper, coal mill capacity air flow, and total air are selected as candidate manipulated variables. Not all the manipulated variables in Table 3 have a significant impact on the boiler efficiency, and the more manipulated variables in the regression equation are not the better. Therefore, there is a problem of how to select manipulated variables that have a significant impact on boiler efficiency. Not all the manipulated variables in Table 3 have a significant impact on the boiler efficiency, and the more manipulated variables in the regression equation are not the better. Therefore, it is necessary to select manipulated variables that have a significant impact on boiler efficiency when the regression model is established.

This paper proposes to adopt SRM to determine the optimal regression variables in each partition; its basic idea is to enter and remove manipulated variables one by one. The specific steps are as follows: each time a variable is introduced, the selected variables are checked one by one, and when the original introduced variable is no longer significant due to the introduction of subsequent variables, it is removed. The introduction or elimination of a variable requires an F test to ensure that the regression equation contains only significant variables before the introduction of the variable [41,42]. In order to prevent the infinite loop of variable introduction and elimination, the significance level needs to meet the following inequality constraints: $\alpha_{\text{entry}} < \alpha_{\text{removal}}$.

4.3.2. Model Performance Criterion

In order to ensure the accuracy of the model, this paper applies the four-fold cross-validation method. In each partition, the backup dataset is randomly divided into four

groups equally; each fold is used for model testing, and the remaining three groups are used for model training. Then, select the model with the smallest root mean square error (RMSE) among the four evaluations. The RMSE equation is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\eta_i^{TE} - \hat{\eta}_i^{TE})^2}{n}} \quad (18)$$

where n represents the number of sampling points in the test set, η_i^{TE} and $\hat{\eta}_i^{TE}$ are, respectively, the i -th real sampled value and the predicted value of the boiler efficiency in the test set.

4.4. Optimization of Manipulated Variables

The main purpose of this paper is to predict and control the boiler efficiency. In order to achieve the purpose of boiler efficiency control, a method to obtain partial derivatives of the MLR equation is proposed to calculate the optimal value of each manipulated variable:

$$\frac{\partial \eta}{\partial x_j} = \beta_j \quad (j = 1, 2, \dots, p) \quad (19)$$

where $\beta_j > 0$ represents the increase in the j -th regression variable, which can improve boiler efficiency, $\beta_j < 0$ represents the decrease in the j -th regression variable, which can improve boiler efficiency, and $\beta_j \approx 0$ represents the adjustment of the j -th regression variable, which has little effect on boiler efficiency.

For a group of real-time control data $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_p\}$ of the boiler, the real-time efficiency of the boiler under this working condition can be calculated by Equation (4). $\hat{\eta}$, obtain the efficiency optimization target value through the boiler efficiency prediction model (13) corresponding to the operating condition zone:

$$\eta_{max} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \dots + \beta_p \hat{x}_p \quad (20)$$

if $\eta_{max} - \hat{\eta} > 0$, there is room for tuning, and Equation (21) can be obtained through Equation (19):

$$\begin{aligned} \frac{\eta_{max} - \hat{\eta}}{\Delta x_j} &= \beta_j \\ \text{s.t. } x_j^{aim} &= \Delta x_j + \hat{x}_j, \quad j = 1, 2, \dots, p \end{aligned} \quad (21)$$

where Δx_j represents the optimized adjustment difference of the manipulated variable and x_j^{aim} represents the optimized target value of the manipulated variable.

5. Test Analysis and Industrial Online Application

5.1. Application Test Analysis

This paper establishes a data-driven MLR model based on the historical operating data in a specific partition to predict the boiler efficiency in the partition and calculates the optimal value of the manipulated variable according to the MLR equation to achieve the purpose of improving boiler efficiency.

5.1.1. Selection of Historical Data

In this paper, COM-Mill is selected as 111110, and the MSFlow clustering [1479.44, 1511.99], the AT clustering [16.38, 19.62], and LHVC clustering [21600, 22758] are used as partitions. Then, carry out a data-driven combustion optimization application test. First, select the sampling points that are better than the average boiler efficiency $\bar{\eta}$ and the average NOx emission $\bar{NO}_x$ in the zone as the candidate dataset for parameter fitting. Then, 800 sampling points are randomly selected from the candidate dataset and evenly divided into four groups according to the number of samples, performing four-fold cross-

validation on the model, including the training dataset (600 samples) and the test dataset (200 samples).

5.1.2. MLR Model and Prediction Performance

Based on four sets of training datasets, the SRM is used to calculate the regression variables. Take the significance level $\alpha_{entry} = 0.3$, $\alpha_{removal} = 0.35$, fit the four sets of the MLR models, and then use the four sets of the test datasets to select the optimal model through RMSE, as in Equation (22):

$$\begin{aligned} \eta = & 19.983 + 0.02x_1 - 0.042x_2 - 0.036x_4 + 0.004x_5 + 0.016x_6 + 0.203x_{10} - 0.026x_{11} \\ & + 0.03x_{13} - 0.033x_{14} + 8.581x_{16} - 0.033x_{17} + 0.07x_{19} + 0.022x_{20} - 0.014x_{21} \\ & - 0.015x_{22} - 0.013x_{23} + 0.008x_{24} + 2.222x_{25} - 1.278x_{26} + 0.03x_{28} + 0.019x_{30} \\ & - 0.02x_{31} + 0.017x_{32} - 0.007x_{34} - 0.088x_{40} + 0.108x_{42} - 0.062x_{41} - 0.017x_{47} \end{aligned} \quad (22)$$

The fitting result shows that the model with the smallest RMSE in this partition includes 28 manipulated variables. Table 8 lists the order of introducing and removing regression variables in the selection process of the SRM. Among them, x_8 , x_7 , x_{38} , x_{29} are removed after being introduced, indicating that there is a correlation between the variables. This method of selecting elements can effectively reduce the multicollinearity problem between the introduced variables and improve the quality of the model.

Table 8. Variable entry and removal order.

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Entry	x_5	x_{34}	x_{26}	x_1	x_{40}	x_{29}	x_{22}	x_7	x_{25}	x_{42}	x_4	x_{20}	x_{16}	x_{21}	x_8	x_{38}	x_6	x_{32}
Removal																		
Order	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Entry	x_{14}	x_{31}	x_{28}	x_{24}	x_{19}			x_2	x_{23}	x_{13}		x_{10}	x_{30}	x_{41}	x_{17}		x_{11}	x_{47}
Removal						x_8	x_7				x_{38}					x_{29}		

The optimal model based on Equation (22) was experimented on the training and test sets, and the prediction results of the model on the training set are shown in Figure 9, which has an RMSE of 0.39%. Figure 10 shows the prediction results of the model on the test set, and its RMSE is 0.65%. It shows that the model prediction accuracy satisfies the application demands.

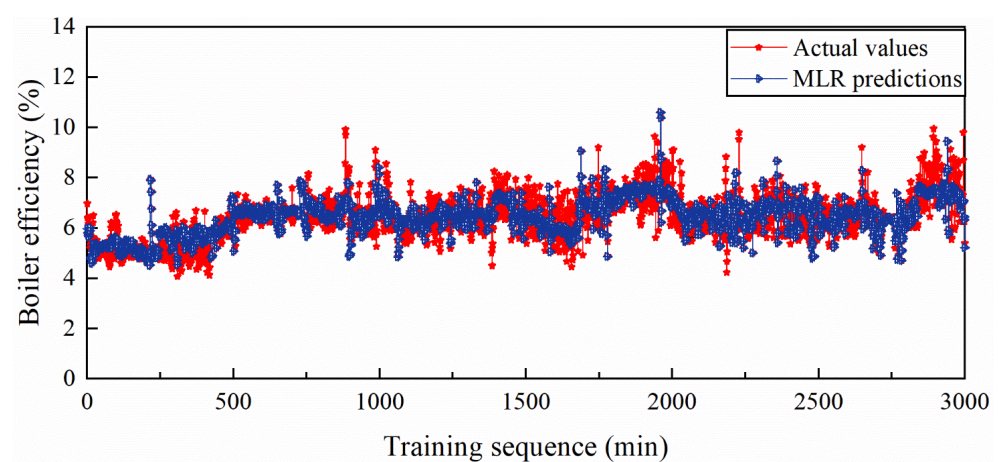


Figure 9. Prediction results on the training set.

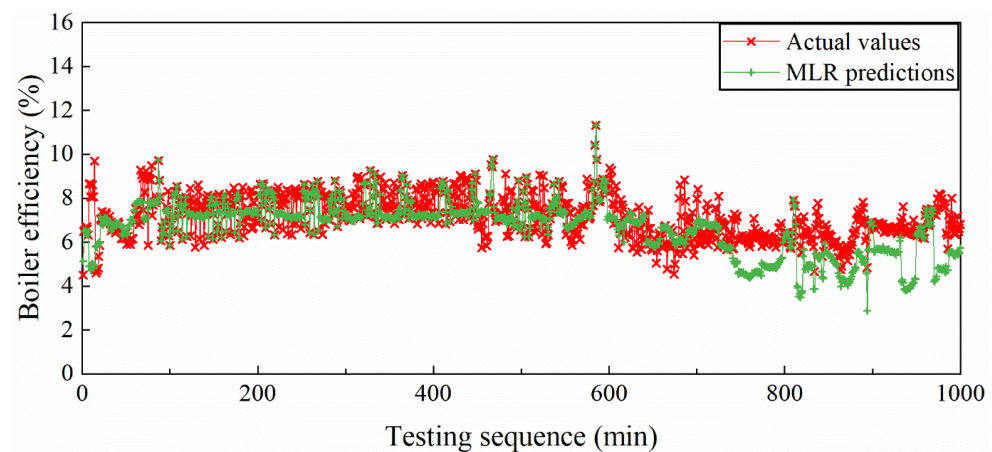


Figure 10. Prediction results on the test set.

5.2. Online Test of Combustion Optimization System

5.2.1. System Description

The combustion optimization system will optimize the control signals for the DCS, both the input of the real-time signals from the unit DCS and the output of the optimization signals from the optimization system. Figure 11 shows the online combustion optimization system. The optimized system hardware is connected to the PI real-time database of the supervisory information system (SIS). The combustion optimization model of each partition is embedded in the optimization system software. The combustion optimization system software obtains the real-time data of the units in the PI database by the utilization of Structured Query Language (SQL). The optimization model calculates the optimal value of the MVs under the current operating state based on the condition variables in the real-time operating state of the unit and transfers the optimized MV signals to the DCS. The combustion optimization system becomes a closed loop by an operator using the ON/OFF switch, and the optimized signals are involved in the boiler combustion online control. The feedback information of the optimized control signals can be reviewed through the history station of the DCS.

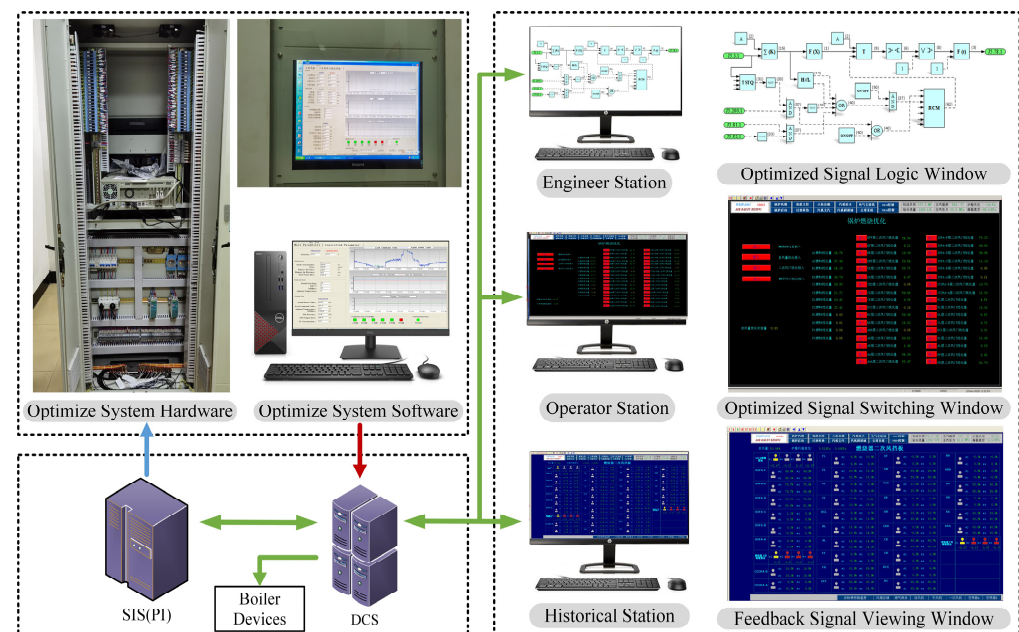


Figure 11. Online combustion optimization system.

Taking the total air optimization signal as an example, a DCS configuration logic of the signal is presented shown in Figure 12. The signal labeled as J5.3.1 is the deviation between the total air optimization control signal and the original one of the DCS. The signal labeled as J5.78.1 will be summed with the output of the original total air signal of the DCS to form a new optimal total air signal to the controlled device, which is the output value of the optimized deviation signal after judgments, speed limit, and amplitude limit. In order to reduce the disturbance of the optimized signal to the DCS and ensure the safe operation of the unit, the optimized deviation signal has a speed limit of 1%/s, and the upper and lower safety limits are $\pm 5\%$ (notes: the limits of the optimization signal will affect the optimization results).

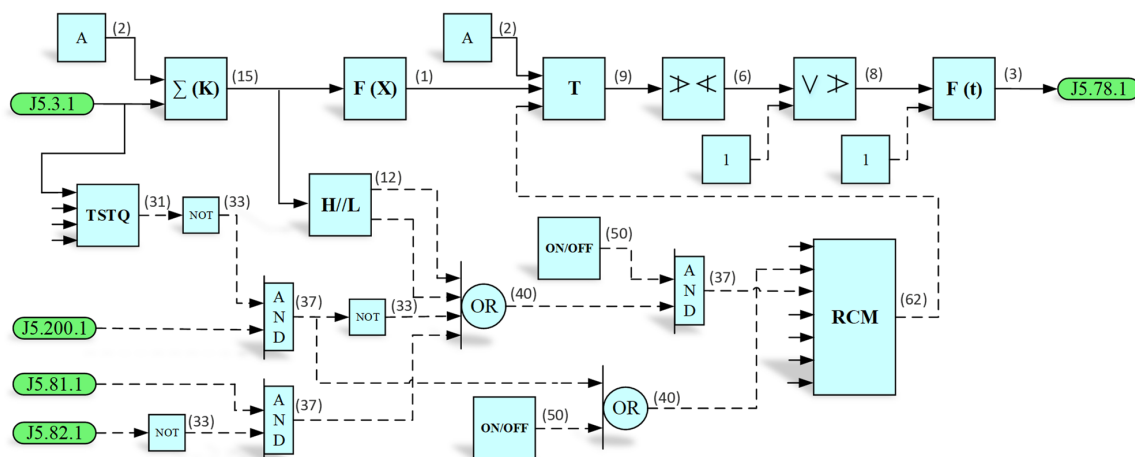


Figure 12. Logic diagram of a signal configuration for total air optimization.

Take the total air optimization signal as an example again. The processes are shown in Figure 13: the dynamic process and stable state of the optimized total air signal in the DCS after the combustion optimization system is ON. Figure 13 shows that the feedback value of the total air flow is a little lower than its setpoint, which is 70% because of the performance degradation of the equipment. After the combustion optimization system is switched ON, the total air flow optimization value (deviation) is 5%. The sum effect of the optimization value that is the dynamic response process of the total air flow can be reviewed by checking the feedback value of the total air flow. This is because the limit rate in the configuration logic is set to 1%/s in Figure 12.

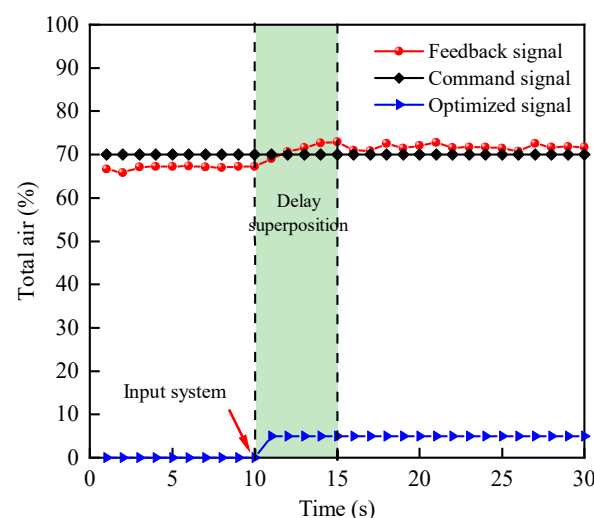


Figure 13. Example of the signal superposition process for total air optimization.

5.2.2. Full-Scale Test

The combustion optimization OCS is applied to the Weifang Power Plant unit 3. In this paper, a full-scale test is designed for the combustion optimization OCS, and the test plan uses the exit/input system for 24 h each. Table 9 shows the test time statistics. After the optimization system is put into use, the system automatically recognizes the working condition zone according to the current unit COM-Mill, MSFlow, AT, and LHVC. Match the combustion optimization model in the corresponding zone, calculate the optimized value corresponding to the controlled parameter, and transmit the optimized signal to the DCS logic to participate in the real-time control of the boiler. Figure 14 shows the main interface of the combustion optimization OCS.

Table 9. Statistics of the test time and MSFlow interval.

Operating Status	Operation Time	MSFlow (t/h)	Samples (min)
Exit	1 December 2022 10:00– 2 December 2022 10:00	[1350, 1950]	1440
Input	2 December 2022 15:00– 3 December 2022 15:00	[1300, 1750]	1440

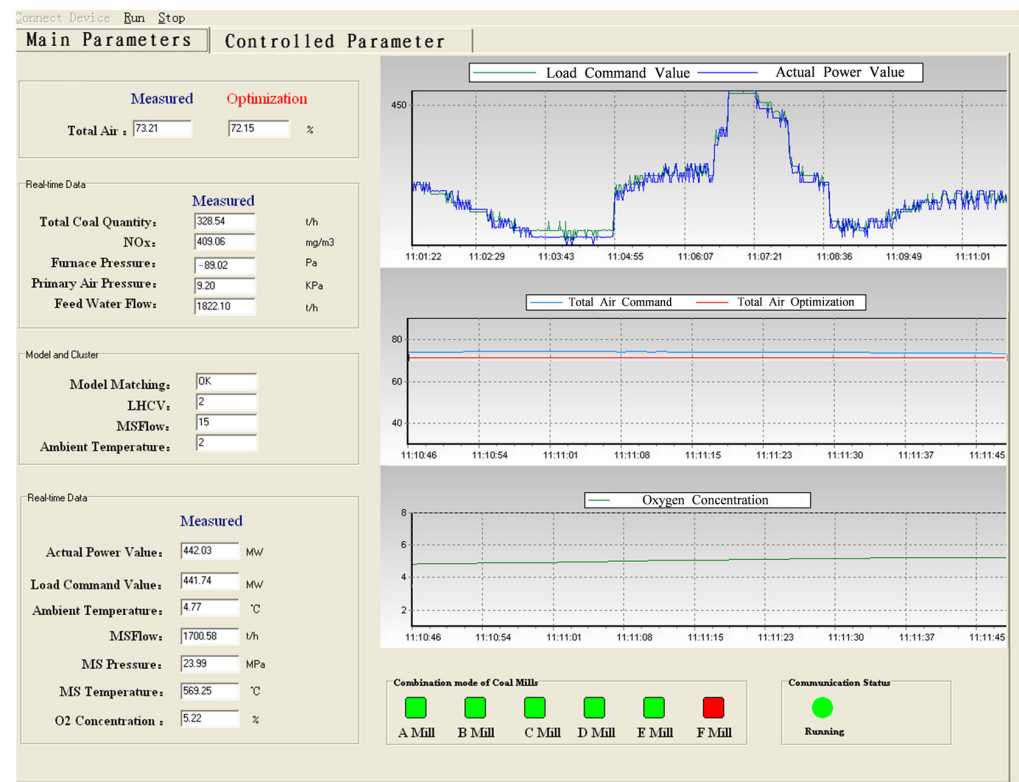


Figure 14. The main interface of the combustion optimization OCS.

5.2.3. Analysis of the Test Results

The coal quality and UBC in fly ash before and after the online control system for combustion optimization is put into use are shown in Table 10. The data come from the average of the data detection of four shifts per day. It is worth noting that UBC in fly ash under optimized operating conditions is reduced by 2.02% compared with unoptimized operating conditions.

This paper uses the MSFlow to characterize the load. Table 9 counts the MSFlow intervals before and after optimization, and the overlap interval is 1350~1750 t/h. This paper only compares and analyzes the total coal quantity, exhaust gas temperature, oxygen content in flue gas, and NO_x emissions in the MSFlow overlapping interval, as shown

in Figure 15. Among them, the total coal quantity has been converted into standard coal based on LHV. When $1350 < \text{MSFlow} < 1525 \text{ t/h}$, the optimized boiler total coal quantity is higher than the unoptimized boiler. For other MSFlow intervals, the optimized boiler's total coal quantity is significantly lower than the unoptimized boiler. When $1350 < \text{MSFlow} < 1550 \text{ t/h}$, with the increase in MSFlow, the boiler exhaust gas temperature rises rapidly under optimized conditions, which is significantly higher than under unoptimized conditions. The main reason is that the proper amount of coal is encountered with excessive air, which leads to an increase in exhaust gas flow and an increase in exhaust gas thermal loss. From the oxygen content in flue gas change trend, it can be seen that the oxygen content in flue gas in this MSFlow interval is obviously higher, which can prove that the "air-to-coal ratio" of this interval deviates from the standard value. When $\text{MSFlow} > 1550 \text{ t/h}$, with the increase in MSFlow, the boiler exhaust gas temperature tends to be stable under the optimized working conditions, and slightly drops. The main reason is that when the oxygen content in flue gas in the MSFlow interval is reasonably controlled, the "air-to-coal ratio" tends to be normal.

Table 10. Boiler combustion coal quality and UBC in fly ash before and after the test.

Operating Status	Proximate Analysis (wt%, as Received)				LHVC (KJ/kg)	UBC in Fly Ash (%)
	Moisture	Volatile Matter	Fixed Carbon	Ash		
Exit	7.23	9.86	56.91	26.00	23,066	11.63
Input	6.35	9.27	59.73	24.65	23,888	9.61

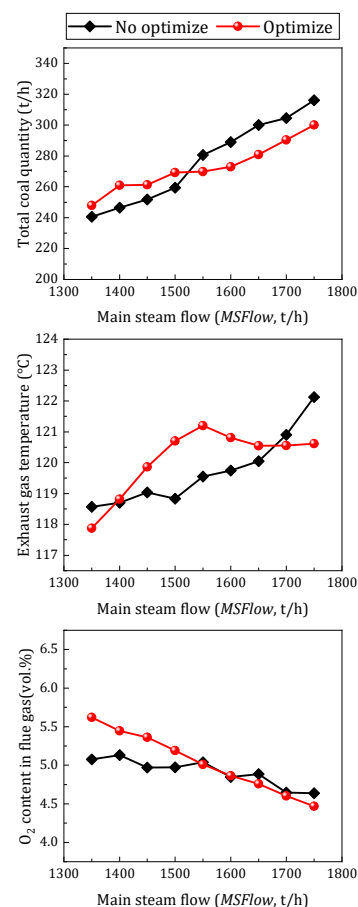


Figure 15. Before and after optimization, the total coal quantity, exhaust gas temperature, and oxygen content in flue gas vary with the MSFlow.

As shown in Figure 16, when $1350 < \text{MSFlow} < 1450 \text{ t/h}$, with the increase in MS-Flow, the NO_x emission under the optimized working condition rapidly increases from 390 mg/m^3 to 427 mg/m^3 . Combined with the analysis of the oxygen content in flue gas in Figure 12, the main reason is that the excess air coefficient in the MSFlow section is large, the strong oxidation and low reduction atmosphere in the combustion zone is enhanced, or the strong reduction and low oxidation atmosphere in the burnout zone is weakened [43]. When $\text{MSFlow} > 1450 \text{ t/h}$, as MSFlow increases, NO_x emissions under optimized conditions tend to be stable and slowly decrease. Combined with the oxygen content in flue gas, the oxygen content in flue gas in the MSFlow interval is reduced, and the excess air coefficient is restored to an appropriate level. The low oxidation atmosphere in the combustion zone inhibits the production of NO_x, and the strong reducing atmosphere in the burnout zone effectively reduces NO_x concentration.

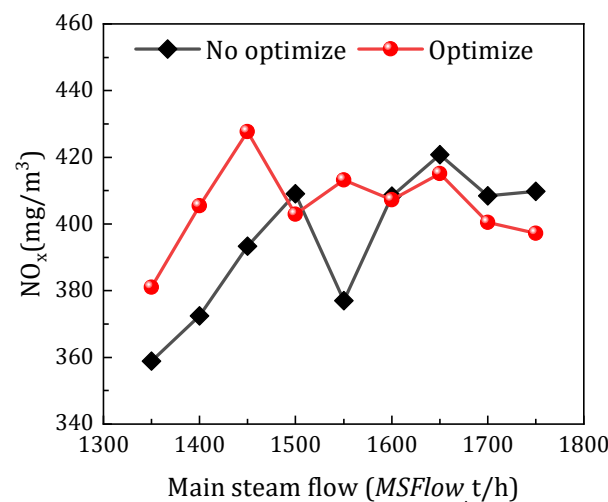


Figure 16. NO_x emissions before and after optimization change with MSFlow.

After calculation, the boiler efficiency before and after optimization increased by 0.69%. The average value of NO_x emissions decreased from 419.2 mg/m^3 to 407.1 mg/m^3 , a decrease of 2.85%. In general, optimizing the input of the system improves the economy and environmental protection of boiler operation.

6. Conclusions

In order to effectively improve boiler combustion efficiency, reduce the co-emission of NO_x and carbon dioxide, and aim at actual industrial application requirements, this paper proposes a data-driven online combustion optimization system. Boiler combustion is a non-linear process with violent fluctuations. The idea of piecewise linearization is used to divide the historical data based on the multi-variable combination of MSFlow, AT, LHVC, and COM-Mill to obtain a lot of partitions, and the linear model was used to construct local linear models for each partition. The multiple local linear models approximate global non-linearity. The prediction error of the data-driven MLR model is small and can meet industrial application requirements. The full-scale industrial test of the combustion optimization online control system based on this model also obtained effective optimization results. The results show that the average value of NO_x emissions (6%O₂) decreased by 2.85% and had an average increase of 0.39% in boiler efficiency.

The main contributions of this paper are as follows:

We quantitatively analyze the influence of ambient temperature on boiler combustion control and adopt the FCM algorithm to divide the historical data by taking the ambient temperature as the criterion with the promotion of the modeling accuracy.

The boiler historical data with non-linear characteristics are divided into a lot of data partitions, which are linear. Taking the industrial operational data of coal-fired boilers in

the Weifang Power Plant as an example, an MLR model is used to establish the mapping relationship between MVs and boiler efficiency, which is suitable for online application.

A partial derivative is used to calculate the optimal control deviation of model variables, and it participates in the online real-time closed-loop control of the boiler.

The optimal values of the MVs are calculated by the partial differential method and are transferred to the DCS of the boiler combustion system to participate in the online closed-loop control.

Author Contributions: Conceptualization, S.C.; investigation, S.X.; methodology, Z.W. and X.P.; project administration, S.C.; software, W.X.; supervision, S.C.; writing—original draft, Z.W. and G.Y.; writing—review and editing, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (51827808).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We state that the data are unavailable due to privacy or ethical restrictions of the power plant and university.

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

Nomenclature

Acronyms

DCS	Distributed Control System
SIS	Supervisory Information System
PI	Plant Information System
CFD	Computational Fluid Dynamic
ANN	Artificial Neural Network
GA	Genetic Algorithm
SVR	Support Vector Regression
MLR	Multiple Linear Regression
CCS	Carbon Capture and Storage
MSF	Main Steam Flow
AT	Ambient Temperature
LHVC	Lower Heating Value of Coal
SRA	Stepwise Regression Algorithm
FCM	Fuzzy C-Means
CV	Controlled Variable
MV	Manipulated Variable
OCS	Online Control System
LNCFS	Low NO _x Concentric Firing System
SOFA	Separated Over Fire Air
CCOFA	Close-Coupled Over Fire Air
UFA	Underfire Air
MCR	Maximum Continuous Rating
BMCR	Boiler Maximum Continuous Rating
SCR	Selective Catalytic Reduction
RMSE	Root Mean Square Error
I/O	Input/Output
SQL	Structured Query Language

Symbols

η	Defined boiler efficiency
$J^{(t)}$	The value of the objective function after the iteration
ε	Convergence condition of objective function iteration
α_{entry}	The significance level of the introduced variable
$\alpha_{removal}$	The significance level of the eliminated variable

References

1. Nakaishi, T. Developing effective CO₂ and SO₂ mitigation strategy based on marginal abatement costs of coal-fired power plants in China. *Appl. Energy* **2021**, *294*, 116978. [CrossRef]
2. United Nations. *Statement by H.E. Xi Jinping President of the People's Republic of China at the General Debate of the 75th Session of the United Nations General Assembly 2020*; Ministry of Foreign Affairs of the People's Republic of China: Beijing, China, 2020.
3. Shan, Y.; Guan, D.; Zheng, H.; Ou, J.; Li, Y.; Meng, J.; Mi, Z.; Liu, Z.; Zhang, Q. China CO₂ emission accounts 1997–2015. *Sci. Data* **2018**, *5*, 170201. [CrossRef] [PubMed]
4. The Central People's Government of the People's Republic of China. Xi Jinping: China Promises to Achieve the Time from Carbon Peak to Carbon Neutralization, Much Shorter than the Time Spent in Developed Countries. 2021. Available online: http://www.gov.cn/xinwen/2021-04/22/content_5601515.htm (accessed on 6 August 2021). (In Chinese)
5. Miličević, A.; Belošević, S.; Crnomarković, N.; Tomanović, I.; Tucaković, D. Mathematical modelling and optimization of lignite and wheat straw co-combustion in 350 MWe boiler furnace. *Appl. Energy* **2020**, *260*, 114206. [CrossRef]
6. Yang, B.; Wei, Y.M.; Hou, Y.; Li, H.; Wang, P. Life cycle environmental impact assessment of fuel mix-based biomass co-firing plants with CO₂ capture and storage. *Appl. Energy* **2019**, *252*, 113483. [CrossRef]
7. Ma, W.; Zhou, H.; Zhang, J.; Zhang, K.; Liu, D.; Zhou, C.; Cen, K. Behavior of Slagging Deposits during Coal and Biomass Co-combustion in a 300 kW Down-Fired Furnace. *Energy Fuels* **2018**, *32*, 4399–4409. [CrossRef]
8. Zhang, Y.; Pan, G.; Chen, B.; Han, J.; Zhao, Y.; Zhang, C. Short-term wind speed prediction model based on GA-ANN improved by VMD. *Renew. Energy* **2020**, *156*, 1373–1388. [CrossRef]
9. Rahat, A.A.; Wang, C.; Everson, R.M.; Fieldsend, J.E. Data-driven multi-objective optimisation of coal-fired boiler combustion systems. *Appl. Energy* **2018**, *229*, 446–458. [CrossRef]
10. Shi, Y.; Zhong, W.; Chen, X.; Yu, A.B.; Li, J. Combustion optimization of ultra supercritical boiler based on artificial intelligence. *Energy* **2019**, *170*, 804–817. [CrossRef]
11. Wang, J.G.; Shieh, S.S.; Jang, S.S.; Wong, D.S.H.; Wu, C.W. A two-tier approach to the data-driven modeling on thermal efficiency of a BFG/coal co-firing boiler. *Fuel* **2013**, *111*, 528–534. [CrossRef]
12. Wang, Q.; Chen, Z.; Han, H.; Zeng, L.; Li, Z. Experimental characterization of anthracite combustion and NO_x emission for a 300-MWe down-fired boiler with a novel combustion system: Influence of primary and vent air distributions. *Appl. Energy* **2019**, *238*, 1551–1562. [CrossRef]
13. Wu, X.; Shen, J.; Wang, M.; Lee, K.Y. Intelligent predictive control of large-scale solvent-based CO₂ capture plant using artificial neural network and particle swarm optimization. *Energy* **2020**, *196*, 117070. [CrossRef]
14. Guo, J.X.; Huang, C. Feasible roadmap for CCS retrofit of coal-based power plants to reduce Chinese carbon emissions by 2050. *Appl. Energy* **2020**, *259*, 114112. [CrossRef]
15. International Energy Agency. Data and Statistics. Available online: <https://www.iea.org/data-and-statistics?country=WORLD&fuel=Energy%20supply&indicator=TPESbySource> (accessed on 3 September 2021).
16. *The 14th Five-Year Plan (2021–2025) for National Economic and Social Development and the Long-Range Objectives through the Year 2035*; The National People's Congress: Beijing, China, 2021. (In Chinese)
17. Energy Administration of Shandong Province. Announcement on Shandong Province's 2020 Coal-Fired Unit Shutdown List (Third Batch). 2021. Available online: <https://huanbao.bjx.com.cn/news/20210113/1129216.shtml> (accessed on 10 July 2023). (In Chinese)
18. Yue, H.; Worrell, E.; Crijns-Graus, W. Impacts of regional industrial electricity savings on the development of future coal capacity per electricity grid and related air pollution emissions—A case study for China. *Appl. Energy* **2021**, *282*, 116241. [CrossRef]
19. Gupta, A.; Davis, M.; Kumar, A. An integrated assessment framework for the decarbonization of the electricity generation sector. *Appl. Energy* **2021**, *288*, 116634. [CrossRef]
20. Chen, Z.; Wang, Q.; Zhang, X.; Zeng, L.; Zhang, X.; He, T.; Liu, T.; Li, Z. Industrial-scale investigations of anthracite combustion characteristics and NO_x emissions in a retrofitted 300 MWe down-fired utility boiler with swirl burners. *Appl. Energy* **2017**, *202*, 169–177. [CrossRef]
21. Ma, L.; Fang, Q.; Tan, P.; Zhang, C.; Chen, G.; Lv, D.; Duan, X.; Chen, Y. Effect of the separated overfire air location on the combustion optimization and NO_x reduction of a 600MWe FW down-fired utility boiler with a novel combustion system. *Appl. Energy* **2016**, *180*, 104–115. [CrossRef]
22. Ma, L.; Fang, Q.; Yin, C.; Wang, H.; Zhang, C.; Chen, G. A novel corner-fired boiler system of improved efficiency and coal flexibility and reduced NO_x emissions. *Appl. Energy* **2019**, *238*, 453–465. [CrossRef]
23. Zhou, H.C.; Lou, C.; Cheng, Q.; Jiang, Z.; He, J.; Huang, B.; Pei, Z.; Lu, C. Experimental investigations on visualization of three-dimensional temperature distributions in a large-scale pulverized-coal-fired boiler furnace. *Proc. Combust. Inst.* **2005**, *30*, 1699–1706. [CrossRef]
24. An, L.S.; Ru, Y.D.; Shen, G.Q. Applications of the GMRES Algorithm in Reconstructing a Three-dimensional Temperature Field by Using the Acoustic Method. *J. Eng. Therm. Energy Power* **2015**, *30*, 88–94. (In Chinese)
25. Dal Secco, S.; Juan, O.; Louis-Louisy, M.; Lucas, J.Y.; Plion, P.; Porcheron, L. Using a genetic algorithm and CFD to identify low NO_x configurations in an industrial boiler. *Fuel* **2015**, *158*, 672–683. [CrossRef]
26. Luan, S.; Ma, Z.; Wang, H.; Zhang, Y.; Lu, P. CFD Modeling and Field Testing of a 600-MW Wall-Fired Boiler Burning Low-Volatile Bituminous Coal. In *International Symposium on Coal Combustion*; Springer: Singapore, 2016.

27. Zhou, H.; Cen, K.F.; Fan, J.R. Multi-objective optimization of the coal combustion performance with artificial neural networks and genetic algorithms. *Int. J. Energy Res.* **2005**, *29*, 499–510. [\[CrossRef\]](#)
28. Wu, F.; Zhou, H.; Ren, T.; Zheng, L.; Cen, K. Combining support vector regression and cellular genetic algorithm for multi-objective optimization of coal-fired utility boilers. *Fuel* **2009**, *88*, 1864–1870. [\[CrossRef\]](#)
29. Shin, H.; Cho, S. Response modeling with support vector machines. *Expert Syst. Appl.* **2006**, *30*, 746–760. [\[CrossRef\]](#)
30. Lv, Y.; Liu, J.; Yang, T.; Zeng, D. A novel least squares support vector machine ensemble model for NO_x emission prediction of a coal-fired boiler. *Energy* **2013**, *55*, 319–329. [\[CrossRef\]](#)
31. Wang, D.F.; Liu, Q.; Han, P.; Zhao, W.J. Combustion optimization in power station based on big data-driven case-matching. *Chin. J. Sci. Instrum.* **2016**, *37*, 420–428. (In Chinese)
32. Slišković, D.; Grbić, R.; Hocenski, Ž. Methods for plant data-based process modeling in soft-sensor development. *Automatika* **2011**, *52*, 306–318. [\[CrossRef\]](#)
33. Athanasopoulou, C.; Chatziathanassiou, V.; Athanasopoulos, G. Control of flue gas emissions based on models derived from historical plant operation data. In Proceedings of the 2011 International Conference on Clean Electrical Power (ICCEP), Ischia, Italy, 14–16 June 2011; IEEE: Piscataway, NJ, USA, 2011.
34. Zhou, H.C.; Fan, H.H.; Zhao, J.; Zeng, X. Analysis of control of total air flowrate following load demand for coal-fired utility boilers. *Clean Coal Technol.* **2019**, *25*, 18–24. (In Chinese)
35. Smrekar, J.; Potocnik, P.; Senegacnik, A. Multi-step-ahead prediction of NO_x emissions for a coal-based boiler. *Appl. Energy* **2013**, *106*, 89–99. [\[CrossRef\]](#)
36. Lv, Y.; Liu, J.; Zhao, W.; Yang, T. Steady-state detecting method based on piecewise curve fitting. *Chin. J. Sci. Instrum.* **2012**, *33*, 194–200. (In Chinese)
37. Jiang, T.; Chen, B.; He, X.; Stuart, P. Application of steady-state detection method based on wavelet transform. *Comput. Chem. Eng.* **2003**, *27*, 569–578. [\[CrossRef\]](#)
38. Gu, Y.; Zhao, W.; Wu, Z. Online adaptive least squares support vector machine and its application in utility boiler combustion optimization systems. *J. Process Control.* **2011**, *21*, 1040–1048. [\[CrossRef\]](#)
39. Zheng, W.; Wang, C.; Yang, Y.; Zhang, Y. Multi-objective combustion optimization based on data-driven hybrid strategy. *Energy* **2019**, *191*, 116478. [\[CrossRef\]](#)
40. Zhao, Y.; Wang, P.H.; Su, Z.G.; Li, Y.G.; Zhu, X.J. TS Modeling Based on Robust Fuzzy C-Regressions and Its Application for Thermal Process. *Proc. CSEE* **2018**, *38*, 2063–2069. (In Chinese)
41. He, X.; Liu, W. *Applied Regression Analysis*; China Renmin University Press: Beijing, China, 2001.
42. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1981; Volume 26.
43. Wang, Z.; Peng, X.; Cao, S.; Zhou, H.; Fan, S.; Li, K.; Huang, W. NO_x emission prediction using a lightweight convolutional neural network for cleaner production in a down-fired boiler. *J. Clean. Prod.* **2023**, *389*, 136060. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.