• **LETTER** •

A hybrid prediction model with a selectively updating strategy for iron removal process in zinc hydrometallurgy

Ning CHEN, Jiayang DAI*, Weihua GUI, Yuqian GUO & Jiaqi ZHOU

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

Dear editor,

In the zinc hydrometallurgy process, iron removal using goethite is an important step [1]. The goethite process involves several reactors that are arranged from high to low. A zinc sulfate solution, which is rich in ferrous ions, overflows from the first reactor and flows into the next one. Oxygen and zinc oxide are added to each reactor, respectively, to remove ferrous ions into goethite. The ferrous ion concentration at the goethite process outlet should be controlled within the desired range. This process aims to remove ferrous and iron ions from the zinc sulfate solution to ensure a high-quality zinc ingot product.

The goethite process comprises three coupled reactions of oxidation, hydrolysis, and neutralization. The ferrous ions are oxidized to ferric ions by oxygen, and the ferric ions are hydrolyzed to form the goethite precipitate in a zinc sulfate solution. To maintain the pH during this process, zinc oxide is added to neutralize the hydrogen ions. The reaction conditions must be strictly controlled. If ferrous ions are oxidized and precipitated too quickly or too slowly, both the iron removal rate and the goethite precipitate quality will be poor. Therefore, establishing a model for the goethite process is important. A mechanism model based on the mass balance and reaction kinetics in a continuous stirred-tank reactor has been established for the goethite process [2]. However, this model does not

consider the dynamics of dissolved oxygen, which influences the accuracy of the model. The mass transfer process of dissolved oxygen is complex, which makes its dynamics difficult to characterize. In addition to mechanism modeling, data-based modeling can also be used for the goethite process. Data-based modeling does not require knowledge of the mechanism of the goethite process and is applicable to a broad class of industrial processes that are difficult to model. However, it relies heavily on the quantity and quality of samples, which are difficult to determine in the goethite process. Herein, to achieve the expected model performance, a hybrid model that combines a mechanism model with a data-based model is proposed. Additionally, the model parameters are identified using the state transition algorithm (STA) [3].

Furthermore, to save computation time and ensure the prediction accuracy, a selectively on-line updating strategy is used to determine whether the parameters should be updated when a new sample is available. Herein, the hybrid model with a selectively updating strategy is verified by simulation using actual industrial data.

Hybrid model. We consider a goethite process, as shown in Figure S1(a). The goethite process involves five reactors. The main reactants of the process include Fe^{2+}, Fe^{3+}, H^+, and dissolved oxygen. On the basis of the mass conservation and reaction kinetics, the mechanism model of the process

* Corresponding author (email: daijiayang@csu.edu.cn)

is established as follows:

$$
\begin{cases}
\dfrac{dc_{Fe^{2+}}}{dt} = \dfrac{F}{V}(c_{Fe^{2+},in} - c_{Fe^{2+}}) \\
\qquad - k_1 c_{Fe^{2+}}^{\alpha} c_{H^+}^{\beta} c_{O_2}^{\gamma}, \\
\dfrac{dc_{Fe^{3+}}}{dt} = \dfrac{F}{V}(c_{Fe^{3+},in} - c_{Fe^{3+}}) \\
\qquad + k_1 c_{Fe^{2+}}^{\alpha} c_{H^+}^{\beta} c_{O_2}^{\gamma} - k_2 c_{Fe^{3+}}, \\
\dfrac{dc_{H^+}}{dt} = \dfrac{F}{V}(c_{H^+,in} - c_{H^+}) - k_1 c_{Fe^{2+}}^{\alpha} c_{H^+}^{\beta} c_{O_2}^{\gamma} \\
\qquad + k_2 c_{Fe^{3+}} - k_3 c_{H^+} m_{ZnO},
\end{cases}
\tag{1}
$$

where F is the flow rate, V is the reactor volume, k_1, k_2, and k_3 are the reaction rate constants, and α, β, and γ are the reaction orders. Parameters $c_{Fe^{2+},in}$, $c_{Fe^{3+},in}$, and $c_{H^+,in}$ are the reactor inlet concentrations of Fe^{2+}, Fe^{3+}, and H^+, respectively. Parameters $c_{Fe^{2+}}$, $c_{Fe^{3+}}$, and c_{H^+} represent the outlet concentrations of Fe^{2+}, Fe^{3+}, and H^+, respectively. Parameter m_{ZnO} represents the mass of zinc oxide, and c_{O_2} is the dissolved oxygen concentration. Parameter c_{O_2} is an important variable that affects the oxidation rate of the ferrous ions. We selected the dissolved oxygen concentration as a new state variable. According to the gas-liquid mass transfer principle, we established the fourth state equation

$$
\frac{dc_{O_2}}{dt} = K_{la} \left(\int \frac{\rho_{O_2} u_{O_2}}{M_{O_2} V} dt - 2c_{O_2} \right.
$$
$$
\left. - \frac{1}{4}(c_{Fe^{2+},in} - c_{Fe^{2+}}) \right),
\tag{2}
$$

where u_{O_2} is the oxygen flow rate, ρ_{O_2} is the oxygen density, M_{O_2} is the molar mass of oxygen, and K_{la} is the mass transfer coefficient of oxygen. The coefficient K_{la} is the output of an unmodeled dynamic to be modeled by the data-based method.

Herein, a hybrid model is established for the goethite process. This hybrid model comprises a mechanism model, data-based model for K_{la}, parameter identification module, and model updating strategy, as shown in Figure S1(b). The mechanism model is characterized by (1) and (2). The data-based model of K_{la} is established by a method that combines the double locally weighted kernel principal component analysis (DLWKPCA) [4] with the least-squares support vector regression (LSSVR) [5]. We name this method DLWKPCA-LSSVR. The STA [3] is used to optimize the hybrid model parameters, and the methods of just-in-time learning (JITL) [6] and approximately linear dependence (ALD) [7] are used to update the parameters in the data-based model.

Methodology. We first describe our construction of a model for the unmodeled dynamics. According to the mass transfer theory of oxygen in an electrolyte solution, the coefficient K_{la} is mainly affected by the goethite precipitate, the concentration of metal ions, and the solution flow rate. The relation between K_{la} and related variables can then be expressed as

$$
K_{la} = f(c_{Fe^{2+}}, c_{Fe^{3+}}, c_{Cu^{2+}}, c_{Zn^{2+}},
$$
$$
F, n_{FeOOH}, m_{ZnO}),
\tag{3}
$$

where f is the function to be determined, and $c_{Cu^{2+}}$ and $c_{Zn^{2+}}$ are the Cu^{2+} and Zn^{2+} concentrations, respectively. n_{FeOOH} represents the mass of the goethite precipitate. The DLWKPCA-LSSVR method is used to establish the model of K_{la} as

$$
K_{la} = \sum_{i=1}^{N} \theta_i K\left(G_i^{Pw,K}, g\right) + b_N,
\tag{4}
$$

where N is the number of local modeling samples, K is a kernel function, and $G_i^{Pw,K}$ and g are the projection of the local modeling samples and the query sample, respectively, which are extracted by DLWKPCA. Parameters θ_i and b_N can be trained by LSSVR. The hybrid model of the goethite process combining (1), (2), and (4) is thus obtained. Other parameters that need to be identified include k_1, k_2, k_3, α, β, and γ in the mechanism model and distance parameter σ for DLWKPCA, kernel width parameters δ_1 and δ_2, and the penalty factor c in LSSVR. The vector of the parameter is defined as $\xi = [k_1, k_2, k_3, \alpha, \beta, \gamma, \sigma, \delta_1, \delta_2, c]$. The hybrid model output is defined as $\hat{y} = c_{Fe^{2+}}$. Obviously, the issue of parameter identification can be considered as an optimization problem $\min J(\xi) = (\sum_{i=1}^{N_{train}}(y_i - \hat{y}_i)^2)/N_{train}$, where y_i is the actual sample value of the Fe^{2+} concentration and N_{train} is the number of training samples.

After the parameter identification, the hybrid model requires selection of the number of modeling samples and construction of a local regression model when a query sample is available. To accommodate the tradeoff between modeling accuracy and efficiency, we propose an update method based on the ALD [7]. The model updating strategy uses the ALD condition to determine when to update the model for a new sample. The specific update conditions are

$$
\begin{cases}
\zeta_q = \min \left\| \sum_{i=1}^{N} a_i x_i - x_q \right\|^2, \\
\zeta_q \leqslant u, \quad \text{do not update the model}, \\
\zeta_q \geqslant u, \quad \text{update the model},
\end{cases}
\tag{5}
$$

where ζ_q is the ALD index, x_q represents a new sample, $\{x_i\}_{i=1}^{N}$ is the set of local modeling samples, a_i is a coefficient, and $\|\sum_{i=1}^{N} a_i x_i - x_q\|^2$ is obtained by optimization.

Simulations and results. We here consider the first reactor as an example. To evaluate the prediction performance of the proposed modeling method, the root-mean-squared error (RMSE), mean absolute error (MAE), and mean relative error (MRE) are used as performance indices. The number of local modeling samples N and principal components d are first determined by cross-validation. Resultantly, N and d are selected as 10 and 3, respectively. Then, the hybrid model parameters are optimized by the STA algorithm.

In the simulations, 50 samples were tested. For performance comparison, a single mechanism model (K_{la} was acquired by STA) and a mechanism-LWKPCR model (K_{la} was predicted by LWKPCR [8]) were used to predict the output. From the simulation results, the RMSE, MAE, and MRE of the proposed hybrid model were 0.2315, 0.1745, and 0.0219, respectively. In comparison with the single mechanism model, these indices were 0.4747, 0.3829, and 0.0458, respectively, for the single mechanism model. The results of the mechanism-LWKPCR model were 0.3816, 0.3259, and 0.0406, respectively. These results demonstrate the effectiveness of the proposed hybrid modeling method.

In addition, the effect of the ALD threshold on modeling accuracy and real-time performance was also studied. Simulations were conducted for different values of the ALD threshold in the range from 0.1 to 0.2. The results show that larger ALD thresholds tended to result in a larger predicted RMSE and that the running time decreased with decreasing ALD threshold. To achieve a better tradeoff between the predictive performance and the modeling efficiency, an appropriate ALD threshold was selected using trial-and-error in the simulations. In fact, when the ALD threshold was increased from 0.17 to 0.185, the predicted RMSE did not change significantly but the calculation time was substantially reduced. When the ALD threshold increased from 0.185 to 0.2, the predicted performance decreased, whereas the running time was not substantially reduced. Therefore, the appropriate value for the ALD threshold was determined as 0.185, which can effectively reduce the modeling time and ensure adequate predictive accuracy.

Conclusion. This study proposes a hybrid modeling method with a mechanism modeling and a DLWKPCA-LSSVR modeling method based on ALD-JITL for the goethite process. In this modeling method, a mechanism model of the goethite process is established by mass conservation and reaction kinetics. Because the mass transfer coefficient of the oxygen in the mechanism model is affected by various factors, accurately obtaining this parameter through mechanism analysis is difficult. The coefficient was modeled by the proposed DLWKPCA-LSSVR on the basis of the process data. By combining the mechanism model and the DLWKPCA-LSSVR model, we obtained a hybrid model of the goethite process. In addition, an update strategy for hybrid modeling was developed using an ALD index. The prediction time was effectively reduced and the prediction accuracy was maintained by setting a reasonable ALD threshold. The effectiveness of this hybrid modeling method was verified against industrial data.

Supporting information Figure S1. The supporting information is available online at info.scichina.com and link. springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Zhou X, Zhou J J, Yang C H, et al. Set-point tracking and multi-objective optimization-based pid control for the goethite process. IEEE Access, 2018, 6: 36683–36698

2 Xie Y F, Xie S W, Chen X F, et al. An integrated predictive model with an on-line updating strategy for iron precipitation in zinc hydrometallurgy. Hydrometallurgy, 2015, 151: 62–72

3 Zhou X J, Yang C H, Gui W H. State transition algorithm. J Ind Manage Optim, 2012, 8: 1039–1056

4 Chen N, Dai J Y, Yuan X F, et al. Temperature prediction model for roller kiln by ALD-based double locally weighted kernel principal component regression. IEEE Trans Instrum Meas, 2018, 67: 2001–2010

5 Chan C L, Chen C L, Ting H W, et al. An agile mortality prediction model: hybrid logarithm least-squares support vector regression with cautious random particle swarm optimization. Int J Comput Intell Syst, 2018, 11: 873–881

6 Yuan X F, Ge Z, Huang B, et al. A probabilistic just-in-time learning framework for soft sensor development with missing data. IEEE Trans Control Syst Technol, 2017, 25: 1124–1132

7 Tang J, Yu W, Chai T Y, et al. On-line principal component analysis with application to process modeling. Neurocomputing, 2012, 82: 167–178

8 Yuan X F, Ge Z, Song Z. Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes. Ind Eng Chem Res, 2014, 53: 13736–13749