Modeling of BOF Steelmaking Based on the Data-driven Approach

Jinna Li¹,a, Yuan Li²,b, Fang Liu³,c, Shengping Bie³,d, Chengchi Tang³,e, Xiaoling Zhang⁴,f

¹Lab of Operation and Control, Shenyang University of Chemical Technology, Liaoning, Shenyang 110142, China
²College of Information Engineering, Shenyang University of Chemical Technology, Liaoning, Shenyang 110142, China
³College of Chemical Engineering, Shenyang University of Chemical Technology, Liaoning, Shenyang 110142, China
⁴Lab of Industrial Control Networks and Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Liaoning, Shenyang 110016, China

a lijinna_721@126.com; b li-yuan@mail.tsinghua.edu.cn; c fangliu@sian.com;
d spb@163.com; e chengct@126.com; f zhangxiaoling@sia.cn

Keyword: Basic oxygen furnace (BOF) steelmaking; Linear regression analysis; Response surface analysis; Particle swarm optimization

Abstract. This paper is concerned with the modeling of both endpoint temperature and carbon content for BOF steelmaking. First, a linear regression predictive model is constructed based on the linear regression analysis method. Next, the response surface analysis method is used to construct a nonlinear predictive model. The significant contribution of this paper is that response surface analysis is proposed for constructing the predictive model of BOF steelmaking. Finally, experiment simulation results show the effectiveness and advantages of the proposed methods.

Introduction

Basic oxygen furnace (BOF) Steelmaking a complexly multi-component and multi-phase Physical and chemical process, where iron and some raw materials are added into the converter, Then the oxygen lance is inserted oxygen points in the furnace mouth, and Oxygen is “blown” into the BOF at supersonic velocities. Until the carbon content and molten steel temperature satisfy the specific standard, the oxygen blowing process stops. The described process is shown in Fig. 1.

![Fig. 1. The basic oxygen furnace (BOF) steelmaking process](image_url)
based on the steelmaking process statistical data, and certain accuracy is obtained. Different from
the above-mentioned literature, response surface analysis method is utilized to construct a
temperature and carbon content of molten steel model. An experiment shows the high hit rate is
possessed by the model obtained in terms of response surface analysis method

Prediction Models

**Linear regression model.** Based on linear regression analysis theory\(^8\), the linear model of the
temperature and carbon content at the endpoint of BOF steelmaking can be described below

\[
y_T = \alpha_{T0} + \alpha_{T1} x_1 + \alpha_{T2} x_2 + \cdots + \alpha_{Tm} x_m + \epsilon
\]

\[
y_C = \alpha_{C0} + \alpha_{C1} x_1 + \alpha_{C2} x_2 + \cdots + \alpha_{Cm} x_m + \epsilon
\]

where \( y_T \) and \( y_C \) denote the endpoint temperature and end-point carbon content, respectively.
\( \alpha_{T0}, \alpha_{T1}, \ldots, \alpha_{Tm} \) and \( \alpha_{C0}, \alpha_{C1}, \ldots, \alpha_{Cm} \) are unknown parameters. Assume \( \epsilon \sim N(0, \delta^2) \), \( \epsilon \) is a statistical
variable (error) with zero expectation and variance \( \delta^2 > 0 \). We give the following algorithm to
calculate the unknown parameters \( \alpha_{T0}, \alpha_{T1}, \ldots, \alpha_{Tm} \).

**Algorithm 1**: Step1. Execute \( n \) \((n \geq 10)\) independent observations to obtain training sampling
\((x_{1i}, x_{2i}, \ldots, x_{ni}, y_i)\). In this case, (1) is rewritten as

\[
\begin{align*}
y_1 &= \alpha_{T0} + \alpha_{T1} x_{11} + \alpha_{T2} x_{12} + \cdots + \alpha_{Tm} x_{1m} + \epsilon_1 \\
y_2 &= \alpha_{T0} + \alpha_{T1} x_{21} + \alpha_{T2} x_{22} + \cdots + \alpha_{Tm} x_{2m} + \epsilon_2 \\
&\vdots \\
y_n &= \alpha_{T0} + \alpha_{T1} x_{ni} + \alpha_{T2} x_{n2} + \cdots + \alpha_{Tm} x_{nm} + \epsilon_n
\end{align*}
\]

where \( \epsilon_1, \epsilon_2, \ldots, \epsilon_n \) are independent and subject to \( N(0, \delta^2) \). Set

\[
X = \begin{bmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1m} \\
1 & x_{21} & x_{22} & \cdots & x_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{ni} & x_{n2} & \cdots & x_{nm}
\end{bmatrix}, \quad
Y = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}, \quad
\alpha_T = \begin{bmatrix}
\alpha_{T0} \\
\alpha_{T1} \\
\vdots \\
\alpha_{Tm}
\end{bmatrix}, \quad
\epsilon = \begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_n
\end{bmatrix}
\]

(3) is simplified as the following matrix form

\[
Y = X \alpha_T + \epsilon
\]

where \( Y \) is the observation vector, \( X \) is the designed matrix. Here, \( X \) is column full rank, \( \alpha_T \) need to be determined, \( \epsilon \) is an observation of the random error vector, (4) is called as the matrix
formulation of the linear regression model. If \( Y \) and \( x_1, x_2, \ldots, x_m \) satisfy model (1), we have

\[
S(\alpha_T) = \sum_{i=1}^{n} \epsilon^2 = \epsilon^T \epsilon = (Y - X \alpha_T)^T (Y - X \alpha_T) = \sum_{i=1}^{n} (y_i - \sum_{j=0}^{m} x_{ij} \alpha_j)^2
\]

(5)

where \( x_{i0} (i = 1, 2, \ldots, n) \). For each \( \alpha_{T0}, \alpha_{T1}, \ldots, \alpha_{Tm} \), we have

\[
\frac{\partial S(\alpha_T)}{\partial \alpha_{T_k}} = - \sum_{i=1}^{n} (y_i - \sum_{j=0}^{m} x_{ij} \alpha_j) x_{ik} = 0, \quad k = 0, 1, \ldots, m
\]

and
Then we derive normal equations $X^T X \alpha_f = X^T Y$. Solve it, we can obtain the least squares estimation of $\alpha_f$:

$$\hat{\alpha}_f = (X^T X)^{-1} X^T Y$$

By (4), we have $E(Y) = X \alpha_f$, then $E(\hat{\alpha}_f) = (X^T X)^{-1} X^T E(Y)$, that is $\hat{\alpha}_f$ is an unbiased estimator of $\alpha_f$. Put $\hat{\alpha}_f = (\hat{\alpha}_{f0}, \hat{\alpha}_{f1}, \ldots, \hat{\alpha}_{fm})$ into (1) and omit the error term, then $y_f = \hat{\alpha}_{f0} + \hat{\alpha}_{f1} x_1 + \hat{\alpha}_{f2} x_2 + \cdots + \hat{\alpha}_{fm} x_m$ is called regression equation.

Step 2. Similarly, we derive the carbon content at the endpoint linear regression model in terms of linear regression theory.

$$\hat{y}_c = \hat{\alpha}_{c0} + \hat{\alpha}_{c1} x_1 + \hat{\alpha}_{c2} x_2 + \cdots + \hat{\alpha}_{cm} x_m$$

Step 3. The $y_f$ and $y_c$ can be predicted from the observed data of variables $x_1$, $x_2$, $\ldots$, $x_m$.

Response surface model\[9, 10\]. The model of $y_f$ and $y_c$ can be expressed as

$$y_f = f(x_1, x_2, \ldots, x_m) + \epsilon$$

$$y_c = f(x_1, x_2, \ldots, x_m) + \epsilon$$

where $f$ is unknown and $\epsilon$ is error term. We make an assumption that $\epsilon$ is independent in different tests, mean is zero and variance is $\delta^2$. Convert $X$ into coding variables $x_1$, $x_2$ $\cdots$ $x_m$, where $x_1$, $x_2$ $\cdots$ $x_m$ with zero mean and the same standard deviation. For the general formula

$$y = f(x_1, x_2, \cdots, x_k) + \epsilon$$

We apply the response surface analysis method. The basic aim of the response surface analysis methods is to maximize or minimize, or achieve the desired response, including testing, modeling, data analysis and optimization. The detailed method is given as follows:

1. Testing: we select a number of important factors, such that the testing can be performed more efficiently and less numbers of tests are required;
2. Modeling, data analysis and optimization.

In the first stage, when the test site is away from the optimal (maximum or minimum) position, we should choose a first-order model:

$$y_f = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \epsilon$$

where $\beta$ is slope or linear effect of Encoding variable $x_i$. The coefficients can be estimated in a rank-order test. The center of the test in the test area added into the first-order test can be carried out on the base surface curvature detection.

The second stage begins until the test area is located close to the optimal area or optimal zone. Near the optimal point of response, curvature effect is the dominant term, and the response surface model should be

$$y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i<j}^{m} \beta_{ij} x_i x_j + \sum_{i=1}^{m} \beta_{ii} x_i^2 + \epsilon$$
where $\beta_i$ is coefficients of linear term and $\beta_{ij}$ is coefficients of mutual term. In the research on response surface, both first-order and second-order test includes the most precipitous climb search method and grid search method. It should be pointed out that the mathematical models of $y_r$ and $y_C$ can be derived by the above method.

**Case Studies**

Experiment produces 89 furnace data, 59 groups of collected data are trained, and other Other 30 groups of collected data are used to test model. **Linear regression model verification.** By algorithm 1, we can get the linear regression models of the endpoint carbon content and temperature as follows

$$y_C = 0.39 - 0.042x_1 - 0.13x_2 - 0.29x_3 + 1.35x_4 + 0.32x_5 - 0.042x_6$$
$$- 0.066x_7 - 0.064x_8 - 0.33x_9$$

$$y_r = -0.14 - 0.12x_1 + 0.023x_2 + 0.71x_3 + 0.039x_4 + 1.61x_5 - 9.806 \times 10^{-3}x_6$$
$$- 0.30x_7 - 0.12x_8 - 0.45x_9$$

(15) (16)

Fig. 2(a) and Fig. 2(b) show the difference between the predictive values and practical values of endpoint carbon content and temperature.

**Response surface prediction model verification.** Fig. 3(a) and Fig. 3(b) illustrate the difference between the predictive value and practical value of endpoint carbon content and temperature.

**Predictive result analysis.** Comparison of the endpoint temperature, the endpoint carbon content and hit rate are presented in Table 1-4. By comparison, the predictive results obtained using response surface analysis method are better than those obtained by the linear regression method.

![Fig. 2 Comparisons between the predicted and experimental of the endpoint carbon content and temperature](image-url)

Fig. 2 Comparisons between the predicted and experimental of the endpoint carbon content and temperature
Fig. 3 Comparisons between the predictive values and experimental values of endpoint carbon content and temperature

Table 1 Comparison of the endpoint temperature

|                      | Error mean  | Error mean square deviation |
|----------------------|-------------|-----------------------------|
| Linear regression method | 0.118743567 | 0.99019327                  |
| Response surface method           | 0.086553842 | 0.626509379                  |

Table 2 Comparison of the endpoint carbon content

|                      | Error mean  | Error mean square deviation |
|----------------------|-------------|-----------------------------|
| Linear regression method | 0.127096546 | 0.994854449                  |
| Response surface method           | 0.093384725 | 0.689269728                  |

Table 3 Hit rate of the endpoint carbon content

|                      |             |
|----------------------|-------------|
| Response surface method | 61.29%     |
| Linear regression method           | 48.39%     |

Table 4 Hit rate of the endpoint temperature

|                      |             |
|----------------------|-------------|
| Response surface method | 67.74%     |
| Linear regression method           | 54.84%     |

Conclusions

In this paper, we develop linear regression analysis and response surface analysis methods to construct a predictive model for the endpoint temperature and carbon content of BOF steelmaking. Compared with the model using the linear regression analysis method, the predictive values are great close to the practical values. Experimental simulations are investigated to verify the effectiveness of the proposed methods.

Acknowledgments

The authors would like to acknowledge the National Natural Science Foundation of China under Grant 61104093, 61174119, 61034006, 61202418, the Scientific Research Project of Liaoning Province of China under Grant L2012141, L2011064 and the Teaching Research Fund Projects of Liaoning Province of China under Grant 2011A017.
References

[1] M. Han and Y. Zhao, “Dynamic control model of BOF steelmaking process based on ANFIS and robust relevance vector machine,” Expert Systems with Applications, vol. 38, no. 12, pp. 14786-14798, 2011.

[2] S.Y. Yun and K.S. Chang, “Dynamic prediction using neural network for automation of BOF process in steel industry,” Iron and Steelmaker, vol. 23, no. 8, pp. 37-42, 1996.

[3] R. Ding and L. Liu, “Artificial intelligence static control model in converter steelmaking,” Iron and Steel, vol. 32, no. 1, pp. 22-26, 1997.

[4] S.M. Xie, J. Tao, and T.Y. Chai, “BOF steelmaking endpoint control based on network,” Control Theory and Applications, vol. 20, no. 6, pp. 903-907, 2003.

[5] Z.J. Wang, T.Y. Chai, and C. Shao, “Slab minerature prediction model based on RBF neural network,” Journal of System Simulation, vol. 11, no. 3, pp. 181-184, 1999.

[6] S.M. Xie, X.W. Gao, and T.Y. Chai, “BOF Endpoint Prediction Based on Grey Model,” Journal of Iron AND Steel Research, vol. 11, no. 4, pp. 9-12, 1999.

[7] H.Y. Wen, Q. Zhao, Y.R. Chen, M.C. Zhou, M. Zhang, and L.F. Xu, “Converter end-point prediction model using spectrum image analysis and improved neural network algorithm,” Optica Application, vol. 38, no. 4, pp. 693-695, 2008.

[8] X.R. Kong, “Research on the endpoint optimization control model of BOF steelmaking,” Hangzhou University of Electronic Science and Technology.

[9] S. Ren, “Modeling the Toxicity of Aromatic Compounds to Tetrahymena pyriformis: The Response Surface Methodology with Nonlinear Methods,” Journal of Chem Inf Comput Sci., vol. 43, no. 5, pp. 1679-1687, 2003.

[10] J.W. Zhuo, Application of MATLAB in mathematical modeling, Beijing: Beijing University of Aeronautics and Astronautics Press, 2011.

[11] R. Liu and C.Y. Lv, The application of adaptively adjustment particle swarm optimization algorithm in catalyzing & cracking fractionating tower, Computers and Applied Chemistry, vol. 27, no. 6, pp. 771-774, 2010.

[12] W.L. Yu, Q.X. Wu, and C.S. Jiang, Application of particle swarm algorithm in 3-D route planning and optimization of air vehicles, Electronics Optic & Control, vol. 15, no. 5, pp. 1-6, 2008.