A Data-Based Optimal Setting Method for the Coking Flue Gas Denitration Process

Yaning Li,1,2 Xuelei Wang,1* Zhenjie Liu,1 Xiwei Bai1,2 and Jie Tan1
1. Institute of Automation, Chinese Academy of Science, Beijing 100190, China
2. University of the Chinese Academy of Sciences, Beijing 100049, China

This study focuses on developing an optimal setting method for the first integrated coking flue gas desulphurization and denitration device in China. Maintaining the denitration process in a state of optimal economic efficiency has become an issue in production optimization control. This paper proposes a data-based two-stage nonparametric optimization method to optimize the operation of the denitration process. A principal component regression (PCR)-based multiple case fusion case-based reasoning (CBR) method is proposed to obtain the initial optimization set points. To overcome the steady-state modelling difficulties associated with the process, a local modelling method for the coking flue gas denitration process is developed using an improved just-in-time learning (JITL) algorithm. Taking the preset values obtained above as the initial value of an active set algorithm, the optimization problem can be solved in a timely and precise manner. The intelligent setting software has been developed for running industrial applications, and the results demonstrate the effectiveness of the proposed optimization approach.

Keywords: coking flue gas denitration, data-based, optimal setting, CBR, JITL

INTRODUCTION

China is a major coke producer in the world, and the new emission standard for coking flue gas pollutants has led to unprecedented pressure for domestic coking enterprises. [1] In view of this increased pressure, Coking & Chemical Co. Ltd. in the Jiangxi province has taken the lead in building and operating an integrated coking flue gas desulphurization and denitration device, which have been proven to exhibit good results after the actual operation, demonstrating that flue gas emissions can meet the national standard while ignoring the economic index. [2]

Considering that the power consumption of the ozone generator in the denitration process accounts for more than 80 % of the total power consumption, our work focuses on optimizing the flue gas denitration process. The outlet flue gas NOx concentration must be controlled within a desired range while also minimizing the energy consumption. Both of these indices are directly related to the set points, including the urea addition and the ozone output of the ozone generator. The denitration process is typically controlled by a distributed control system (DCS) with human supervision in industrial fields. Therefore, the set points of the control loop are determined by engineers and operators based largely on their experience, making it difficult to achieve satisfactory results.

The difficulties arising from attempting to minimize the energy consumption and achieving satisfactory NOx removal efficiency stem from two considerations: (1) In the coking flue gas denitration process, one-off offline optimization (constant set points) cannot meet the needs of the denitration process for two reasons. First, the set points should be adjusted in response to changes in working conditions, such as the concentration or volume of raw flue gas. Second, the persistent random disturbances in the practical operation of the unit can cause the system operating point to drift, preventing the current optimized point from maintaining its optimality permanently. (2) Characteristics of the denitration process and upstream coking production are highly complex. [3–7] The denitration process involves nonlinearities, and complex mass transfer and heat transfer processes, as well as numerous physical and chemical reactions, and no relevant mechanism research has been presented in the literature. The reversing process is a special production operation for coking production; this process hinders the current working condition information from being obtained. Therefore, frequently used model-based optimal controls are difficult to design and implement.

Since the basic concept of real-time optimization (RTO) was proposed in the 1970s, [8,9] there have been three main adaptive optimization approaches that differ in the way in which adaptation is performed, namely, model-parameter adaptation, modifier adaptation, and direct input adaptation. [10] However, process data must satisfy the continuous incentive conditions in model-parameter adaptation, which is difficult to realize in the actual process. Modifier adaptation aims to match the Kuhn-Tucker conditions of the model and the actual process, and is too idealistic to be applied in complex industrial processes. Direct input adaptation is the least practical because of its complexity. In addition to these practicability problems, RTO methods are all based on an accurate mechanism model of the process, which is not available for the complex industrial processes that are difficult to model. Recently, economic model predictive control (EMPC) has received considerable attention in both academia and industry. [11–14] EMPC is advantageous in that the real-time control process and economic performance optimization are integrated in an optimal control framework. However, this “economic” performance optimization does not refer in particular to an economic performance index; instead, it refers to a class of non-positive or non-convex arbitrary performance functions. [15] In addition, the optimization problem stated in EMPC is a fully dynamic

* Author to whom correspondence may be addressed.
E-mail address: xlwang98@sina.com
© 2018 Canadian Society for Chemical Engineering
DOI 10.1002/cjce.23226
Published online 17 April 2018 in Wiley Online Library (wileyonlinelibrary.com).
optimal control problem, which means that the objective function and the process model are time-related aspects. Therefore, the above optimization methods cannot be used directly in the process under consideration. A novel RTO setting method is urgently needed for the coking flue gas denitration process.

Recently, with the improved basic automation level of factories, copious production data are stored in the production database. Making effective use of these industrial data has become a focus of research. Yacoub and Macgregor\[16\] adopted nonlinear data-based PLS to model the production process, and the optimization was entirely performed based on the data model. Macgregor et al.\[17,18\] presented the advantages of latent variable methods for extracting and using information from big data in the process industries, particularly for process analysis, monitoring, optimization, and control. Zhao et al., Chai et al., Li et al., and Zhou et al.,\[19–23\] proposed a hybrid intelligent optimal setting control system framework and applied it to the metallurgical process. In general, hybrid intelligent optimal control combines one or more advanced intelligent technologies, such as expert system, neural network, evolutionary computation, and fuzzy reasoning, as well as traditional optimization and control techniques to solve optimal control problems. However, the stability of this hybrid system should be further investigated. Peng et al., Gui et al., Wu et al., and Yang et al.,\[24–27\] proposed a data-driven operational-pattern optimization method that can extract the operation pattern from the data and set the parameters according to the operation pattern of the current production conditions. The method has been successfully used in the copper flash smelting process, the cobalt removal process, and the metallurgical industry. This method is essentially an improvement and promotion of CBR. Li et al.,\[28\] proposed a Q-learning-based suboptimal setting algorithm by utilizing only data; this is the most typical illustration of data-driven optimization. However, the author provided only theoretical solutions for general industrial problems, which are not applicable to actual industrial processes.

Inspired by the successful data-driven optimal control techniques applied in previous research, in this paper, based on the specific industrial background of the coking flue gas denitration process, we present a data-based two-stage operation optimal method for the operational optimization of the coking flue gas denitration process. The pre-setting values are solved by an improved CBR technology and taken as the initial value of the optimization calculation after the operation of the feedback adjustor. Then the optimized operating parameters are obtained via an active set algorithm. An improved just-in-time learning (JITL)-based local modelling method is proposed to overcome the difficult modelling problem of the flue gas denitration process. As a result, we extend this approach to solve the operation optimization problem arising in the coking flue gas denitration process. The main contributions and novelty of this paper are summarized as follows. First, we propose a PCR multiple case fusion CBR method, which fuses case solutions at different time scales using PCR to calculate preset values and overcome the shortcoming by which the traditional method may result in a biased solution using a single case description. Second, we introduce time order into the basic JIT algorithm and propose a local polynomial estimation model for the local modelling. The experimental results demonstrate the effectiveness of the proposed algorithm. Further RTO can be performed based on this model. Finally, the proposed strategy is successfully applied to a real-world process, demonstrating the superior performance of the system.

The remainder of the paper is organized as follows. In the Optimization Model of the Coking Flue Gas Denitration Process section, the flue gas desulphurization and denitration process and the current operation method are introduced, and the optimization problem is established. In the Data-Based Optimal Setting Method section, the data-based optimal setting framework is developed, and the algorithm is introduced in detail. An industrial application of the proposed approach is described in the Industrial Application section. Finally, concluding remarks are presented in the Conclusions section.

OPTIMIZATION MODEL OF THE COKING FLUE GAS DENITRATION PROCESS

Process Description

The wet ammonia ultra-intense turbulent technique and two-section forced oxidation with urea technique are adopted for desulphurization and denitration, respectively. The specific process is shown in Figure 1.

Flue gas generated from the coke oven is fed into a heat-recovery boiler through a draught fan, and the temperature is reduced from 300–160 °C. Then, the flue gas is mixed with ozone before entering the desulphurization tower; a portion of the nitric oxide (NO) in the flue gas can react rapidly with ozone, producing nitrogen dioxide (NO2). Then, the flue gas enters the concentration section of the desulphurization tower, where spraying and washing reduce its temperature to 60 °C. Finally, the flue gas enters the absorption section of the desulphurization tower through gas caps and a countercurrent in contact with desulphurization absorption liquid sprayed from the top of the tower, and SO2 in the flue gas reacts with ammonium sulphite in the absorption liquid, converting it into ammonium hydrogen sulphite. As a result, SO2 is removed and purified.

The flue gas pipeline is connected to the ozone input pipe before entering the denitration tower. Ozone is mixed into the flue gas at approximately 60 °C, and a portion of the NO in the flue gas is converted into NO2. In the countercurrent, the denitration absorption liquid is sprayed from the top of the denitration tower after entering the lower part of the tower; NO, NO2, and urea in the liquid participate in a reduction reaction, generating nitrogen, carbon dioxide, and water, and thus, the denitration process is completed. The flue gas reaches the emission standard and is discharged into the atmosphere from the top of the denitration tower.

As described in the Introduction, the purpose of operation optimization is to control the outlet NOx concentration strictly within its desired range \((C_d - C_\Delta, C_d + C_\Delta)\) based on safe running conditions, where \(C_d\) is the desired NOx concentration, and \(C_\Delta\) is the allowed fluctuation range that satisfies \(C_d + C_\Delta = C_{NT}\) (\(C_{NT}\) is the emission standard and typically ranges from 150–500 mg/ Nm\(^3\)). Furthermore, power and urea consumption must be minimized.

The reversing process is a special working condition that exists in coke ovens and occurs once every 30 min. When the coke oven is undergoing a reversing process, the inlet flue gas indices, such as the SO2/NOx concentration and O2 content, change dramatically. Details regarding this process are provided by Yao et al.\[29\] and Li et al.\[30\] In this paper, we focus only on the optimization of the denitration process under non-reversing process.

Current Operation Status

At present, the denitration process depends mainly on the DCS control under human supervision, as shown in Figure 2, where \(U\) (\(U_{r}, \Delta t_{1}, \Delta t_{2}\) denotes the urea solution addition (m\(^3\)/h), duration...
(min), and interval time (h); $D_r$ and $d_r$ are the unknown disturbance in the coking flue gas denitration process and basic control loop caused by measurement error and mechanical wear and tear, respectively.

Operation engineers are always required to achieve a satisfactory effect. The proper set points for the lower-level controllers are determined according to their perceived operating conditions, as shown in Figure 2. The purpose of such human supervision can be summarized as follows:

1) to obtain and process the data, such as the original flue gas concentration and flow rate, from the DCS monitoring system;
2) to provide set points ($F_{ts}$, $F_{tn}$, and $O_3^*$) for the ozone output control loop according to human experience as well as the original flue gas indices from the monitoring system, desired and outlet concentrations, and the basic control loop tracking and device operation status;
3) to reduce the degree of freedom of manipulated variables, the ozone flow $F_{ts}$ and $F_{tn}$ of the two towers and the absorbed liquid circulation amount typically remain unchanged;
4) to determine the set points $U$ of the urea solution control loop according to experience.

Due to the variable upstream coking conditions, large process disturbances, and complex process mechanism, the operating conditions of the coking flue gas denitration process often vary with time. Furthermore, a lack of adequate operational experience also prevents the operator from being 100% effective. Thus, it is difficult to set the set points properly in a timely manner under the operating mode shown in Figure 2, which may cause the outlet flue gas concentration to be exceeded. At the present stage, the ozone unit is often kept in high-power operation to ensure the outlet flue gas indices reach the standard, although this comes with a substantial loss of energy. Therefore, the manual-based control methods have stability, economy, and reliability concerns, and an optimization model for the effective regulation of the operating parameters in the denitration process must be established.

![Figure 1. Desulphurization and denitration integrated device for coking flue gas.](image1)

![Figure 2. DCS-controlled system with human supervision.](image2)
Optimization Problem Establishment

The denitrification process involves the following main reactions:

\[ O_3 + NO \rightarrow NO_2 + O_2 \]  
\[ NO_2 + O_3 \rightarrow NO_3 + O_2 \]  
\[ NO_2 + O_3 \rightarrow N_2O_5 \]  
\[ CO(NH_2)_2 + NO_2 + NO \rightarrow CO_2 + 2N_2 + 2H_2O \]  
\[ 2NH_3 + NO + NO_2 + H_2O \rightarrow 2NH_4NO_2 \]  
\[ 2NH_3 + 2NO_2 + H_2O \rightarrow NH_4NO_2 + NH_4NO_3 \]  

Under normal circumstances, the main operational variables that affect the removal efficiency are the output NOx (kg/h), absorption liquid (desulfurization and denitrification) concentration and absorbed liquid spray quantity. According to the actual situation of the denitrification process, the desulfurization absorption liquid is always adequate for Reactions (5) and (6). In addition, the cost of the urea absorption liquid is negligible compared to the power consumption of the ozone generating unit, thus the set points relevant to the urea absorption liquid are all set via CBR. Therefore, the operating variable that needs to be adjusted is NOx. According to Equation (7), there are two operating variables, namely, the ozone for the desulfurization tower \( (N_1 \text{ kg/h}) \) and for the denitrification tower \( (N_2 \text{ kg/h}) \).

\[ N_{Ox} = N_1 + N_2 = F_{flue} C_{Ox} + F_{In} C_{Ox} \]  

where \( F_{flue} \) and \( F_{In} \) denote the ozone flow (m³/h, measurable) for the desulfurization and denitrification tower, respectively; \( C_{Ox} \) denotes the ozone concentration (kg/m³).

The power consumption cost of the denitrification process can be described as follows:

\[ B_{T0} = \sum_{i=1}^{5} E_i(T_0) W_e \times E_i(T_0) \]  

where \( B \) denotes the power consumption cost (RMB), \( T_0 \) denotes the optimization cycle (h), \( E_i(i = 1, 5) \) is the energy consumption (kw·h) of the ozone generator, screw compressor, freeze dryer, adsorption dryer, and water pump, respectively, and \( W_e \) is the industrial electricity price (RMB/kw·h). Because the entire ozone generator unit is equipped with a SIEMENS automatic control system, the operating powers of the last four devices are all automatically adjusted according to the ozone generator, thus, \( B \) is positively related to the energy consumption of the ozone generator during \( T_0 \).

\( E_i \) can be expressed as follows, according to Equation (7):

\[ E_i = N_{Ox} \sigma = (N_1 + N_2) \sigma \]

where \( \sigma \) is the power factor (kw·h/kgO₃) acquired from the technical parameter table of the ozone generator, which is 16 in this paper.

The power cost to be minimized during each optimization cycle can now be substituted into the following index:

\[ \min_{N_1, N_2} = N_1 + N_2 \]

The equality condition is the model of the denitrification process, which can be expressed as follows:

\[ C_{out} = C_{in}, (1 - \eta_1^N) (1 - \eta_2^N) = f_N(T_{in}, C_{O2}, V_{in}, C_{den}, C_{in}, N_1, N_2) \]

The variables in the above equations are described as follows:

\( C_{out} \) is the NOx concentration (mg/m³, measurable) of the outlet flue gas. \( \eta_1^N \) and \( \eta_2^N \) are the denitration efficiencies (%) in the desulfurization tower and denitrification tower, respectively. \( f_N \) is the unknown nonlinear function of the process. \( T_{in} \) is the temperature (°C, measurable) of the inlet flue gas. \( C_{O2} \) and \( V_{in} \) are the O₂ content (%), measurable and velocity (m/s, measurable), respectively, of the inlet flue gas. \( C_{den} \) is the urea absorption solution concentration (unmeasurable). Finally, \( C_{in} \) is the NOx concentration (mg/m³, measurable) of the inlet flue gas.

The inequality conditions include operation index requirements and equipment limitations as follows:

\[ C_d - C_{\Delta} \leq C_{out} \leq C_d + C_{\Delta} \]
\[ 0 \leq N_1 + N_2 \leq N_{Orm} \]
\[ 0 \leq N_1, N_2 \]

where \( C_d \) and \( C_{\Delta} \) are determined by the technical workers, and \( N_{Orm} \) (kg/h) is the maximum output of the ozone generator, which is 50 in this paper.

In summary, the optimization problem of the coking flue gas denitrification process can be described as follows:

Based on the steady-state model of the flue gas denitrification process (Equation (11)), a pair of optimal solutions of the decision variables \( N_1 \) and \( N_2 \) that minimize the power consumption cost (Equation (10)) must be determined, and constraint Equations (12–14) need to be satisfied.

As a typical nonlinear optimization problem including complex equality and inequality constraints, the steady-state model of the process (Equation (11)) cannot be based on the process mechanism. Even if the model can be established by data-driven approaches, such as neural network or support vector machine, the real-time demand cannot be satisfied based on the above large-scale static model; thus, the problem of optimizing the coking flue gas denitrification process cannot be solved easily using traditional optimization methods.

**DATA-BASED OPTIMAL SETTING METHOD**

Framework of the Data-based Optimal Setting Method

A framework for the optimal setting system was established based on the lower-level control layer, as shown in Figure 3. The framework consists of a CBR-based pre-setting controller, a time-order JITL (T-JITL)-based real-time optimization module, and a feedback adjustor. This optimal setting system is designed to automatically adjust the set points for the basic control layer under the variations of working conditions. The major components of the system are as follows:

- CBR-based pre-setting controller: this module provides the initial optimization value according to the description of the coking condition at different time scales.
• Feedback adjustor: the feedback adjustor responds to the error \( e = C_{out} - C_{lim} \) to obtain the compensating increments \( \Delta \Theta_0 \) for the initial set points, where \( C_{lim} \) is the NOx output boundary.

• T-JITL-based real time optimization module: this module is applied to solve the optimization problem stated previously.

In Figure 3, \( \Omega_1 \) and \( \Omega_H \) denote the input conditions used for T-JITL-based local modelling and CBR-based pre-setting, respectively. \( \Theta^* = [N_1, N_2, U_r, \Delta t] \) is the optimal setting value obtained by CBR, \( \Theta_0 = [N_1, N_2] \) is a subset of \( \Theta^* \). \( \Delta \Theta_0 \) is the adjusted increments of \( \Theta_0 \) generated by the feedback adjustor, \( \Theta_{ok} = \Theta_0 + \Delta \Theta_0 \) is adopted as the initial value of optimization, and \( \Theta^{opt} = [N^*_1, N^*_2] \) is the final optimized set point of the current optimization cycle.

In each running cycle, the CBR-based pre-setting controller will check the most similar case from the case database based on the characteristic variables of the current coking condition. The set points related to the absorption solution are directly provided for the lower-level controllers. The set points related to ozone (\( \Theta_0 \)) are regarded as the suboptimal values (initial values) and adopted to drive the process at this operating point to facilitate the following further optimization. Once the actual value \( C_{out} \) is outside the desired range, the feedback adjustor provides compensating increments \( \Delta \Theta_0 \) in response to the error. Then, the T-JITL-based real-time optimization module can perform the calculation of the optimization problem at the initial value of \( \Theta_{ok} \).

The innovative combination of the two-stage operation optimization method combined with a feedback adjuster poses a data-driven and real-time optimal setting methodology for basic controllers of the denitrification process, which is difficult for traditional optimal controls to achieve.

CBR-Based Pre-Setting Controller

Because actual industrial processes exhibit compounded dynamics and complexity, it is difficult to establish and adapt the process model. However, experienced operators can provide desired quality settings to some extent. Therefore, CBR was developed to provide a good remedy to this problem. CBR can easily facilitate the combination of expert experience or knowledge and enhance self-reasoning and decision-making, hence allowing the performance of the concerned systems to improve gradually. In this context, the purpose of CBR is to employ existing similar cases to solve current problems. In addition, considering the abrupt change in certain correlation description indices caused by the coke oven reversing process, deviation from the results may occur because a single feature is used to describe the current working conditions in the traditional case reuse method.[32, 33] Therefore, a case retrieval and reuse method is further proposed from PCR multiple case fusion, as shown in Figure 4, which includes four components: case generation, case retrieval and matching, case reuse and fusion, and case revision and retention.[33]

Case generation

This unit generates the initial cases by examining industrial process operations. The case representation can be described as follows:

\[
M_k : \left\{ T_k, (x_{k,1}, x_{k,2}, x_{k,3}, x_{k,4}, x_{k,5}) \right\} \rightarrow (s_{k,1}, s_{k,2}, s_{k,3}, s_{k,4}, s_{k,5}) \]

\[
(15)
\]

where \( k = 1, \ldots, n \), \( n \) is the number of cases, \( T \) denotes the time when a case is created, \( X \) denotes the case descriptor, which is expressed in Equation (16) based on an understanding of the denitrification process, and \( S \) is the case solution, as shown in Equation (17).

\[
X = (x_{k,1}, x_{k,2}, x_{k,3}, x_{k,4}, x_{k,5}) = (C_d, C_m, T_m, C_{O2}, V_m)
\]

\[
(16)
\]

\[
S = (s_{k,1}, s_{k,2}, s_{k,3}, s_{k,4}, s_{k,5}) = (N^*_1, N^*_2, U^*_r, \Delta t^*_1, \Delta t^*_2)
\]

\[
(17)
\]

The meaning of each variable is provided above. All variables are numeric.
Case retrieval and matching

Suppose that the case under the current operation condition is \( M \). The similarity value between \( M \) and the stored case \( M_k \) in the case database is given by:

\[
F(M, M_k) = \frac{\sum_{i=1}^{5} \omega_i \text{sim}(x_i, x_{k,i})}{\sum_{i=1}^{5} \omega_i} = 1 - \frac{|x_{k,i} - x_i|}{\max(x_{k,i})}
\]

where \( \omega_i \) represents the weight of each characteristic, and is typically determined by experience; \( \text{sim}(x_i, x_{k,i}) \) is the similarity between the characteristic \( x_{k,i} \) in the stored case and the characteristic description \( x_i \) of the current operating condition.

The \( F(M, M_k) \) values are ranked in descending order, and the cases that satisfy \( F(M, M_k) \geq F_{th} \) are retrieved as the matching cases, where \( F_{th} \) is the threshold and is initially determined by expert experience (e.g., \( F_{th} = 0.85 \)) and then determined by the following criteria:

\[
F_{th} = \min\left(F_{th}', \max_{k=1 \ldots n} F(M, M_k)\right)
\]

Case reuse

If \( r \) matching cases \( \{M_1, M_2, \ldots, M_r\} \) are retrieved from the case database, the similarities between these cases and the current coking condition are \( \{F_1, F_2, \ldots, F_r\} \) \( F_1 \leq F_2 \leq \ldots \leq F_r \), and the solutions of each case are \( \{S_1, S_2, \ldots, S_r\} \). The final case solution \( S' \) is obtained by:

\[
S' = \sum_{k=1}^{r} F(M, M_k)S_k / \sum_{k=1}^{r} F(M, M_k)
\]

Equations (18–20) indicate that the final case solution \( S' \) depends only on the current case descriptor \( x \) and the coefficients, such as weight \( \omega \) and \( F_{th} \). However, Figure 5 shows that the flue gas concentrations fluctuate abruptly due to the coke oven reversing operation, which means that a single randomly extracted case description may not be sufficient to reflect the current working conditions. Even if the coefficients are appropriately determined, the solution obtained by a single case description is often biased from the ideal value. Thus, a case reuse method based on PCR multi-case fusion is proposed to address the shortcomings of the above traditional method.

Considering the different time scales, multigroup case description data under the non-reversing process are collected at an equal interval \( \Delta t_c \). The sub-cases under different time scales are defined as \( M_1 \sim M_m \), where \( m \in \mathbb{N} \) is the number of sub-cases, and the case description of \( M_m \) is \( X_{m,i} = \{x_{m_1,i}, x_{m_2,i}, x_{m_3,i}, x_{m_4,i}, x_{m_5,i}\} \). For each case description \( X_{m,i} \), \( r \) matching cases \( M_j(j = 1, 2 \ldots r) \) can be retrieved. The size of \( r \) can be determined by the first case description \( X_{m_1,i} \) and the threshold \( F_{th} \), and the solutions that correspond to \( M_j(j = 1, 2 \ldots r) \) for each case description \( X_{m,i} \) are \( \{S_{m_1}, S_{m_2}, \ldots, S_{m_r}\} \). The final solution of the proposed method is the weighted connection of the solutions \( \{S_{m_1}, S_{m_2}, \ldots, S_{m_r}\} \), which are the solutions of the matching cases \( M_j(j = 1, 2 \ldots r) \) retrieved by the last case description \( X_{m,i} \). The final solution can be represented as follows:

\[
S_f = S'_1W^T = w_1S_{m_1} + w_2S_{m_2} + \ldots + w_rS_{m_r}
\]

where \( S_f \) is the final solution of the \( \Theta' \), \( S'_1 \) is the vector of the solutions \( \{S_{m_1}, S_{m_2}, \ldots, S_{m_r}\} \), \( W^T \) is the connection weight vector and \( W^T(r \times 1) = [w_1, w_2, \ldots, w_r], \) and \( w_r \) is the connection weight of the \( r \)th case solution and final solution \( S_f \).

Figure 5. Fluctuation in coking flue gas indices.
The key to this method is to determine the connection weight $W$, as described below.

For $m$ case descriptions $X^0_m \sim X^0_0$, $m$ case solutions that corresponding to $r$ matching cases can be obtained by similarity calculations. Taking $X^0_1$ as an example, we can obtain $S_{11}, S_{12}, \ldots, S_{1r}$, composing $H^{(m,r)}$, that is:

$$H = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1r} \\ S_{21} & S_{22} & \cdots & S_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mr} \end{bmatrix}.$$  

Vector $S_{mca}$ is composed of case solutions $S_m$, which are the final solution of each case description $X^0_m$. $W^T$ can be solved by the least square method, i.e.,

$$W^T = H^{T}S_{max} \text{ where } H = (H^T H)^{-1}H^T.$$  

To eliminate the correlation between different case solutions, $H$ can be expressed as the following equation using the PCR method:

$$H = T^{(m,r)}P^{(r,c)} = [t_1^T \ t_2^T \ \cdots \ t_r^T] [p_1^T \ p_2^T \ \cdots \ p_r^T]$$  

where $t_i^{(1:r)} = z_i/\sqrt{\sigma_i}$ is the ratio of the normalized data to the corresponding eigenvalues of the $i$th principal component, and $p_i^{(1:r)}$ is the load component of the $i$th principal component and is also a unit orthogonal vector, $i = 1, 2, \ldots, r$.

By setting a principal component contribution rate $\eta$, the principal component analysis is used to calculate the first $r$ items that can reflect the main features of the object:

$$H \approx H_c = TP^T$$  

where $H_c$ is the matrix $H$ after principal component analysis, $T^{(m,r)} = [t_1^T \ t_2^T \ \cdots \ t_r^T]$, and $P^{(r,c)} = [p_1^T \ p_2^T \ \cdots \ p_r^T]$.

The following equation can be obtained by Equation (21):

$$S_{max} = HW^T = TP^TW_c$$  

Let $W_{c}^{r+1} = P^TW_c$. Then, $W_c$ can be calculated by the least squares method:

$$W_{c}^{r+1} = (T^T T)^{-1}T^T S_{max}$$  

Thus, the connection weight $W_c$ can be obtained by

$$W_c = (PP^T)^{-1}PW_{c}^{r+1} = (PP^T)^{-1}P(T^T T)^{-1}T^T S_{max}$$  

The weights calculated by PCR are converted into the weights of each original component as follows:

$$W = W_{c}^{T}P/\sqrt{\sigma_i} = W_{c}^{T} [p_1^T \ p_2^T \ \cdots \ p_r^T]^T/\sqrt{\sigma_i}$$  

The final output of this step is:

$$S_l = \sum_{k=1}^{r} w_k S_{mk}$$  

Case revision and retention

After obtaining the solution for the pre-setting points $\Theta^*$ of the denitration process, the formerly provided $\Theta_O$ is considered reasonable, and case revision is not needed if the outlet concentration is within the desired range. Otherwise, the feedback adjuster generates the increments $\Delta \Theta_O$ and then takes $\Theta_O = \Theta_O + \Delta \Theta_O$ as the initial value of the optimization calculation instead of $\Theta_O$. Finally, this revised new case is stored.

Feedback Adjustor

Various factors and disturbances affect the denitration process, thus, its operating condition fluctuates greatly, which makes the original operating points drift away from their pre-specified positions. To either reduce or eliminate the effects of the drifting of the operating points on the preset operational indices, a tuning mechanism that can compensate for the set points generated by the pre-setting controller must be established. For this purpose, a dynamic tuning mechanism is constructed.

The feedback compensated value is updated as follows:

$$\Delta \Theta_O(t) = k_f(p(t) - e(t - 1)) + k_1e(t)$$  

$$\Theta_O(t) = \Theta_O(t - 1) + \Delta \Theta_O(t)$$

where $k_f$ and $k_1$ are the integration and proportional coefficients of the feedback PI compensators, respectively, and they are determined by experience.

T-JITL-Based Real-Time Optimization

JITL (also called lazy learning) originated in the machine learning area. In the JITL model structure, a local model is built using the most relevant samples from the historical data set around a query sample. Thus, JITL overcomes the problem of maintaining a large-scale static model for the denitration process, and enables the real-time optimization of the process.

In this section, we introduce time order into the selection rules of JITL to determine a modelling region at the current operating point and to improve the modelling accuracy. Based on a similar set, a local nonlinear model can be built and used for the optimization calculation and implementation. Then, these steps can be repeated in the next optimization cycle. The above optimization steps are illustrated in Figure 6.

Basic JITL algorithm

In basic JITL, the similarity factor and the distance and angle information between the current working point $H_i$ and the sample $H_j$ in the historical database are represented as follows:

$$D(H_c, H_i) = \rho \exp(-d(H_c, H_i)) + (1 - \rho)\cos \theta(H_c, H_i)$$

$$d(H_c, H_i) = \sqrt{|H_c - H_i|^2}$$

$$\cos \theta(H_c, H_i) = \frac{H_c^T H_i}{||H_c||^2 ||H_i||^2}$$

where $d(H_c, H_i)$ and $\theta(H_c, H_i)$ are the distance similarity and angle similarity between $H_c$ and $H_i$ in the database, respectively, and $0 \leq \rho \leq 1$ is a weight parameter. The value of $D(H_c, H_i)$ is bounded between 0 and 1, and when it approaches 1, $H_i$ closely resembles $H_c$. Note that the similarity of $D(H_c, H_i)$ cannot be calculated using Equation (31) if $\cos \theta$ is negative.

Improved JITL set determination

In industrial production, the equipment characteristics and operating conditions change with time, thus, the time characteristics of data must be considered when determining the learning
The historical data are divided into \( q \) sub-blocks. The time weights within the data blocks are the same, and the weights between different blocks are different. The \( k_{\text{max}} \) group data are selected to form the initial learning set according to the improved learning set selection criteria, this step can be called rough selection. Then, another set of time weights is determined, and the final learning set is selected according to the improved selection criteria again: this step is called careful selection.

The time weight vectors are defined as:

\[
W_A = \begin{pmatrix} w_{A1}, w_{A2}, \ldots, w_{Aq} \end{pmatrix}^T, \sum_{k=1}^{q} w_{Ak} = 1, w_{Ak} \in [0,1] \tag{32}
\]

\[
W_B = \begin{pmatrix} w_{B1}, w_{B2}, \ldots, w_{Bp} \end{pmatrix}^T, \sum_{k=1}^{p} w_{Bk} = 1, w_{Bk} \in [0,1] \tag{33}
\]

where \( W_A \) and \( W_B \) are the weight vectors of the rough selection and careful selection processes, respectively, \( q \) is the number of blocks in the historical data, and \( p \) is the number of data in the careful selection process and is often equal to \( k_{\text{max}} \), as denoted in Equation (41).

We set \( W_A \) (\( W_B \)) to be incremental vectors (Equation (34)) after considering the actual characteristics of the process; thus, the data closer to the current time should be taken more seriously.

\[
w_k < w_{k+1}, k = 1, 2, \ldots, q(p) - 1 \tag{34}
\]

The historical data are divided into \( q \) sub-blocks. The time weights within the data blocks are the same, and the weights between different blocks are different. The \( k_{\text{max}} \) group data are selected to form the initial learning set according to the improved learning set selection criteria, this step can be called rough selection. Then, another set of time weights is determined, and the final learning set is selected according to the improved selection criteria again: this step is called careful selection.

\[
W_{A_k} = w_{A(k-1)} + \Delta_1, k = 2, 3, \ldots, q \tag{35}
\]

where \( \Delta_1 \) is the interval, \( W_A \) can be determined after the given \( \Delta_1 \) or \( w_{A1} \), and the following two conditions need to be guaranteed: (i) \( \Delta_1 = \frac{(1-q^2)}{q(q-1)} > 0 \) and (ii) \( w_{A1} \) should be as small as possible to ensure that the weights between different blocks are as large as possible. The first \( k_{\text{max}} \) group data with the largest similarity according to Equation (39) are selected as rough selection results.

Suppose that the ordinal numbers of data in the original database are \( r_1, r_2, \ldots, r_{k_{\text{max}}} \). These data are taken as a new data set, and the weight vector of the careful selection process with an interval of \( \Delta_2 \) is constructed. The following equation is obtained according to Equation (33):

\[
w_{B1} + (w_{B1} + (r_2 - r_1)\Delta_2) + \ldots + (w_{Bp} + (r_{k_{\text{max}}} - r_1)\Delta_2) = 1 \tag{36}
\]

The above formula can be simplified as follows:

\[
w_{Bn} + (w_{B(n-1)} + (rn - r_1)\Delta_2), n = 2, 3, \ldots, k_{\text{max}} \tag{37}
\]

\[
k_{\text{max}} w_{B1} + \sum_{i=2}^{k_{\text{max}}} r_i \Delta_2 - (k_{\text{max}} - 1) r_1 \Delta_2 = 1 \tag{38}
\]

Thus, \( W_B \) can be determined after being given \( \Delta_2 \) or \( w_{B1} \), the condition \( \Delta_2 = \frac{1 - k_{\text{max}} w_{B1}}{\sum_{i=2}^{k_{\text{max}}} r_i - (k_{\text{max}} - 1) r_1} > 0 \) also needs to be guaranteed.

We set \( W_A \) (\( W_B \)) to be incremental vectors (Equation (34)) after considering the actual characteristics of the process; thus, the data closer to the current time should be taken more seriously.

\[
w_k < w_{k+1}, k = 1, 2, \ldots, q(p) - 1 \tag{34}
\]
directly given \( k \). This factor can also reduce the size of the learning set to a certain extent, thus improving the real-time model performance.

**T-JITL-based nonparametric modelling**

Assuming that the input sample set is \( G = \{G_1 \ldots G_k\} \) and the output set is \( Y = \{Y_1 \ldots Y_k\} \) (\( k \) is the sample size) in the final learning set, where \( G_1 = [g_{i1} \ldots g_{im}'] \) (\( m' \) is dimension of input), then a local polynomial approximation estimation model can be expressed under the hypothesis that \( n' \)-order derivatives of the model exist at \( G_{op} \):

\[
f(G) = \beta_0 + \sum_{j=1}^{m'} \sum_{i=1}^{n'} [\beta_{ji} \cdot (g - g_{op,i})]'
\]

where \( f(\cdot) \) denotes the unknown relation between the input and output; \( \beta_0 \) and \( \beta_{ji} \) are the parameters to be identified. The parameter estimation problem can be formulated as follows:

\[
\begin{align*}
\min J &= \sum_{c=1}^{k} [Y_c - f(G_c)]^2 \omega_c \\
\text{s.t. } f(G) &= \beta_0 + \sum_{j=1}^{m'} \sum_{i=1}^{n'} [\beta_{ji} \cdot (g - g_{op,i})]'
\end{align*}
\]

\[
\omega_c \geq 0
\]

where \( \omega_c \) is the weight coefficient. To obtain \( \beta_0 \) and \( \beta_{ji} (j = 1, \ldots, m'; i = 1, \ldots, m') \), the following set of equations are used:

\[
\begin{align*}
\tilde{Y} &= [Y_1 \ldots Y_k]^T \\
G^T &= \begin{bmatrix}
1 & g_{i1} - g_{op,1} & \ldots & g_{im'} - g_{op,m'} \\
\vdots & \vdots & \ddots & \vdots \\
1 & g_{i1} - g_{op,1} & \ldots & g_{im'} - g_{op,m'}
\end{bmatrix}
\end{align*}
\]

\[
W = \text{diag}(\omega_c)
\]

Thus, the parameter matrix \( \Psi \) can be estimated by the least square method as:

\[
\Psi = (GWG^T)^{-1} GWY
\]

In summary, the T-JITL-based modelling steps are as follows:

1. Calculate the distance and angle information between the current working point \( H_1 \) and the input \( H_i \) in the historical database. Discard the data with angle information less than zero.
2. Set the number of data blocks and the time weight vectors \( W_A \).
3. Rank the similarity factors calculated by Equation (39) in descending order and select \( k_{max} \) group data, which constitutes the initial learning set.
4. Determine the time weight vectors \( W_B \) according to the initial learning set.
5. Calculate the similarity factors using Equation (39) and rank them in descending order. The final learning set is determined according to the cumulative similarity factor.
6. Identify the model parameters and complete the modelling.

**INDUSTRIAL APPLICATION**

To validate the proposed method for the coking flue gas denitration process, an industrial application of the developed system in a coking plant in Jiangxi province, China, is introduced in this section. First, the development of an optimal setting system is introduced, followed by a description of the details of the industrial application. Figure 8a and b show the actual scene of the 55-hole and 6 m top charging coke oven and integrated desulphurization and denitration device, respectively.

Development of an Optimal Setting System

The general architecture of the system is shown in Figure 9. The STEP7 DCS from Siemens Co. is applied for lower-level loop control, logic control, alarm setting, network communication, and

![Figure 8](image-url)
I/O data acquisition. The SIMATIC WinCC6.0 software is used to develop the human-computer interaction monitoring platform for human supervision. The higher-level optimal setting system was developed by the Microsoft Visual Studio 2010 (VS 2010). VS 2010 provides an expedient way to realize communication and data exchange with the monitoring system and other application languages, such as real-time database pSpace and MATLAB. pSpace is used to develop the case database and other databases for the system, and MATLAB is used to execute certain necessary calculations online.

Industrial Application

Control objectives

During the three months of operation from April to June 2017, according to the technical requirements, the desired NOx concentration was $C_d = 200 \text{ mg/Nm}^3$, and the allowed fluctuation range was $C_4 = 30 \text{ mg/Nm}^3$.

System parameter designing

The case database contains 4103 cases, which are all from 1.5 years of running data combined with operator experience after the start of the device in October 2015; thus, $n = 4103$. The other parameters include $\omega_{v1-d} = [0.3, 0.4, 0.12, 0.1, 0.08]$, $\Delta t_3 = 15 \text{ min}$, and $m = 24$; thus, the optimization cycle is 6 h. The principal component contribution is $\eta = 87 \%$.

For the feedback adjustor, $k_{bp} = 4.1$ and $k_{df} = 1.3$. Similarly, all of the lower-level controllers of the process are PI-controlled and turned by the method presented in Cooper et al. [38]

The parameters of T-JITL-based RTO are determined as follows: $k_{min} = 500$, $k_{max} = 2500$, $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$, $w_{d1} = 0.01$, $w_{p2} = 0.002$, and $\omega_3 = 1$. The order of the model is an inherent property of the process and is determined to be 2 through experiments; thus, the optimization problem stated above is a quadratic programming problem and can be solved by such algorithms as the active set method. The block scale factor $b$ and cumulative similarity factor $z_k$ are set to 800 and 0.8 based on the grid search method and 5-fold cross validation, respectively.

Application results

The T-JITL-based modelling results are shown as Figure 10. The model performance is evaluated based on the average relative error (ARE), maximum relative error (MRE), root-mean-square error (RMSE), and relative root-mean-square error (RRMSE). The model performance evaluation is shown in Table 1.

The error indicators of the T-JITL-based model are all less than that of the original JITL approach. This result indicates the effectiveness of the proposed T-JITL method and that a good foundation is laid for the following optimization. The T-JITL-based real-time optimization is performed after the CBR-based pre-setting operation is completed. A time threshold $T_{co}$ is set (typically 30 min); thus, when the feedback adjustor remains non-functioning for $T_{co}$, the preset values are considered suitable for further optimization.

The application results of the optimal setting system are shown in Figure 11, which indicates the time-varying trends of the manipulated variable $N_1$, $N_2$, and ozone output and outlet NOx concentration over 5 h.

![Figure 9. Architecture of the developed system.](image)

![Figure 10. T-JITL-based local modelling results.](image)
In Figure 11c, the dotted red lines denote the upper and lower values of the desired range. The dotted green lines denote the duration of each reversing process. The reversing operation occurs every 15 and 45 min past an hour (e.g., 03:15, 04:15, 07:15, and 07:45). The feedback adjustor remains functioning only under the non-reversing process. As indicated by the curve before 04:40, the NOx concentration exceeds the desired range because of the varied coking condition; thus, the corresponding set points must be re-calculated and optimized at this time.

As shown in Figure 11a, before 04:40, the ozone generator worked at the optimal values $\Theta_{opt}^{last} = [N_1, N_2] = [16.12, 6.07]$ of the prior optimization cycle. At 04:40, the preset points $\Theta_0(04:40) = [N_1, N_2] = [18.34, 7.93]$ were calculated by the CBR-based pre-setting controller and implemented for the process, and the feedback adjustor produced the initial optimization value $\Theta_0(05:10) = \Theta_0 + \Delta \Theta_0 = [18.7, 8.39]$. Until approximately 06:11, the feedback adjustor remained non-functioning for 30 min. Therefore, the T-JITL-based real-time optimization module produced the final optimized set points $\Theta_0 = [N_1^{opt}, N_2^{opt}] = [17.21, 8.97]$. As shown in Figure 11b, the proposed optimal setting method enables the set points to be automatically adjusted as the working condition varies while minimizing the ozone output under the premise of the outlet NOx concentration being controlled within the desired range.

The application effects of human supervision and the optimal setting system are compared in Figure 12. To illustrate the results of the entire operation, two groups of data were collected (from April to June 2017) and then averaged into daily values. Another two groups of data were also collected (from April to June 2016, the same operation requirements as the former) from the database before the system started running. These data are used to illustrate the advantages of the developed system compared to human operation.

In Figure 12a, compared with human supervision, the NOx concentrations with the optimal setting system are further within the desired range and fluctuate less. Only 66.6 % of the data with manual control are within the desired range. Figure 12b shows that the ozone output with the optimal setting system can be set appropriately according to different working conditions. However, the ozone output with manual control is typically set at a high value and adjusted in an untimely and inappropriate manner.

To further illustrate the economic benefits obtained from adopting the proposed method, the power consumption cost per week with the optimal setting system and with human supervision are compared in Figure 13. The weekly cost savings increase by 29.01, 24.64, 24.98, 18.94, 26.47, 29.28, 25.54, 14.69, 13.78, 8.78, 11.57, and 13.66 % in sequence, and the total savings over the 3 months are 192 900 RMB. Thus, annual savings of 0.772 million RMB can be obtained by extrapolating these results. As is evident, the proposed approach yields an improved operation performance and remarkable economic benefits for the coking flue gas denitrification process.

### Table 1. Model performance evaluation

<table>
<thead>
<tr>
<th>Item</th>
<th>ARE</th>
<th>MRE</th>
<th>RMSE</th>
<th>RRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>JIT</td>
<td>0.0068</td>
<td>0.0207</td>
<td>2.1249</td>
<td>0.1325</td>
</tr>
<tr>
<td>T-JITL</td>
<td>0.0022</td>
<td>0.0106</td>
<td>0.7535</td>
<td>0.0470</td>
</tr>
</tbody>
</table>

**Figure 11.** Application result of the proposed method. The sampling interval is 3 s. (a) Ozone content of $N_1$ and $N_2$, (b) ozone output $(N_1 + N_2)$, and (c) outlet NOx concentration.

**Figure 12.** Comparison of control effects with human supervision and with the optimal setting system.

April to June 2017) and then averaged into daily values. Another two groups of data were also collected (from April to June 2016, the same operation requirements as the former) from the database before the system started running. These data are used to illustrate the advantages of the developed system compared to human operation.

In Figure 12a, compared with human supervision, the NOx concentrations with the optimal setting system are further within the desired range and fluctuate less. Only 66.6 % of the data with manual control are within the desired range. Figure 12b shows that the ozone output with the optimal setting system can be set appropriately according to different working conditions. However, the ozone output with manual control is typically set at a high value and adjusted in an untimely and inappropriate manner.

To further illustrate the economic benefits obtained from adopting the proposed method, the power consumption cost per week with the optimal setting system and with human supervision are compared in Figure 13. The weekly cost savings increase by 29.01, 24.64, 24.98, 18.94, 26.47, 29.28, 25.54, 14.69, 13.78, 8.78, 11.57, and 13.66 % in sequence, and the total savings over the 3 months are 192 900 RMB. Thus, annual savings of 0.772 million RMB can be obtained by extrapolating these results. As is evident, the proposed approach yields an improved operation performance and remarkable economic benefits for the coking flue gas denitrification process.

**Figure 13.** Power consumption cost with human supervision and with the optimal setting system.
CONCLUSIONS

We proposed a data-based optimal setting strategy for the coking flue gas denitration process. This methodology can determine the proper set points for the coking flue gas denitration process under non-reversing conditions without accurate mathematical models. The approach was applied to an integrated desulphurization and denitration device at a coking plant in China. The successful application illustrated the validity and effectiveness of the proposed approach, which is also promising for solving other optimization problems in industrial processes. The main contributions of this paper are summarized as follows:

- A data-based two-stage optimization strategy was proposed to solve the optimal setting problem of the coking flue gas denitration process and applied to domestic desulphurization and denitration equipment in China.
- Regarding the data-based optimization strategy, a PCR multiple case fusion case-based reasoning method was proposed to calculate the preset values. A T-JITL algorithm was developed for the local modelling of the denitration process to perform further RTO.
- A real-world application of the system indicated that the system can control the outlet NOx concentration within its desired range, yielding an annual cost saving of nearly 1 M RMB. The system can provide effective control and yield notable social and economic benefits.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China [U1701262] and the 2016 intelligent manufacturing project of the Ministry of Industry and Information Technology of China [2016ZXFMO6005].

REFERENCES

[33] M. Han, Z. J. Cao, Neurocomputing 2015, 149, 1245.