


Article

A Multi-Output Regression Model for Energy Consumption Prediction Based on Optimized Multi-Kernel Learning: A Case Study of Tin Smelting Process

Zhenglang Wang¹, Zao Feng^{1,2,3,*} , Zhaojun Ma⁴ and Jubo Peng⁴

¹ Faculty of Information and Automation, Kunming University of Science and Technology, Kunming 650500, China; 20222104025@stu.kust.edu.cn

² Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China

³ Yunnan International Joint Laboratory of Intelligent Control and Application of Advanced Equipment, Kunming University of Science and Technology, Kunming 650500, China

⁴ Yunnan Tin Group (Holding) Company Limited, Kunming 650126, China; zhaojunma11@163.com (Z.M.); jubopeng@ytc.cn (J.P.)

* Correspondence: zaofeng@kust.edu.cn

Abstract: Energy consumption forecasting plays an important role in energy management, conservation, and optimization in manufacturing companies. Aiming at the tin smelting process with multiple types of energy consumption and a strong coupling with energy consumption, the traditional prediction model cannot be applied to the multi-output problem. Moreover, the data collection frequency of different processes is inconsistent, resulting in few effective data samples and strong nonlinearity. In this paper, we propose a multi-kernel multi-output support vector regression model optimized based on a differential evolutionary algorithm for the prediction of multiple types of energy consumption in tin smelting. Redundant feature variables are eliminated using the distance correlation coefficient method, multi-kernel learning is introduced to improve the multi-output support vector regression model, and a differential evolutionary algorithm is used to optimize the model hyperparameters. The validity and superiority of the model was verified using the energy consumption data of a non-ferrous metal producer in Southwest China. The experimental results show that the proposed model outperformed multi-output Gaussian process regression (MGPR) and a multi-layer perceptron neural network (MLPNN) in terms of measurement capability. Finally, this paper uses a grey correlation analysis model to discuss the influencing factors on the integrated energy consumption of the tin smelting process and gives corresponding energy-saving suggestions.

Keywords: multi-kernel learning; multi-output support vector regression; differential evolutionary algorithm; energy consumption prediction



Citation: Wang, Z.; Feng, Z.; Ma, Z.; Peng, J. A Multi-Output Regression Model for Energy Consumption Prediction Based on Optimized Multi-Kernel Learning: A Case Study of Tin Smelting Process. *Processes* **2024**, *12*, 32. <https://doi.org/10.3390/pr12010032>

Academic Editors: Chao Min, Xin Ma, Xiaogang Li, Huohai Yang and Yi Man

Received: 20 November 2023

Revised: 14 December 2023

Accepted: 20 December 2023

Published: 22 December 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

The rapid development of society has been accompanied by an increasing demand for energy, and the problem of energy consumption has become increasingly serious. According to the World Energy Outlook 2022 published by the International Energy Agency (IEA), the industrial sector accounts for about 38% of the total global energy consumption and 45% of the total global CO₂ emissions, and improving energy efficiency in the industrial sector is of great significance for low-carbon sustainable development of industry [1]. In China, industrial energy consumption accounts for 67% of the total energy consumption, and the energy consumption of metal smelting accounts for more than 27% of the energy consumption of the entire manufacturing industry [2]. As a high-emission, high-energy-consumption industry, decarbonization of non-ferrous metal smelting is key to its sustainable development, and this is one of the key industries in China's efforts to

achieve the 2030 carbon peak target. Non-ferrous metals in the smelting process consume a large amount of coal, ore, and other natural mineral resources, and energy consumption is one of the most important costs in the enterprise, where energy consumption prediction to tap the potential for energy saving has a very important role [3]. In recent years, people have become more and more interested in the study of energy consumption prediction [4]. There are also many research methods in the field of energy consumption prediction, which can be broadly classified into mechanistic modelling, data-driven modelling, and hybrid modelling (combining mechanistic analysis and data-driven modelling).

Mechanism-based modelling methods, mostly for domain experts with a wealth of domain knowledge, are based on the reaction operation mechanism within the object process, using the laws involved in the process such as the laws of chemical reaction, thermodynamics, hydrodynamics, the law of conservation of energy and mass, etc., to establish a process model method [5]. This method is generally complex, and the accuracy of the model is high once the model is built correctly. At the same time, the mechanism modelling calculation process is clear, the physical meaning of the results obtained is clear, and the model is highly interpretable. K. Liddell et al. [6] conducted simulation experiments on a metal smelting process using IDEAS simulation software. Their chemical reactions were modelled through thermodynamic and chemical analyses, and the consumption of water, steam, fuel, and electricity throughout the metallurgical process was estimated on the basis of energy balance and mass balance. Umit Unver et al. [7] used AMPL software to simulate the overall production of a high steel forging plant, to calculate its minimum production energy consumption. Peng Jin et al. [8] analyzed the energy consumption as well as the carbon emissions of a steel mill roof gas recovery oxygen blast furnace based on material and energy flows. Hongming Na et al. [9] analyzed the energy consumption and carbon emission of a typical steel production process with the constraints on material parameters, process parameters, and reaction conditions of the steel production process, and with the optimization objective of maximum energy efficiency. Wenjing Wei et al. [10] analyzed the primary energy consumption and greenhouse gas emissions of nickel smelting products using a process model based on mass and energy balances. P. Coursol et al. [11] calculated the energy consumption of the smelting process of copper sulfide concentrates using thermochemical modelling and industrial data. Lei Zhang et al. [12] obtained the minimum fire loss, as well as the corresponding production cost and fire efficiency, by developing an optimization model based on the material balance, thermodynamic law, and reaction mechanism of a steel manufacturing process. However, most of the modelling work carried out by domain experts on this process has only addressed parts of the system and established local relationships between variables. These models can help to a certain extent to make qualitative judgements, whereas quantitative analysis is difficult to achieve. In the face of the high temperatures and dust in the tin smelting process, which involves complex physicochemical reactions and energy–mass conversion of multiphase flows, the key state parameters of the smelting process cannot be accurately sensed, and the establishment of a global model throughout the process, in order to achieve the provision of more valuable information for the production process, is still difficult with a mechanism-based modelling approach.

Data-driven approaches do not need to be overly reliant on process mechanisms and knowledge, and they only require an understanding of the system and data characteristics, in order to use the high-value data accumulated in the process for process modelling. In recent years, with the development of sensor and computer technologies, data-driven energy consumption prediction models have been widely used in power grids, buildings, metal smelting, and other fields. For example, in the field of power grids, A. Di Piazza et al. [13] proposed an artificial neural network-based energy prediction model for grid management, to be used for predicting hourly wind speed, solar radiation, and power demand. And their simulation analysis proved that the method had good prediction performance in the short-term time periods. Nada Mounir et al. [14] combined a modal decomposition algorithm with a bidirectional long and short-term memory network model to achieve short-term

power load forecasting for smart grid energy management systems. The superiority of the model's prediction performance was experimentally verified. Wang Yi et al. [15] used an integrated learning approach to achieve short-term nodal voltage prediction for the grid. A case study was conducted on a real distribution network to verify the effectiveness of the proposed method. In the field of buildings, Zengxi Feng et al. [16] proposed a combined prediction model for energy consumption prediction in office buildings and verified the superiority of the model with building data. Lucia Cascone et al. [17] combined short-term memory with convolutional neural networks to predict household electricity consumption using data read from smart meters. Aseel Hussien et al. [18] used the random forest algorithm to predict the thermal energy consumption of building envelope materials, which was shown using a large number of simulation results to outperform other traditional methods. In the field of metal smelting, Yang Hongtao et al. [19] proposed a dual-wavelet neural-network-based energy consumption prediction model for manganese-silicon alloy smelting and used real data to predict the electricity consumption of the smelting process. Experiments showed that the model had a higher accuracy in electricity consumption prediction. Zhaoke Huang et al. [20] proposed a hybrid support vector regression model with an adaptive state transition algorithm for predicting energy consumption in the non-ferrous metal industry. Experiments showed that this method outperformed other energy consumption prediction models. Zhen Cheng et al. [21] proposed a back propagation neural network based on genetic algorithm optimization for oxygen demand prediction model of iron and steel enterprises, and experimentally proved that the prediction accuracy of the model was better than that of the ARIMA model. Shenglong Jiang et al. [22] proposed a hybrid model integrating multivariate linear regression and Gaussian process regression for the prediction of oxygen consumption in the converter steel training process, and verified the accuracy of the model with real data. Experiments showed that the model not only achieved point prediction, but could also accurately estimate the probability interval. Zhang Qi et al. [23] proposed an artificial-neural-network-based prediction model for the supply and demand of blast furnace gas in iron and steel mills. The results showed that the established prediction model had high accuracy and small error, and it could effectively solve the prediction problem of blast furnace gas in actual production. Xiao Xiong et al. [24] proposed a random forest prediction model based on principal component degradation and artificial bee colony dynamic search fusion for the prediction of the power loss of multi-size locomotives in the control section of a strip steel hot finishing mill. The feasibility of the method was verified using real-time data at the mill level, and the experimental results showed that the method could accurately predict the power loss of multi-size rolling pieces, with a short calculation time and high prediction accuracy. Angelika Morgoeva et al. [25] proposed a machine-learning-based energy consumption prediction model. Experiments on electricity consumption prediction in metallurgical companies showed that the gradient boosting model based on the CatBoost library predicted the best results. Their data-driven modelling approach did not rely excessively on the mechanistic knowledge of the reaction process, and the energy consumption prediction model was built by analyzing the process characteristics and data features, which had high precision and fast response, but the method suffered from poor interpretability and the performance of the model was also dependent on the quality of the collected data [26]. The parameters of a data-driven model have a significant impact on the model's predictive performance; hence, optimization algorithms are often combined with the model in the modelling process to improve the model's predictions. Xu Yuanjin et al. [27] explored the effectiveness of various optimization algorithms to optimize the parameters of a multilayer perceptron model and to predict the cooling and heating loads of a building. The experimental results showed that a biogeography-based optimization (BBO) algorithm was the most applicable. In addition, multi-kernel learning is often applied to data-driven models, in order to better describe the complex patterns of data. Xian Huafeng et al. [28] proposed a multi-kernel support vector machine integration model based on unified optimization and whale optimization algorithms, and they confirmed the superiority of the model with real data. Zhang Yingda et al. [29] proposed a multi-kernel extreme learning

machine model integrating radial-based kernels and polynomial kernels, and combined this with an optimization algorithm to optimize the model parameters and finally successfully applied the model to the life prediction of batteries.

By introducing a priori knowledge into the modelling and analysis process, mechanism and data-driven hybrid models can, not only greatly improve the efficiency of modelling optimization, but also solve the problem of poor model generalization. Chengzhu Wang et al. [30] proposed a digital twin for a zinc roaster based on knowledge-guided variable mass thermodynamics. Based on the mechanism analysis of mass and energy balance, a particle swarm optimization algorithm was introduced to optimize the parameters, from which the digital twin of the roaster was constructed, and then the control strategy of the roaster was optimized. Pourya Azadi et al. [31] developed a hybrid dynamic model for the prediction of iron silica content and slag alkalinity in the blast furnace process by analyzing the principles of the blast furnace operation process. Wu Zhiwei et al. [32] proposed an energy consumption prediction model for electrofused magnesia products, consisting of a single tonne energy consumption master model for mechanistic analysis and a neural-network-based compensation model. Jie Yang et al. [33] combined a mechanistic model with a data-driven approach to achieve power demand forecasting for the electrofusion magnesium smelting process. Simulation and industrial application results showed that the effectiveness of the proposed intelligent demand forecasting method was validated.

In the face of the high temperature and dust levels in the tin smelting process, which involves complex physicochemical reactions and energy–mass conversion of multiphase flows, it is difficult to use mechanistic analysis modelling when the key state parameters of the smelting process cannot be accurately sensed, and the selection of data-driven modelling methods is more suitable for the analysis of the energy consumption of the whole process in tin smelting. Despite the limitations of data-driven modelling, these models have been heavily researched in recent years and can achieve satisfactory accuracy. Many current energy forecasting models only analyze a production process or a single energy source, but a single model cannot meet the demand for multi-output forecasting. The non-ferrous metal smelting process is accompanied by multiple types of energy consumption, and the smelting process has many processes and couplings between the energy consumption of each process; if a single model is used for prediction, potential cross-correlations between multiple outputs will be ignored. Based on this, this paper proposes a multi-kernel multi-output support vector regression prediction model optimized based on a differential evolutionary algorithm for solving the problem of predicting multiple types of energy consumption for multi-process production using a small sample dataset for the tin smelting process. Due to the limited processing power of the model algorithm, redundant variable information will affect the model performance, so a distance correlation coefficient matrix is introduced to remove redundant feature variables. The collected data are multidimensional and highly nonlinear, multi-kernel learning is combined with multi-output support vector regression to improve the model fit, and a differential evolutionary algorithm is used to find the optimal model hyperparameters. Finally, a grey correlation analysis model is applied to analyze the contribution of each energy consumption influencing factor and the comprehensive energy consumption in the tin smelting process, and corresponding energy-saving suggestions are put forward for the tin smelting process. The innovations of this study are as follows: (1) Metal smelting, as a high-energy consumption industry, consumes different types of energy, while strong coupling effects exist during the process; production data collection is also challenging due to the high-temperature environment, which results in a relatively small amount of data that is also highly non-linear. This paper proposes a multi-output support vector regression model for energy consumption prediction based on optimized multi-kernel learning and a differential evolutionary algorithm. The model overcomes the shortcoming of traditional models that only predict a single type of energy consumption. (2) By introducing multi-kernel learning into a multi-output support vector regression model, the improved model is able to maintain a satisfying prediction performance even with a small data volume. The model was validated with the production data of a tin smelting enterprise in Southwest China, and the experiment results showed that the energy

consumption prediction model proposed in this paper achieved a high prediction accuracy, as well as satisfying performance stability. This study also provides targeted guidance, according to the research conclusions, on energy planning and adjustment for enterprises.

2. Methodological and Theoretical Foundations

2.1. Data Preprocessing

Data preprocessing is the removal of outliers, missing values, and data dimensionlessness from a dataset. Missing values or outliers affect the performance of predictive models [34]. A boxplot is a commonly used method of outlier detection, which was proposed by the American statistician John Tukey in 1977 as a statistical method for displaying the characteristics of a data distribution, and outlier detection does not require the data to obey a specific distribution. When outliers or missing values are present in a dataset, common treatments include mean replacement, Lagrange interpolation, and random forest filling [35,36].

In this paper, Lagrange interpolation is chosen to fill the outliers and missing values. For a certain polynomial function, known to have given $k + 1$ values $(x_0, y_0), (x_1, y_1), \dots, (x_k, y_k)$, and assuming that any two distinct values are mutually exclusive, the polynomial obtained by applying the Lagrange interpolation formula is

$$L(x) = \sum_{j=0}^k y_j l_j(x) \quad (1)$$

$$l_j(x) = \prod_{i=0, i \neq j}^k \frac{x - x_i}{x_j - x_i} \quad (2)$$

where y_j denotes the value of the j th independent variable position, and $l_j(x)$ denotes the interpolation function.

As different input features have different dimensions, standardization is required to remove the effect of dimensionality. In addition, singular sample data in the sample may increase the computational complexity of the model; moreover, the role of features with smaller variations in a predictive model may be overwhelmed by features with larger variations [37]. Data standardization is the transformation of data into a standard state distribution with a mean of 0 and variance of 1. The transformation function is as follows:

$$x = \frac{(x - \mu)}{\sigma} \quad (3)$$

where μ in the formula denotes the mean and σ denotes the standard deviation.

2.2. Distance Correlation Coefficient Based Feature Dimensionality Reduction Method

The high dimensionality of the features leads to a more computationally intensive model, and redundant features lead to a lower prediction accuracy of the model [38]. The Pearson correlation coefficient method [39] is one of the commonly used feature selection methods, but this method can only be applied to data obeying a normal distribution and requires that the variables are linearly correlated. The distance correlation coefficient method [40] precisely compensates for the shortcomings of Pearson's algorithm and can be used to assess the correlation between linear variables, as well as the correlation between non-linear variables [41]. It is calculated as follows:

$$R^2(x, y) = \frac{c^2(x, y)}{\sqrt{c^2(x, x)c^2(y, y)}} \quad (4)$$

$$c^2(x, y) = \frac{1}{n^2} \sum_{i,j=1}^n M_{i,j} N_{i,j} \quad (5)$$

$$M_{i,j} = \|x_i - x_j\|_2 - \frac{1}{n} \sum_{k=1}^n \|x_k - x_j\|_2 - \frac{1}{n} \sum_{l=1}^n \|x_i - x_l\|_2 + \frac{1}{n^2} \sum_{k,l=1}^n \|x_k - x_l\|_2 \quad (6)$$

$$N_{i,j} = \|y_i - y_j\|_2 - \frac{1}{n} \sum_{k=1}^n \|y_k - y_j\|_2 - \frac{1}{n} \sum_{l=1}^n \|y_i - y_l\|_2 + \frac{1}{n^2} \sum_{k,l=1}^n \|y_k - y_l\|_2 \quad (7)$$

where x, y denote variables; n denotes the total number of samples.

The distance correlation method allows the calculation of the ratio of distance correlation coefficients for each feature; a larger coefficient ratio indicates a stronger correlation between the variables, while a coefficient ratio of 0 indicates that the two variables are independent of each other.

2.3. Multi-Kernel Support Vector Regression

When training data using multi-output support vector regression, the kernel function makes it easier to fit the mathematical model, but in the face of multi-dimensional, non-linear, and strongly coupled data structures, a single kernel function cannot satisfy this demand [42]. In order to fit the data better and obtain more accurate predictive values, researchers combined the existing kernel functions to obtain multi-kernel support vector regression. The main multi-kernel learning methods include infinite kernel, multiscale kernel, and synthesis kernel methods [43]. In this paper, we use a multiple kernel linear combination synthesis method in synthetic kernels, where multiple kernel matrices are given respective weights [44]. All the weighting coefficients are summed to get 1. The principle of composition is shown in Figure 1.

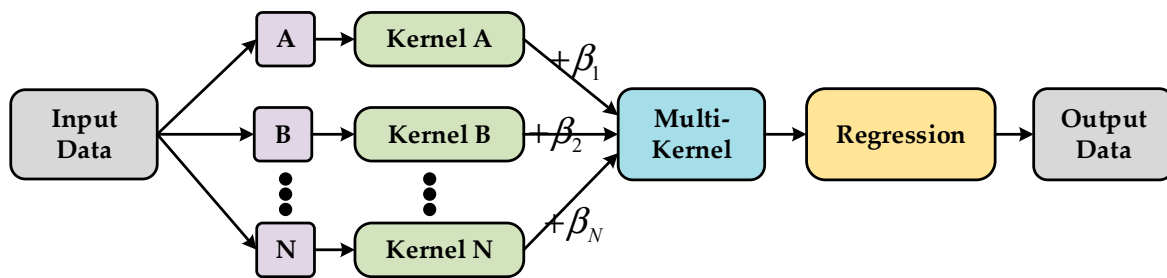


Figure 1. Principles of multicore building.

The multi-kernel function is constructed as follows:

$$\kappa^*(x_i, x_j) = \sum_{i=1}^L \beta_i \kappa(x_i, x_j) \quad (8)$$

$$s.t. \begin{cases} \beta_i \geq 0 \\ \sum_{i=1}^L \beta_i = 1 \end{cases}$$

where $\kappa^*(x_i, x_j)$ denotes the combined multiplicative kernel function and β_i denotes the weight of each kernel function.

There are four commonly used kernel functions, as shown below. The basic idea of the linear kernel function is to classify and fit the data by directly calculating the inner product of the two input parameters; the linear kernel function is simple and convenient, but only applies to linear relationships. The polynomial kernel function has more parameters than the other kernel functions for mapping data to a higher dimensional space with a polynomial function. Radial basis kernel functions are easier to compute than polynomial kernel functions but are prone to overfitting. The sigmoid kernel function is a kernel function similar to that of a multilayer perceptual neural network, whose individual layers are determined automatically in the computation.

Linear kernel function:

$$\kappa(x_i, x_j) = x_i \cdot x_j \quad (9)$$

Polynomial kernel function:

$$\kappa(x_i, x_j) = ((x_i, x_j) + c)^d, c \geq 0, d \in X^+ \quad (10)$$

Radial basis kernel function:

$$\kappa(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

Sigmoid kernel function:

$$\kappa(x_i, x_j) = \tanh(v \cdot (x_i, x_j) + c), v > 0, c < 0 \quad (12)$$

In order to reduce the dependence of the multi-kernel function on the individual base kernel functions and to reduce the computational complexity of the base kernel weights β_i , the value of the weights for each base kernel function can be determined based on the magnitude of the root mean square error (RMSE) obtained from modelling each base kernel function. This means that a basis kernel function with a smaller root mean square error will receive larger weights. The specific calculation formula is as follows:

$$\mu_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

where n is the amount of original training sample data, y_i denotes the i -th true value, and \hat{y}_i denotes the i -th predicted value.

$$\beta_i = \frac{\sum_{i=1}^L \mu_L - \mu_i}{(L-1) \sum_{i=1}^M \mu_L} \quad (14)$$

where μ_i denotes the RMSE predicted by the i -th kernel; $\sum_{i=1}^L \mu_L$ denotes the sum of the RMSEs obtained from modelling all the base kernel functions.

2.4. Multi-Output Support Vector Regression

Multi-output regression aims to learn the mapping from a multivariate input feature space to a multivariate output space [45]. The multi-output support vector regression algorithm is a new SVM algorithm proposed for the system function, whose output y is a multi-dimensional vector [46]. For a function fitting problem with input dimension M and output dimension N , let the training samples be $S = \{(x_i, y_i), i = 1, 2, 3, \dots, L\}$, where $x_i \in R^M, y_i \in R^N$. Construct the regression function as follows:

$$F(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_N(x) \end{bmatrix} = \begin{bmatrix} w^1 \Phi(x) + b^1 \\ \vdots \\ w^N \Phi(x) + b^N \end{bmatrix} = W \Phi(x)^T + B \quad (15)$$

where $\Phi(\cdot)$ is a nonlinear mapping in higher-dimensional space; W, B are regression coefficients. $W = [w^1, w^2, \dots, w^N], B = [b^1, b^2, \dots, b^N]$.

Based on the structural risk minimization principle, the regression problem is equated to the following constrained optimization problem:

$$\min L(W, B) = \frac{1}{2} \sum_{i=1}^N \|w^i\|^2 + C \sum_{i=1}^L L(u_i) \quad (16)$$

where $L(u)$ is the loss function defined on the hypersphere with the expression:

$$L(u) = \begin{cases} 0, & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2, & u > \varepsilon \end{cases} \quad (17)$$

where $u_i = \|e_i\| = \sqrt{e_i^T e_i} = y_i^T - W\Phi^T(x_i) - B$, ε is the hyperspherical insensitivity domain. When $\varepsilon = 0$, this is a least squares regression for each output component. When $\varepsilon \neq 0$, each of the regressors w_j and b_j will be solved taking into account the fit of the other output components, so that the resulting solution will be the overall best-fitting solution.

Based on the objective function and constraints, the following Lagrangian function can be obtained:

$$L(W, B) = \frac{1}{2} \sum_{i=1}^N \|w^i\|^2 + C \sum_{i=1}^L L(u_i) - \sum_{i=1}^L \alpha_i (u_i^2 - \|y_i - W\Phi(x_i)^T - B\|^2) \quad (18)$$

At the extreme points of the function, for the variables w^j, b^j, u_i, a_i , the partial derivatives are 0, and so it follows:

$$\begin{bmatrix} \Phi^T D_\alpha \Phi + I & \Phi^T \alpha \\ \alpha^T \Phi & I^T \alpha \end{bmatrix} \begin{bmatrix} w^j \\ b^j \end{bmatrix} = \begin{bmatrix} \Phi^T D_\alpha y^j \\ \alpha^T y^j \end{bmatrix} \quad (19)$$

where $\Phi = [(\phi(x_1), \dots, \phi(x_n))]^T$, $D_\alpha = \text{diag}\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, $\alpha = [\alpha_1, \dots, \alpha_n]^T$, $I = (1, 1, \dots, 1)^T$

Denoting w^j as a linear combination of the feature space and setting $w^j = \sum_i \phi(x_i) \beta^j = \Phi^T \beta^j$, Equation (19) can be expressed as

$$\begin{bmatrix} K + D_\alpha^{-1} & I \\ \alpha^T K & I^T \alpha \end{bmatrix} \begin{bmatrix} \beta^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ \alpha^T y^j \end{bmatrix} \quad (20)$$

where $K = \kappa(x_i, x_j) = \phi^T(x_i) \phi(x_j)$.

If β^j is solved, for each x one obtains $y^j = \phi^T(x) \phi(x) \beta^j$. Defining $\beta = [\beta^1, \beta^2, \dots, \beta^N]$, the N outputs can be expressed as

$$y = \phi^T(x) \phi(x) \beta = K_x \beta \quad (21)$$

2.5. Differential Evolutionary Algorithm

Differential evolutionary algorithms are optimization algorithms based on the theory of group intelligence and are intelligent optimization search algorithms that arise through cooperation and competition between individuals within a group [47]. These algorithms are very similar to genetic algorithms, in that they include mutation, crossover, and selection operations, but these operations are defined differently, and this reduces the complexity of the evolutionary computational operations using real-number coding, simple difference-based mutation operations, and a "one-to-one" competitive survival strategy [48]. The flow of a differential evolution algorithm is shown in Figure 2, which mainly includes four parts: population initialization, mutation, crossover, and selection.

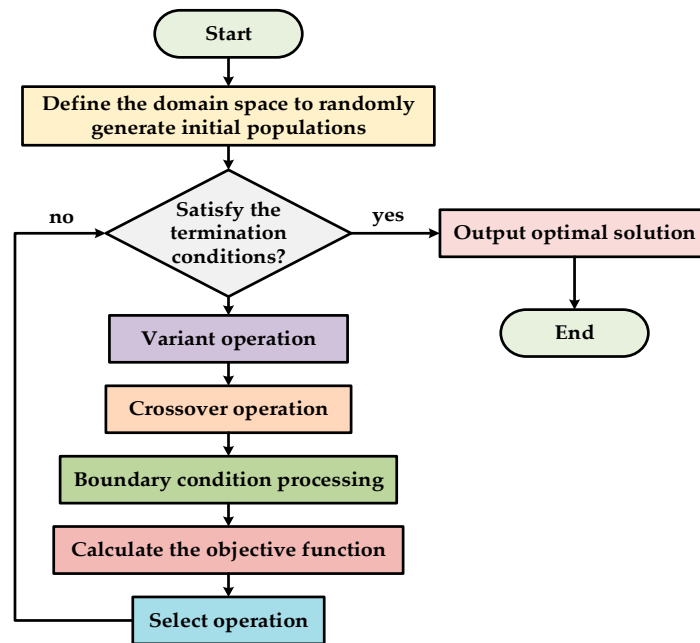


Figure 2. Differential evolution algorithm flow.

2.5.1. Population Initialization

The population initialization process is represented by Equation (22), which generally initializes individual attribute values set as random numbers between the upper and lower bounds.

$$x_{ij,0} = rand[0, 1] \times (x_j^U - x_j^L) + x_j^L \quad (22)$$

where i denotes the number of individuals in the population; j denotes the number of individual attributes; N_p denotes population size; D denotes the individual dimension; x_j^U denotes the upper bound of j -th variable; and x_j^L denotes the lower bound of the j -th variable.

2.5.2. Variation

Mutation is the operation of generating new individuals from the original individuals, and the new vector of variables is generated using the following equation:

$$v_{i,G+1} = x_{r1,G} + F \times (x_{r2,G} - x_{r3,G}) \quad (23)$$

where r_1, r_2, r_3 denote the random individual ordinal number, which are all different; F denotes the variation operator, taking the value in the range of $[0, 2]$; and G denotes the number of evolutionary generations.

2.5.3. Cross-Cutting

This is the operation of generating new individuals from mutated and current individuals according to certain rules, mainly to increase the diversity of interference parameter vectors. The operation process is as follows:

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1}, & rand(j) \leq C_R \text{ or } j = rnbr(i) \\ x_{ji,G+1}, & rand(j) \geq C_R \text{ or } j \neq rnbr(i) \end{cases} \quad (24)$$

where $rand(j)$ denotes the j -th estimate of a random number generator producing a random number between $[0, 1]$; $rnbr(i)$ denotes a randomly selected sequence; and C_R denotes the crossover operator, which takes values in the range $[0, 1]$.

2.5.4. Option

The selection operation focuses on screening the new individuals generated by the crossover operation, to select those that will go into the next generation. In this algorithm, the value of the optimization function is referred to as the adaptation value, and the screening rule functions by comparing the adaptation value of the new individual with that of the current individual, and the individual with the smallest adaptation value enters into the next generation, which ensures that the adaptation value of the individual is continuously and iteratively reduced.

2.6. Grey Correlation Analysis

Grey correlation analysis is a method of multi-factor statistical analysis. The grey correlation method is often used when the amount of data for the research object is small [49]. This method makes it possible to determine the degree of influence of each factor on the results. The specific calculation process is as follows:

Dimensionless processing of data. Due to the different physical significance of the selected influencing factors, it is not convenient and can be difficult to compare them when performing grey correlation analysis [50]. Thus, it is necessary to perform dimensionless processing first, and there are many ways to deal with dimensionless data. In this paper, we use the homogenization of each column of the data, and the calculation formula is as follows:

$$x'_i = \frac{x_i}{\bar{x}_i} \quad (25)$$

where x_i denotes the value of the i th data and \bar{x}_i denotes the mean value.

Solving Absolute Difference Sequences. Let $\Delta_i(k)$ represent the absolute difference between the respective sequence of variables and the dependent variable, which is calculated as

$$\Delta_i(k) = |Y'(k) - x'_i(k)| \quad (26)$$

Solving the sequence of correlation coefficients. Let $\zeta_i(k)$ denote the relative difference between the observed values in each period of the series of the independent variable and the observed values in the dependent variable, and the correlation coefficient is calculated as follows:

$$\zeta_i(k) = \frac{\min(\min\Delta_i(k)) + \rho\max(\max\Delta_i(k))}{\rho\max(\max(\Delta_i(k))) + \Delta_i} \quad (27)$$

where ρ denotes the resolution coefficient, the value range is $(0, 1)$, in general $\rho = 0.5$.

Solving for correlation. The degree of association at different moments in the sequence of correlation coefficients is concentrated into a single value by averaging. The formula is as follows:

$$r_i = \frac{1}{n} \sum_{k=1}^n \zeta_i(k) \quad (28)$$

The correlation degree indicates the degree of similarity and association between each evaluation item and the "reference value" (parent series), and the value of correlation degree ranges from 0 to 1. The larger the value is, the stronger the correlation between the evaluation item and the "reference value"; with a higher correlation degree, this means that the relationship between the evaluation item and the "reference value" is closer, and thus the higher the evaluation of the evaluation item [51].

2.7. Sensitivity Analysis

The Sobol sensitivity analysis method is a quantitative global sensitivity analysis method based on Monte Carlo sampling and model decomposition technology. This method can easily calculate and analyze the first-order, high-order sensitivity coefficients and total sensitivity coefficients of each input parameter on the output result, and distinguish the effects of parameter independence and parameter interaction on the output result. The calculation steps are as follows:

1. Using the Sobol sequence sampling principle, select the number of samples N and the number of independent variables D .
2. Generate an $N \times 2D$ sample matrix, set the first D columns of the matrix to matrix A , and set the last D columns to matrix B .
3. Construct an $N \times D$ matrix AB^i ($i = 1, 2, \dots, D$), and replace the j -th column in matrix A with the i -th column in matrix B . Bring the constructed set of input data into the trained model to obtain the corresponding output matrix Y .
4. Calculate the first-order sensitivity coefficient S_i and the total sensitivity coefficient S_{Ti} according to the following formula:

$$\text{Var}(Y) = \frac{1}{N-1} \sum_{i=1}^n (X_i - \bar{X}) \quad (29)$$

where N represents the number of variables; X_i represents the elements in the Y matrix; and \bar{X} represents the mean of the elements X_i .

$$\text{Var}_{X_i}[E_{X \sim i}(Y|X_i)] \approx \frac{1}{N} \sum_{j=1}^N \left\{ f(B)_j \left[f(AB^i)_j - f(A)_j \right] \right\} \quad (30)$$

where $f(X)_j$ represents the value obtained by bringing the X matrix into the model. The trained model can be regarded as a “function” between input and output.

$$E_{X \sim i}[\text{Var}_{X_i}(Y|x_{\sim i})] \approx \frac{1}{2N} \sum_{j=1}^N [f(A)_j - f(AB^i)_j]^2 \quad (31)$$

$$\text{Var}(Y) = \text{Var}(Y_A + Y_B) \quad (32)$$

$$S_i = \frac{\text{Var}[E_{X \sim i}(Y|X_i)]}{\text{Var}(Y)} \quad (33)$$

S_i is called the first-order sensitivity index, which reflects the degree of contribution of the variable X_i to the total variance of the function Y , and its value range is $[0, 1]$. The larger the index, the greater the impact of the change on the final output. In order to control changes in the final output, we must focus on controlling input variables with larger first-order sensitivity indexes.

$$S_{Ti} = \frac{E_{X \sim i}[\text{Var}_{X_i}(Y|X_{\sim i})]}{\text{Var}(Y)} \quad (34)$$

S_{Ti} is defined as the total sensitivity index of variable X_i , which reflects the degree of influence of the first-order sensitivity index of variable X_i and the cross-effect with other variables on the variance of function Y . The value range is $[0, 1]$. The total sensitivity index includes the cross-effects between each variable. A smaller total effect of the input variable indicates that the change of the variable has little impact on the change of the output, and the cross-effect between the variable and other variables has a small impact on the output. In actual calculations, in order to simplify the calculation model, variables with a small total sensitivity index can be reduced.

3. Analysis of Energy Consumption and Influencing Factors of Tin Smelting Process

3.1. Principles of Tin Smelting Process

Tin is a silver-white metal with low melting point, good ductility, soft texture, and five toxic properties, and that easily forms alloys with many metals. Tin and its alloys have good oil film retention ability and are mainly used in the production of tin-plated products, tin solder, tin alloys, tin chemical products, and float glass carriers. It has a very

wide range of uses in the food, machinery, electrical appliances, automotive, aerospace, and other industrial sectors.

Tin ingot production is mainly divided into the roasting process, melting process, refining process, and waste heat recovery process, and a flow chart is shown in Figure 3. The process includes tin concentrate, coal, and other mineral resources for roasting; with roasting sand after cooling and solvent, coal, tin slag, and other raw materials added to the top blowing furnace melting, smelting of tin slag using the smelting furnace recycling process, the smelting of the crude tin through the refining process to remove impurities, the casting process to obtain the product of tin ingots, smelting process through the recovery of waste heat, which can be used for power generation and flue gas acid production.

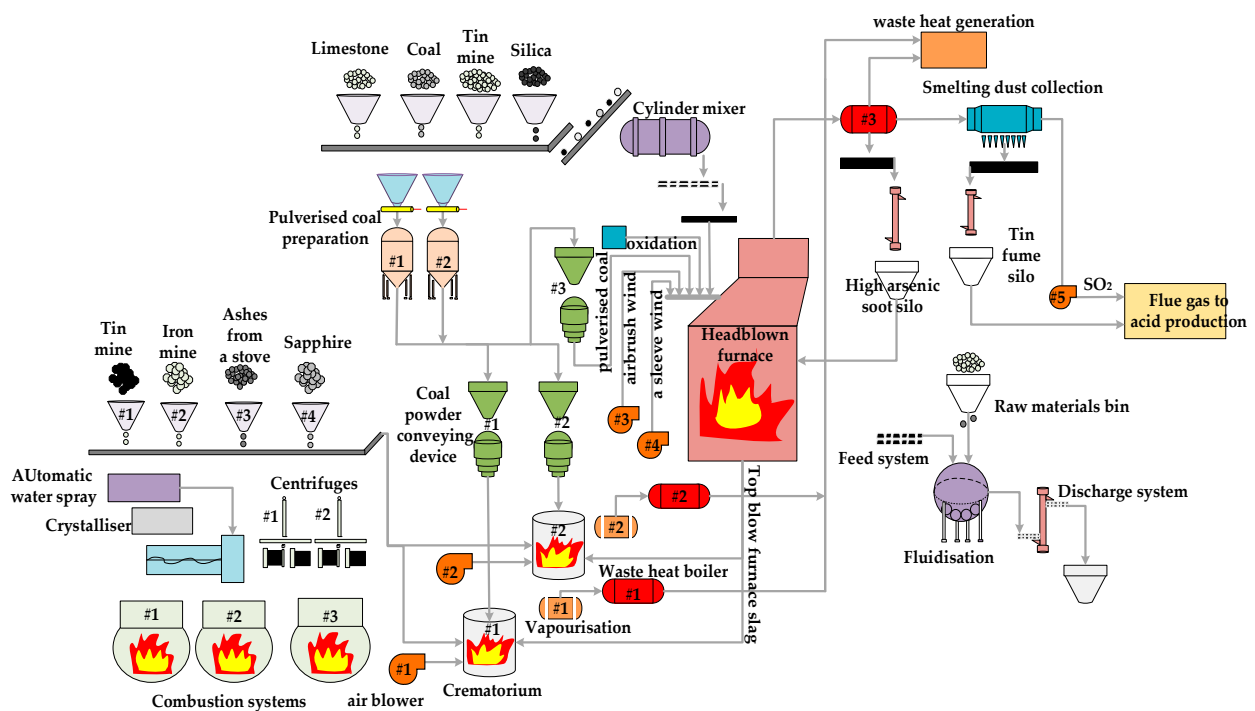


Figure 3. Tin ingot production process.

3.2. Tin Smelting Energy Consumption and Influencing Factors

The energy consumption involved in the whole process of tin smelting mainly consists of electricity, coal, water, natural gas, and oxygen; we collated the energy consumption data of each step in the whole process, analyzing and deriving a total of 20 main influencing factors on energy consumption, and the statistical results are shown in Table 1.

Table 1. Factors affecting energy consumption and types of energy consumption.

Description	Unit of Measure	Variable Name
Compressed air for roasting process	m ³	x1
Roasted sand in roasting process	t	x2
Roasting process air inlet speed	Nm ³ /h	x3
Roasting process air inlet pressure	Pa	x4
Compressed air for austenitic melting process	m ³	x5
Furnace pressure in the smelting process	Pa	x6
Crude tin in the smelting process	t	x7
Oven melting process air inlet speed	Nm ³ /h	x8
Oven melting process air inlet pressure	Pa	x9
Furnace melting process slag temperature	°C	x10
Refining process compressed air	m ³	x11

Table 1. Cont.

Description	Unit of Measure	Variable Name
Refining process solder	t	x12
Refining process air inlet speed	Nm ³ /h	x13
Refining process air inlet pressure	Pa	x14
Refining process tin ingot	t	x15
Charcoal dross in refining process	t	x16
Refining process aluminum dross	t	x17
Waste heat recovery process flue gas pressure	Pa	x18
Flue gas temperature of waste heat recovery process	°C	x19
Total tin ingot smelted	t	x20
Total electricity consumption in tin smelting process	Kwh	y1
Total coal consumption in tin smelting process	Kg	y2
Total water consumption in tin smelting process	m ³	y3
Total natural gas consumption in the tin smelting process	m ³	y4
Total oxygen consumption in the tin smelting process	m ³	y5

4. Experiments and Conclusions

This paper took a tin smelting enterprise located in southwest China as the research object. Due to the inconsistent data collection frequency of each process and in order to facilitate establishment of the model, this paper collated a total of 120 sets of production data and energy consumption data from the enterprise on a monthly basis. Aiming at the problems of multiple energy consumption in the tin smelting process, the multi-process production, and the small amount of available data, and considering the coupled relationship between different energy uses, a multi-output support vector regression prediction model was constructed. The multi-output support vector regression model was improved by introducing multi-kernel learning to improve the fitting effect of the model, and the model hyper-parameters were optimized using a differential evolutionary algorithm to further analyze the energy consumption of the smelting process as well as its potential for energy saving, and the overall framework of the experiment is shown in Figure 4.

4.1. Data Preprocessing

Directly collected process data cannot be used directly for modelling, as sensor anomalies, abnormal working conditions, data transmission failures, etc. may cause data anomalies in the production process. In addition, differences in the structure and scale of the data can also affect the prediction of the model. So, reasonable data preprocessing is extremely important [52].

The raw data were analyzed for missing values and data descriptive information; and the data were checked for outliers using box-and-line diagrams. The outlier detection results are shown in Figure 5. The vertical axis in Figure 5 represents the normalized variable values, while the horizontal axis represents the variables. The red circles represent the outliers. Outliers were not directly removed but interpolated. The detailed number of outliers and missing values, and the statistical information of the data are shown in Table 2. As can be seen from Table 2, the range of values of the variables varied too much. Some variables varied too little and some varied too much. For example, the variable x10 varied very little, with a minimum value of 1100 and a maximum value of 1200, while the variable x14 varied greatly, with a minimum value of 540,000 and a maximum value of 762,000. There were missing values for the input variables x2, x3, x6, x7, x10, x14, x16, x17, and for the target variables y2 and y4.

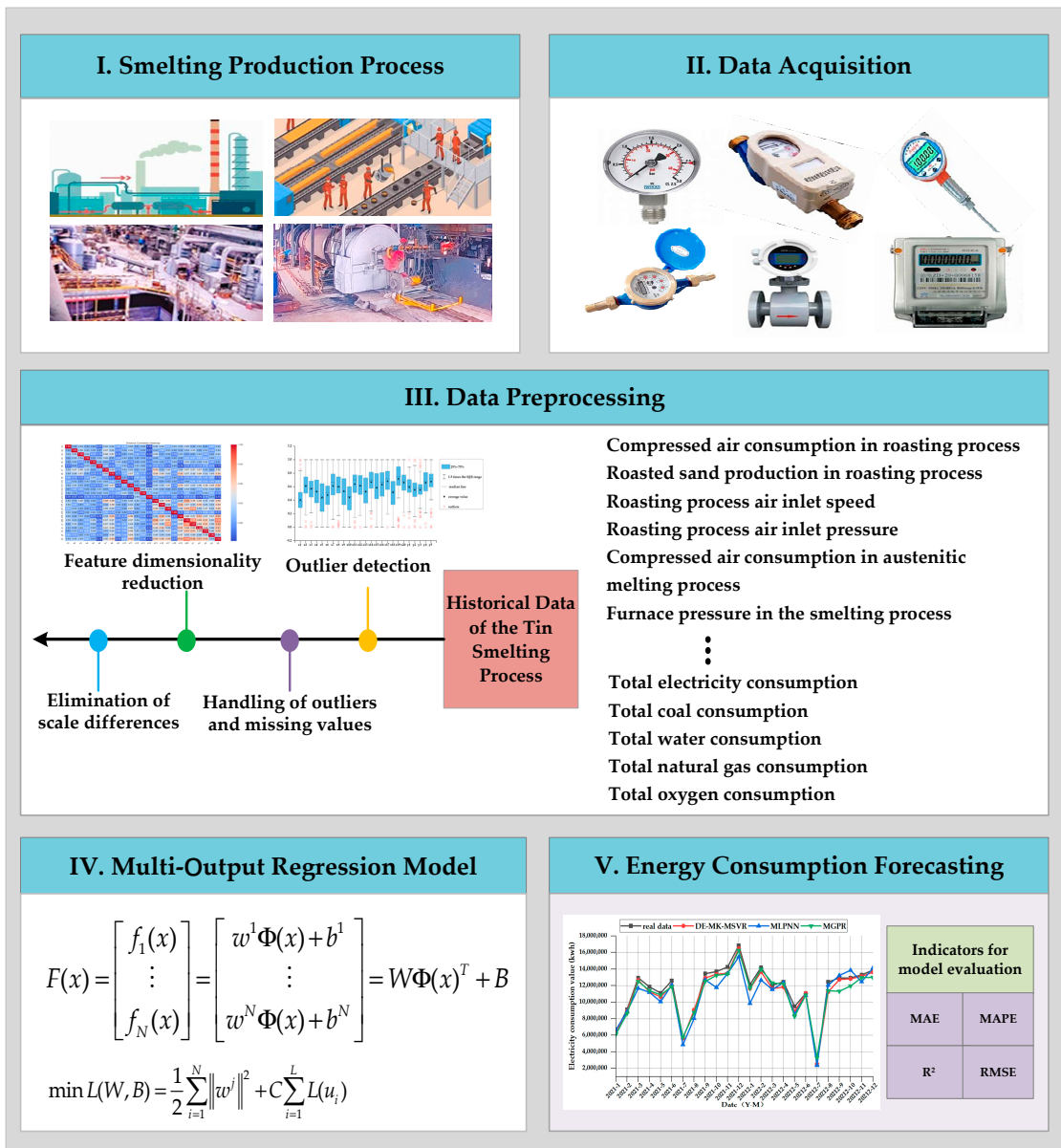


Figure 4. Overall process framework.

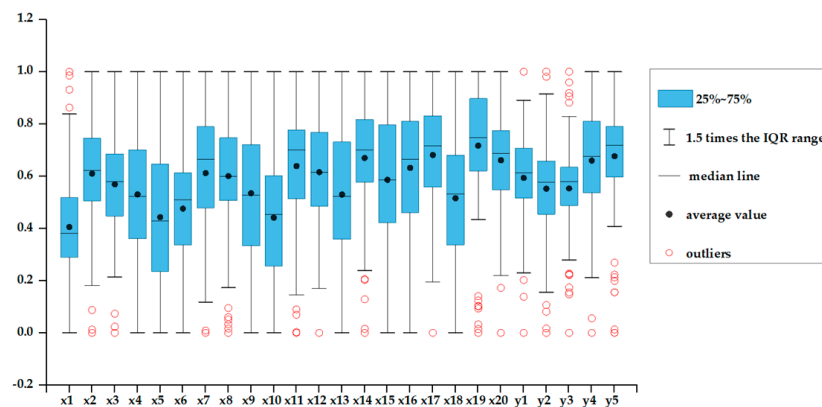


Figure 5. Box plot outlier detection.

Table 2. Statistical information for data.

Variable Name	Average Value	Standard Deviation	Minimum Value	Maximum Value	Number of Missing Values	Number of Outliers
x1	399,294	145,667	119,296	809,984	-	4
x2	6904	1785	1309	10,485	5	3
x3	6777	410	5413	7813	8	3
x4	16,056	800	14,200	17,700	-	-
x5	290,802	67,142	170,560	442,132	-	-
x6	-17	4	-27	-5	3	-
x7	6998	794	4802	8394	7	2
x8	13,345	1253	9600	15,839	-	7
x9	186,033	7031	170,000	200,000	-	-
x10	1144	23	1100	1200	4	-
x11	2305	414	1087	2994	-	3
x12	530	128	142	773	-	1
x13	490	80	319	641	-	-
x14	688,591	43,212	540,000	762,000	6	4
x15	4562	1304	1231	6920	-	-
x16	69	21	7	105	3	-
x17	221	62	13	318	9	1
x18	530,226	40,886	440,000	615,000	-	-
x19	128	32	27	168	-	7
x20	5155	1088	1318	7123	-	2
y1	10,975,904	2,361,977	2,434,472	16,837,375	-	4
y2	4,484,116	1,368,535	205,864	7,960,266	3	6
y3	9105	1716	3537	13,610	-	11
y4	311,559	80,609	39,267	451,994	5	2
y5	4,222,810	1,277,492	503	6,244,448	-	7

In this paper, Lagrange interpolation was used to fill in the anomalous and missing data. Due to the different units of measurement between variables, the range of values of each variable varied too much. In order to eliminate the influence of the scale between variables, z-score standardization was used to scale the range of values for each variable.

The predicted energy consumption in the tin smelting process includes electricity, coal, water, natural gas, and oxygen. To facilitate the subsequent analysis of the correlation between the comprehensive energy consumption and the influencing factors, this paper converted values into a unified unit of measurement. Referring to GB/T 2589-2020 “General Rules for Calculating Comprehensive Energy Consumption” issued by China in 2020 to convert each energy consumption into standard coal [53], the conversion coefficients were as shown in Table 3. The conversion factor refers to the physical amount of energy per unit of energy or the physical amount of energy consumed in the production of a unit for an energy-consuming workpiece, which was converted into the amount of standard coal. Fuel with a low-level heat value equal to 29,307.6 kilojoules (KJ) is defined as 1 kg of standard coal (1 kgce).

Table 3. Types of energy source and discounted standard coal coefficients.

Type of Energy	Variable Name	Unit of Measure	Discount Factor for Standard Coal
Electronic	y1	kwh	0.1229 (kgce/kwh)
Coal	y2	kg	0.9000 (kgce/kg)
Water	y3	m ³	0.4857 (kgce/m ³)
Natural Gas	y4	m ³	1.3300 (kgce/m ³)
Oxidation	y5	m ³	0.4000 (kgce/m ³)

4.2. Feature Analysis

Feature dimensionality reduction helps to reduce model computation and model runtime. Analyzing redundant features helps to improve the model prediction accuracy. There were 20 input variables (x1~x20) and 5 output variables (y1~y5) in this paper. By constructing the distance correlation coefficient matrix, as shown in Figure 6, the correlation coefficient between the variables was calculated; its correlation coefficient ratio ranged [0, 1]. If the correlation coefficient ratio is 0, this means that the variables are independent of each other; the closer the correlation coefficient ratio is to 1, the stronger the correlation between the variables.

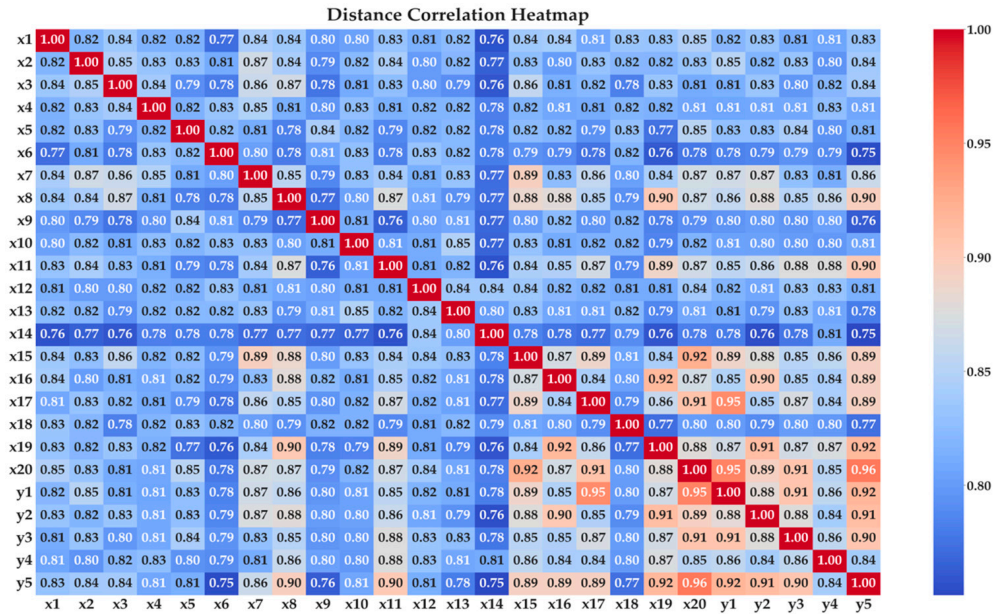


Figure 6. Distance correlation coefficient matrix.

In this paper, a correlation coefficient ratio of 0.9 was used as a threshold, and a correlation coefficient higher than 0.9 indicated a strong correlation between the variables. The specific screening method for redundant features was to compare the correlation between two strongly correlated variables with the correlation between the output variables separately and retain the set of variables that had a higher correlation with the output variables.

As can be seen in Figure 6, in the correlation matrix, darker colors indicate stronger correlation between the variables, and the ratio of correlation coefficients calculated between the variables is shown as a numerical value. The correlation coefficient ratios between variables x19 and x8, x19 and x16, and x15 and x20 were higher than 0.9, which indicated that the correlation between the variables was very strong, and for this reason, the redundant variables had to be removed. From the viewpoint of the correlation between the more redundant variables and the output variables (y1~y5), the correlation between x19 and the output variables was higher than that between x8 and x16. Therefore, the variable x19 was retained, and x8 and x16 were removed; Similarly, the variable x15 was removed and the variable x20 was retained.

4.3. Building Predictive Models

Support vector regression is suitable for solving problems such as non-linearity, small samples, and high-dimensional modelling [54]. In the tin smelting process, there are many production processes, each process has multiple types of energy consumption, and there is a coupling relationship between the energy sources. In a high-temperature environment, due to the high frequency of damage to sensors and the difficulty of maintenance, there are often only a handful of sensors installed in the field, resulting in a small number of samples of collected data. The data from the production process exhibited characteristics such as multi-

dimensional non-linearity. Thus, a multi-output support vector regression prediction model was established. In the face of a complex data structure, using single-kernel multi-output support vector regression to train the data struggles to meet the accuracy requirements of multi-output predictor variables. In order to fit the data better and obtain more accurate predictive values, the concept of multi-kernel learning was introduced, which gave the model better fitting for the multi-output problem using a linear combination of individual kernel functions. Considering that the accuracy of the prediction model was affected by the penalty coefficient C , as well as the kernel parameters, the model hyper-parameters were optimized using a differential evolutionary algorithm, the algorithmic framework of which is shown in Figure 7.

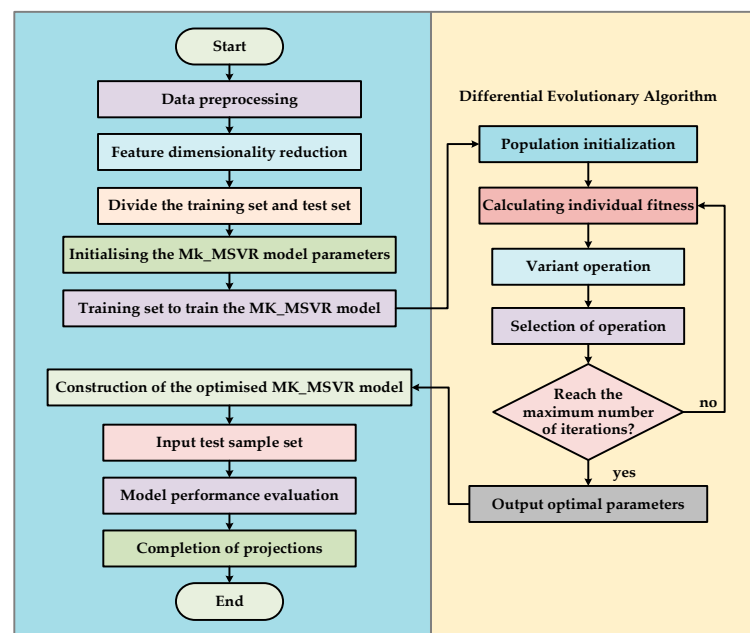


Figure 7. DE_MK_MSVM predictive modelling framework.

The algorithm framework mainly included four parts: data preprocessing, model training, parameter optimization, and model evaluation. Data preprocessing was mainly to deal with the missing values, outliers, and the data scale. Lagrange interpolation was used to deal with missing values and outliers, and the z-score algorithm was used to standardize the scales of the variables. Redundant variables were eliminated using correlation analysis of each variable, to improve the computational speed of the model. Next, 80% of the data were divided into a training set and trained on the model, with 20% of the data used for model testing. Considering that the hyperparameters of a prediction model and the parameters of the kernel function have a great impact on the performance of the model, a differential evolutionary algorithm was introduced to optimize the model parameters. The prediction performance of the model before and after optimization is then compared based on the results trained on the test set.

4.4. Projected Results

In order to demonstrate the superiority of the proposed multi-kernel multi-output support vector regression prediction model, three sets of comparison experiments were set up in this paper. The first set of experiments compared the prediction effect of MSVM under different kernel functions, considering the influence of model parameters on the prediction accuracy, all of which used a differential evolutionary algorithm to optimize the model hyperparameters. On the basis of the first set of experiments, the prediction effect of the different optimization algorithms for optimizing the hyperparameters of MK_MSVM was chosen to compare with the prediction effect of MK_MSVM, which included a particle

swarm optimization algorithm (PSO) and a Bayesian optimization algorithm (BOA). The third set of experiments investigated the prediction effect of the different models to compare with the prediction effect of the different models, and the chosen comparison models were the multi-output Gaussian process regression (MGPR) model and multi-layer perceptual machine neural network model (MLPNN). The model evaluation indexes chosen in this paper included the coefficient of determination (R^2), root mean square error (RMSE), mean error (MAE), and percentage error (MAPE).

The coefficient of determination, R^2 , is the proportion of variation in the dependent variable that can be predicted from the independent variable and is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=0}^m (y_i - \hat{y}_i)^2}{\sum_{i=0}^m (y_i - \bar{y})^2} \quad (35)$$

where m denotes the total number of samples; y_i and \hat{y}_i denote the measured and predicted values, respectively; and \bar{y} denotes the mean of the measured values.

The RMSE is the standard deviation of the residuals (prediction error). The residuals are a measure of the distance of the data points from the regression line, so the RMSE is a measure of the degree of distribution of those residuals. The formula is calculated as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (36)$$

where m denotes the total number of samples; y_i and \hat{y}_i denote the measured and predicted values, respectively.

MAE is the average size of the measurement error and is calculated as follows:

$$MAE = \frac{1}{m} \sum_{i=0}^m |y_i - \hat{y}_i| \quad (37)$$

where m denotes the total number of samples; y_i and \hat{y}_i denote the measured and predicted values, respectively.

MAPE is a measure of the predictive accuracy of forecasting methods in statistics that produces a measure of relative overall fit, calculated as follows:

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (38)$$

where m denotes the total number of samples; y_i and \hat{y}_i denote the measured and predicted values, respectively.

4.4.1. Effect of Different Kernel Functions on Predictive Models

First, this paper analyzed the energy consumption prediction effect of multi-output support vector regression with different kernel functions selected. The types of kernel function can be classified into global and local kernels, and the commonly used kernel functions are linear kernel (Lin), polynomial kernel (Poly), radial basis kernel (RBF), and sigmoid kernel. The Lin kernel is only suitable for linear relationships. The Poly kernel maps data onto high dimensional space and is suitable for nonlinear data. The RBF kernel can achieve nonlinear mapping of data but is prone to overfitting. The Sigmoid kernel function has a similar performance to the RBF kernel. Currently, there is a lack of a theoretical basis for a specific selection of the kernel function, which can only be verified through experiments. For the data situation in this paper, the polynomial kernel (Poly) with global properties and the radial basis kernel (RBF) with local properties were selected for convex linear combination. The model was trained using 80% of the data, and the remaining 20% was used as a test set for comparing the effectiveness of the prediction models. The

evaluation metrics are shown in Table 4, and the energy consumption prediction results are shown in Figure 8.

Table 4. Evaluation metrics for predictive models with different kernel functions.

Type of Energy	Different Kernel Functions	Evaluation Indicators			
		MAPE	MAE	MSE	R^2
Electronic	DE_RBF_MSVR	0.3987	0.3109	0.1172	0.9456
	DE_Poly_MSVR	0.5542	0.3256	0.1298	0.9397
	DE_MK_MSVR	0.331	0.2545	0.0829	0.9615
Coal	DE_RBF_MSVR	0.3665	0.1481	0.0449	0.9736
	DE_Poly_MSVR	0.6945	0.3253	0.1489	0.9125
	DE_MK_MSVR	0.317	0.1269	0.0268	0.9843
Water	DE_RBF_MSVR	0.3562	0.3328	0.214	0.9021
	DE_Poly_MSVR	0.8012	0.4936	0.3426	0.8433
	DE_MK_MSVR	0.3096	0.2305	0.0837	0.9617
Natural Gas	DE_RBF_MSVR	4.1449	0.3294	0.133	0.9033
	DE_Poly_MSVR	4.6763	0.3512	0.1876	0.8636
	DE_MK_MSVR	3.3537	0.2613	0.0952	0.9308
Oxidation	DE_RBF_MSVR	0.22	0.2182	0.0781	0.9655
	DE_Poly_MSVR	0.4722	0.3073	0.1267	0.9441
	DE_MK_MSVR	0.1703	0.1527	0.0364	0.9839

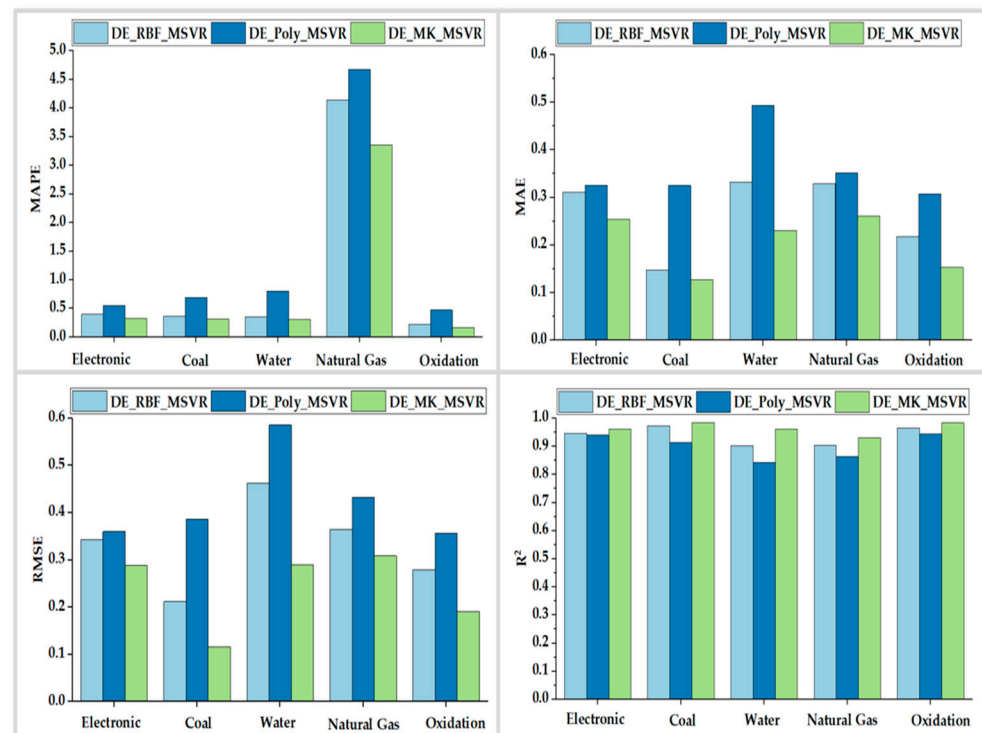


Figure 8. Comparison of evaluation indicators under different kernel functions.

As shown in Table 4, among the multi-energy prediction, the DE_MK_MSVR model was lower than the single-kernel model for the evaluation indexes MAPE, MAE, and RMSE, and its R^2 was higher than the single-kernel prediction model. Figure 8 shows the comparison results of the model evaluation indexes under different kernel functions. Smaller values of the model evaluation indexes MAPE, MAE, and RMSE indicate that the model performance was better. The evaluation index R^2 represents the model's fitting effect

on the data, and the closer the value is to 1, the better the model performance. As can be seen from Figure 8, the DE_MK_MSVM model prediction evaluation metrics were the best among all the metrics, indicating that multi-kernel learning had a better fitting effect for multi-output processing problems. Considering the prediction accuracy of the model, DE_MK_MSVM was selected for training in the subsequent work in this paper.

4.4.2. Effect of Different Optimization Algorithms on Predictive Models

As an efficient heuristic parallel search technology, the differential evolution (DE) algorithm has the characteristics of a fast convergence speed, few control parameters and simple settings, and robust optimization results [55]. The DE algorithm has excellent optimization capabilities and performs well for high-dimensional spaces and nonlinear relationships. As a typical swarm intelligence optimization algorithm, particle swarm optimization algorithm (PSO) has the characteristics of few parameters, simple principle, and easy implementation. This algorithm is more suitable for continuous optimization problems. The Bayesian optimization (BOA) algorithm based on the probability model is a very effective global optimization algorithm. It can effectively use complete historical information to improve search efficiency. It is often used in black box optimization and sequence optimization problems. Considering that the data used in this article had high-dimensional nonlinear characteristics, the DE algorithm was used to optimize the hyperparameters of the model, and the PSO algorithm and BOA algorithm were compared with them. A total of 80% of the data were used as a training set to train the model, and 20% were used as a test set to compare the performance of MK_MSVM prediction model under different optimization algorithms. The model evaluation metrics are shown in Table 5, and the energy consumption prediction results of MK_MSVM based on the different optimization algorithms are shown in Figure 9. As shown in Table 5, the DE_MK_MSVM model had the best evaluation metrics for each energy consumption prediction. The R^2 of the prediction models were all greater than 0.9, with the R^2 for coal and oxygen reaching more than 0.98. Figure 9 shows the statistical results of the evaluation indexes for the prediction of the MK_MSVM model with different optimization algorithms, from which it can be seen that the DE_MK_MSVM model had the best evaluation indexes compared to the PSO_MK_MSVM and BOA_MK_MSVM prediction models. The experimental results show that the best prediction performance was achieved using the multi-kernel multi-output regression model optimized based on the DE algorithm, which confirmed the applicability of the DE algorithm to this problem.

Table 5. Evaluation metrics of prediction models under different optimization algorithms.

Type of Energy	Different Optimization Algorithms	Evaluation Indicators			
		MAPE	MAE	RMSE	R^2
Electronic	PSO_MK_MSVM	0.5035	0.3464	0.4089	0.9224
	BOA_MK_MSVM	0.4802	0.3423	0.4089	0.9224
	DE_MK_MSVM	0.331	0.2545	0.2879	0.9615
Coal	PSO_MK_MSVM	0.8931	0.3259	0.4273	0.8927
	BOA_MK_MSVM	0.4994	0.2884	0.4119	0.9003
	DE_MK_MSVM	0.367	0.1269	0.1164	0.9843
Water	PSO_MK_MSVM	0.3876	0.2542	0.4151	0.9211
	BOA_MK_MSVM	0.406	0.3153	0.4632	0.9018
	DE_MK_MSVM	0.3096	0.2305	0.2893	0.9617
Natural Gas	PSO_MK_MSVM	4.8385	0.4489	0.5106	0.8105
	BOA_MK_MSVM	5.0288	0.4951	0.6411	0.7012
	DE_MK_MSVM	3.3537	0.2613	0.3084	0.9308
Oxidation	PSO_MK_MSVM	0.5855	0.4372	0.5135	0.8837
	BOA_MK_MSVM	0.5019	0.503	0.5948	0.8439
	DE_MK_MSVM	0.1703	0.1527	0.1909	0.9839

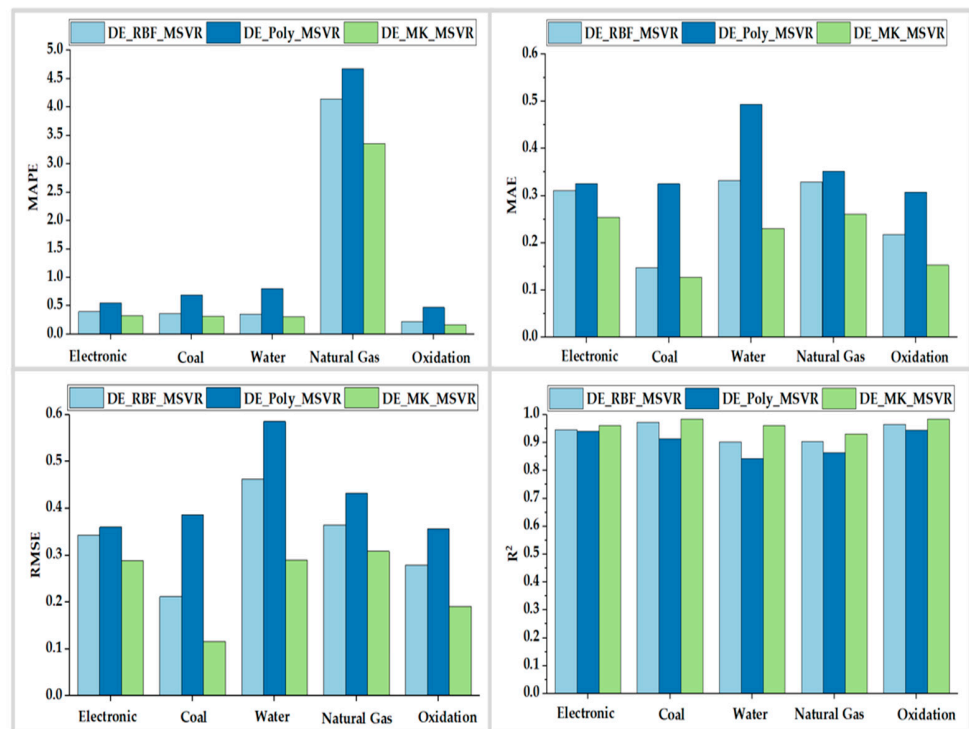


Figure 9. Comparison of evaluation metrics under different optimisation algorithms.

4.4.3. Comparison of DE-MK-MSVR with Other Multi-Output Prediction Models

Figure 10 shows the statistical evaluation indexes, and it can be seen that the R^2 of the prediction methods used in this paper were higher than 0.93, which was the highest among all the models, indicating that the model is feasible for the prediction of multi-energy consumption with small amounts of sample data. Moreover, the prediction model proposed in this paper had the smallest MAPE, MAE, and RMSE among all the prediction models.

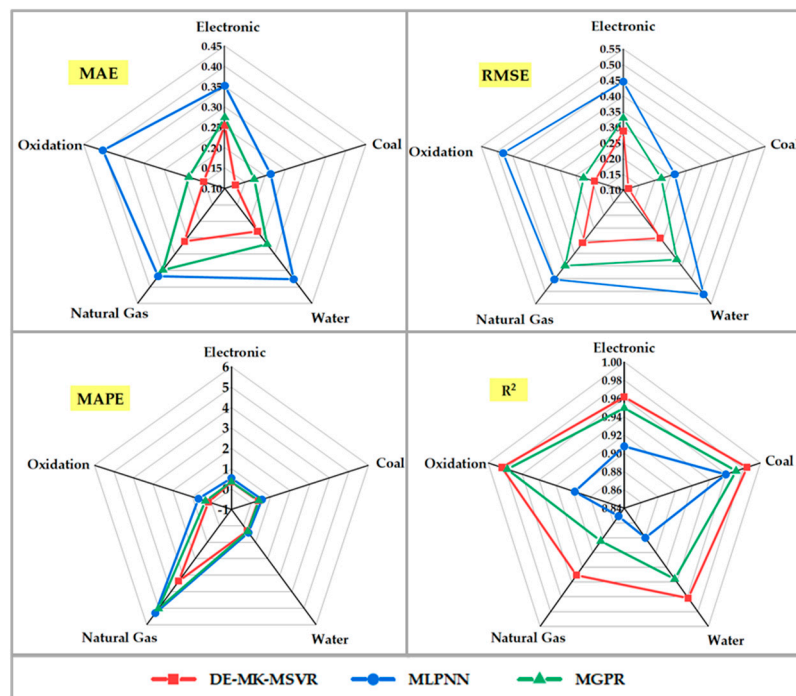


Figure 10. Comparison of evaluation indicators under different forecasting methods.

The MGPR model and MLPNN model were experimentally compared with the DE-MK-MSVR model proposed in this paper. The evaluation indexes of each model are shown in Table 6, and the energy consumption results predicted by different methods are shown in Figure 11.

Table 6. Evaluation indicators under different models.

Type of Energy	Different Forecasting Models	Evaluation Indicators			
		MAPE	MAE	RMSE	R^2
Electronic	DE-MK-MSVR	0.3310	0.2545	0.2879	0.9615
	MLPNN	0.5367	0.3523	0.4460	0.9076
	MGPR	0.3434	0.2742	0.3299	0.9495
Coal	DE-MK-MSVR	0.3670	0.1269	0.1164	0.9843
	MLPNN	0.5552	0.2138	0.2630	0.9593
	MGPR	0.3979	0.1729	0.2205	0.9714
Water	DE-MK-MSVR	0.3096	0.2305	0.2893	0.9617
	MLPNN	0.4133	0.3763	0.5117	0.8802
	MGPR	0.3436	0.2698	0.3741	0.9360
Natural Gas	DE-MK-MSVR	3.3537	0.2613	0.3084	0.9308
	MLPNN	5.2915	0.3672	0.4530	0.8509
	MGPR	5.0207	0.3480	0.3985	0.8845
Oxidation	DE-MK-MSVR	0.1703	0.1527	0.1909	0.9839
	MLPNN	0.6806	0.4022	0.4805	0.8982
	MGPR	0.3285	0.1888	0.2258	0.9775

As shown in Table 6, the DE-MK-MSVR model had the best prediction performance under multiple energy consumption. The R^2 of coal and oxygen reached more than 0.98, the R^2 of electricity and water reached more than 0.96, and the R^2 of natural gas was more than 0.93. Among all models, the DE-MK-MSVR model had the best evaluation index. As can be seen from Figure 11, among the prediction models, the DE-MK-MSVR model predicted values closer to the real values and had the best prediction performance, and the stability of the model was also the best.

Due to the high-temperature environment of the tin smelting process, sensor damage is frequent and data acquisition is expensive. In this study, there were many production processes and the data acquisition frequency of the different processes is not consistent, while the data differed in structure and scale. In addition, some process data collected were abnormal or missing due to the complexity of the smelting conditions. The data samples for modelling were small, there were strong non-linear relationships between the data, and the actual data collected did not meet a Gaussian distribution. So, the data did not predict well using the MLPNN model and MGPR model.

4.4.4. Grey Correlation Analysis and Sensitivity Analysis

The comprehensive energy consumption of the smelting process was calculated by converting each type of energy consumption into standard coal. The grey correlation analysis model was used to analyze the correlation between the factors influencing the energy consumption of the smelting process and the comprehensive energy consumption. The order of the degree of correlation between each influencing factor and the comprehensive energy consumption is shown in Table 7.

As can be seen from Table 7, the degrees of correlation between the energy consumption influencing factors and the comprehensive energy consumption were all between 0.6 and 0.8. In terms of the degree of correlation, there were seven variables with a high contribution to the energy consumption, all of which were above 0.8. They were charcoal slag in the refining process (x16), total tin ingot in the smelting process (x20), tin ingot in

the refining process (x15), flue gas temperature in the waste heat recovery process (x19), crude tin in the refining process (x7), and roasted sand in the roasting process (x2).

In order to further analyze the impact of the inputs on outputs, this paper used the Sobolj sensitivity analysis method to analyze the sensitivity of the input parameters. The results are shown in Figure 12.

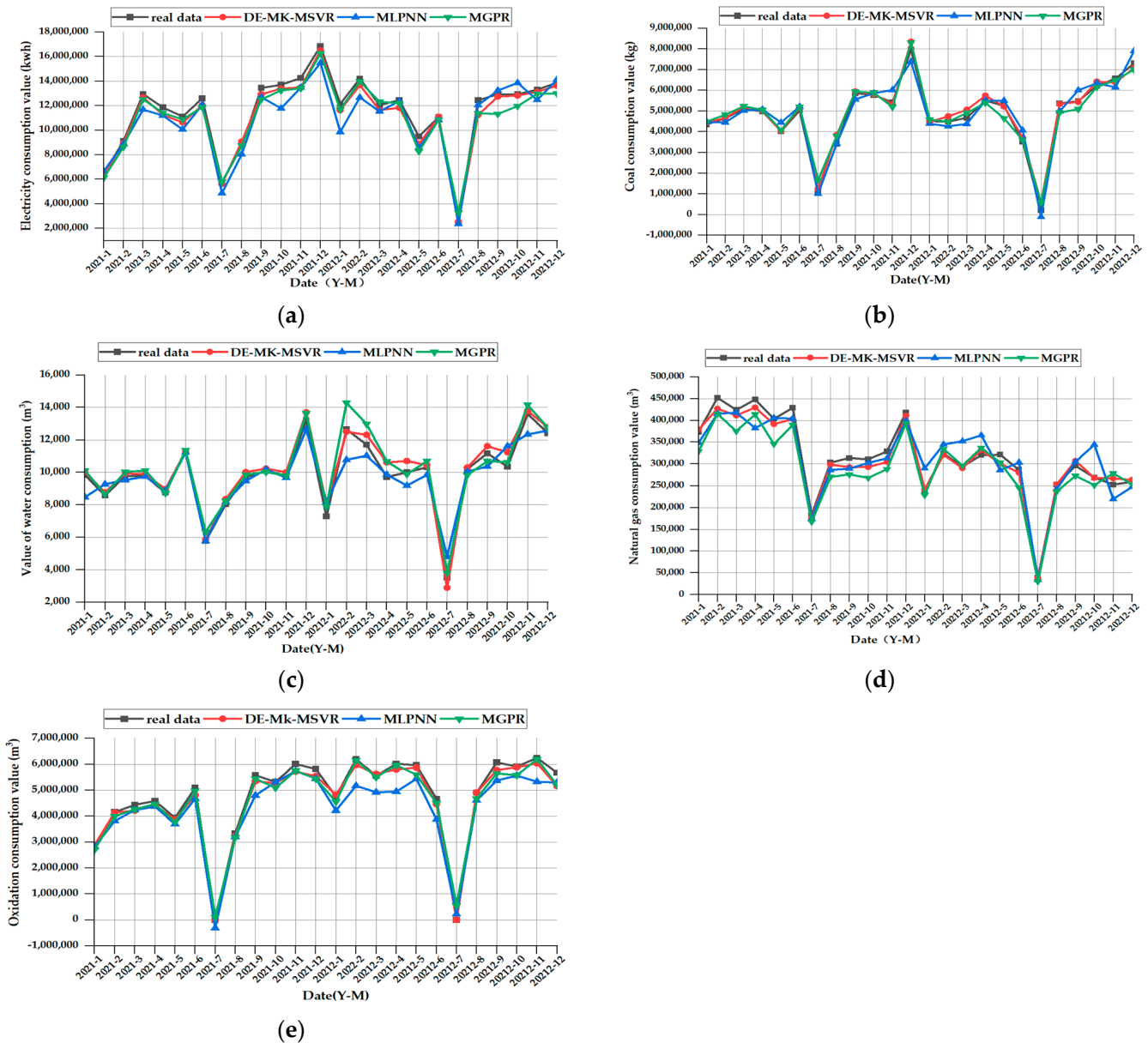


Figure 11. (a) Electricity consumption prediction under different methods; (b) coal consumption prediction under different methods; (c) prediction of water consumption under different methods; (d) prediction of natural gas consumption under different methods; (e) oxygen consumption prediction under different methods.

It can be seen from Figure 12 that total tin ingot smelted was the most sensitive variable. Its $S_i = 0.198$ and $St_i = 0.2$, this variable had a significant impact on comprehensive energy consumption; followed by the flue gas temperature of waste, where this variable had a very important effect on energy consumption, with $S_i = 0.137$ and $St_i = 0.14$. In addition, variables like the refining process for tin ingots, refining process of aluminium dross, refining process of compressed air, and the roasting sand in the roasting process played a non-negligible role in the energy consumption. The oven melting process air inlet pressure

had little impact on energy consumption, mainly because the variable value was between 0.17 and 0.2 Mpa.

Table 7. Correlation between the various factors influencing energy consumption and the overall energy consumption.

Description	Variable Name	Correlation	Order of Importance
Refining process carbon slag	x16	0.865	1
Smelting total tin ingot	x20	0.860	2
Refining process tin ingot	x15	0.850	3
Waste heat recovery process flue gas temperature	x19	0.842	4
Crude tin in the smelting process	x7	0.836	5
Roasted sand in roasting process	x2	0.818	6
Refining process aluminum slag	x17	0.801	7
Slag temperature in AUS furnace smelting process	x10	0.790	8
Inlet air speed of the smelting process in the furnace	x8	0.779	9
Roasting process air inlet speed	x3	0.777	10
Inlet air pressure in the smelting process	x9	0.775	11
Compressed air consumption in refining process	x11	0.766	12
Waste heat recovery process flue gas pressure	x18	0.760	13
Compressed air consumption in the smelting process	x5	0.754	14
Roasting process air inlet pressure	x4	0.754	15
Refining process air inlet pressure	x14	0.750	16
Roasting process compressed air consumption	x1	0.705	17
Refining process solder	x12	0.665	18
Refining process air inlet speed	x13	0.663	19
Furnace pressure for melting process	x6	0.652	20

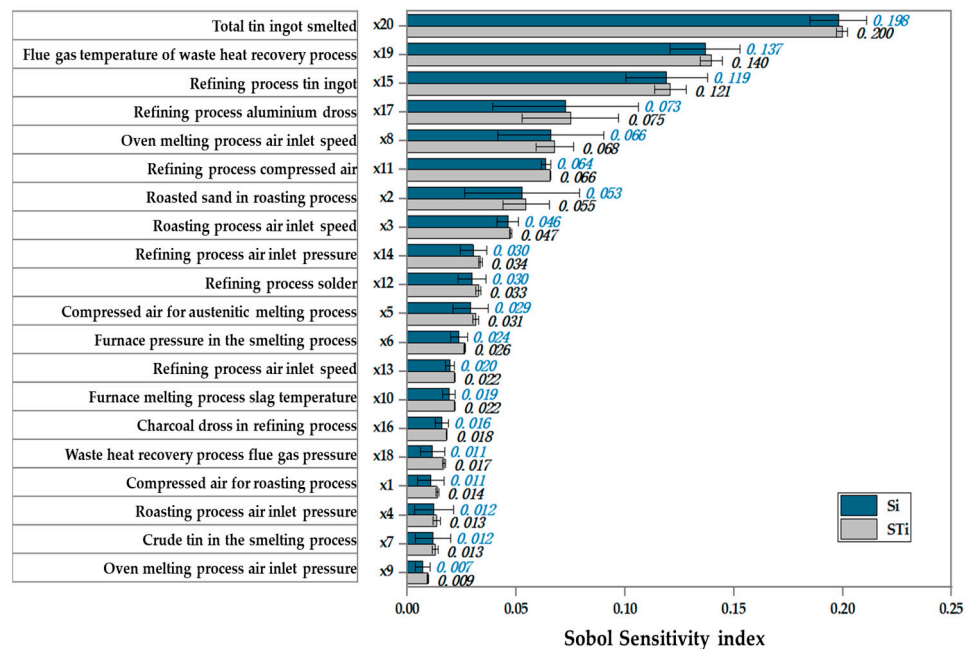


Figure 12. Sensitivity index of input parameters.

In the whole tin smelting process, the consumption of electricity was the highest; the consumption of coal and oxygen was the second highest; and the consumption of water and natural gas was low. Natural gas is a clean and environmentally friendly high-quality energy source; almost free of sulfur, dust, and other harmful substances; and produces less carbon dioxide than other fossil fuels when burned. However, enterprises still rely heavily on coal in the tin smelting process, and to achieve energy saving and carbon reduction, they should use more clean energy, rather than fossil fuels. Electricity is a high-demand energy

source for the smelting process, and the use of renewable energy sources will greatly reduce carbon emissions, such as the use of renewable energy sources such as wind and solar energy for energy supply. In addition, charcoal residues from the refining process have a significant impact on energy consumption and can be treated as an auxiliary material in the next process to improve material utilization, based on an analysis of the material and energy flows of the smelting process.

5. Energy Saving Advice

Industry associations should strengthen their guidance and assessment of energy efficiency benchmarking activities for non-ferrous metal enterprises and further improve energy-efficiency benchmarking management mechanisms. They should actively implement an energy manager system for the whole industry, standardize and establish energy management accounts, diagnose and analyze the energy application status of enterprises, study and propose energy-saving measures, explore the energy-saving potential of various production links between enterprises, measure and verify the energy-saving capacity, and establish an all-round energy management system covering all production links.

The optimization level of production scheduling in the tin smelting process greatly affects the material and energy consumption of enterprises, and production planning and scheduling ability is directly related to whether the resources of enterprises can be reasonably utilized, thus affecting the production, operation, and management efficiency of enterprises. Enterprises should pay attention to the optimization of the smelting process, in order to achieve the purpose of saving energy, consumption reduction, and efficiency increases.

The recycling of energy and materials from the smelting process is a practice worth promoting. As certain enterprises have conducted, the soot produced by each process can be collected for further reuse as an auxiliary material for the process; and the waste heat from the waste heat boiler can be recovered for power generation, making full use of existing resources. The recycling aspect of the smelting process can improve the resource utilization rate and is of great significance for reducing energy consumption. However, from the results of the parameter sensitivity analyses, the flue gas temperature of waste heat recovery had a very important impact on energy consumption. To achieve energy-efficient production, the energy consumption in the waste heat recovery process should not be neglected.

6. Conclusions

The non-ferrous metal smelting process involves multiple types of energy use, and energy consumptions are often coupled with each other. The traditional single energy consumption prediction model cannot be applied to the prediction of multiple energy consumption. Facing the problems of complex production processes, multiple types of energy consumption, and small available data samples in the tin smelting process, this paper proposed a multi-kernel multi-output support vector regression prediction model based on the optimization of a differential evolutionary algorithm. The grey correlation analysis model was used to analyze the contribution of the factors affecting the energy consumption in the tin smelting process with comprehensive energy consumption, and corresponding energy-saving recommendations were put forward based on the results of the analysis for the tin smelting process. In this paper, a DE_MK_MSVM methodology for multi-process and whole-process multi-energy consumption prediction for metal smelting processes was proposed. The main conclusions are as follows:

Aiming at multi-energy use in the smelting process, less effective data samples, strong data nonlinearity, and other characteristics, the multi-output support vector regression (MSVM) model was adopted as a benchmark. The concept of multi-kernel learning was introduced, and the kernel function of MSVM was improved using the method of linear combination. Compared with the MSVM model with single kernel function, the MSVM

model based on multiple kernel function had better prediction ability, and the multiple kernel function could better discover the relationships hidden in the data.

The hyperparameters of the prediction model were obtained using the optimization algorithm. Comparing different optimization algorithms, DE had better tuning ability for the MK_MSVM model and the DE_MK_MSVM prediction model had the highest accuracy. To demonstrate the prediction performance of the model, DE_MK_MSVM was compared with other multi-output prediction models. The experiments showed that the DE_MK_MSVM model had the best evaluation index, which proved the superiority of this model in multi-energy prediction.

A grey correlation analysis model explored the importance of the influencing factors on energy consumption in each process on the comprehensive energy consumption. The sensitivity of the input parameters was discussed using the Sobol sensitivity analysis method, giving corresponding energy-saving suggestions for the tin smelting process. The use of clean energy for smelting, such as natural gas, wind energy, solar energy, and other resources is conducive to achieving energy savings and efficiency; in addition, recycling and processing of soot and dust in various stages of smelting and generating electricity from waste heat is a practice worthy of being advocated, and technological inputs to the recycling process should be increased to improve energy efficiency.

In future studies, we will work on the following problems that may be of interest for industrial applications and scientific research: (1) The development of energy-intensive processes in process industries towards the direction of massiveness, integration, and scalability. A hybrid approach of mechanism analysis and data-driven modeling will be introduced into the modelling and analysis process, which may significantly improve not only the modelling efficiency, but also solve the problem of poor model generalization ability. (2) Appropriate virtual samples will be generated by combining domain prior knowledge, and they will be added to the training samples to achieve data expansion and feature enhancement, which in turn may improve the generalization ability of the model.

Author Contributions: Conceptualization, Z.F. and Z.W.; formal analysis, Z.M., J.P. and Z.W.; methodology, Z.W. and Z.F.; Software, Z.W. and J.P.; validation, Z.F., Z.M. and J.P.; writing—original draft, Z.W. and Z.F.; writing—review and editing, Z.W. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61563024).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Zhaojun Ma and Jubo Peng were employed by the Yunnan Tin Group (Holding) Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. IEA. World Energy Outlook 2022. Available online: <https://www.iea.org/reports/world-energy-outlook-2022> (accessed on 15 August 2023).
2. Bureau of Statistics of the People's Republic of China. China Statistical Yearbook 2020. Available online: <http://www.stats.gov.cn/sj/ndsj/2020/indexch.htm> (accessed on 18 August 2023).
3. Liu, C.; Su, Z.; Zhang, X. A data-driven evidential regression model for building hourly energy consumption prediction with feature selection and parameters learning. *J. Build. Eng.* **2023**, *80*, 107956. [CrossRef]
4. Sun, Y.; Haghighat, F.; Fung, B.C. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* **2020**, *221*, 110022. [CrossRef]
5. Weihua, G.; Chunhua, Y.; Xiaofang, C.; Yalin, W. Selected issues and challenges in the modelling and optimisation of non-ferrous metallurgical processes. *J. Autom.* **2013**, *39*, 197–207. (In Chinese)
6. Liddell, K.; Newton, T.; Adams, M.; Muller, B. Energy consumption for kelle hydrometallurgical refining versus conventional pyrometallurgical smelting and refining of pgm concentrates. *J. S. Afr. Inst. Min. Metall.* **2011**, *111*, 127–132.
7. Unver, U.; Kara, O. Energy efficiency by determining the production process with the lowest energy consumption in a steel forging facility. *J. Clean. Prod.* **2019**, *215*, 1362–1370. [CrossRef]

8. Jin, P.; Jiang, Z.; Bao, C.; Hao, S.; Zhang, X. The energy consumption and carbon emission of the integrated steel mill with oxygen blast furnace. *Resour. Conserv. Recycl.* **2017**, *117*, 58–65. [[CrossRef](#)]
9. Na, H.; Sun, J.; Qiu, Z.; Yuan, Y.; Du, T. Optimization of energy efficiency, energy consumption and CO₂ emission in typical iron and steel manufacturing process. *Energy* **2022**, *257*, 124822. [[CrossRef](#)]
10. Wei, W.; Samuelsson, P.B.; Tilliander, A.; Gyllenram, R.; Jönsson, P. Energy consumption and greenhouse gas emissions of nickel products. *Energies* **2020**, *13*, 5664. [[CrossRef](#)]
11. Coursol, P.; Mackey, P.J.; Díaz, C.M. Energy consumption in copper sulphide smelting. *Proc. Copp.* **2010**, *2*, 649–668.
12. Zhang, L.; Na, H.; Yuan, Y.; Sun, J.; Yang, Y.; Qiu, Z.; Che, Z.; Du, T. Integrated optimization for utilizing iron and steel industry's waste heat with urban heating based on exergy analysis. *Energy Conv. Manag.* **2023**, *295*, 117593. [[CrossRef](#)]
13. Di Piazza, A.; Di Piazza, M.C.; La Tona, G.; Luna, M. An artificial neural network-based forecasting model of energy-related time series for electrical grid management. *Math. Comput. Simul.* **2021**, *184*, 294–305. [[CrossRef](#)]
14. Mounir, N.; Ouadi, H.; Jrhilifa, I. Short-term electric load forecasting using an emd-bi-lstm approach for smart grid energy management system. *Energy Build.* **2023**, *288*, 113022. [[CrossRef](#)]
15. Wang, Y.; Von Krannichfeldt, L.; Zufferey, T.; Toubeau, J. Short-term nodal voltage forecasting for power distribution grids: An ensemble learning approach. *Appl. Energy* **2021**, *304*, 117880. [[CrossRef](#)]
16. Feng, Z.; Zhang, M.; Wei, N.; Zhao, J.; Zhang, T.; He, X. An office building energy consumption forecasting model with dynamically combined residual error correction based on the optimal model. *Energy Rep.* **2022**, *8*, 12442–12455. [[CrossRef](#)]
17. Cascone, L.; Sadiq, S.; Ullah, S.; Mirjalili, S.; Siddiqui, H.U.R.; Umer, M. Predicting household electric power consumption using multi-step time series with convolutional lstm. *Big Data Res.* **2023**, *31*, 100360. [[CrossRef](#)]
18. Hussien, A.; Khan, W.; Hussain, A.; Liatsis, P.; Al-Shamma'A, A.; Al-Jumeily, D. Predicting energy performances of buildings' envelope wall materials via the random forest algorithm. *J. Build. Eng.* **2023**, *69*, 106263. [[CrossRef](#)]
19. Yang, H.; Li, X.; Zhang, N. Predictive model of Mn-Si Alloy Smelting Energy Consumption based on Double Wavelet Neural Network. In Proceedings of the 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering, Changchun, China, 24–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; Volume 3, pp. 267–270.
20. Huang, Z.; Yang, C.; Zhou, X.; Yang, S. Energy consumption forecasting for the nonferrous metallurgy industry using hybrid support vector regression with an adaptive state transition algorithm. *Cogn. Comput.* **2020**, *12*, 357–368. [[CrossRef](#)]
21. Cheng, Z.; Zhang, P.; Wang, L. Oxygen demand forecasting and optimal scheduling of the oxygen gas systems in iron-and steel-making enterprises. *Appl. Sci.* **2023**, *13*, 11618. [[CrossRef](#)]
22. Jiang, S.; Shen, X.; Zheng, Z. Gaussian process-based hybrid model for predicting oxygen consumption in the converter steelmaking process. *Processes* **2019**, *7*, 352. [[CrossRef](#)]
23. Zhang, Q.; Gu, Y.L.; Ti, W.; Cai, J.J. Supply and demand forecasting of blast furnace gas based on artificial neural network in iron and steel works. *Adv. Mater. Res.* **2012**, *443*, 183–188. [[CrossRef](#)]
24. Xiong, X.; Daoming, D.; Yuxiong, X.; Qiang, G.; Yongjun, Z. Research on prediction method of finish rolling power consumption of multi-specific strip steel based on random forest optimization model. In Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 5977–5984.
25. Morgoeva, A.; Turluev, R.; Madaeva, M. Short-term electricity consumption forecasting for a steel enterprise. In Proceedings of the 2023 International Russian Automation Conference (RusAutoCon), Sochi, Russia, 10–16 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 843–847.
26. Yan, F.; Zhang, X.; Yang, C.; Hu, B.; Qian, W.; Song, Z. Data-driven modelling methods in sintering process: Current research status and perspectives. *Can. J. Chem. Eng.* **2023**, *101*, 4506–4522. [[CrossRef](#)]
27. Xu, Y.; Li, F.; Asgari, A. Prediction and optimization of heating and cooling loads in a residential building based on multi-layer perceptron neural network and different optimization algorithms. *Energy* **2022**, *240*, 122692. [[CrossRef](#)]
28. Xian, H.; Che, J. Unified whale optimization algorithm based multi-kernel svr ensemble learning for wind speed forecasting. *Appl. Soft. Comput.* **2022**, *130*, 109690. [[CrossRef](#)]
29. Zhang, Y.; Ma, H.; Wang, S.; Li, S.; Guo, R. Indirect prediction of remaining useful life for lithium-ion batteries based on improved multiple kernel extreme learning machine. *J. Energy Storage* **2023**, *64*, 107181. [[CrossRef](#)]
30. Wang, C.; Wang, Z.; Huang, K.; Li, Y.; Yang, C. Digital twin for zinc roaster furnace based on knowledge-guided variable-mass thermodynamics: Modeling and application. *Process Saf. Environ. Protect.* **2023**, *173*, 39–50. [[CrossRef](#)]
31. Azadi, P.; Winz, J.; Leo, E.; Klock, R.; Engell, S. A hybrid dynamic model for the prediction of molten iron and slag quality indices of a large-scale blast furnace. *Comput. Chem. Eng.* **2022**, *156*, 107573. [[CrossRef](#)]
32. Wu, Z.; Chai, T.; Wu, Y. Hybrid forecasting model for energy consumption per tonne of electrofused magnesium sand products. *J. Autom.* **2013**, *39*, 2002–2011. (In Chinese)
33. Yang, J.; Chai, T.; Luo, C.; Yu, W. Intelligent demand forecasting of smelting process using data-driven and mechanism model. *IEEE Trans. Ind. Electron.* **2018**, *66*, 9745–9755. [[CrossRef](#)]
34. Johnson, B.; Munch, S.B. An empirical dynamic modeling framework for missing or irregular samples. *Ecol. Model.* **2022**, *468*, 109948. [[CrossRef](#)]
35. Walker, S.; Khan, W.; Katic, K.; Maassen, W.; Zeiler, W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy Build.* **2020**, *209*, 109705. [[CrossRef](#)]

36. Xu, X.; Wang, W.; Hong, T.; Chen, J. Incorporating machine learning with building network analysis to predict multi-building energy use. *Energy Build.* **2019**, *186*, 80–97. [CrossRef]
37. Liu, Y.; Chen, H.; Zhang, L.; Wu, X.; Wang, X. Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in china. *J. Clean. Prod.* **2020**, *272*, 122542. [CrossRef]
38. Norouziasl, S.; Jafari, A. Identifying the most influential parameters in predicting lighting energy consumption in office buildings using data-driven method. *J. Build. Eng.* **2023**, *72*, 106590. [CrossRef]
39. Li, G.; Zhang, A.; Zhang, Q.; Wu, D.; Zhan, C. Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 2413–2417. [CrossRef]
40. Székely, G.J.; Rizzo, M.L. Partial distance correlation with methods for dissimilarities. *arXiv* **2014**, arXiv:1310.2926. [CrossRef]
41. Ruan, S.; Chen, B.; Song, K.; Li, H. Weighted naïve bayes text classification algorithm based on improved distance correlation coefficient. *Neural Comput. Appl.* **2022**, *34*, 2729–2738. [CrossRef]
42. Han, M.; Zhang, H. Multiple kernel learning for label relation and class imbalance in multi-label learning. *Inf. Sci.* **2022**, *613*, 344–356. [CrossRef]
43. Xing, M.; Zhang, Y.; Yu, H.; Yang, Z.; Li, X.; Li, Q.; Zhao, Y.; Zhao, Z.; Luo, Y. Predict dlbc patients' recurrence within two years with gaussian mixture model cluster oversampling and multi-kernel learning. *Comput. Meth. Programs Biomed.* **2022**, *226*, 107103. [CrossRef]
44. Chao, S.; Qiang, L. Multi-kernel support vector machine based on feature weighting. *J. Xi'an Univ. Posts Telecommun.* **2017**, *22*, 84–88. (In Chinese)
45. Xu, S.; An, X.; Qiao, X.; Zhu, L.; Li, L. Multi-output least-squares support vector regression machines. *Pattern Recognit. Lett.* **2013**, *34*, 1078–1084. [CrossRef]
46. Hu, Z.; Bao, Y.; Chiong, R.; Xiong, T. Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. *Energy* **2015**, *84*, 419–431. [CrossRef]
47. Price, K.; Storn, R.M.; Lampinen, J.A. *Differential Evolution: A Practical Approach to Global Optimization*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
48. Han, Y.; Peng, H.; Mei, C.; Cao, L.; Deng, C.; Wang, H.; Wu, Z. Multi-strategy multi-objective differential evolutionary algorithm with reinforcement learning. *Knowl.-Based Syst.* **2023**, *277*, 110801. [CrossRef]
49. Huang, M.; Wang, B. Evaluating green performance of building products based on gray relational analysis and analytic hierarchy process. *Environ. Prog. Sustain. Energy* **2014**, *33*, 1389–1395. [CrossRef]
50. Zhu, H.; Liao, Q.; Qu, B.; Hu, L.; Wang, H.; Gao, R.; Zhang, Y. Relationship between the main functional groups and complex permittivity in pre-oxidised lignite at terahertz frequencies based on grey correlation analysis. *Energy* **2023**, *278*, 127821. [CrossRef]
51. Xu, H.; Hu, S.; Yao, X.; Chu, Z. Research on the composition of glass artefacts based on k-means clustering and grey correlation analysis. *J. Xinjiang Norm. Univ. (Nat. Sci. Ed.)* **2023**, *42*, 66–73. (In Chinese)
52. Oluwasakin, E.; Torku, T.; Tingting, S.; Yinusa, A.; Hamdan, S.; Poudel, S.; Hasan, N.; Vargas, J.; Poudel, K. Minimization of high computational cost in data preprocessing and modeling using mpi4py. *Mach. Learn. Appl.* **2023**, *13*, 100483. [CrossRef]
53. Sac, S.A.O.C. General Principles for the Calculation of Integrated Energy Consumption. Available online: <https://openstd.samr.gov.cn/bz/gk/gb/newGbInfo?hcno=53D1440B68E6D50B8BA0CCAB619B6B3E> (accessed on 30 August 2023).
54. Peng, C.; Che, Z.; Liao, T.W.; Zhang, Z. Prediction using multi-objective slime mould algorithm optimized support vector regression model. *Appl. Soft. Comput.* **2023**, *145*, 110580. [CrossRef]
55. Zeng, Z.; Zhang, H. An evolutionary-state-based selection strategy for enhancing differential evolution algorithm. *Inf. Sci.* **2022**, *617*, 373–394. [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.