ANODE EFFECT PREDICTION BASED ON COLLABORATIVE TWO-DIMENSIONAL FORECAST MODEL IN ALUMINUM ELECTROLYSIS PRODUCTION

ZUGUO CHEN, YONGGANG LI*
XIAOFANG CHEN, CHUNHUA YANG AND WEIHUA GUIL

School of Information Science and Engineering, Central South University
Changsha 410083, China

(Communicated by Cheng-Chew Lim)

ABSTRACT. In this study, a new prediction algorithm is proposed, based on the collaborative two-dimensional forecast model (CTFM) that combines the traditional method and similarity search technique. The main idea of the algorithm is that the prediction of the change trend of the slope and the accumulated slope of the cell resistance as well as the useful knowledge obtained using the similarity search technique are used as the main criteria to calculate anode effect (AE)-prediction reliability. The accumulated mass deviation value is used as an auxiliary criterion to adjust the AE-prediction reliability. Finally, the current AE-process is marked according to the current AE-prediction reliability. The prediction model based on CTFM is tested on a real situation, in which multiple samples are extracted from the production of a 400 kA aluminum electrolysis cell. We observe that when the time advance of AE-prediction is within 20 ∼ 40 min, the accuracy rate of the CTFM algorithm is greater than 95% and the applicability of the method is excellent, showing a high prediction accuracy for different aluminum electrolysis cells.

1. Introduction. The anode effect (AE) is a special phenomenon that occurs in an anode under a molten salt electrolysis process [1, 2]. It is conducive to cleaning the sediment of hearth bottom and the carbon residue of electrolyte [3]. But when an AE occurs, it reduces the cleaning efficiency and lowers the quality of aluminium products. In addition, AE is also accompanied by a high cell temperature, extensive material waste, high energy consumption, and harmful fluoride gas emission [4]. Thus, the frequency with which AE occurs must be controlled to an acceptable level. Therefore, predicting AE accurately, which is critical in aluminum electrolysis production, is essential. However, the aluminum electrolysis production is part of a time-varying nonlinear system with time delays, and it involves multiple variables that are coupled and strongly correlated. Thus, an accurate mathematical model
is difficult to be established. In addition, for aluminum electrolysis, this process is conducted in high-temperature, high-magnetic, and strongly corrosive conditions. Thus, estimating the values of system parameters accurately is difficult when done online. However, off-line measuring is discontinuous and unstable [5]. All of these undesirable factors have contributed to difficulties in AE-prediction accuracy.

Many studies have been conducted on AE-prediction of aluminium electrolysis in the past 20 years. Usually, an AE-prediction method based on the slope and the accumulated slope of cell resistance is a classical mechanism analytical method [6]. It is simple and fast. However, the accuracy of predicting values when using this module is only approximately 60 to 70% and its lead time is short. To improve the accuracy and lead time of AE-prediction, many prediction techniques have been proposed, including fuzzy recognition, wavelet packet transform, expert systems, neural networks, and data-driven methods [7, 8, 9, 10, 11, 12]. A AE-prediction system for aluminium electrolytic cells is designed using fuzzy recognition technology, which can predict AE in a period of 14 to 18 min with an accuracy rate is greater than 89% [7, 8]. The wavelet packet transform method was used in [9] to analyze anode guide rod equidistant pressure drop and cell resistance signals. In this method, the norm value of the wavelet coefficients was used to construct the characteristic vector, which can reveal different feature information between the AE and normal cell states. Then the feature information is used to predict AE. This method can obtain higher prediction accuracy, but it has poor anti-interference performance. In [10], a neural network was combined with an expert system to establish an identification model. This approach was proven effective in AE-prediction and the accuracy of the predicted values was 88.3%. In [11], the data-driven method was adopted to design and develop an aluminum electrolysis fault diagnosis system, which could effectively diagnose anode failure in addition to other types of failure.

In [8], an AE predictor was designed, in which a higher-order statistical function was applied in order to analyze cell impedance data and in turn produce a spectrum representation of cell behavior. The AE prediction can detect the changing spectrum patterns, which are then used to determine the oncoming AE. This method can increase the accuracy rate of AE-prediction. However, its lead time is shorter. In [12], a method was proposed based on the Holt-Winters forecasting algorithm. It was implemented on a process control computer, that is capable of inferring from electrical readings whenever the alumina concentration in a cell approaches the level at which an AE will occur. For this approach, the AE-prediction accuracy rate can be increased, but the method can only estimate a characteristic parameter.

However, the methods previously mentioned use only some characteristic parameters prior to AE, and do not cover all types of AE. In addition, neither is poor anti-interference ability nor small lead time of an AE covered. Clearly, much work remains in developing more effective AE-prediction techniques. However, achieving better effects from AE-prediction techniques based on the classical mechanism analysis, traditional numerical analysis method, and intelligent prediction algorithm is difficult. Thus, a similarity search technique is used to obtain useful information from a set containing extensive historical data related to cell resistance and alumina feeding. It is known that hundreds of electrolytic cells exist in every aluminium electrolytic plants and that every electrolytic cell will produce more than ten million pieces of data related to cell resistance and alumina feeding in a single year. The amount of data is much greater than what we can imagine. If the stored data cannot
be understood and analyzed by effective techniques, they cannot be fully utilized to predict AE.

To improve the lead time and accuracy of AE-prediction, the collaborative two-dimensional forecast model (CTFM) algorithm is proposed. The CTFM algorithm undergoes two concurrent prediction stages: (i) a historical data sequence that is similar to the current data sequence is searched from the historical database for cell resistance and alumina feeding using the similarity search algorithm based on dynamic pattern matching distance (DPMD). Then, the voting mechanism is used to calculate the AE-prediction reliability; (ii) the slope and accumulated slope of cell resistance are calculated and the change trend is predicted by the square-root cubature Kalman filter based on the Levenberg and Marquardt algorithm (LMSR-CKF). The AE-prediction reliability is then determined according to change trend of the slope and accumulated slope of cell resistance. The AE-prediction reliability results obtained from these two stages are combined using fuzzy rule evaluation, and the syncretic AE-prediction reliability is adjusted based on the value of the accumulated mass deviation. Finally, the current AE-process is marked as different classes according to the current AE-prediction reliability.

The paper aims to contribute to AE-prediction research and practice in two specific ways. First, an AE-prediction model based on the CTFM algorithm for AE-prediction is proposed, which can improve the lead time and accuracy of AE-prediction. This study proposes. Second, the effectiveness of the AE-prediction model is verified using a sufficiently large number of samples extracted from the production data of the Jinglian Aluminum Electrolytic Plant in Inner Mongolia.

The remainder of this paper is organized as follows. Section 2 describes the aluminum electrolysis process, AE process, and explains the difficulties involved in AE-prediction. Section 3 illustrates the basic principles of the CTFM algorithm. Section 4 presents the results of an experiment conducted using samples extracted from production data. Section 5 provides a conclusion to the study and offers suggestions for future research.

2. Electrolysis process and problem.

2.1. Aluminum electrolysis process. The Hall-Heroult process is used to produce aluminium ingots in the primary aluminum industry. Alumina \((\text{Al}_2\text{O}_3)\) is used as a raw material consisting of high currents passing through an electrolytic bath with cryolite \((\text{Na}_3\text{AlF}_6)\) as the main ingredient, and the process includes other chemical additives [13]. Electrolysis occurs in electrolytic reduction cells, (also known as pots), in which the alumina molecule is broken into aluminum \((\text{Al}^{3+})\) and carbon dioxide \((\text{CO}_2)\). These cells have two electrodes: the anode (positive pole) in its upper part; and the cathode (positive pole) on the bottom. They contain mainly carbonaceous materials [14]. The electrolytic voltage is approximately 4 V. A current of approximately 400 kA is conducted through these electrodes, whereby the electrolytic bath is heated to a temperature of approximately 950 °C [15]. The craft process of the cryolite-alumina molten salt electrolysis aluminum is shown in Figure 1.

2.2. The AE process and problem description. In the aluminum electrolysis process, the AE is a major operation fault. Mechanism analysis shows that AE is caused by very low levels of alumina concentration. In generally, when alumina concentration decreases to less than 2.0%, the AE will occur [15]. When the AE
occurs, the cell voltage will suddenly increase from the normal target value of 3.8 ∼ 4.5 V to 30.0 ∼ 50.0 V, or even higher [16]. In addition, the anode gas compositions have also obvious changes. For example, the amount of carbon monoxide can increase to approximately 60% or that of carbon dioxide can decrease to approximately 20%. Approximately 20% of carbon tetrafluoride (CF$_4$) exists [17]. Simultaneously, a bright arc discharge phenomenon occurs in surrounding areas of the anode. the temperature of the electrolyte rises sharply, causing the furnace wall to melt and the carbon-block of side wall to erode [18]. These changes caused by AE will reduce current efficiency and the quantity and quality of aluminum. They will also increase power consumption and pollute the environment. Therefore, the AE is not conducive to energy-saving, cost-reduction, and production safety. However, to some extent, it can also bring some benefits for the aluminum electrolytic cell, such as promoting the separation between carbon residues and electrolyte, cleaning the cell hearth, and accelerating the alumina dissolution. Therefore, the AE should be neither too little nor too much. The frequency with which AE occurs should be controlled to within a reasonable range.

To improve production efficiency of electrolytic aluminium, plants should ensure that normal AE occurs on time and should also prevent abnormal AE from appearing. However, the aluminum electrolysis production process is a system that is non-linear time-changing, multi-variable coupling, and has a large delay. Therefore, detecting alumina concentration on-line in real-time is difficult. Fortunately, in this study, we found that the real time measure value of cell resistance shows abnormal fluctuations before AE occurs. In addition, alumina feeding directly influences changes of cell resistance. To improve the lead time and accuracy of AE-prediction, the CTFM algorithm is proposed. The main idea of the algorithm that a search algorithm (developed based on the DPMD algorithm) is used to obtain useful knowledge from historical data related to cell resistance and alumina feeding. This useful knowledge is then used to predict AE by combining the change trend of the slope and
accumulated slope of cell resistance. Finally, to increase AE-prediction reliability, the reliability is adjusted based on the value of the accumulated mass deviation.

3. CTFM algorithm. The previous discussion shows that predicting AE accurately is difficult. To improve the lead time and accuracy of AE-prediction, the CTFM algorithm is proposed. Figure 2 gives an overview of the CTFM algorithm. The process is described as follows:

1) Data related to cell resistance and alumina feeding is pre-processed. Missing data is then filled in the database and incorrect data is eliminated.
2) The algorithm is divided into two concurrent prediction branches, A and B. In Branch A, sliding window is used to divide subsequences, and the historical data sequence that is similar to the current data sequence is searched for and obtained from the historical database related to cell resistance and alumina feeding using the similarity search algorithm developed based on DPMD algorithm.
3) Select the first five search results having the greatest similarity. Then, AE-prediction credibility is calculated using the voting mechanism based on the occurrence of the AE phenomenon in each selected historical data sequence.
4) Handle the AE-prediction credibility in Branch A using a fuzzy method.
5) In Branch B, the slope and accumulated slope of cell resistance are calculated and their change trends are predicted using the LMSRCKF algorithm.
6) Calculate the credibility of AE-prediction according to the change trends of the slope and accumulated slope and handle the AE-predicting credibility in Branch B using a fuzzy way.
7) The credibility results of two concurrent prediction branches is combined using fuzzy rule evaluation, and a fuzzy reasoning table of credibility is generated.
8) Adjust AE-prediction reliability based on value of the accumulated mass deviation and deal with the course flags for AE-prediction.

3.1. Data pre-processing. In industrial production, some data (including consecutive data) are missing or incorrect in the historical data set related to cell resistance and alumina feeding. This could be caused by many factors, including transport process failure, equipment failure and human faults [20]. The data pre-processing technologies can improve data quality, and help improve the results of the similarity search and the accuracy of AE-prediction [21]. Two main types of abnormal data related to cell resistance and alumina feeding exist: incorrect values and missing values.

The usual methods for removing incorrect values include: the criterion of $3\sigma$ [22], Chauvenel criteria [23], Grubbs criterion [24], and Dixon criterion [25]. All of the methods have been established by experts' subjective experience, and thus far, no unifying rules [26] exist. The premise of these methods is that the data should obey normal distribution. Therefore, when the data deviate from normal distribution, these methods are unreliable [27]. To improve the reliability of these methods, this study uses the criterion of $3\sigma$ in the Dixon criterion. Note that whether the doubtful value is abnormal can be determined from judging the results of both the criterion of $3\sigma$ and the Dixon criterion.

On the other hand, missing data will also severely affect the accuracy of AE-prediction. Therefore, handling missing data properly is major part of data pre-processing. This type of data mainly includes includes a value or continuous value [28]. If a value is lost, it can be replaced by the mean value obtained from several sets of adjacent data. If a continuous set of data is lost, the mean value of cell
resistance and alumina feeding in the same period of several adjacent cells can be used to replace it.

3.2. **Similarity search based on DPMD algorithm.** Similarity search has become increasingly important in large-scale and high-dimensional databases in which the contained objects do not possess any natural order [29]. The most common similarity search uses the mathematical notion of metric space, which enables efficient index structures to be built. Thus, similarity search is a crucial task in many multimedia and data mining applications, and extensive studies have been conducted in this area of research [30]. In general, a similarity search is performed based on a similarity measurement. However, the relationship between similarity and distance measurements can be described by the following equation.

\[ S(x, y) = \frac{1}{d(x, y) + 1}, \]  

(1)

where \( d(x, y) \) is distance measurement. \( x \) and \( y \) are coordinates data, respectively.

As equation (1) reveals, the key aspect of similarity measurement is whether the distance measurement can be calculated accurately. Currently, many distance calculation methods are used for similarity measurement. For example, the Minkowski distance was adopted in [31] to calculate similarity measurement in a similarity search of big data. However, when time series are affected by self-noise and fluctuation characteristics, multiple deformations will generate similar time series such as amplitude shift and scaling, linear drift, and discontinuity. In these cases, the
Minkowski distance between the times series is very big and is opposed to the shape similarity between two time series [42]. Therefore, it cannot accurately describe the similarity between two time series.

To address the aforementioned problem, the DPMD algorithm is used to calculate the similarity measurement. The pattern is described by a line segment and its two characteristics (length and slope) are extracted. It is known that the pattern length is irrelevant to the amplitude of time series and that the pattern slope can reflect the relative size of amplitude of the time series. Therefore, the DPMD algorithm can solve the problem when a time series is affected by self-noise and fluctuation characteristics.

Consider a time series \( X = (x_1, x_2, x_3, \ldots, x_m) \). Its pattern representation is: \( P(X) = (px_1, px_2, px_3, \ldots, px_v) \), where \( px_i = (lx_i, \kappa x_i) \), \( 1 \leq i \leq v \), \( lx_i \) is the length of the pattern and \( \kappa x_i \) is the slope of the pattern. Consider another time series \( Y = (y_1, y_2, y_3, \ldots, y_n) \). Its pattern representation is \( P(Y) = (py_1, py_2, py_3, \ldots, py_u) \), where \( py_i = (ly_i, \kappa y_i), 1 \leq j \leq u \), with \( ly_j \) and \( \kappa y_j \) being the length and the slope of the pattern, respectively. The calculating equation of the DPMD is given by [32]:

\[
D(X, Y) = d(px_1, py_1) + \min \left\{ \begin{array}{l}
D(P(X) - px_1, P(Y)) \\
D(P(X), P(Y) - py_1) \\
D(P(X) - px_1, P(Y) - py_1)
\end{array} \right.,
\]  

(2)

where \( P(X) - p(x_1) \) and \( P(Y) - p(y_1) \) are subsequence whose first element of the \( P(X) \) and \( P(Y) \) is removed, respectively. \( d(px_1, py_1) \) is the pattern distance between \( px_1 \) and \( py_1 \) and its calculating equation is [32]:

\[
d(px_1, py_1) = \frac{|lx_1 - ly_1|}{\min\{lx_1, ly_1\}} + \frac{|\kappa x_1 - \kappa y_1|}{\min\{\kappa x_1, \kappa y_1\}},
\]  

(3)

where \( lx_1 \) is the length of the pattern \( P(X) \); \( ly_1 \) is the length of the pattern \( P(Y) \). \( \kappa x_1 \) is the slope of the pattern \( P(X) \); \( \kappa y_1 \) is the slope of the pattern \( P(Y) \). The calculating equation of similarity is given by equation (1).

When the similarity between two time series is calculated, the corresponding elements in the times series must be aligned for comparison. The similarity can determine whether the two time series are similar. The similarity search not only can reveal the original characteristics, but also can obtain useful knowledge. During the searching process, the times series should be divided into many subsequences using sliding window in order to enhance the extracting pattern characteristics [33].

3.3. Calculate the AE-predicting credibility. After the historical data sequence that is similar to the current data sequence is searched for and obtained the historical database related to cell resistance and alumina feeding using the DPMD based similarity search algorithm, we select the first five search results having the greatest similarity. Then, the AE-prediction credibility is calculated using the voting mechanism based on the occurrence of the AE phenomenon in each selected historical data sequence. For example, if the AE phenomenon occurs in one of the five selected historical data sequences, the AE-prediction credibility \( T_r \) is 1, and the range of credibility \( T_a \) is \([0, 5] \). Similarly, the range of the AE-prediction credibility \( T_a \) obtained based on search results of the alumina feeding is also \([0, 5]) \). The total AE-prediction credibility obtained based on similarity search will be mapped to \([0, 1]\). It can be expressed as

\[
CB = \frac{T_r + T_a}{10},
\]  

(4)
where the range of $CB$ is $[0,1]$.

3.4. **Calculate the slope and the accumulated slope of cell resistance.** The slope of cell resistance is the average change rate of the filter resistance in the last 2 min. The accumulated slope of cell resistance is the cumulative increment of filter resistance in the last 8 min. The sample period of the filter resistance is 2 s. The slope and the accumulated slope of cell resistance can be obtained as [7]

$$S_R(k) = \frac{\left\{R_w(k-1) - R_w(k-3)\right\} + 2 \times [R_w(k) - R_w(k-4)]}{5}$$  \hspace{1cm} (5)

$$CS = \frac{(15/16) \times A_R(k-1) + S_R(k)}{2}, A_R(0) = S_R(0)$$ \hspace{1cm} (6)

where $R_w(k)$ is the filter resistance at time $k$, $S_R(k)$ is the slope of cell resistance, and $A_R(k)$ is the accumulated slope of cell resistance. $S_R(k)$ is the slope of cell resistance. $CS$ is the accumulated slope of cell resistance.

3.5. **LMSRCKF algorithm.** The slope and the accumulated slope of cell resistance will happen obvious change before AE occur. During AE-prediction, if we merely use the traditional method (the slope and accumulated slope of cell resistance) to predict AE, obtaining a longer AE-prediction lead time is difficult. In industrial production, if the AE-prediction lead time is too short, it will be of little use in guiding production. The LMSRCKF algorithm is used to the change trends of the slope and the accumulated slope of cell resistance, which can obtain earlier their the change situations. So LMSRCKF algorithm can obtain a longer lead time.

The state and measurement equations must be deduced correctly before running the LMSRCKF algorithm. The two equations are deduced by the auto-regressive and moving average model (ARMA) in [34, 35]. The LMSRCKF algorithm is an improvement on the square-root cubature Kalman filter (SRCKF) algorithm using the Levenberg and Marquardt (LM) algorithms. The specific procedure of the SRCKF algorithm is detailed in [36]. For real application in an actual engineering project, if uncertainties exist such as a simplified model state, inaccuracies, or mutation of the initial state, the performance of the SRCKF algorithm will be affected [37]. In addition, the SRCKF algorithm must obtain a priori statistical properties of both the system and measurement noise. If the statistical properties of the noises are unknown or time-varying, the updated value of state observation cannot ensure that the estimation error is decline. In fact, it will cause the estimated covariance value to be smaller than the real covariance value, which will affect the utilization efficiency of the observed information. These problems may also cause non-convergence of the SRCKF algorithm.

To improve robustness, convergence speed, and estimation accuracy of the SRCKF algorithm, it is essential that some optimal algorithms are used to improve it. The LM algorithm combines the steepest descent and Newton methods. Not only does it possess high learning efficiency and fast convergence speed but also has a high recognition rate [38]. The LM algorithm is used to adjust the predicted error covariance and state updating value of the SRCKF algorithm, which can enhance the robustness, and improve convergence speed of the algorithm. The LMSRCKF algorithm for estimating the state vectors $H_t$ in a nonlinear discrete-time dynamical system can be summarized as follows.

(a) Time updating
The cubature points are obtained by

\[ H_{a,t-1} = s_{t-1}^{\alpha} + \hat{H}_{t-1|t-1}, \alpha = 1, 2, \ldots, L, \]  

where \( L \) is the number of cubature points. If a third order cubature principle is adopted in the algorithm, then \( L = 2n_x \), where \( n_x \) is the state dimension. In addition, \( s_{t-1} \) is the triangular square-root matrix of the error covariance matrix and \( t \) is the number of iterations. \( \hat{H}_{t-1|t-1} \) is used to estimate the value of the \((t-1)\)th time. Finally, \( \xi_\alpha \) is a basic cubature point, which is be generated by the method in [36].

The propagated cubature points are obtained by

\[ H^*_{a,t|t-1} = f(H_{a,t-1}, U_{t-1}), \]  

where \( U_{t-1} \) is the known control input.

The Cholesky division factor of the state prediction vector is given by

\[ \hat{H}_{t|t-1} = \sum_{\alpha=1}^{L} \omega_\alpha H^*_{a,t|t-1}. \]  

The triangular square-root matrix of the predicted covariance is calculated as given below:

\[ \hat{W}_{t|t-1} = Tria([H^*_{a,t|t-1}, Q_{t-1}]), \]  

thus, \( Q_{t-1} = J_{Q,t-1}^{T}Q_{t-1}^{T} \), and \( Q_{t-1} \) is the covariance of the process noise at the \((t-1)\)th time. \( Tria() \) is the function through which the square matrices of the matrices are obtained using matrix diagonalization. \( H^*_{a,t|t-1} \) is defined as

\[ H^*_{t|t-1} = \frac{1}{L}[H_{1,t|t-1} - \hat{H}_{t|t-1}, H_{2,t|t-1} - \hat{H}_{t|t-1}, \ldots, H_{L,t|t-1} - \hat{H}_{t|t-1}]. \]  

(b) measurement-update based on LM algorithm

The Levenberg-Marquardt method is adopted to optimize predicted covariance

\[ \hat{R}_{t}^{(0)} = \hat{R}_{t|t-1}; \quad \hat{W}_{t}^{(0)} = \hat{W}_{t|t-1}, \]  

\[ \hat{W}_{k|t-1} = [I - \hat{W}_{k|t-1}^{(k)} \times (\hat{W}_{t|t-1}^{(k)} + \mu_1 I)^{-1}]\hat{W}_{t|t-1}^{(k)}, \]  

where \( k = 1, 2, \ldots, N \), \( N \) is maximum iterations. \( \mu_1 \) is weighting coefficient. When \( \mu_1 \) is large, LM is the steepest descent method; when \( \mu_1 \) is zero, LM is Newton method. \( \mu_1 \) is 0.5 to 0.8 according to experience.

The Cholesky decomposition can be replaced by the following equation:

\[ \hat{C}_{t|t-1}^{(k)} = chol(\hat{W}_{t|t-1}^{(k)}), \]  

where, the \( chol() \) is Cholesky decomposition.

The cubature point is obtained by

\[ H^{(k)}_{a,t|t-1} = \hat{C}_{t|t-1}^{(k)} \xi_\alpha + \hat{H}_{t|t-1}, \]  

The transmission cubature point is obtained by

\[ Z^{(k)}_{a,t|t-1} = f(H^{(k)}_{a,t|t-1}). \]
The factor of the measurement prediction, the covariance of the Cholesky division of the measurement prediction, and the new information variance are given by

\[
\tilde{Z}_{a,t|t-1}^{(k)} = \sum_{a=1}^{L} \omega_a Z_{a,t|t-1}^{(k)*},
\]

\[
\eta_{zz,t|t-1}^{(k)} = \sum_{a=1}^{L} \omega_a Z_{a,t|t-1}^{(k)*} (Z_{a,t|t-1}^{(k)*})^T - \tilde{Z}_{a,t|t-1}^{(k)} (\tilde{Z}_{a,t|t-1}^{(k)})^T,
\]

\[
S_{zz,t|t-1}^{(k)} = qr\{\eta_{zz,t|t-1}^{(k)} \sqrt{R_t} \},
\]

where \( R_t \) is the covariance of the measurement noise.

\[
\chi_{zz,t|t-1}^{(k)} = \sum_{a=1}^{L} \omega_a H_{a,t|t-1}^{(k)} (Z_{a,t|t-1}^{(k)*})^T - \tilde{H}_{t|t-1}^{(k)} (\tilde{Z}_{a,t|t-1}^{(k)})^T,
\]

\[
\beta_{zz,t|t-1}^{(k)} = \chi_{zz,t|t-1}^{(k)} (\eta_{zz,t|t-1}^{(k)})^T.
\]

To obtain a more accurate LMSRCKF state estimate, the LM algorithm is used again to optimize the filtering gain of the LMSRCKF. The LMSRCKF state estimate and its covariance can be calculated by the following equation.

\[
K_{t|t}^{(k)} = [(\beta_{zz,t|t-1}^{(k)}/S_{zz,t|t-1}^{(k)} T)/S_{zz,t|t-1}^{(k)}] \times \{ I - [(\beta_{zz,t|t-1}^{(k)}/S_{zz,t|t-1}^{(k)} T)/S_{zz,t|t-1}^{(k)} + \mu_2 I] \}^{-1},
\]

\[
H_{t|t}^{(k)} = \tilde{H}_{t|t-1} + K_{t|t}^{(k)} (Z_{a,t|t-1}^{(k)*} - \tilde{Z}_{a,t|t-1}^{(k)}),
\]

\[
W_{t|t}^{(k)} = \text{chol}([\tilde{C}_{t|t-1}^{(k)}, K_{t|t}^{(k)} S_{zz,t|t-1}^{(k)} - 1]),
\]

where \( \mu_2 \) is the weighting coefficient. When it is sufficiently large, LM behaves like the steepest descent method. When it approximates zero, LM behaves like the Newton method \([38]\). \( \mu_2 \) is a value from between 0.5 to 0.8 based on experience. Termination conditions for the iteration is

\[
H_{t|t}^{(k+1)} - H_{t|t}^{(k)} \leq \varepsilon \quad \text{or} \quad k \leq N_{max},
\]

where \( \varepsilon \) and \( N_{max} \) are pre-set threshold value.

3.6 Calculate the AE-predicting credibility according to the change trends of the slope and the accumulated slope. To calculate the credibility of AE-prediction using change trends of the slope and accumulated slope of cell resistance, two variables are mapped to the range \([0, 5]\). The process is similar to variable fuzzing. The algorithm selects trapezium distribution as a membership function. It can be expressed as

\[
u(b) = \begin{cases} 
0 & b \leq e_1 \\
\frac{5(b - e_1)}{(e_2 - e_1)} & e_1 < b \leq e_2 \\
5 & b > e_2
\end{cases}
\]

\[
u(\lambda) = \begin{cases} 
0 & \lambda \leq e_1 \\
\frac{5(\lambda - e_1)}{(e_2 - e_1)} & e_1 < \lambda \leq e_2 \\
5 & \lambda > e_2
\end{cases}
\]

where \( b \) and \( \lambda \) are change trends of the slope and accumulated slope, respectively. The slope and accumulated slope of cell resistance can be obtained by equation 5 and 6. The change trends of the slope and accumulated slope can be predicted by
the LMSRCKF algorithm. $u(b)$ and $u(\lambda)$ are values mapped to the range $[0, 5]$, $e_1$ is the lower limit of AE-occurrence, and $e_2$ is the upper limit of AE-occurrence. The AE-prediction credibility $CA$ can be expressed as

$$CA = (u(b) + u(\lambda))/10,$$

(28)

where the range of $CA$ is $[0,1]$.

### 3.7. Fuse AE-predicting credibility using fuzzy rule evaluation.

The ranges for both $CA$ and $CB$ are $[0, 1]$. $CA$ and $CB$ are divided into seven levels using the ideas of variable fuzzing. The seven levels are defined as $BT$ (positive big), $MT$ (middle big), $ST$ (small big), $Z$ (zero), $SF$ (negative small), $MF$ (negative middle), and $BF$ (negative big). To achieve more accurate credibility, the fuzzy rule-union operator is used to increase the AE-predicting credibility. The fuzzy rule-union operators [39] are as follows:

$$CAB = \vee(CA, CB) = \max(CA, CB),$$

(29)

where $CAB$ is the result for which AE-predicting credibility is fused. Its type is as shown in Figure 3.

| $\begin{array}{c|cccccc}
\text{result type} \\
\hline 
\text{CA} & \text{CB} & \text{BT} & \text{MT} & \text{ST} & \text{Z} & \text{SF} & \text{MF} & \text{BF} \\
\hline 
\text{BT} & \text{BT} & \text{BT} & \text{BT} & \text{BT} & \text{BT} & \text{BT} & \text{BT} & \text{BT} \\
\text{MT} & \text{BT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} \\
\text{ST} & \text{BT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} & \text{MT} \\
\text{Z} & \text{BT} & \text{MT} & \text{ST} & \text{ST} & \text{ST} & \text{ST} & \text{ST} & \text{ST} \\
\text{SF} & \text{BT} & \text{MT} & \text{ST} & \text{Z} & \text{Z} & \text{Z} & \text{Z} & \text{Z} \\
\text{MF} & \text{BT} & \text{MT} & \text{ST} & \text{Z} & \text{SF} & \text{SF} & \text{SF} & \text{SF} \\
\text{BF} & \text{BT} & \text{MT} & \text{ST} & \text{Z} & \text{SF} & \text{MF} & \text{MF} & \text{BF} \\
\end{array}$

**Figure 3.** The overview of CTFM algorithm

### 3.8. Calculate material accumulation deviation.

Many researchers suggest that AE occurs because of low alumina concentration [17, 18, 40, 41]. Therefore, the material accumulation deviation, which is the difference between the addition of alumina in the electrolytic cell and theoretical consumption of alumina within the nearest two hours, is more important for AE-prediction. This information can be obtained using the material balance method. This can be calculated as [7, 40]

$$CW = E(k) = (119/200) \times E(k-1) + J(k) - c \times I(k),$$

(30)

where $E(k)$ and $E(k-1)$ are material accumulation deviations at the present and the previous periods, respectively; $J(k)$ is the addition of alumina at the present period; $I(k)$ is the average current within the current period; $c$ is the conversion coefficient (kg·kA$^{-1}$) between the consumption of alumina and the current. $CW$ is material accumulation deviation.
3.9. **Adjust reliability and obtain AE-predicting results.** In the industrial production process, many factors can cause the change of cell resistance. Moreover, the alumina concentration in electrolyte is a principal factor. Therefore, the curve of the alumina feeding is associated with the curve of cell resistance [42]. The high accuracy of AE-prediction that is based only on alumina feeding and cell resistance is difficult to obtain. To improve the accuracy of AE-prediction, the curve of the alumina feeding and cell resistance is used as the main criterion, and the material accumulation deviation is used as an auxiliary criterion. In other words, the credibility of AE-prediction is adjusted according to the material accumulation deviation after the AE-prediction result is obtained using the curve of the alumina feeding and cell resistance. The specific adjustment procedures are described as follows:

1) The material accumulation deviation ($CW$) is divided into seven levels by fuzzy language. The seven levels are defined as $BT$ (positive big), $MT$ (middle big), $ST$ (small big), $Z$ (zero), $SF$ (negative small), $MF$ (negative middle) and $BF$ (negative big).

2) A rule which is used to adjust reliability is established. For example, if $CW$ is $BF$, increase $CAB$. However, if $CW$ is $BT$, reduce $CAB$. The adjustment amount is also divided into seven levels by a fuzzy language.

3) A time of 1 min is used as a prediction period. If the reliability is greater than the setting value of 0.5, there will be a trend of AE occurrence. Otherwise, the AE will not occur.

4) The process of AE-prediction is marked and produces some conclusive predictions, such as “start,” “continue,” “fade away,” and “no occur.”

4. **Experimental results and analysis.** A set of experiments was conducted to verify the feasibility and high accuracy of the CTFM algorithm. Experimental data were obtained from an aluminium electrolytic plant. The sampling interval of data was 2s. The experimental environment was a Windows PC with AMD A10-6700 APU and radeon (tm) HD Graphics CPU 4G RAM. The MATLAB language was used to implement a similarity search and to execute the LMSRCKF algorithm.

To verify the feasibility of the DPMD algorithm, an experiment was designed to use the algorithm to search for a data segment that is similar to the sample date segment from the 46 GB data set. In the experimental process, the number of the sample data segment was 200. The sliding window was used to capture equal-length data segments from the historical data set related to cell resistance and alumina feeding. There are 200 numerical value in the sliding window. The sliding interval of the sliding window was 10. Figure 4 shows the results of cell resistance obtained by the searching algorithm at different times. In the actual search process, when the sliding window was moved, the similarity between the historical data and current data was calculated and displayed in the test window. Figure 4(a) and 4(b) show the results of cell resistance obtained by the search algorithm at the 990th and at the 3040th similarity calculations, respectively. The red, green, and blue curves represent historical data, current data, and similarity, respectively.

Figure 5 presents the five most similar curves of cell resistance with current data, which were obtained from the historical data set. They are the red, the green, the blue, the turquoise and the carmine curves in the first box. The green curve in the second box represents current data curve. Table 1 is the similarity corresponding to each curve shown in Figure 5. We can see from the experimental results that use the DPMD algorithm for searching data relate to cell resistance is feasible.
Figure 4. The results of the cell resistance at different times are obtained by the search algorithm.

Figure 5. The five most similar curves of the cell resistance with current data are obtained from historical data set

Table 1. The corresponding similarity to each curve in Figure 5.

| Curve types     | Similarity |
|-----------------|------------|
| Red curve       | 0.9419     |
| Green curve     | 0.9526     |
| Blue curve      | 0.9661     |
| Turquoise curve | 0.9628     |
| Carmine curve   | 0.9587     |
Figure 6 shows the results of the alumina feeding obtained by the search algorithm at different times. As previously mentioned, when the sliding window was moved, the similarity between the obtained historical and current data were calculated and displayed in the test window. The similarity calculation results are shown as a blue curve in Figure 6. Figure 6(a) and (b) show the results of the alumina feeding obtained by the search algorithm at the 215th and at the 780th similarity calculations, respectively. Likewise, the red, green, and blue curves represent historical data, current data, and similarity, respectively.

Figure 7 shows the five most similar curves of the alumina feeding with current data, which were obtained from the historical data set. They are the red, green, blue, turquoise, and carmine curves in the first box. The green curve in the second box is the current data curve. Table 2 presents the corresponding similarity to each curve shown in Figure 7. We can see from the experimental results that using DPMD algorithm for searching data related to alumina feeding is also feasible.

To verify the accuracy of the DPMD algorithm for AE-prediction during the similarity search, three similarity search algorithms (DPMD, Minkowski [43] and Euclidean [44]) were used to search the same data set related to cell resistance and alumina feeding. Each similarity search algorithm was used on the data of

![Figure 6](image1.png)  
(a) 215th similarity calculation  
(b) 780th similarity calculation

![Figure 7](image2.png)
Figure 7. The five most similar curves the alumina feeding with current are obtained from historical data set (the red curve, the green curve, the blue curve the turquoise curve and the carmine curve in the first box are five most similar curves, respectively. The green curve in the second box is current data curve)

Table 3. Mean accuracy and time advance of three similarity search methods

| Time advance (min) | DPMD Mean accuracy | Minkowski distance Mean accuracy | Euclidean distance Mean accuracy |
|-------------------|-------------------|---------------------------------|--------------------------------|
| 0 ~ 40            | 55.4%             | 47.9%                           | 45.9%                           |
| 5 ~ 40            | 51.1%             | 42.3%                           | 38.5%                           |
| 10 ~ 40           | 46.8%             | 38.3%                           | 35.6%                           |
| 15 ~ 40           | 43.9%             | 32.6%                           | 31.2%                           |
| 20 ~ 40           | 40.4%             | 26.1%                           | 25.6%                           |
| 25 ~ 40           | 38.4%             | 21.4%                           | 22.1%                           |
| 30 ~ 40           | 36.5%             | 15.2%                           | 18.3%                           |

10 electrolytic cells. 50 data segments related to the cell resistance and alumina feeding of each electrolytic cell were selected as sample data.

Mean accuracies of the three similarity search algorithms for AE-prediction at different time advances were calculated, see Table 3. Table 3 shows that the mean accuracy of AE-prediction decreases with each time advance. However, the mean accuracy of AE-prediction by the DPMD algorithm is the highest. The mean accuracy of the Minkowski distance algorithm is higher than that of the Euclidean distance algorithm, the latter thus being the worst of the three. Therefore, the DPMD algorithm achieves the best results for AE-prediction when compared with the other two algorithms.
To verify the validity and accuracy of the LMSRCKF algorithm for change trend, three prediction algorithms (extended Kalman filter (EKF), SRCKF and LMSRCKF) were used to predict change trend of the slope and accumulated slope of cell resistance and calculate AE-prediction accuracy based on the change trend. Each prediction algorithm was used with data of 10 electrolytic cells during the experiment. We considered 50 data segments of the slope and accumulated slope of cell resistance of each electrolytic cell as test data. In addition, each data segment contained 641 numerical values. To verify accuracy of the LMSRCKF algorithm, the errors (including absolute and relative) for prediction results and actual values were analyzed. The calculation of these absolute errors is:

\[ E_{\text{absolute error}} = Y_P - Y_T, \]  

where \( Y_P \) is prediction value and \( Y_T \) is actual value. The calculation formula of relative errors is:

\[ E_{\text{relative error}} = \left( \frac{E_{\text{absolute error}}}{Y_T} \right) \times 100\%, \]

where \( E_{\text{relative error}} \) is the relative error and \( E_{\text{absolute error}} \) is the absolute error.

Figure 8 shows the prediction results of the three algorithms for the slope of cell resistance. We can see that the predictive effect of the LMSRCKF algorithm is the best, the SRCKF algorithm ranks second, and the EKF algorithm is the worst.

---

**Figure 8.** The prediction results of the three algorithms for the slope of cell resistance. (Figure 8(a) presents prediction result of the EKF algorithm for the slope of cell resistance. Figure 8(b) shows prediction result of the SRCKF algorithm for the slope of cell resistance. Figure 8(c) gives Prediction result of the LMSRCKF algorithm for the slope of cell resistance.)

Figure 9 shows absolute errors between actual and prediction values for the slope of cell resistance by the three algorithms. We can see that the absolute error obtained by the LMSRCKF algorithm is the smallest, and that obtained by the EKF algorithm is the largest.

By contrast, Figure 10 shows relative errors between actual and prediction values for the slope of cell resistance as obtained by the three algorithms. The relative error obtained by the LMSRCKF algorithm is the smallest and the prediction result is the most stable. The relative error obtained by the EKF algorithm is the largest and its prediction result is the most unstable. In summary, not only are the absolute and
relative errors obtained by the LMSRCKF algorithm the smallest, but its prediction result is also the most stable. In addition, because of the influence of the noises, some abnormal jumping appear in the prediction results of the other two algorithms, which means that satisfying the requirements of AE-prediction is difficult.

Figure 11 shows prediction results of the three algorithms for the accumulated slope of cell resistance. We can see that the predictive effect of the LMSRCKF algorithm is the best. In addition, the predictive effect of the SRCKF algorithm is better than that of the EKF algorithm.

Figure 12 shows absolute errors between actual and prediction values obtained by the three algorithms for the accumulated slope of cell resistance. The same conclusion Figure 9 can be drawn for Figure 12.

Figure 13 shows relative errors between actual and prediction values obtained by the three algorithms for the accumulated slope of cell resistance. Here, the same conclusion as that of Figure 10 can be drawn. To verify the validity and accuracy of the LMSRCKF algorithm for AE-prediction, the obtained change trend of the slope and accumulated slope of cell resistance by three prediction algorithms (EKF,
Figure 11. The prediction results obtained by the three algorithms for the accumulated slope of cell resistance (Figure 11(a) is on prediction result obtained by the EKF algorithm for the accumulated slope of cell resistance. Figure 11(b) is on prediction result obtained by the SRCKF algorithm for the accumulated slope of cell resistance. Figure 11(c) is on prediction result obtained by the LMSRCKF algorithm for the accumulated slope of cell resistance.)

Figure 12. The absolute errors between actual value and prediction value obtained by the three algorithms for the accumulated slope of cell resistance.

SRCKF and LMSRCKF) is used to predict AE. Mean accuracies of the three prediction algorithms for AE-prediction at different time periods were calculated, as shown in Table 4. Table 4 shows that the mean accuracies of the three prediction algorithms for AE-prediction all decreased with each time advance. However, the mean accuracy of the LMSRCKF algorithm for AE-prediction was better than those of the EKF and SRCKF algorithm at different time advances. The mean accuracy of the SRCKF algorithm for AE-prediction was better than that of the EKF algorithm at different time advances. The mean accuracy of the EKF algorithm for AE-prediction was the worst. In summary, the results show that the LMSRCKF algorithm for AE-prediction was the best.

In experiment, the three input variables $CA$, $CB$ and $E(k)$ used the trapezoid membership function, as shown in Figure 14. Figure 14(a), (b), and (c) show the membership function of the input variables $CA$, $CB$, and $E(k)$, respectively.
Figure 13. The relative error between actual value and prediction value obtained by the three algorithms for the accumulated slope of cell resistance.

Table 4. Mean accuracy and time advance of three prediction algorithms

| Time advance (min) | LMSRCKF | SRCKF | EKF  |
|-------------------|---------|-------|------|
| 0 ~ 40            | 75.6%   | 57.1% | 51.6%|
| 5 ~ 40            | 72.3%   | 51.2% | 45.3%|
| 10 ~ 40           | 68.8%   | 44.7% | 40.2%|
| 15 ~ 40           | 62.9%   | 35.3% | 32.7%|
| 20 ~ 40           | 59.8%   | 31.2% | 28.6%|
| 25 ~ 40           | 55.4%   | 25.7% | 23.4%|
| 30 ~ 40           | 51.2%   | 19.3% | 17.6%|

Table 5. The result statistics of AE-predicting using the fused results of fuzzy variable CA and fuzzy variable CB.

| Results of AE-predicting of the fused CA and CB | Mean accuracy | Time advance (min) |
|------------------------------------------------|---------------|-------------------|
| 93.3%                                         | 0 ~ 40        |
| 91.5%                                         | 5 ~ 40        |
| 89.2%                                         | 10 ~ 40       |
| 88.6%                                         | 15 ~ 40       |
| 85.1%                                         | 20 ~ 40       |
| 83.9%                                         | 25 ~ 40       |
| 80.2%                                         | 30 ~ 40       |

In the CTFM algorithm, the input variables CA and CB were fused by the fuzzy rules previously mentioned. The results were transformed into the range [0, 1] using the defuzzification method of center of area (COA), as shown in Figure 15, and were used to predict the AE. Statistical results of AE-prediction are shown in Table 5. Table 5 shows that the results are apparently better than those shown in Tables 3 and 4.
To verify the feasibility and accuracy of the CTFM algorithm for AE-prediction, the data from industrial field was used to conduct prediction experiments. The combined results of fuzzy variables CA and CB were adjusted using material accumulation deviation. Then, the adjusted results were transformed into the range [0, 1] using the defuzzification method of COA, as shown in Figure 16, and were used to predict the AE. In industrial production, there is little use in guiding production if the response time is too short. The time advance of AE-prediction within the period
of 20 ∼ 40 min is more reasonable based on field operation experience. When big production data of the 400kA aluminum electrolysis cell is used to test the CTFM algorithm, the AE occurs in a period of 20 ∼ 40 min after AE-prediction. The AE-prediction is then considered a success. Otherwise, it is regarded as a failure. When the AE occurs within 20 min after AE-prediction, then AE-prediction is regarded as being under-reported. When the AE occurs after 40 min following AE-prediction, then AE-prediction is regarded as being mis-reported. Table 6 shows the experimental statistical results of AE-occurrence and AE-prediction. We can see that the prediction accuracies of the CTFM algorithm were all higher than 93% for every electrolytic cell. The mean accuracy was 95.3%. Compared to those shown in Table 5, the prediction results as shown in Table 6 were apparently better and thus satisfy the requirements of efficiency and energy saving.

5. Conclusions. In this study, a new prediction algorithm (CTFM) for the AE-prediction was proposed. The method contains three steps for prediction. First, the
method obtains useful knowledge from historical data sets related to cell resistance and alumina feeding using the similarity searching method developed based on the DPMD algorithm. Then, this knowledge is used to predict AE by combining with the change trend of the slope and accumulated slope of cell resistance. Finally, the AE-prediction reliability is adjusted according to the value of the accumulated mass deviation.

This study developed a new method to predict AE in real time by analyzing historical data in industrial production. The method goes beyond the typical data-driven system based on discretization and nonlinear process data and includes analysis of the dynamic process for the working conditions of the electrolysis, anode changes and alumina feeding. The method effectively realized a prediction method based on knowledge fusion. Our study showed that when the time advance of AE-prediction was within a period of 20 ~ 40 min, the mean accuracy of the CTFM algorithm for AE-prediction was greater than 95%. These prediction results satisfy the requirements of efficiency and energy saving.

Acknowledgments. The authors would like to extend sincere gratitude to Professor Kok Lay Teo, Fanbiao Li, Jie Li, and Biao Huang for their instructive advice and guidance.

REFERENCES

[1] G. Bearne, M. Dupuis and G. Tarcy, On the anode effect in aluminum electrolysis, in Essential Readings in Light Metals: Aluminum Reduction Technology, (eds. J. Thonstad, T. A. Utigard and H. Vogt), Metals and Alloys, 2 (2013), 131–138.
[2] B. Bardet, T. Foetisch, S. Renaudier, J. Rappaz, M. Flueck and M. Picasso, Alumina dissolution modeling in aluminium electrolysis cell considering MHD driven convection and thermal impact, in Light Metals, Springer International Publishing, (2016), 315–319.
[3] D. S. Wong, P. Fraser, P. Lavoie and J. Kim, PFC emissions from detected versus nondetected anode effects in the aluminum industry, JOM, 67 (2015), 342–353.
[4] L. Dion, L. I. Kiss, S. Ponskisk and C. L. Lagacé, Prediction of low-voltage tetrafluoromethane emissions based on the operating conditions of an aluminium electrolysis cell, JOM, 68 (2016), 2472–2482.
[5] L. Kong, C. Yu, K. L. Teo and C. Yang, Robust real-time optimization for blending operation of alumina production, Journal of Industrial and Management Optimization, 13 (2017), 1149–1167.
[6] M. Farrow, Prediction of AEs in aluminum reduction cells, JOM, 36 (2013), 33–34.
[7] J. Li, F. Q. Ding, M. J. Li, J. Xiao and Z. Zou, Intelligent anode effect prediction method for prebaked-anode aluminum reduction cells, Journal of Central South University of Technology, 32 (2001), 29–32.
[8] D. G. Bell, System for predicting impending anode effects in aluminum cells, US, US. 6192571[P], (2001).
[9] Y. Zhang, Study on anode effect prediction of aluminum reduction applying wavelet packet transform, International Conference on Intelligent Computing, Springer Berlin Heidelberg, (2010), 477–484.
[10] J. Xing and D. Y. Xiao, Ordered neural network and its application to prediction of anode effect, Control Engineering of China, 14 (2007), 27–36.
[11] J. B. Harley and J. M. F. Moura, Data-driven matched field processing for Lamb wave structural health monitoring, The Journal of the Acoustical Society of America, 135 (2014), 1231–1244.
[12] V. Y. Bazhin, A. A. Vlasov and A. V. Lupenkov, Controlling the AE in an aluminum reduction cell, Metallurgist, 55 (2011), 463–468.
[13] Y. Song, J. P. Peng, Y. W. Wang, Y. Z. Di, B. K. Li and N. X. Feng, Magneto-hydrodynamics simulation of 300 KA novel cell for aluminum electrolysis, Metalurgiya, 55 (2016), 22–24.
[14] J. Yi, D. Huang, S. Fu, H. He and T. Li, Multi-objective bacterial foraging optimization algorithm based on parallel cell entropy for aluminum electrolysis production process, *IEEE Transactions on Industrial Electronics*, 63 (2016), 2488–2500.

[15] H. Viumdal and S. Mylvaganam, System identification of a non-uniformly sampled multirate system in aluminum electrolysis cells, *Modeling Identification and Control*, 35 (2014), 127–146.

[16] A. Solheim, Entropic heat effects in aluminum electrolysis cells with inert anodes, *Metallurgical and Materials Transactions - B*, 47 (2016), 1274–1279.

[17] F. Allard, G. Soucy and L. Rivoal, Formation of deposits on the cathode surface of aluminum electrolysis cells, *Modeling Identification and Control*, 35 (2014), 127–146.

[18] A. Solheim, Entropic heat effects in aluminum electrolysis cells with inert anodes, *Metallurgical and Materials Transactions - B*, 47 (2016), 1274–1279.

[19] J. J. Li, Z. J. Wang and J. L. Zhu, Aluminum electrolysis multi-objective control system based on quantum optimized, *Advanced Materials, Technology and Application: Proceedings of the 2016 International Conference on Advanced Materials, Technology and Application (AMTA2016). World Scientific*, (2016), 417–423.

[20] H. Zhang, T. Li, J. Li, S. Yang and Z. Zou, Progress in aluminum electrolysis control and future direction for smart aluminum electrolysis plant, *JOM*, 69 (2017), 292–300.

[21] C. K. Hu, F. B. Liu and C. F. Hu, Efficiency measures in fuzzy data envelopment analysis with common weights, *Journal of Industrial and Management Optimization*, 13 (2017), 237–249.

[22] H. Z. Haghighi, S. Adeli and F. H. Lotfi, Revenue congestion: An application of data envelopment analysis, *Journal of Industrial and Management Optimization*, 12 (2016), 1311–1322.

[23] A. Klos, J. Bogusz, M. Figurski and W. Kosek, On the handling of outliers in the GNSS time series by means of the noise and probability analysis, *Springer Berlin Heidelberg*, 143 (2015), 657–664.

[24] A. Katayev, J. K. Fleming, D. Luo, A. H. Fisher and T. M. Sharp, Reference intervals data mining, *American Journal of Clinical Pathology*, 143 (2015), 134–142.

[25] Q. X. Chi and X. C. Si, Discussion for radar signal sorting method based on the grubb's' criterion, *Chinese Journal of Sensors and Actuators*, 6 (2006), 2625–2629.

[26] Z. N. Qu and J. L. Xie, Long-term periodicity variations of the solar radius, *Astrophysical Journal*, 762 (2012), 23–28.

[27] F. Gürbüz and P. M. Pardalos, A decision making process application for the slurry production in ceramics via fuzzy cluster and data mining, *Journal of Industrial and Management Optimization*, 8 (2013), 285–297.

[28] M. Kato, H. Masuyama, S. Kasahara and Y. Takahashi, Effect of energy-saving server scheduling on power consumption for large-scale data centers, *Journal of Industrial and Management Optimization*, 12 (2016), 667–685.

[29] Z. Gong, C. Liu and Y. Wang, Optimal control of switched systems with multiple time-delays and a cost on changing control, *Journal of Industrial and Management Optimization*, 14 (2018), 183–198.

[30] F. M. Amour, R. Sethi and Y. K. Lai, Semantic retrieval of trademarks based on conceptual similarity, *IEEE Transactions on Systems Man and Cybernetics Systems*, 46 (2016), 220–233.

[31] Y. Xia, Convex hull of the orthogonal similarity set with applications in quadratic assignment problems, *Journal of Industrial and Management Optimization*, 9 (2013), 689–701.

[32] V. Satuluri and S. Parthasarathy, Bayesian locality sensitive hashing for fast similarity search, *Proceedings of the VLDB Endowment*, 5 (2012), 439–441.

[33] H. Xiao, Similarity Search and Outlier Detection in Time Series, Department of Computer and Information Technique, Ph.D thesis, Fudan University in shanghai, 2005.

[34] L. Zhang, J. Lin and R. Karim, Sliding window-based fault detection from high-dimensional data streams, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47 (2017), 289–303.

[35] R. Faragher, Understanding the basis of the Kalman filter via a simple and intuitive derivation, *IEEE Signal processing magazine*, 29 (2012), 128–132.

[36] T. Schuhmann, W. Hofmann and R. Werner, Improving operational performance of active magnetic bearings using Kalman filter and state feedback control, *IEEE Transactions on Industrial Electronics*, 59 (2012), 821–829.

[37] V. F. De, A. Brandl, M. Battipede and P. Gili, Joseph covariance formula adaptation to square-root sigma-point Kalman filters, *Nonlinear Dynamics*, 88 (2017), 1969–1986.

[38] B. Jin, M. Xin and Y. Cheng, High-degree cubature Kalman filter, *Automatica*, 49 (2013), 510–518.
[38] J. Shawash and D. R. Selviah, Real-time nonlinear parameter estimation using the Levenberg-Marquardt algorithm on field programmable gate arrays, IEEE Transactions on Industrial Electronics, 60 (2013), 170–176.

[39] V. López, S. delRío, J. M. Benítez and F. Herrera, Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data, Fuzzy Sets and Systems, 258 (2015), 5–38.

[40] C. C. Jiang, R. F. Zhu, G. Y. Xiao, L. L. Wang, Y. Z. Zheng and Y. P. Lu, Communication-effect of nano-alumina concentration on the microstructure and corrosion resistance of phosphate chemical conversion coating, Journal of The Electrochemical Society, 163 (2016), C339–C341.

[41] S. Zhang, X. Chen and Y. Yin, An ELM based online soft sensing approach for alumina concentration detection, Mathematical Problems in Engineering, 2015 (2015), Article ID 268132, 8 pages.

[42] G. Bearne, M. Dupuis and G. Tarcy, Pseudo resistance curves for aluminium cell control - alumina dissolution and cell dynamics, in Essential Readings in Light Metals: Aluminium Reduction Technology, Volume 2 (eds. H. Kvande, B. P. Moxnes, J. Skaar and P. A. Solli), Metals and Alloys, (2013), 760–766.

[43] Q. Zhai, J. Yang, M. Xie and Y. Zhao, Generalized moment-independent importance measures based on Minkowski distance, European Journal of Operational Research, 239 (2014), 449–455.

[44] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ò. Belmonte and J. Huerta, Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems, Expert Systems with Applications, 42 (2015), 9263–9278.

[45] G. H. B. Foo, X. Zhang and D. M. Vilathgamuwa, A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended Kalman filter, IEEE Transactions on Industrial Electronics, 60 (2013), 3485–3495.

Received October 2017; revised December 2017.

E-mail address: zuguochen@csu.edu.cn
E-mail address: liyonggang@csu.edu.cn
E-mail address: xiaofangchen@csu.edu.cn
E-mail address: ychh@csu.edu.cn
E-mail address: gwh@csu.edu.cn