Control Theory Tech, Vol. 17, No. 1, pp. 24–36, February 2019

Control Theory and Technology

http://link.springer.com/journal/11768

Prediction method for energy consumption per ton of fused magnesium furnaces using data driven and mechanism model

Dan GUO [1], Zhiwei WU [1,2†], Tianyou CHAI [1,2], Jie YANG [1], Jinliang DING [1,2]

1.*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang Liaoning 110819, China;*

2.*National Engineering Research Center of Metallurgical Automation (Shenyang), Shenyang Liaoning 110819, China*

Received 5 September 2018; revised 4 October 2018; accepted 8 October 2018

Abstract

The electric energy consumed in every ton of acceptable product, namely energy consumption per ton (ECT), is an important overall index for the production process of a fused magnesium furnace. The furnace is the equipment for producing the fused magnesia. The ECT value depends on the current in the smelting process. The optimal operation for a fused magnesium furnace is supposed to have the ECT as low as possible, where the key is to predict ECT accurately. By introducing an unknown high-order nonlinear term, this paper builds a dynamic ECT model for different production batches based on the static ECT model for one batch. The average current is taken as the input of the dynamic ECT model, which is composed of the unknown high-order nonlinear term and a nonlinear model with unknown parameters. The order of the nonlinear term is determined by the distance correlation and the nonlinear term is estimated by the stochastic configuration network, while the parameters of the nonlinear model is identified by the least square method. The estimation of the nonlinear term alternates with the parameter identification. This paper proposes a prediction method for ECT, which is composed of the order identification of the nonlinear term, the alternating identification of the model and the ECT prediction model. The simulation experiments are conducted by the on-site data, and the results verify the effectiveness of the proposed prediction method.

Keywords: Fused magnesia, energy consumption per ton, alternating identification, stochastic configuration network, distance correlation

DOI https://doi.org/10.1007/s11768-019-8191-9

†Corresponding author.

E-mail: wuzhiwei@mail.neu.edu.cn.

This work was supported in part by the National Natural Science Foundation of China (Nos. 61525302, 61590922, 61503066, 61533007), in part by the Project of Industry and Information Technology Ministry (No. 20171122-6), and in part by the Projects of Shenyang (No. Y17-0-004).

1 Introduction

The fused magnesia is a kind of high-purified crystal MgO, and it is the main raw material of refractory in the fields of aerospace industry and manufacturing industry because of its good characteristics, such as a high melting point, anti-oxidation, good electric insulation, structural integrity and so on. The equipment for producing the fused magnesia is the three-phase alternating current fused magnesium furnace. The smelting temperature for the fused magnesia is higher than that for the steel-making, and it is as high as about 3000 degrees Celsius. The fused magnesium furnace is different from the electric-arc furnace for steel-making, and it is a kind of submerged arc furnace where the three-phase electrodes are buried in the raw material and the electric arcs are developed among the raw material and three-phase electrodes [1]. The raw material is heated by the electric arcs, and the molten pool is formed in the furnace. After the molten liquid cooling down naturally and crystallizes, the fused magnesia product will be got [2].

The fused magnesium furnace is a type of high energy-consuming facilities, and the cost of electric energy accounts for more than 60 percent of the whole cost in the magnesia production. Lowering the electricity cost and raising the electric energy efficiency have become the major issues for the magnesia enterprises [3]. The overall index for the production process of a fused magnesium furnace is the electric energy consumed in every ton of acceptable product, namely the energy consumption per ton (ECT), which can reflect the electric energy efficiency in the smelting [4]. The electric energy consumed in one batch depends on the average current in the smelting process because the voltage and the production time change slightly in different batches. There is a close relationship between ECT and the current: If the current is too low, the raw material will not be melted completely and the yield of the acceptable product will be low, so the ECT value will be high; If the current is too high, the electric energy will be wasted as there is an upper limit for the yield in one batch, so the ECT value will also be high [5]. As a result, the optimum current corresponds to the lowest ECT. The aim of the optimal operation for a fused magnesia furnace is to have the ECT as low as possible under the demand constraint [6]. Thus, an accurate ECT prediction model is helpful to decide the optimum current and to lower the ECT value.

In recent years, the hybird modeling approaches based on models and data have become the hotspot in the modeling of industrial process: The known dynamic properties of the complex system are described by the mechanism models, while the dynamic properties that are not included in the models are described by the collected data [7]. Since wastewater treatment plants are cascaded processes, the mechanism model is connected with the hierarchical neural networks, and the combination of the model and the neural networks provide water quality in inner reactors [8]. To accurately predict the matte grade in copper flash smelting process, a self-adaptive combination technology is adopted to integrate the mathematic model and a fuzzy neural network [9]. Three kinetic models of p-xylene oxidation process are studied in [10], and then a comparative study of the six hybrid models combining kinetic and regression module parts is carried out. A hybrid modeling approach for centrifugal compressor is proposed, and the hybrid models consist of a three-layer artificial neural network and a mechanistic component [11]. An assumption in the above studies is that the parameters of the mechanism models are fixed and changeless. In the production of the fused magnesia, however, the ingredient and the particle size of the raw material fluctuate frequently, and the parameters within the furnace are uncertain and undetectable, like the arc length, the resistance of the molten pool. As a result, it is hard for the above methods to obtain an accurate ECT prediction model. The mechanism prediction model of ECT for one production batch has been constructed by the principle of the conservation of energy [12]. However, the ECT prediction model is of low quality, and there are some reasons. Above all, the parameters of the mechanism model are considered as the constants and they are chosen by the experience, but in fact these parameters are unknown nonlinear functions. In addition, the raw material, the operations and the loss of equipment change with production batches, which are not taken into account. Last but not the least, the ECT values for different production batches are assumed to be independent and irrelevant. Thus, both the correlation and the difference among different batches need to be considered for an accurate ECT prediction model.

The main motivation of this paper is to take advantage of the ECT mechanism model for one batch [12] to build an accurate dynamic ECT prediction model for different batches, and a prediction method for ECT is proposed. The main contributions of this paper are as follows:

1) A dynamic ECT model for different production batches is built, and the dynamic model is described

by a nonlinear model with unknown parameters and an unknown high-order nonlinear term.

2) The nonlinear model is transformed into a model whose parameters can be identified.

3) A prediction method for ECT using data driven and mechanism model is proposed, which is composed of three parts, i.e., the order identification of the nonlinear term, the alternating identification of the model and the ECT prediction model.

4) The simulation experiments are conducted by the on-site data, and the results demonstrate that the proposed prediction method can significantly improve the prediction accuracy of ECT.

The rest of this paper is organized as follows. Section 2 briefly introduces the production process of the fused magnesia. A dynamic ECT model for different production batches is built in Section 3. Section 4 provides the ECT prediction model and the transformation of the nonlinear model for the identification of its parameters. The main components of our proposed prediction method are explained in Section 5. Section 6 presents the simulation results. Section 7 concludes this paper.

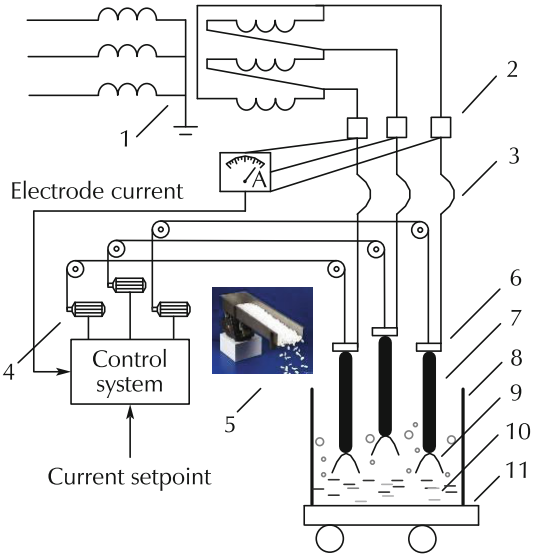
2 Production process of fused magnesia

A schematic diagram of the production process of the fused magnesia is illustrated in Fig. 1, and the whole process in every batch lasts about 10 hours. The control system of the fused magnesium furnace is designed based on stable currents, and the manipulated variables are the currents of the three-phase electrodes. The transformer voltage fluctuates within a small range during the production, so the current value depends on the arc resistance which has a close relationship with the length of the arc. The control system of the fused magnesium furnace adjusts the distance between the electrodes and the surface of the molten pool by the positive-negative rotating of motors, so the currents of electrodes are able to follow the current setpoint [13]. The arcs among electrodes and raw materials convert the electric energy to the thermal energy. Certain amount of caustic calcined magnesia powder will be charged to the furnace every 8-12 minutes, and the main reactions for caustic calcined magnesia are [14]

$$\mathrm{MgO(non-crystal)} \xrightarrow{800\sim1700^{\circ}\mathrm{C}} \mathrm{MgO(crystal)}, \quad (1)$$

$$\mathrm{MgO(crystal)} \xrightarrow{\geqslant 2800^{\circ}\mathrm{C}} \mathrm{MgO(liquid)}, \quad (2)$$

where equation (1) is the process of high-temperature calcination, and the product of high density is obtained, while equation (2) is the liquefaction process, and the product are purified. As the melting points of NaOH, CaO, Fe_2O_3 are lower than that of MgO, these impurities will first melt and migrate out. The surface of the molten pool gradually raises with the charge and the fusion of the raw material. Power supply is cut off and the car is dragged away when the pool surface approaches to the furnace mouth. The liquid in the furnace crystallizes after a natural cooling, and then the crystal is broke and sorted for the final product. The ECT value of the furnace in one production batch will be finally counted based on the electric energy consumption and the yield of the acceptable product.



1 Single-furnace transformer 2 AC Current transformer
3 Short net 4 Motor 5 Automatic feeder 6 Electrode holder
7 Electrode 8 Furnace shell 9 Arc 10 Weld pool 11 Car

Fig. 1 A schematic diagram of the production process of the fused magnesia.

Fig. 2 shows the nonlinear relationship between the average current and ECT. The key to the optimal operation for a fused magnesia furnace is to build an accurate ECT prediction model. However, there are a few reasons why it is hard to build an accurate ECT model: First, there exist many physical and chemical processes in the production, such as calcination, fusion, impurity precipitation and crystallization; in addition, a blend of solid, liquid and gas in the environment of about 3000 degrees Celsius leads to the failure of all sensors, and it is unable to detect the parameters in the furnace, like the arc resistance, the depth of the molten pool; Last but not least, the raw material is charged intermittently, and

the ingredient and the particle size of the raw material fluctuate, so the parameters inside the furnace change with time.

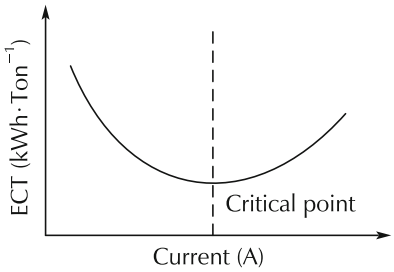


Fig. 2 The nonlinear relationship between the average current and ECT.

3 Dynamic ECT model for different production batches

Taking the average current as the input, the mechanism model of ECT for one batch has been constructed by the principle of the conservation of energy [12], and it has the following expression:

$$r = \frac{\sqrt{3}U\bar{y}\cos\phi}{\dfrac{p}{r_{\text{theory}}}\left\{\dfrac{[A\bar{y}+BC]^2}{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}\bar{y}} - k_p\left[\dfrac{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}\bar{y}^3}{[A\bar{y}+BC]^2}\right]^2\dfrac{C}{\bar{y}}\right\}}, \tag{3}$$

where $\bar{y}$ is the average current in the smelting process,

$$\bar{y} = \frac{1}{L}\sum_{K=1}^{L} y(t), \quad y_{\min} \leqslant y(t) \leqslant y_{\max}, \tag{4}$$

where $y_{\min}$ and $y_{\max}$ are the lower limit and the upper limit for the current in the smelting process, respectively, and $y(t)$ is the average current of the three electrodes at time t. The related parameters to the mechanism model (equation (3)) and their meanings are listed in Table 1.

In [12], the parameters of the mechanism models are assumed to be fixed and changeless, so equation (3) is a static model. During one production batch, however, the parameters in Table 1 are acturally the unknown variables. One reason is that the ingredient and the particle size of the raw material fluctuate frequently. In addition, the powder raw material is charged by automatic feeder intermittently, and it is easy for the raw material to fall into the different regions because of the looseness of the powder and the motion of the electrodes, so the length and the resistance of the arcs change with the thickness of the material layer in the furnace.

Table 1 The list of the parameters in the ECT mechanism model.

Parameter	Meaning
U	Secondary phase voltage of the transformer.
$\cos\phi$	Power factor.
r_{theory}	Theoretical ECT value when the raw material is the caustic calcined magnesia powder.
A	Voltage drop between arc cathode and arc anode.
B	Voltage gradient of arc column.
C	Coefficient of arc length.
f_1	Arc resistivity.
f_2	Resistivity of molten pool.
k_p	Tunable parameter.
p	Overall thermal efficiency, which represents the percentage that the heat absorbed by the fusion of the raw materials accounts for the electric energy input by the arcs.

Furthermore, the melting rate of raw material is related to the distance to the arcs, and the change of the melting rate will lead to the fluctuation of the resistance of the molten pool. Another reason is that there actually exist three loads because of three-phase electrodes, and the change of the arc resistance or the pool resistance in any phase will affect the currents of the three-phase electrodes, so there is a continuous fluctuation for the thermal loss of the power supply system and the overall thermal efficiency. Overall, the parameters in equation (3) are unknown nonlinear functions, so the accurate expression for equation (3) is

$$r = \frac{\sqrt{3}U(\cdot)\bar{y}\cos\phi(\cdot)}{\dfrac{p(\cdot)}{r_{\text{theory}}(\cdot)}\left\{\dfrac{[A(\cdot)\bar{y}+B(\cdot)C(\cdot)]^2}{\frac{U(\cdot)f_1(\cdot)}{\sqrt{3}[f_1(\cdot)+0.92f_2(\cdot)]}\bar{y}} - k_p(\cdot)\left[\dfrac{\frac{U(\cdot)f_1(\cdot)}{\sqrt{3}[f_1(\cdot)+0.92f_2(\cdot)]}\bar{y}^3}{[A(\cdot)\bar{y}+B(\cdot)C(\cdot)]^2}\right]^2\dfrac{C(\cdot)}{\bar{y}}\right\}}. \tag{5}$$

The above expression is the ECT model for just one production batch, and a ECT prediction model should be suitable for different batches. Considering that the raw material, the operations and so on change with batches, a

nonlinear term $\Delta r_1(S)$ is introduced to represent the differences among different batches, so the ECT model for any batch S can be formulated as

$$r(S)=\frac{\sqrt{3}U(\cdot)y(S)\cos\phi(\cdot)}{\dfrac{p(\cdot)}{r_{\text{theory}}(\cdot)}\left\{\dfrac{[A(\cdot)y(S)+B(\cdot)C(\cdot)]^2}{\frac{U(\cdot)f_1(\cdot)}{\sqrt{3}[f_1(\cdot)+0.92f_2(\cdot)]}y(S)}-k_p(\cdot)\left[\dfrac{\frac{U(\cdot)f_1(\cdot)}{\sqrt{3}[f_1(\cdot)+0.92f_2(\cdot)]}y(S)^3}{[A(\cdot)y(S)+B(\cdot)C(\cdot)]^2}\right]^2\dfrac{C(\cdot)}{y(S)}\right\}}+\Delta r_1(S). \tag{6}$$

The ECT values can not be calculated according to the above equation as the unknown variables are included, thus the unknown constants are used to replace the variables. The errors caused by the replacement is denoted as the unknown nonlinear term $\Delta r_2(S)$. Overall, the dynamic ECT model for different production batches can be expressed as

$$r(S)=\frac{\sqrt{3}Uy(S)\cos\phi}{\dfrac{p}{r_{\text{theory}}}\left\{\dfrac{[Ay(S)+BC]^2}{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}y(S)}-k_p\left[\dfrac{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}y(S)^3}{[Ay(S)+BC]^2}\right]^2\dfrac{C}{y(S)}\right\}}+\Delta r(S), \tag{7}$$

where the nonlinear term $\Delta r(S)=\Delta r_1(S)+\Delta r_2(S)$.

4 Prediction model of ECT

4.1 ECT prediction model

The dynamic ECT model in equation (7) is composed of a nonlinear model with unknown parameters and an unknown high-order nonlinear term. If the nonlinear model is denoted as $\tilde{r}(S)$, then

$$\tilde{r}(S)=\frac{\sqrt{3}Uy(S)\cos\phi}{\dfrac{p}{r_{\text{theory}}}\left\{\dfrac{[Ay(S)+BC]^2}{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}y(S)}-k_p\left[\dfrac{\frac{Uf_1}{\sqrt{3}[f_1+0.92f_2]}y(S)^3}{[Ay(S)+BC]^2}\right]^2\dfrac{C}{y(S)}\right\}}, \tag{8}$$

and so

$$r(S)=\tilde{r}(S)+\Delta r(S). \tag{9}$$

There is no big difference for some working conditions in two successive batches, for example, the loss of the equipment, so it is assumed that the ECT value of the $(S+1)$th batch has a close relationship with that of the Sth batch,

$$r(S+1)=r(S)+\Delta r'(S+1), \tag{10}$$

where the nonlinear term $\Delta r'(S+1)$ represents the difference in ECT values between the two batches. Combining equation (9) and equation (10), the new equation is

$$\begin{aligned} r(S+1)&=\tilde{r}(S)+\Delta r(S)+\Delta r'(S+1)\\ &=\tilde{r}(S)+\Delta\bar{r}(S+1), \end{aligned} \tag{11}$$

where $\tilde{r}(S)$ is derived from equation (8) and the high-order nonlinear term $\Delta\bar{r}(S+1)$ is unknown. As the parameters in equation (8) can not be detected, the estimation $\hat{\tilde{r}}(S)$ will be got based on the approximation of the parameters. $\Delta\bar{r}(S+1)$ is estimated based on the ECT and the average current in the previous production batches, that is $\Delta\hat{\bar{r}}(S+1)=f(r(S),r(S-1),\ldots,r(S-d_r),y(S),y(S-1),\ldots,y(S-d_y))$, where d_r and d_y are unknown. Overall, the ECT prediction model is formulated as

$$\hat{r}(S+1)=\hat{\tilde{r}}(S)+\Delta\hat{\bar{r}}(S+1). \tag{12}$$

The parameters in equation (8) need to be identified before ECT is predicted based on equation (12), and thus the nonlinear model (equation (8)) is supposed to be transformed into the model with the identifiable parameters.

4.2 Parameter identification in ECT prediction model

New variables $F_0\sim F_{11}$ is introduced to transform equation (8) into the model with the identifiable parameters, and the expressions for $F_0\sim F_{11}$ is in Table 2. The new equation based on $F_1\sim F_{11}$ is

$$\tilde{r}(S-1)=\frac{F_7+F_8y(S-1)^{-1}+F_9y(S-1)^{-2}+F_{10}y(S-1)^{-3}+F_{11}y(S-1)^{-4}}{1-F_1y(S-1)^{-1}-F_2y(S-1)^{-2}-F_3y(S-1)^{-3}-F_4y(S-1)^{-4}-F_5y(S-1)^{-5}-F_6y(S-1)^{-6}}. \tag{13}$$

From equation (11), we can get $\tilde{r}(S-1)=r(S)-\Delta\bar{r}(S)$. $r(S)-\Delta\bar{r}(S)$ is made use of to replace $\tilde{r}(S-1)$ in equation (13), and the equation for parameter identification finally is

$$r(S)-\Delta\bar{r}(S)=[F_1,F_2,F_3,F_4,F_5,F_6,F_7,F_8,F_9,F_{10},F_{11}]\times\begin{bmatrix}y(S-1)^{-1}(r(S)-\Delta\bar{r}(S))\\ y(S-1)^{-2}(r(S)-\Delta\bar{r}(S))\\ y(S-1)^{-3}(r(S)-\Delta\bar{r}(S))\\ y(S-1)^{-4}(r(S)-\Delta\bar{r}(S))\\ y(S-1)^{-5}(r(S)-\Delta\bar{r}(S))\\ y(S-1)^{-6}(r(S)-\Delta\bar{r}(S))\\ 1\\ y(S-1)^{-1}\\ y(S-1)^{-2}\\ y(S-1)^{-3}\\ y(S-1)^{-4}\end{bmatrix}. \tag{14}$$

Table 2 The expressions for $F_0\sim F_{11}$.

Parameter	Expression
F_0	$A^6[(f_1+0.92f_2)/U/f_1]^3/(k_pC)-1/(3\sqrt{3})$
F_1	$-6A^5B[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_2	$-15A^4B^2C[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_3	$-20A^3B^3C^2[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_4	$-15A^2B^4C^3[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_5	$-6AB^5C^4[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_6	$-B^6C^5[(f_1+0.92f_2)/U/f_1]^3/(k_pF_0)$
F_7	$A^4r_{\text{theory}}[f_1+0.92f_2]^2\cos\phi/(pk_pCUf_1^2F_0)$
F_8	$4A^3BCr_{\text{theory}}[f_1+0.92f_2]^2\cos\phi/(pk_pCUf_1^2F_0)$
F_9	$6A^2B^2C^2r_{\text{theory}}[f_1+0.92f_2]^2\cos\phi/(pk_pCUf_1^2F_0)$
F_{10}	$4AB^3C^3r_{\text{theory}}[f_1+0.92f_2]^2\cos\phi/(pk_pCUf_1^2F_0)$
F_{11}	$B^4C^4r_{\text{theory}}[f_1+0.92f_2]^2\cos\phi/(pk_pCUf_1^2F_0)$

The nonlinear term $\Delta\bar{r}(S)$ needs to be known before equation (14) is applied to identify $F_1\sim F_{11}$ by the least square method, but $F_1\sim F_{11}$ also need to be known before $\Delta\bar{r}(S)$ is calculated according to equation (13) and equation (11), so $\Delta\bar{r}(S)$ and $F_1\sim F_{11}$ affect each other.

5 Prediction method of ECT

This paper proposes a prediction method based on the ECT prediction model (equation (12)), which is composed of the order identification of the nonlinear term, the alternating identification of the model and the ECT prediction model. The detailed ECT prediction strategies and algorithms will be described as follows.

5.1 Prediction strategy of ECT

As shown in Fig. 3, the proposed structure diagram for the ECT prediction method is made up of three parts.

. Order identification of the nonlinear term d_r and d_y are determined by the distance correlation method based on the collected data. d_r is the number of batches where historical ECT values are related to the nonlinear term $\Delta\bar{r}(S+1)$, while d_y is the number of batches where historical average currents are related to the nonlinear term.

. Alternating identification of the model The nonlinear term $\Delta\bar{r}(S+1)$ is estimated by the stochastic configuration network, and the parameters of the nonlinear model (equation (8)) is identified by the least square method. The estimation of the nonlinear term alternates with the parameter identification.

. ECT prediction model The prediction value of ECT is calculated by equation (12) based on the average current $y(S)$, the estimations $\hat{F}_1\sim\hat{F}_{11}$ and $\Delta\hat{\bar{r}}(S+1)$.

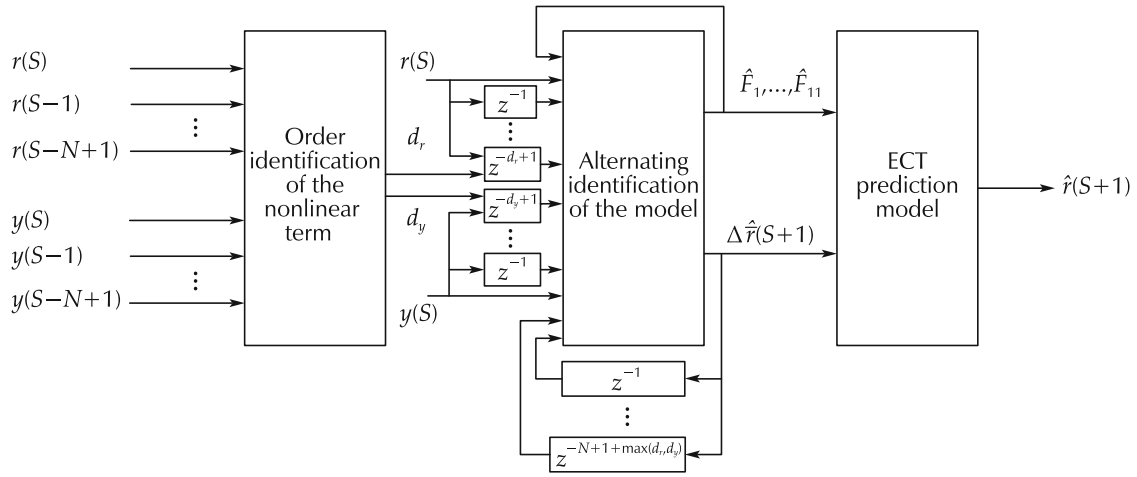


Fig. 3 The structure diagram for the ECT prediction method.

5.2 Prediction algorithm of ECT

5.2.1 Order identification of nonlinear term

The distance correlation method suggested in [15] is applied to describe the degree of relevance between random variables. Both linear and nonlinear association can be measured by the distance correlation. The measurement of distance correlation is zero if and only if two random variables are independent.

It is assumed that there are N observations for two variables X ($X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]^{\mathrm{T}}$) and Y ($Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N]^{\mathrm{T}}$), and then two $K \times K$ matrices (A and B) are obtained by

$$\begin{cases} a_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2, \quad b_{i,j} = \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2, \\ A_{i,j} = a_{i,j} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} - \bar{a}_{\cdot\cdot}, \\ B_{i,j} = b_{i,j} - \bar{b}_{i\cdot} - \bar{b}_{\cdot j} - \bar{b}_{\cdot\cdot}, \\ i, j = 1, 2, \ldots, N, \end{cases} \tag{15}$$

where $\bar{a}_{i\cdot}$ and $\bar{a}_{\cdot j}$ are the mean value of ith row and jth column respectively, and $\bar{a}_{\cdot\cdot}$ is the grand mean. A scalar (distance covariance) can be obtained by an average of the product of A and B

$$\mathrm{dCov}^2(X, Y) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{i,j} B_{i,j}. \tag{16}$$

The association measurement of X and Y by the distance correlation method is

$$\mathrm{dCor}(X, Y) = \frac{\mathrm{dCov}(X, Y)}{\sqrt{\mathrm{dCov}(X, X)\mathrm{dCov}(Y, Y)}}, \tag{17}$$

where the expressions for $\mathrm{dCov}(X, X)$ and $\mathrm{dCov}(Y, Y)$ can refer to equation (16). From the above, the steps to determine the orders d_r and d_y are as follows (suppose that there are N sets of collected data):

1) According to the empirical value of the parameters [12], like the resistance of molten pool, voltage drop and so on, calculate the initial estimations $\hat{F}_1 \sim \hat{F}_{11}$ based on Table 2.

2) According to equation (13), equation (11) and $\hat{F}_1 \sim \hat{F}_{11}$, estimate $\Delta\hat{\bar{r}}(S), \Delta\hat{\bar{r}}(S-1), \ldots, \Delta\hat{\bar{r}}(S-N+2)$.

3) Set $d_r = d_y = 0$.

4) Measure the association between $[\Delta\hat{\bar{r}}(S), \Delta\hat{\bar{r}}(S-1), \ldots, \Delta\hat{\bar{r}}(S-N+2+d_r)]^{\mathrm{T}}$ and $[r(S-1-d_r), r(S-2-d_r), \ldots, r(S-N+1)]^{\mathrm{T}}$ according to equations (15)–(17). If the distance correlation is smaller than the threshold value δ, then d_r is the number of batches where historical ECT values are related to the nonlinear term, otherwise set $d_r = d_r + 1$ and repeat step 3).

5) Measure the association between $[\Delta\hat{\bar{r}}(S), \Delta\hat{\bar{r}}(S-1), \ldots, \Delta\hat{\bar{r}}(S-N+2+d_y)]^{\mathrm{T}}$ and $[y(S-1-d_y), y(S-2-d_y), \ldots, y(S-N+1)]^{\mathrm{T}}$ according to equations (15) ~ (17). If the distance correlation is smaller than the threshold value δ, then stop and d_y is the number of batches where historical average currents are related to the nonlinear term, otherwise set $d_y = d_y + 1$ and repeat step 4).

5.2.2 Alternating identification

. Estimation of nonlinear term

d_r and d_y have been determined by the distance correlation in the last section, so $\Delta\hat{\bar{r}}(S+1)$ can be expressed by a nonlinear function $f_{\Delta r}(\cdot)$ whose inputs are $d_r + d_y$

variables, that is

$$\begin{aligned}&\Delta\hat{\bar{r}}(S+1)\\&=f_{\Delta r}(r(S),r(S-1),\ldots,r(S-d_r+1),\\&\quad y(S),y(S-1),\ldots,y(S-d_y+1)).\end{aligned}\tag{18}$$

For the sake of brevity, let

$$\begin{aligned}&\Theta(S,d_r,d_y)\\&=[r(S),\ldots,r(S-d_r+1),y(S),\ldots,y(S-d_y+1)],\end{aligned}\tag{19}$$

then equation (18) can be written as

$$\Delta\hat{\bar{r}}(S+1)=f_{\Delta r}(\Theta(S,d_r,d_y)).\tag{20}$$

In this paper, the stochastic configuration network (SCN) [16, 17] is used to estimate $f_{\Delta r}(\cdot)$. Suppose that there are N sets of historical ECTs and average currents, that is $[r(S),r(S-1),\ldots,r(S-N+1)]$ and $[y(S),y(S-1),\ldots,y(S-N+1)]$, then the training data of SCN are

$$\begin{cases}X=[\Theta(S-1,d_r,d_y),\ldots,\Theta(S-N+\max(d_r,d_y),\\\quad d_r,d_y)]^{\mathrm{T}},\\Y=[\Delta\hat{\bar{r}}(S),\ldots,\Delta\hat{\bar{r}}(S-N+\max(d_r,d_y),d_r,d_y)]^{\mathrm{T}},\end{cases}\tag{21}$$

where $X\in\mathbb{R}^{d_r+d_y}$, $Y\in\mathbb{R}$, and $\Delta\hat{\bar{r}}(S),\ldots,\Delta\hat{\bar{r}}(S-N+\max(d_r,d_y),d_r,d_y)$ are derived from step 1) in Section 5.2.1.

The training of SCN is fast. In addition, SCN has a satisfied generalization ability even if there is no prior knowledge about the model complexity. Starting from a small-scale network, SCN adds the hidden neurons gradually until the stopping criterion is reached. It has been proved that SCN has the universal approximation property.

The training process of SCN is illustrated as Fig. 4. SCN generates the initial weights and biases within a small range. If the generated weights and biases can not provide satisfied neurons, the range will be extended. The positive scalars $\Lambda=\{\lambda_{\min}:\Delta\lambda:\lambda_{\max}\}$ are used to define the range of weights and biases. An increasing sequence Γ ($\gamma\in\Gamma$) is defined in SCN, and every element in Γ is smaller than 1 but approaches to 1. If no neurons satisfies the condition which is related to γ, the next element in Γ will be chosen as the new value for γ. Suppose that there are currently $L-1$ hidden neurons, SCN will choose the most suitable neuron from many ($\leqslant T_{\max}$) candidates to be the Lth hidden neuron. The output of one neuron can be expressed as

$$\begin{aligned}h_L(X)=&[g_L(w_L^{\mathrm{T}}\Theta(S-1,d_r,d_y)+b_L),\ldots,\\&g_L(w_L^{\mathrm{T}}\Theta(S-N+\max(d_r,d_y),d_r,d_y)+b_L)],\end{aligned}\tag{22}$$

where w_L, b_L and g_L are the weights, biases and activation function of the neuron, respectively. Denoted the current residual error by $E_{L-1}(X)$, and the condition which determines whether the neuron will be the candidates of the Lth hidden neuron is

$$\xi_L=\frac{(E_{L-1}^{\mathrm{T}}\cdot h_L)^2}{h_L^{\mathrm{T}}\cdot h_L}-(1-\gamma)(E_{L-1}^{\mathrm{T}}\cdot E_{L-1}).\tag{23}$$

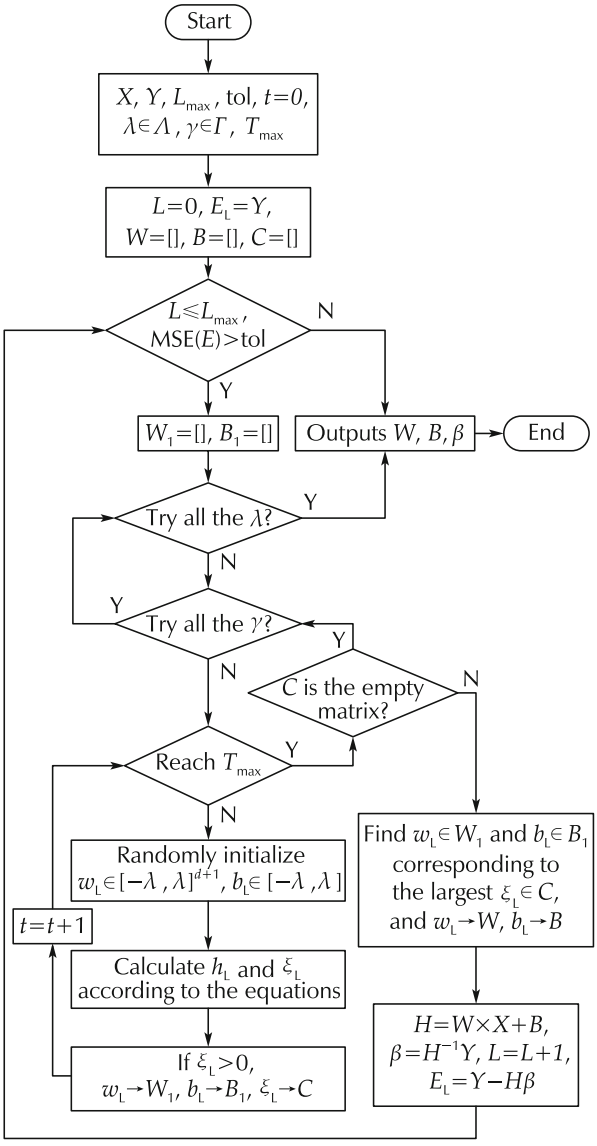


Fig. 4 The training process of the stochastic configuration network.

If $\xi_L > 0$, the neuron will be candidates, and w_L, b_L and ξ_L are restored in W_1, B_1 and C, respectively. The candidate with the largest ξ_L will finally be the Lth hidden neuron. The stopping criterion is that either the upper limit of the number of hidden neurons $L_{\max}$ is reached or the root mean square error is smaller than the predifined parameter Tol. According to the weight and the bias matrixes of the hidden layer and the weights of the output layer, the prediction for the high-order nonlinear term can be formulated as

$$\begin{aligned}\Delta\hat{\bar{r}}(S+1) = [&g_1(W \times \Theta(S, d_r, d_y) + B), \ldots, \\ &g_L(W \times \Theta(S, d_r, d_y) + B)] \times \beta. \end{aligned} \tag{24}$$

. Alternating identification

The alternating identification of the nonlinear term and parameters is adopted for the accurate ECT prediction [18]. After $\Delta\hat{\bar{r}}(S+1)$ is calculated (equation (24)), the parameters $\boldsymbol{F} = [F_1, \ldots, F_{11}]^{\mathrm{T}}$ can be estimated again by the least squre method based on equation (14),

$$\hat{\boldsymbol{F}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(R - \Delta\hat{\bar{R}}), \tag{25}$$

where

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}(S), \boldsymbol{\phi}(S-1), \ldots, \boldsymbol{\phi}(S-N+2+\max(d_r, d_y))],$$

$$\boldsymbol{\phi}(S-i) = \begin{bmatrix} y(S-i-1)^{-1}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ y(S-i-1)^{-2}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ y(S-i-1)^{-3}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ y(S-i-1)^{-4}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ y(S-i-1)^{-5}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ y(S-i-1)^{-6}(r(S-i) - \Delta\hat{\bar{r}}(S-i)) \\ 1 \\ y(S-i-1)^{-1} \\ y(S-i-1)^{-2} \\ y(S-i-1)^{-3} \\ y(S-i-1)^{-4} \end{bmatrix},$$

$$R = [R(S), R(S-1), \ldots, R(S-N+2+\max(d_r, d_y))]^{\mathrm{T}},$$

$$\Delta\hat{\bar{R}} = [\Delta\hat{\bar{r}}(S), \Delta\hat{\bar{r}}(S-1), \ldots, \Delta\hat{\bar{r}}(S-N+2+\max(d_r, d_y))]^{\mathrm{T}},$$

where $0 \leqslant i \leqslant N - 2 - \max(d_r, d_y)$. After the new values of $\hat{\boldsymbol{F}}$ is calculated, the new tutor signal for SCN can be obtained according to equations (11) and (13). As a result, SCN will be trained further and more accurate estimation of the nonlinear term will be got. Overall, the steps of the alternating identification for ECT prediction model are as follows:

1) Intialize SCN and calculate the initial estimations $\hat{F}_1 \sim \hat{F}_{11}$ based on the empirical values of the parameters [12] and Table 2.

2) Calculate $\hat{\bar{r}}(S), \hat{\bar{r}}(S-1), \ldots, \hat{\bar{r}}(S-N+1)$ according to equation (13), and then calculate the tutor signal $\Delta\hat{\bar{R}} = \Delta\hat{\bar{r}}(S), \ldots, \Delta\hat{\bar{r}}(S-N+1+\max(d_r, d_y))$ according to equation (11).

3) Train SCN and calculate the new estimations $\Delta\hat{\bar{R}}$ according to equation (24).

4) Calculate the ECT prediction (equation (12)) based on $\hat{\bar{r}}(S-1), \ldots, \hat{\bar{r}}(S-N+1)$ in step 2) and $\Delta\hat{\bar{R}}$ in step 3). If the root mean square error of the ECT predictions is smaller than Tol, then stop, otherwise go on to the next step.

5) Calculate the new values of $\hat{\boldsymbol{F}}$ based on $\Delta\hat{\bar{R}}$ in step 3) according to equation (25), then go to step 2).

Overall, the ECT prediction model is given in equation (12), where the estimations of the parameters and the high-order nonlinear term are given in equation (25) and equation (24), respectively.

6 Simulation experiments

The data collected from a factory in Liaoning Province (China) are taken for example. The fused magnesium furnace has a designed production capacity of 30 ton. After preprocessing and eliminating outliers, 382 sets of oberservations have been finally utilized. The training set is of size 300 while the test set is of size 82.

The current of one electrode in one production batch is plotted in Fig. 5. As the current fluctuates dramatically, equation (4) is used to calculate the average current in the smelting process, and $y_{\min}$ and $y_{\max}$ are set to 12000 A and 17000 A, respectively. The parameters in Table 1 are estimated by the experience as in [12]: U can be detected and it is about 190 V; $\cos\phi = 0.8$ which is estimated by the power consumption, the average current and U; $r_{\text{theory}} = 1577$ kWh/ton is estimated by the energy absorbed by the fusion of MgO and the latent heat of fusion [5]; $p = 0.7$ which is estimated by the collected data, that is dividing the product of the actual yield and r_{theory} by the actual energy consumption; The values for A, B and C are respectively 34 V, 1000 V/m and 0.35 m·kA [19]; $f_1 = 0.005\,\Omega$m, $f_2 = 0.001\,\Omega$m, and $k_p = 0.2305\,(\text{kW}\cdot\text{m}\Omega^{-\beta}\cdot\text{m}^{-1})$ [20]. The initial estimations of $F_1 \sim F_{11}$ can be obtained based on these empirical values of the parameters according to Table 2.

The threshold $\delta = 0.5$ when the order of the nonlinear term is identified. Table 3 is the distance correlation of the random variables $\Delta\hat{\hat{R}}_i = [\Delta\hat{\hat{r}}(S), \Delta\hat{\hat{r}}(S-1), \ldots, \Delta\hat{\hat{r}}(S-N+1+i)]^{\mathrm{T}}$ and $R_i = [r(S-i), r(S-1-i), \ldots, r(S-N+1)]^{\mathrm{T}}$, while Table 4 is the distance correlation of the random variables $\Delta\hat{\hat{R}}_i$ and $Y_i = [y(S-i), y(S-1-i), \ldots, y(S-N+1)]^{\mathrm{T}}$. As a result, $d_r = 8, d_y = 4$ and the inputs of SCN are $\Theta(S, d_r, d_y) = [r(S), \ldots, r(S-7), y(S), \ldots, y(S-3)]$ (equation (19)).

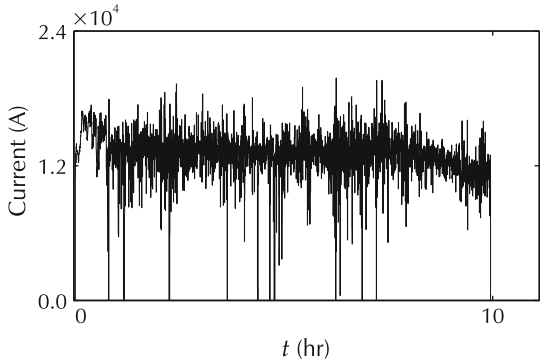


Fig. 5 The sampled current of the fused magnesium furnace.

Table 3 The distance correlation of the nonlinear term and the historical ECT values.

Random variables	Distance correlation
$\Delta\hat{\hat{R}}_1, R_1$	0.7853
$\Delta\hat{\hat{R}}_2, R_2$	0.7424
$\Delta\hat{\hat{R}}_3, R_3$	0.6986
$\Delta\hat{\hat{R}}_4, R_4$	0.6534
$\Delta\hat{\hat{R}}_5, R_5$	0.6388
$\Delta\hat{\hat{R}}_6, R_6$	0.5627
$\Delta\hat{\hat{R}}_7, R_7$	0.5271
$\Delta\hat{\hat{R}}_8, R_8$	0.4639

Table 4 The distance correlation of the nonlinear term and the historical average currents.

Random variables	Distance correlation
$\Delta\hat{\hat{R}}_1, Y_1$	0.6785
$\Delta\hat{\hat{R}}_2, Y_2$	0.5772
$\Delta\hat{\hat{R}}_3, Y_3$	0.5361
$\Delta\hat{\hat{R}}_4, Y_4$	0.4155

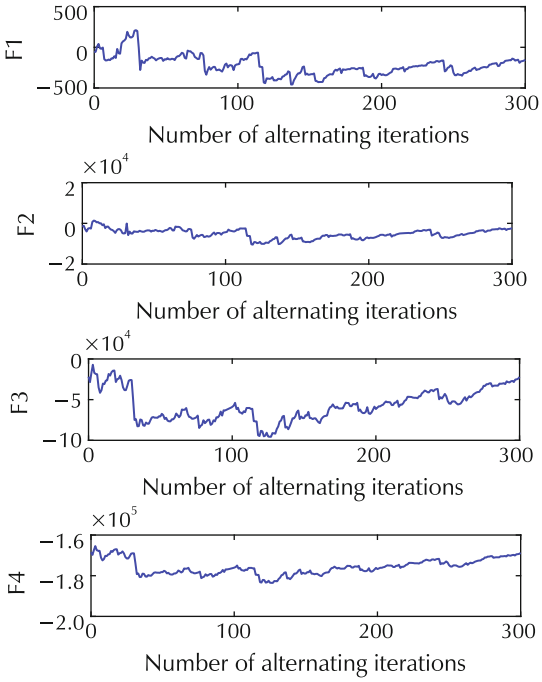
The settings of SCN is shown in Table 5, and Tol is set to 15 for the stopping criterion in the alternating identification. The root mean square error (RMSE) and the mean absolute percent error (MAPE) are adopted as the performance indicators,

$$\begin{cases} \mathrm{RMSE} = \sqrt{\dfrac{\sum\limits_{i=S-N+2}^{S}(r(i)-\hat{r}(i))^2}{N-1}} \times 100\%, \\ \mathrm{MAPE} = \dfrac{1}{N-1}\sum\limits_{i=S-N+2}^{S}\left|\dfrac{r(i)-\hat{r}(i)}{\hat{r}(i)}\right| \times 100\%. \end{cases} \tag{26}$$

Table 5 The settings of the stochastic configuration network.

Parameter	Values
$L_{\max}$	250
Λ	[0.5,1.5,10,30,50,100,150,200,250]
Γ	[0.9,0.99,0.999,0.9999,0.99999]
$T_{\max}$	100

Fig. 6 shows the change of the parameters in the process of alternating identification. The predicted ECT values by our method and the method of [12] are plotted in Fig. 7 and Fig. 8, respectively. RMSE and MAPE for the two prediction methods are shown in Table 6. It can be seen clearly that our prediction method can improve the prediction accracy significantly. The histogram of the error distribution and the probability density function of the error are shown in Fig. 9, which demonstrates that the prediction results are satisfactory.



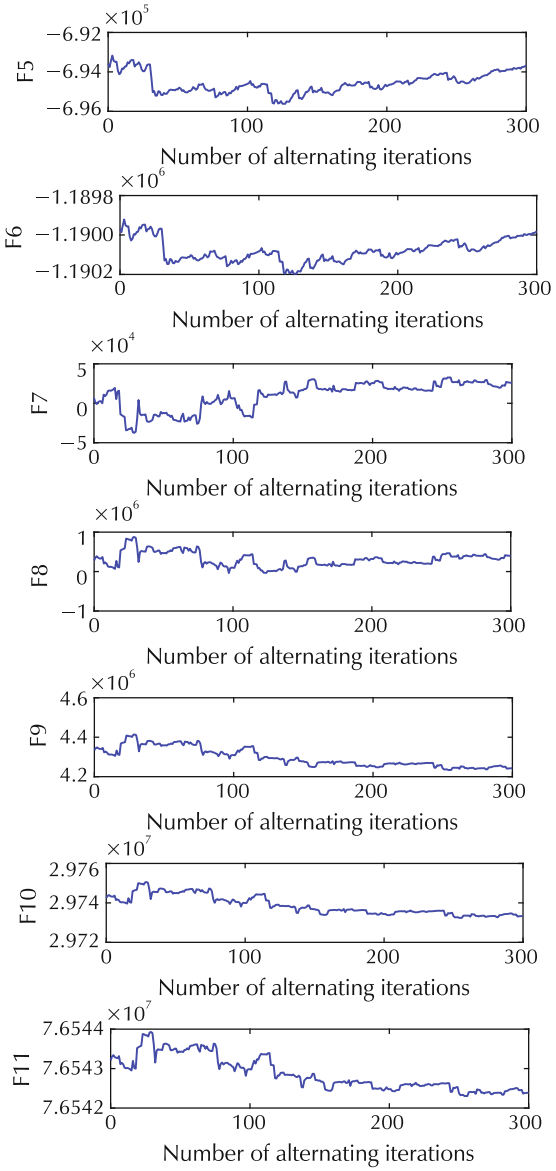


Fig. 6 The change of the parameters in the process of alternating identification.

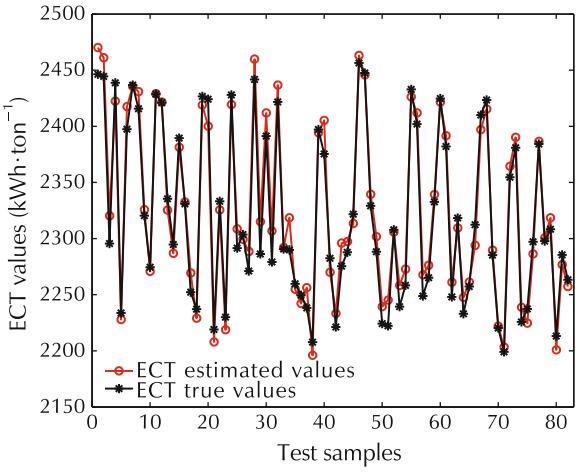


Fig. 7 The predicted ECT values by our method.

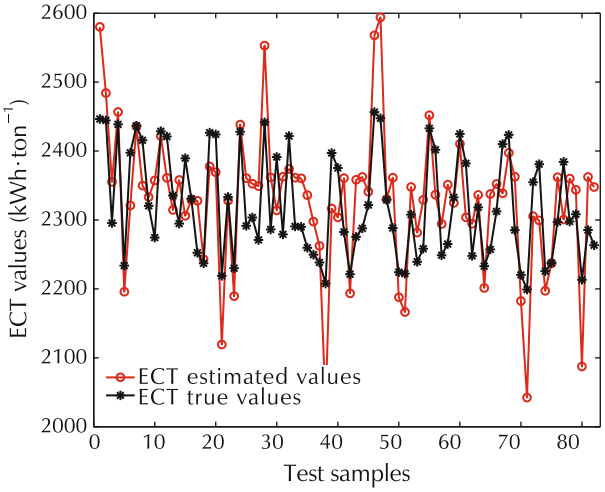


Fig. 8 The predicted ECT values by [12].

Table 6 RMSE and MAPE for two prediction methods.

Prediction method	RMSE	MAPE
Our method	13.4761	0.49%
Method in [12]	68.4483	2.51%

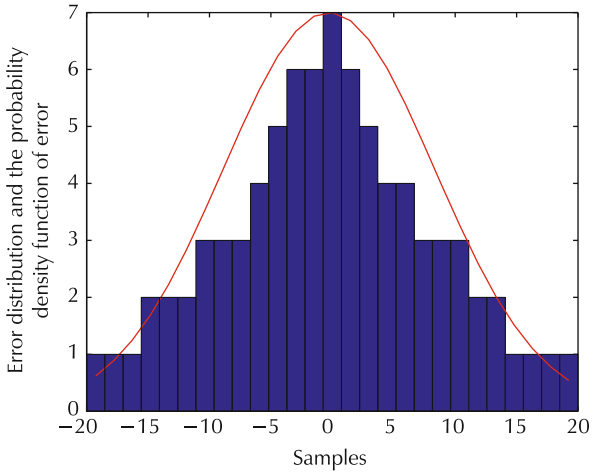


Fig. 9 The histogram of the error distribution and the probability density function of the error.

7 Conclusions

This paper builds a dynamic ECT prediction model for different production batches, which is made up of a nonlinear model with unknown parameters and an unknown high-order nonlinear term. The nonlinear model is transformed for the identification of the parameters. A prediction method for ECT is proposed, which includes the order identification of the nonlinear term, the alternating identification of the model and the ECT prediction

model. Experiment results based on the real data show that the proposed method can provide accurate ECT predictions, which lay the foundation for the optimal operation for the fused magnesium furnace.

References

[1] D. B. Paolo, F. Dughiero, M. Dusi, et al. 3D FE analysis and control of a submerged arc electric furnace. *International Journal of Applied Electromagnetics and Mechanics*, 2012, 39(1–4): 555 – 561.

[2] Z. Wang, N. H. Wang, T. Li. Transient 3D simulation of a submerged-arc furnace for production of MgO single crystal. *Proceedings of the 7th International Forum on Advanced Material Science and Technology*, Zurich: Trans. Tech. Publications Ltd., 2011: 995 – 998.

[3] T. Y. Chai, Z. W. Wu, H. Wang. A CPS based optimal operational control system for fused magnesium furnace. *IFAC-PapersOnLine*, 2017, 50(1): 14992 – 14999.

[4] N. H. Tang, N. Nirmalakhandan, R. E. Speece. Weir aeration: models and unit energy consumption. *Journal of Environmental Engineering*, 1995, 121(2): 196 – 199.

[5] G. H. Gu, C. X. Lin, M. X. Guo. Determination of the rational operating current in melt-down period of electric arc furnace. *Iron and Steel*, 1994, 29(6): 71 – 75 (in Chinese).

[6] W. J. Kong, J. L. Ding, T. Y. Chai, et al. A multiobjective particle swarm optimization algorithm for load scheduling in electric smelting furnaces. *IEEE Symposium on Computational Intelligence for Engineering Solutions*, Piscataway: IEEE, 2013: 188 – 195.

[7] J. H. Liu, W. H. Gui, Y. F. Xie, et al. Dynamic modeling of copper flash smelting process at a Smelter in China. *Applied Mathematical Modelling*, 2014, 38(7/8): 2206 – 2213.

[8] Q. M. Cong, W. Yu, T. Y. Chai. Hierarchical neural network model for water quality prediction in wastewater treatment plants. *Advances in Computational Intelligence*. Poland: Springer, 2009: 155 – 166.

[9] W. H. Gui, L. Y. Wang, C. H. Yang, et al. Intelligent prediction model of matte grade in copper flash smelting process. *Transactions of Nonferrous Metals Society of China*, 2007, 17(5): 1075 – 1081.

[10] Y. Dong, X. F. Yan. A comparative study of hybrid models combining various kinetic and regression models for p-xylene oxidation. *Korean Journal of Chemical Engineering*, 2014, 31(10): 1746 – 1756.

[11] F. Chu, F. L. Wang, X. G. Wang, S. N. Zhang. A hybrid artificial neural network-mechanistic model for centrifugal compressor. *Neural Computing and Applications*, 2014, 24(6): 1259 – 1268.

[12] Z. W. Wu, T. Y. Chai, Y. J. Wu. A hybrid prediction model of energy consumption per ton for fused magnesia. *Acta Automatica Sinica*, 2013, 39(2): 2002 – 2011 (in Chinese).

[13] Z. W. Wu, T. F. Liu, J.-P. Jiang, et al. Nonlinear control tools for fused magnesium furnaces: design and implementation. *IEEE Transactions on Industrial Electronics*, 2018, 65(9): 7248 – 7257.

[14] Z. W. Wu, Y. J. Wu, T. Y. Chai, et al. Data-driven abnormal condition identification and self-healing control system for fused magnesium furnace. *IEEE Transactions on Industrial Electronics*, 2015, 62(3): 1703 – 1715.

[15] G. J. Szekely, M. L. Rizzo, N. K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 2007, 35(6): 2769 – 2794.

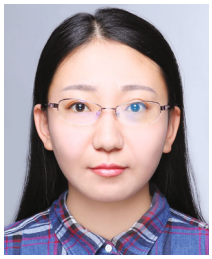
[16] M. Li, D. Wang. Insights into randomized algorithms for neural networks: practical issues and common pitfalls. *Information Sciences*, 2017, 382: 170 – 178.

[17] D. H. Wang, M. Li. Stochastic configuration networks: fundamentals and algorithms. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3466 – 3479.

[18] Y. J. Zhang, T. Y. Chai, D. H. Wang. An alternating identification algorithm for a class of nonlinear dynamical systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(7): 1606 – 1617.

[19] M. X. Guo. *Industry Furnace*. Beijing: Metallurgical Industry Press, 2002 (in Chinese).

[20] Y. Wang, Z. Z. Mao; Y. Li, et al. Study on modeling and coupling simulation of power supply system for AC electric arc furnace. *Journal of System Simulation*, 2010, 22(4): 841 – 844 (in Chinese).



Dan GUO received the B.Sc. degree in Automation from Northeastern University, Shenyang, China, in 2012, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Synthetical Automation for Process Industry. Her current research interests include data driven modeling and optimization. E-mail: guodan717@163.com.

Zhiwei WU received the B.Sc. degree in Electronic and Information Engineering from Dalian Nationalities University, Dalian, China, in 2004, the M.Sc. degree in Control Theory and Engineering from Shenyang University of Chemical Technology, Shenyang, China, in 2007, and the Ph.D. degree in Control Theory and Engineering from Northeastern University, Shenyang, China, in 2015. He is currently a Lecturer with the State Key Laboratory of Synthetical Automation for Process Industries (Northeastern University), China. His current research interests include operational control for complex industry process and industrial embedded control system. E-mail: wuzhiwei@mail.neu.edu.cn.

Tianyou CHAI received the Ph.D. degree in Control Theory and Engineering from Northeastern University, Shenyang, China, in 1985. He is the Founder and Director of the Center of Automation, which became a National Engineering and Technology Research Center and a State Key Laboratory. He became a Professor at Northeastern University in 1988. He is the Director of the Department of Information Science, National Natural Science Foundation of China. He has authored 150 peer-reviewed international journal

papers. He has developed control technologies with applications to various industrial processes. His research interests include modeling, control, optimization, and integrated automation of complex industrial processes. Dr. Chai is a Member of the Chinese Academy of Engineering and is an International Federation of Automatic Control (IFAC) Fellow. His paper titled ybrid intelligent control for optimal operation of shaft furnace roasting process as selected as one of three Best Papers for the Control Engineering Practice Paper Prize for 2011–2013. For his contributions, he has won four prestigious awards from the National Science and Technology Progress and National Technological Innovation and the 2007 Industry Award for Excellence in Transitional Control Research from the IEEE Multiple Conference on Systems and Control. E-mail: tychai@mail.neu.edu.cn.

Jie YANG is a Ph.D. candidate at the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. His research interest covers data driven modeling technology and application for industrial process. E-mail: yjercou@126.com.

Jinliang DING received the Ph.D. degree in Control Theory and Control Engineering from Northeastern University, Shenyang, China, in 2012. He is a Professor with the State Key Laboratory of Synthetical Automation for Process Industry, Northeastern University. He has authored or co-authored over 90 refereed journal papers and refereed papers at international conferences. He is also the inventor or co-inventor of 17 patents. His current research interests include modeling, plant-wide control and optimization for the complex industrial systems, stochastic distribution control, and multi-objective evolutionary algorithms and its application. Dr. Ding was a recipient of the Young Scholars Science and Technology Award of China in 2016, the National Science Fund for Distinguished Young Scholars in 2015, the National Technological Invention Award in 2013, two First-Prize of Science and Technology Award of the Ministry of Education in 2006 and 2012, respectively, and the Best Paper Award of 2011–2013 for Control Engineering Practice. E-mail: jlding@mail.neu.edu.cn.