Mathematical Modeling for the Industrial 2-Mercaptobenzothiazole Batch Production Process

Enzhi Liang, Song Zhang, Bin Liu, Bujin Qi, Yanpei Nie, and Zhihong Yuan*

ABSTRACT: As an important chemical intermediate, 2-mercaptobenzothiazole (MBT) is widely used in various processes, especially in the rubber industry. However, there is no first-principles model that describes the synthetic process of MBT. This paper focuses on the formulation of a reliable mathematical model represented by a series of differential and algebraic equations for the industrial batch MBT production process. It is difficult to estimate all of the unknown parameters in the model because of the lack of sufficient industrial/experimental data. Thus, a reduced estimable parameter set is derived by performing estimability analysis on the original estimation problem. A multiple-starting-point strategy is then applied to numerically solve the non-convex parameter estimation problem with the weighted least-squares approach. Subsequently, a cross-validation technique is employed to evaluate the generalizability of the proposed model. Finally, it is confirmed that the proposed model produces encouraging prediction performance with regard to independent test data.

1. INTRODUCTION

The chemical intermediate 2-mercaptobenzothiazole (MBT) is a crucial product in the rubber and automobile manufacturing industries. Traditionally, MBT was extensively used as an accelerator in the rubber vulcanization process. Later, the capability of MBT for inhibiting metal corrosion was discovered, and biological applications of MBT as a fungicide were also investigated. In recent research studies, new functions of MBT and its derivatives have been disclosed that suggest it can be used in the modified Julia olefination reaction, which is an effective tool for advance fragment linkage. MBT is also an essential precursor of many other popular rubber additives, such as 2,2'-dibenzothiazyl disulfide (MBTS) and N-cyclohexylbenzothiazole-2-sulphenamide (CBS). Furthermore, several widely used antibiotics, such as cephalosporins, can also be synthesized from the derivatives of MBT.

As the batch process is more flexible and closer to the market requirements than its continuous counterpart, batch production of MBT with aniline as a reactant, usually referred to as the Kelly process, remains a popular choice. The Kelly process charges the reactants, i.e., aniline, sulfur, and carbon disulfide, in a closed autoclave. The temperature is then raised to 220−250 °C and maintained for about 4 h. During the reaction, the pressure of the autoclave increases to around 10 MPa as a result of the production of H₂S. Under these tough reaction conditions, major manipulations of the industrial MBT production process still rely on the prior experiences of workers, which may lead to significant reductions in profits and could even cause safety problems. Therefore, it is essential to build a reliable process model for MBT and estimate the relevant process parameters for the industrial production process in preparation for further optimization, automatic control, and monitoring.

Existing approaches for process modeling can be categorized into various levels according to the capabilities of the model in describing the relevant process; common examples include white-box models, gray-box models, and black-box models. Among the various types of process models, the first-principles models, or fundamental mathematical models, theoretically provide the best representations of the dynamic characteristics and nonlinear behaviors of the chemical process. Unfortunately, the rigorous reaction mechanism of MBT synthesis is very complicated, with at least 50 species separated and identified from the generated mixture. The inter-reactions between the minor components have not yet been fully identified. Therefore, the establishment of a full and rigorous
first-principles model that includes all reaction components is a huge challenge. Fortunately, the major reaction steps of MBT synthesis have been disclosed in recent research studies.\textsuperscript{5,15} Thereby, kinetics of these key reactions can be mathematically formulated. Additionally, the less important reactions can be represented by a lump reaction. In this way, a complete theoretical model describing the kinetics, mass/energy balances, and equations of state can be established.

Parameter estimation is one of the key components in establishing any model. In certain linear systems, where the output variables can be easily interpreted as explicit functions of the input variables and model parameters, parameter estimation is often referred to as explicit estimation.\textsuperscript{16,17} In contrast, implicit estimation refers to estimation problems in which the outputs are implicitly dependent on the inputs and parameters.\textsuperscript{18} Implicit parameter estimation is usually performed by solving an optimization problem that minimizes the mismatches between the model predictions and real datasets, with the parameters serving as the decision variables.\textsuperscript{19,20} The implicit estimation strategy is applied in this study due to the nonlinearity of the reaction kinetics of the MBT production process.

In recent decades, a great number of methods have been developed for implicit parameter estimation. Among these, the most popular one is the least-squares (LS) approach, in which the objective function is generally defined as a summation of the 2-norm residuals of output variables between model prediction and measurement.\textsuperscript{13,17,21} However, the standard LS approach assumes that different output data have same contributions to the estimation of unknown parameters.\textsuperscript{22} Therefore, the weighted least-squares (WLS) method was developed, whereby suitable weights are assigned to the relevant residuals.\textsuperscript{19,23} The relatively compact form and high computational efficiency of the WLS approach have led to its widespread application to parameter estimation problems in various chemical processes.\textsuperscript{23–25} The WLS approach assumes that the input variables can be precisely measured and that there are no errors in the input data. Nevertheless, errors in the input measurements cannot be ignored in a few industrial processes, such as the measurement errors of jacket temperature in the solid–liquid reactions.\textsuperscript{26} In such cases, the errors-in-variables (EIV) method can better describe these processes, because all of the errors in the measured data, including the input data, are considered in the EIV formulation.\textsuperscript{27,28} However, the EIV approach requires far more computational resources than WLS.\textsuperscript{29} For example, in the parameter estimation problem of the batch MBT process where 18 experiments are available in the training dataset and each experiment records around 60 temperature points as input...
variables, the EIV approach has to consider the extra $18 \times 60$ measured temperature values as additional decision variables, which leads to unnecessary use of computational resources. Moreover, treating all the inputs as decision variables can lead to an ill-condition problem especially when these inputs are not informative enough.32

Industrial batch MBT production is essentially a dynamic process that has no stable operation point. Hence, the corresponding mathematical model should incorporate ordinary differential equations and algebraic equations, giving a differential and algebraic equation (DAE) system. The parameter estimation for such a nonlinear DAE system will often encounter challenges such as non-convexity and ill-conditioning36,38 where the solution is very sensitive to the variations in the data. The non-convexity of an optimization problem can lead to multiple local optima. Therefore, the multiple-starting-point (MSP) approach is applied to the estimation problem to improve global optimality. On the other hand, the difficulty of optimizing an ill-conditioned and over-fitted problem can be properly addressed by performing estimability analysis and reducing the number of parameters to be estimated.33,35 In general, parameters that are coupled with each other and parameters that have no significant influence on the output data cannot be estimated, and are therefore removed from the estimation problem to avoid numerical intractability and unreliable solutions.

Several effective techniques for improving ill-conditioned estimation problems and selecting the most informative parameters have been developed. For instance, the ridge regression approach regularizes the parameter estimation problem by adding a specific term to the LS term of the objective function. Nevertheless, ridge regression cannot provide enough extra constructive information. Another approach is principal component analysis (PCA), a model-reduction technique that can identify and choose the most influential parameters to be estimated.35 Specifically, PCA applies eigenvalue decomposition to the information matrix of the corresponding process and selects the largest eigenvalue that makes the greatest contribution to the system equations.36 The original matrix is then projected into a space spanned by the selected eigenvectors. After repeating the eigenvalue decomposition and projection steps, the full space of the original matrix is replaced by the reduced space spanned by the eigenvectors with the largest eigenvalues. Hence, the computational burden of the original problem can be significantly reduced. Nevertheless, PCA can only be applied to linear or almost-linear systems, while the estimation problem for the MBT synthesis process is highly nonlinear. A third technique is global sensitivity analysis (GSA). GSA can naturally incorporate the parameter information into the sensitivity analysis to describe the parameter behaviors over a large range.33,37 Hence, GSA can evaluate the parameter estimability over the whole space. However, the evaluation of the global sensitivity matrix requires the computation of a complicated multi-dimensional integral term. Furthermore, the experimental data cannot be used directly and need to be reformulated before applying the global sensitivity technique.35 Another technique is parameter estimability analysis based on the local sensitivity matrix, which has been widely studied and applied in the past decades.36,32,38 and is used in this work.

The overall research methodology of this study is displayed in Figure 1. Based on the prior experiments and existing literature, the reaction mechanism for parameter estimation is disclosed. Then, the dynamics of the process and corresponding parameter estimation problem are formulated. The estimability analysis is conducted to improve the condition number of the estimation problem. Subsequently, the estimation problem is solved and the selected unknown parameters are revealed. Finally, the established model is validated by independent experiment data and industrial data.

We concentrate on developing a reliable first-principles mathematical model and producing optimal estimates of the relevant parameters for the industrial MBT batch production process. The main scientific contributions can be described as follows:

1. The mathematical model is formulated for the industrial MBT batch production process.
2. Parameter estimability analysis is conducted to select the estimable parameters and improve the condition number of the parameter estimation model for MBT production process.
3. The nonconvex parameter estimation problem is solved by the MSP approach to get the optimal parameters for the model of MBT production process.
4. The quality of the proposed model is evaluated using independent industrial/experimental datasets.

The remainder of this paper is arranged as follows. In Section 2, the reaction mechanism and the model equations are introduced. Section 3 then discusses the parameter estimation model and estimability analysis approach. In Section 4, the unknown parameters are estimated and the model is validated. Finally, a brief conclusion to this study is presented in Section 5.

2. REACTION MECHANISM AND MODEL DEVELOPMENT

The overall reaction mechanism of MBT synthesis can be expressed as follows:

\[
aniline + S + CS_2 \rightarrow \text{MBT} + H_2S
\]

The MBT yield in the current industrial batch process is usually 80–90% based on aniline. At present, the complete steps of the elementary reactions are still under exploration.5 In general, the intermediates associated with MBT synthesis can be categorized into two groups, namely, the desired group that generates MBT and the undesired group. Specifically, the most crucial intermediates in the desired group are N,N-diphenylthiourea (DPTU) and N-phenyl-2-benzothiazolamine (ABT). Both intermediates take part in the major reaction steps for synthesizing MBT.15,39 The undesired group includes those byproducts that finally and irreversibly form tar. As the real composition of tar is very complex, a virtual component is used to represent it in our work, the formula for which is derived based on element conservation. Note that the most abundant byproduct, 1,3-benzothiazole (BT) can also be categorized as an undesired product because the generation of more BT usually leads to less production of MBT.

Based on previous studies on the major reaction steps of MBT synthesis,5,15,39 the dominant reaction path can be described as shown in Figure 2. For brevity, the “smaller” components, such as sulfur, CS2 and H2S, are not displayed in Figure 2.

According to Figure 2, the MBT synthesis pathway can be decomposed into the following three cascade reactions:

\[
\begin{align*}
\text{(i) } & 2\text{aniline} + CS_2 \rightarrow \text{DPTU} + H_2S \\
\text{(ii) } & \text{DPTU} + S \rightarrow \text{ABT} + H_2S
\end{align*}
\]
Figure 2. Main pathway of the MBT synthesis reaction.

\[ \text{Aniline} \xrightarrow{(i)} \text{DPTU} \xrightarrow{(ii)} \text{ABT} \xrightarrow{(iii)} \text{MBT} \xrightarrow{(iv)} \text{Tar} \]

where:

- Aniline \( \rightarrow \) DPTU
- DPTU \( \rightarrow \) ABT
- ABT \( \rightarrow \) MBT
- MBT \( \rightarrow \) Tar

\[ \text{ABT} + \text{S} + \text{CS}_2 \rightarrow 2\text{MBT} \quad (\text{iii}) \]

The byproduct BT can be generated from the following side reaction, which is much slower than the main reactions:

\[ \text{ABT} + \text{H}_2\text{S} \rightarrow \text{BT} + \text{S} + \text{Aniline} \quad (\text{iv}) \]

The lump reaction (v) represents the complex inter-reactions between the undesired byproducts, which will irreversibly generate tar:

\[ 2\text{MBT} + a\text{S} + b\text{H}_2\text{S} + c\text{CS}_2 + \text{daniline} \rightarrow \text{tar} \quad (\text{v}) \]

where the molar weight of the virtual tar component and the stoichiometric coefficients \( a, b, c \) and \( d \) in reaction (v) can be derived by conservation: \( a = 0.41, b = 0.37, c = 0.82, \) and \( d = -0.168. \)

The traditional apparent Arrhenius equation is defined as follows:

\[ r_i = A_0^i \exp \left( -\frac{E_{a,i}}{RT} \right) \prod_j c_j^{\alpha_{ij}} \quad (1) \]

where \( i = 1, \ldots, 5 \) indicate the indexes of the reaction, \( j = 1, \ldots, 9 \) represent the indexes of the considered components, \( A_0^i \) denotes the Arrhenius coefficient, \( E_{a,i} \) represents the activation energy, and \( \alpha_{ij} \) is the reaction order of component \( j \).

In this work, the equivalent re-parameterized form of the apparent Arrhenius equation is given in eq 2. This is used to reduce the computational complexity because the orders of magnitude of the kinetic constants \( A_0^i \) and \( E_{a,i} \) in the traditional expressions are unbalanced.

\[ r_i = \exp \left( a_i + b_i \left( 1 - \frac{T_f}{T} \right) \right) \prod_j c_j^{\alpha_{ij}} \quad (2a) \]

where \( T_f \) is the reference temperature and \( a_i \) and \( b_i \) are the re-parameterized kinetic parameters that can be defined as:

\[ a_i = \ln A_0^i - \frac{E_{a,i}}{RT_f} \quad (2b) \]

\[ b_i = \frac{E_{a,i}}{RT_f} \quad (2c) \]

Note that \( b_i > 0 \) holds in most cases because \( E_{a,i} > 0 \), while the sign of \( a_i \) is often dependent on the choice of reference temperature \( T_f \). In this work, the reference temperature is selected as 273.15 K to let \( a_i < 0 \) hold and avoid the change of \( a_i \) from plus to minus. In this way, the sensitivities of model variables with respect to \( a_i \) can be evaluated without extra consideration of the sign change. Moreover, an increase in \( b_i \) will lead to a faster reaction rate because \( T_f \) is much lower than the reaction temperature \( T \). In addition, \( a_i \) and \( b_i \) are dimensionless parameters.

Importantly, notwithstanding that the batch reactor has vapor and liquid phases, reactions can only take place in the liquid phase. Based on the empirical equations of state, the average density of the liquid reactant shows no significant change before and after the reaction. As massive amounts of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) dissolve in the liquid phase under high pressure, the change of mass in the liquid phase is negligible compared with the overall mass. As a consequence, the liquid volume \( V_{liq} \) can also be seen as a constant during the batch process. Therefore, the concentrations of the heavy components, including aniline, sulfur, BT, MBT, ABT, DPTU, and tar, can be expressed as follows:

\[ c_{i,h} = \frac{n_{i,h}}{V_{liq}} \quad (3) \]

where \( n_{i,h} \) is the molar amount of the heavy component.

The light components, i.e., \( \text{CS}_2 \) and \( \text{H}_2\text{S} \), exist in both the liquid and vapor phases. Hence, the effective molar amounts of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in the liquid phase that can participate in the reactions are lower than their total abundances. In general, the molar fractions of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in a liquid−vapor system change with different compositions, reactor pressures, and temperatures. Thus, the relationships between the fractions of light components in the vapor phase, \( x_{CS_2, vap}, x_{H_2S, vap} \) and related variables, can be described as:

\[ x_{CS_2, vap} = f(n_{CS_2}, n_{H_2S}, n_{H, vap}, P, T) \]

\[ x_{H_2S, vap} = f(n_{CS_2}, n_{H_2S}, n_{H, vap}, P, T) \quad (4) \]

where \( n_{CS_2}, n_{H_2S} \in [0, 1] \) and \( n_{H, vap} \in [0, 1] \) represent the molar fractions of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in the vapor phase, respectively. The molar amounts of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in the vapor phase can be computed from \( n_{CS_2, vap} = n_{CS_2} x_{CS_2, vap} \) and \( n_{H_2S, vap} = n_{H_2S} x_{H_2S, vap} \). Note that \( x_{CS_2, vap} + x_{H_2S, vap} = 1 \) and \( x_{CS_2, vap} + x_{CS_2, liq} = 1 \) always hold during the reaction process.

As the overall compositions, pressures, and temperatures change with time, the molar fractions of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in the vapor phase also change over time. Hence, the rigorous form of eq 4 is difficult to measure. Fortunately, the change in the overall composition is related with the reactor pressure, because an increase in the reactor pressure indicates the generation of the terminal product \( \text{H}_2\text{S} \). A higher pressure is correlated with more products and less remaining reactants. In this way, the compositions of all the components can be expressed as an implicit function related to pressure. Hence, eq 4 can be simplified as:

\[ x_{CS_2, vap} = f(n_{CS_2}(P), n_{H_2S}(P), n_{H, vap}(P), P, T) \]

\[ x_{H_2S, vap} = f(n_{CS_2}(P), n_{H_2S}(P), n_{H, vap}(P), P, T) \quad (5) \]

Furthermore, eq 5 can be approximated by the following linear expressions:

\[ x_{CS_2, vap} = k_{p1}P + k_{T1}T + x_1 \]

\[ x_{H_2S, vap} = k_{p2}P + k_{T2}T + x_2 \quad (6) \]

where \( x_1 \) and \( x_2 \) indicate the intercepts of the affine approximation, while \( k_{p1}, k_{p2}, k_{T1}, \) and \( k_{T2} \) denote the slopes. Note that the values of the slopes and intercepts can be obtained offline via the equations of state and the empirical formula. Therefore, the concentrations of \( \text{CS}_2 \) and \( \text{H}_2\text{S} \) in the liquid phase are defined by
As mentioned above, the concentrations of the components change over the reaction process. Taking a mass balance, the change of concentration with respect to time can be defined as

$$\frac{dn_j}{dt} = V_{liq} \sum_{i=1}^{\phi} r_{fi}$$  

(8)

where $r_{fi}$ are the stoichiometric coefficients and $n_j$ denotes the molar amount of component $j$. In addition, the total reactor pressure is equal to the pressure of the vapor phase, which can be calculated by the pressure–volume–temperature (PVT) relationship as

$$P = \frac{Z_R T n_{CS_{vap}} + n_{H_2S_{vap}}}{V_{vap}}$$  

(9)

where $Z_R$ is the compression factor and $V_{vap}$ denotes the volume of the vapor phase. As the total volume of the reactor is fixed and the volume of the liquid phase can be regarded as a constant, the vapor volume can also be assumed to be constant.

The mathematical model of the industrial MBT batch production process is summarized in Table 1. The model contains 9 ordinary differential equations with nine initial conditions and 21 algebraic equations. The model has 31 time-varying variables, as listed in Table 2. Note that $n_{j,k}$ is a sub-vector of $n_j$ excluding CS$_2$ and H$_2$S. As a result, the mathematical model has one degree of freedom, which is equal to the number of independent variables to subtract the number of independent parameters to be estimated. For example, once the temperature profile is given, the whole reaction process should reach the same termination point under the same initial conditions. Importantly, the energy balance is not considered in the mathematical model because the batch reactor is a non-adiabatic reactor and no measured data relative to energy balance is available. In addition, the parameters of the proposed model are summarized in Table 3.

### Table 1. Mathematical Model of the MBT Batch Production Process

| physical meaning | equations |
|------------------|-----------|
| kinetics         | $r_{fi} = \exp(a_{fi} + b_{fi}(1 - r_{fi}^\theta)) \prod_{j}^{n_j} r_{fi}^\theta$ |
| mass balances    | $\frac{dn_j}{dt} = V_{liq} \sum_{i=1}^{\phi} r_{fi}$ |
| concentrations of heavy components | $\phi = \frac{n_j}{n_{CS_{vap}}} \left(\text{except CS}_2 \text{ and } H_2S\right)$ |
| concentrations of light components | $\phi = \frac{n_j}{n_{CS_{vap}}}$ |
| PVT relationship | $P = \frac{Z_R T (n_{CS_{vap}} + n_{H_2S_{vap}})}{V_{vap}}$ |
| molar fractions of CS$_2$ and H$_2$S in the vapor phase | $x_{CS_{vap}} \cdot k_{p}P + k_{T}T + x_{CS_{vap}} + k_{p}P + k_{T}T + x_{H_2S_{vap}}$ |
| molar amounts of CS$_2$ and H$_2$S in the vapor phase | $n_{CS_{vap}} = n_{CS_{vap}} + n_{CS_{vap}}$ |
| total molar amounts of CS$_2$ and H$_2$S | $n_{CS_{vap}} + n_{CS_{vap}}$ |

To be specific, the reactor temperature is controlled by the power of an electrical heating belt and the heat transfer between the reactor and environment cannot be neglected since the reactions happen at around 250 $^\circ$C.

### Table 2. Variables in the Proposed Model

| variables | units | description | number of variables |
|-----------|-------|-------------|---------------------|
| $r_{fi}$  | mol·min$^{-1}$ | reaction rates | 5                   |
| $n_j$     | mol   | molar amounts of reaction component $j$ | 9                   |
| $n_{j,k}$ | mol   | molar amounts of CS$_2$ and H$_2$S in the liquid phase | 2                   |
| $x_{vap}$ | /     | molar fractions of CS$_2$ and H$_2$S in the vapor phase | 2                   |
| $\phi_j$  | mol·L$^{-1}$ | concentrations of components in the liquid phase | 9                   |
| $P$       | MPa   | reactor pressure | 1                   |
| $T$       | $^\circ$C | reactor temperature | 1                   |
| $n_{j,k}$ | mol   | molar amounts of the heavy component | 7                   |

To be specific, the reactor temperature is controlled by the power of an electrical heating belt and the heat transfer between the reactor and environment cannot be neglected since the reactions happen at around 250 $^\circ$C.

### 3. PARAMETER ESTIMATION MODEL AND ESTIMABILITY ANALYSIS

#### 3.1. Parameter Estimation Model

The overall parameter estimation problem of the DAE system representing the MBT production process can be defined as

$$\min_\theta \theta^{WLS} = \sum_{m=1}^{M} \left\| y_m(t_{m}, \theta) - y_m^* \right\|^2$$

s. t. $k_m = f(x_m, u_m, \theta, t)$

$$y_m(t_{m}, \theta) = h(x_m, \theta, t)$$

$$x_m(0) = x_{m0}, \forall m \in I$$

$$\theta^{LB} \leq \theta \leq \theta^{UB}$$

(10)

where $y_m^*$ denotes the measured value of output variable $y$ at sampling time $t_{m}$ in the $m$th experiment, $y_m(t_{m}, \theta)$ denotes the model prediction for $y_m^*$, $\Sigma_{m}$ is the inverse of the covariance matrix, and $x_{m0}$ and $u_m$ represent the state and input variables in the $m$th experiment, respectively. $\theta \in \mathbb{R}^n$ denotes unknown parameters whose lower and upper bounds are defined as $\theta^{LB}$. 

---

This content is a fragment of a scientific article discussing the mathematical modeling and parameter estimation of a batch production process. The text includes detailed equations and tables summarizing the physical meanings, parameters, and variables involved in the model. The model contains 9 ordinary differential equations with nine initial conditions and 21 algebraic equations. The model has 31 time-varying variables, as listed in Table 2. The parameters of the proposed model are summarized in Table 3. The reactor temperature is controlled by the power of an electrical heating belt and the heat transfer between the reactor and environment cannot be neglected since the reactions happen at around 250 $^\circ$C. The overall parameter estimation problem of the DAE system representing the MBT production process can be defined as a minimization problem using the weighted least squares method, where the objective is to minimize the sum of squared residuals between the model predictions and the measured data, subject to the given constraints on the parameters.
and \( \theta^{UB} \). The whole DAE system contains \( n_m n_x \) differential equations and \( n_m n_y \) algebraic equations, where \( n_m \) denotes the number of batch experiments in dataset \( I \).

Solving the parameter estimation problem often gives point estimation. To further evaluate the accuracy of estimation, the confidence interval (CI) can also be evaluated. The CI of the \( n \)th parameter can be derived by the following definition \(^{21,41} \) for a given significance level \( \alpha \):

\[
\theta_{CL,n} = \theta^*_n \pm F^{-1}(P/\nu) \sqrt{d_m^2 WLS \nu^{-1}}
\]

where \( \theta^*_n \) denotes the estimated value, \( F^{-1}(P/\nu) \) represents the value of the Student's cumulative distribution function, \( P = 1 - \alpha/2 \) is the probability, \( \nu \) is the number of degrees of freedom, and \( d_m^2 \) is the \( n \)th diagonal element of \((Z^T Z)^{-1}\) where \( Z \) is the sensitivity matrix.

### 3.2. Estimability Analysis via a Local Sensitivity Matrix.

The estimability analysis reflects whether the fitted model with the selected parameter subset can properly depict the correlations between the output and input data. \(^{22} \) To illustrate the estimability analysis procedure, an explicit DAE system is defined here as

\[
\dot{x} = f(x, u, \theta, t)
\]

\[
y = h(x, \theta, t)
\]

where \( x \in \mathbb{R}^n_x \) denotes the state variable, \( y \in \mathbb{R}^n_y \) denotes the output variable, \( u \in \mathbb{R}^n_u \) is the input variable, and \( \theta \in \mathbb{R}^n_\theta \) denotes the unknown parameter. The initial conditions are defined as \( x(0) = x_0 \) and \( y(0) = y_0 \).

The sensitivity of the \( k \)th state variable \( x_k \) with respect to the \( p \)th unknown parameter \( \theta_p \) at sampling time \( t_i \) is defined as

\[
\left( \frac{\partial x_k}{\partial \theta_p} \right)_{t_i} = \left( \frac{\partial y_k}{\partial \theta_p} \right)_{t_i} + \sum_{k=1}^{n_x} \left( \frac{\partial x_k}{\partial x_i} \right) \left( \frac{\partial x_i}{\partial \theta_p} \right)_{t_i}
\]

The analytical form of the sensitivity coefficient can be given as adjoint sensitivity \(^{43} \) or direct sensitivity formulation. \(^{44,45} \) In this work, the direct sensitivity is chosen:

\[
\frac{d}{dt} \frac{\partial x_k}{\partial \theta_p} = \sum_{i=1}^{n_x} \left( \frac{\partial x_k}{\partial x_i} \right) \left( \frac{\partial x_i}{\partial \theta_p} \right) + \sum_{k=1}^{n_x} \left( \frac{\partial x_k}{\partial x_i} \right) \left( \frac{\partial x_i}{\partial \theta_p} \right)_{t_i}
\]

Based on the chain rule, the sensitivity coefficient that reveals the relationship between the \( l \)th output variable and the \( n \)th unknown parameter can be derived as

\[
\left( \frac{\partial y_l}{\partial \theta_p} \right)_{t_i} = \left( \frac{\partial y_l}{\partial \theta_p} \right)_{t_i} + \sum_{k=1}^{n_x} \left( \frac{\partial y_l}{\partial x_k} \right) \left( \frac{\partial x_k}{\partial \theta_p} \right)_{t_i}
\]

The sensitivity coefficients are usually scaled because the magnitudes of the parameters may be different. The scaled sensitivity can be expressed as

\[
s_{l,n} (t_i) = \frac{\bar{y}_l (t_i)}{y_l (t_i)} \left( \frac{\partial y_l}{\partial \theta_p} \right)_{t_i}
\]

where \( \bar{y}_p \) denotes the parameter point that is used for sensitivity evaluation, while \( y_l (t_i) \) represents the measured data of the output variable at time \( t_i \). The absolute value function \( l \cdot l \) is used in eq 16 because the scale factor \( \frac{\bar{y}_l}{y_l (t_i)} \) might be negative.

Thus, the scaled sensitivity matrix of output variable vectors with respect to parameter vectors at \( t_i \) can be expressed as

\[
S (t_i) = \begin{bmatrix}
s_{1,1} (t_i) & \cdots & s_{1,n_y} (t_i) \\
\vdots & \ddots & \vdots \\
0 & \cdots & s_{n_y,n_y} (t_i)
\end{bmatrix}
\]

where \( S (t_i) \in \mathbb{R}^{n_y \times n_\theta} \) represents the scaled sensitivity. Note that the outputs may not be measurable at every sampling moment. In this case, the corresponding sensitivity should be fixed to zero. The entire sensitivity matrix \( Z \) at every sampling moment \( t_i \in \{ t_0, \ldots, t_N \} \) can be defined as

\[
Z = \begin{bmatrix}
S (t_0) \\
\vdots \\
S (t_N)
\end{bmatrix}
\]

The full sensitivity matrix \( Z \) has \( n_0 \) columns and \( (N + 1) n_y \) rows. Importantly, the column rank of matrix \( Z \) indicates the maximum number of parameters that can theoretically be identified. \(^{33} \) Furthermore, the well-conditioned property of the parameter estimation problem can be ascertained by the condition number of \( Z^T Z \). If the condition number of \( Z^T Z \) is very large, the estimation problem is ill-conditioned and will usually not be estimable.

The estimability of the parameters can be sorted based on their influence on the measured data and whether these parameters are correlated with each other. \(^{26} \) Therefore, the parameter ranking policy mainly concentrates on the magnitudes and linear dependencies of the columns of \( Z \). In the following, two parameter ranking strategies are introduced. Both ranking methods will be used in the estimability analysis in this study.

The first approach iteratively regresses the full sensitivity matrix \( Z \) with the columns of the most predictable parameters to determine the estimable rankings of the parameters. \(^{32} \) The parameter rankings interpret the estimability of the corresponding parameters, from that with the greatest influence on the model outputs to that with the least influence. The detailed procedure is displayed in Figure 3, and the corresponding steps are described as follows: \(^{32} \)

1. Initialize the loop with \( k = 1 \), residual matrix \( R_0 := Z \), and define the initial selected column \( X_k \) as an empty matrix.
2. Calculate the 2-norm for every column of the residual matrix \( R_k - 1 \).
3. Search for the column with the largest 2-norm and remark the column as \( C_k \). The corresponding parameter is selected and ranked after the parameter chosen in the previous iteration.
4. The matrix denoting the already selected columns is augmented as \( X_k = [X_k - 1 \ C_k] \).
5. Calculate the prediction matrix \( \hat{Z}_k \) of the original matrix \( Z \) such that \( \hat{Z}_k = X_k (X_k^T X_k)^{-1} X_k^T Z \).
6. Calculate the updated residual matrix \( R_k := Z - \hat{Z}_k \).
The output of the order of columns in \( X_k \) if the norm of residual matrix \( R_k \) is below the stop criteria \( \epsilon \). Else, repeat the steps (3)−(6).

The second approach is to apply individual variance contribution analysis based on QR decomposition.\(^{38} \) In detail, the sensitivity matrix \( Z \) is decomposed by QR decomposition with column permutation as

\[
ZP = QR
\]

where \( P \) is the column permutation matrix, \( R \) is an upper triangular matrix, and \( Q \) is an orthogonal matrix. The matrix \( R \) is further decomposed as

\[
R = DR
\]

where \( D = \text{diag} [d_1, \ldots, d_n] \) and \( \hat{R} \) is a unit triangular matrix in which all diagonal elements are 1. With the supposition that all the measurement errors obey Gaussian distribution, the individual variance contributions of the unknown parameters can be written as

\[
\Delta \text{var} \theta_n = \frac{||r_n||^2}{d_n^2}
\]

where \( r_n \) is the \( n \)th column of the inverse matrix of \( \hat{R} \). The workflow of this approach is displayed in Figure 4.

**4. INDUSTRIAL CASE STUDIES**

The numerical problem was programmed on a personal computer with 16 GB RAM and an i7-8750H CPU @2.20 GHz. The estimation problem was implemented on the MATLAB 2018a platform. The control vector parameterization approach\(^{47} \) was used to discretize the DAE system with a MATLAB ode45 solver for integration, as this approach generates a relatively small nonlinear programming (NLP) problem when incorporating experimental data. The resulting NLP was then solved by the IPOPT solver\(^{48} \) with a MATLAB interface. The absolute and relative function evaluation tolerances of the IPOPT solver are set as \( 1 \times 10^{-7} \), while the other settings are set by default. The optimization problem with multiple starting points was solved by the parallel computing toolbox of MATLAB with a parallel pool (12 workers) on local computers.

**4.1. Estimation Model of the MBT Batch Production Process.** The experiments were conducted in a medium-sized autoclave (volume 4.0 L) with the experimental devices shown in Figure 5. Industrial data were collected from a larger industrial autoclave (volume = 1.0 m\(^3\)). The overall dataset contains 38 independent batches, including 28 batches of experimental data and 10 batches of industrial data. The whole dataset was divided into three subsets. The first subset, which served as the training dataset, contained 18 batches of experimental data for determining the values of the unknown parameters. The second subset, named test dataset I, contained 10 batches of experimental data, and the third subset, named test dataset II, contained 10 batches of industrial data.

In both the experimental and industrial batches, the charged molar amounts of aniline and sulfur/CS\(_2\) solution in the autoclave were fixed. The reactor temperature and pressure were recorded every 5 min. The weight fraction of MBT in the liquid crude product was only measured after the whole batch process had terminated. Online sampling of the reactant is highly dangerous because of the toxic H\(_2\)S in a high-pressure environment (~10 MPa). Therefore, sampling for the outlet crude product could only be conducted after the vapor phase was completely exhausted. The weight fraction of MBT in the crude product was detected by high-performance liquid
chromatography (HPLC). Note that the total mass of crude product from the industrial autoclave could not be measured. Additionally, because the vapor phase is exhausted in the industrial plant for treatment, the composition and molar amount of vapor could not be detected.

The types of accessible data are summarized in Table 4. Because the temperature of the autoclave can be manipulated by adjusting the electric heating power directly, the whole temperature curve is regarded as an input variable. The other measured data should be regarded as output variables because they cannot be manipulated directly. Hence, the system has only one degree of freedom.

The measured outputs can be related with the state variables in the DAE system. Based on mass balance, the total mass of the charged raw material should equal the summation of the liquid mass and vapor mass during the reaction process. Moreover, no CS₂ and H₂S exists in the crude product. Therefore, the mass of the discharged crude product can be calculated by

\[
\begin{align*}
    m_p(t_f) &= \sum_{i \in \{CS_2, H_2S\}} M_i n_i(t_f) - \sum_{i \in \{CS_2, H_2S\}} M_i n_i(t_0) \\
    &= \sum_{i \in \{CS_2, H_2S\}} M_{MBT} n_{MBT}(t_f) - \sum_{i \in \{CS_2, H_2S\}} M_i n_i(t_f)
\end{align*}
\]  

(22)

where \(m_p(t_f)\) denotes the mass of crude product, \(M_i\) denotes the molar weight of raw materials, and \(M_{MBT}\) denotes the vapor components.

Moreover, the weight fraction of MBT in the crude product can be defined as

\[
\begin{align*}
    x_{MBT}(t_f) &= \frac{M_{MBT} n_{MBT}(t_f)}{\sum_{i \in \{CS_2, H_2S\}} M_i n_i(t_f)} \\
    &= \frac{M_{MBT} n_{MBT}(t_f)}{\sum_{i \in \{CS_2, H_2S\}} M_i n_i(t_f)}
\end{align*}
\]  

(23)

where \(x_{MBT}(t_f)\) denotes the weight fraction of MBT.

Finally, the objective function was derived using the WLS approach, with the gap between the model predictions and real output values minimized. The objective function is defined as

\[
\begin{align*}
    \phi_i^\text{MBT} &= \sum_{m \in I} w_m (x_{MBT,m}(t_f) - x^*_{MBT,m,t_f})^2 \\
    &+ \sum_{m \in I} w_m (m_{p,m}(t_f) - m^*_{p,m,t_f})^2 \\
    &+ \sum_{m \in I} w_m \left[ P(m, i) - P^*_{m,i} \right]^2
\end{align*}
\]  

(24)

In the MBT model, the molar fractions of CS₂ and H₂S in the liquid—vapor phases and the compression factor \(Z\) in the PVT equation can be obtained in advance from empirical formulas. In addition, the stoichiometric coefficients of the reactions are known. Furthermore, the apparent reaction orders are assumed to be one based on the results of prior experiments. Therefore, only the unknown kinetic parameters, i.e., \(a_i\) and \(b_i\), have to be estimated in the full estimation problem.

In the MBT production process, the average reaction rates of the main reactions are around 0.001–0.1 mol/L/min because the reaction duration is around 250 min and the MBT produced in the experiment is around 20 mol. In addition, the temperature should be higher than 200 °C; otherwise, the reaction cannot take place. Therefore, the apparent activation energy of the main reactions should be within \([80, 160]\) kJ/mol. Correspondingly, the pre-exponential Arrhenius factor should be bounded within \([10^5, 10^{12}]\). The reference temperature is chosen as 273.15 K. As a result, the re-parameterized kinetic parameters should have the following ranges: \(a_i \in [-50, -10]\) and \(b_i \in [35, 70]\). In summary, the whole parameter estimation model for the MBT batch production process under the whole unknown parameter set is defined as

\[
\begin{align*}
    \min_{a_i, b_i} \phi_i^\text{MBT} \\
    s. t. \quad r_i &= \exp \left( a_i + b_i \left( 1 \frac{T_i}{T} \right) \right) \prod_{j \in I} c^\text{MBT}_j \\
    \frac{d\gamma_j}{dt} &= V_{liq} \sum_{i = 1}^s \gamma_j \gamma_i \\
    c_{CS_2, liq} &= n_{CS_2}(1 - x_{CS_2, vap}) \\
    c_{H_2S, liq} &= n_{H_2S}(1 - x_{H_2S, vap}) \\
    n_{CS_2} &= n_{CS_2, vap} + n_{CS_2, liq} \\
    n_{H_2S} &= n_{H_2S, vap} + n_{H_2S, liq} \\
    n_{CS_2, vap} &= n_{CS_2} x_{CS_2, vap} \\
    n_{H_2S, vap} &= n_{H_2S} x_{H_2S, vap} \\
    x_{CS_2, vap} &= k_{p1} P + k_{T1} T + x_1 \\
    x_{H_2S, vap} &= k_{p2} P + k_{T2} T + x_2 \\
    P &= Z_T (n_{CS_2, vap} + n_{H_2S, vap}) \\
    a_i \in [-50, -10], \quad b_i \in [35, 70]
\end{align*}
\]  

(25)

where the index \(i\) denotes the reactions (\(i = 1\)–\(v\)) and the decision variables are the unknown kinetic parameters \(a_i\) and \(b_i\). The estimability analysis would be performed later to decide which unknown parameters could be estimated with higher accuracy.

The parameter estimation problem for a nonlinear model usually has multiple local optima, although only the global optimum can guarantee the most precise approximation to the true values of the unknown parameters. Therefore, the MSP method is applied to solve the estimation problem for better global performance. Each start point can reach an independent
extremum on the local convex region under a gradient-based approach.\textsuperscript{30} Although the global optimum cannot be rigorously guaranteed, this approach is often sufficient to reach a satisfying solution within an acceptable computation budget.\textsuperscript{49} Moreover, compared with meta-heuristic approaches, the MSP approach is much easier to program and parallelize to further enhance computational efficiency.\textsuperscript{30}

4.2. Estimability Analysis for the MBT Batch Production Process. The local sensitivities were evaluated at the predefined parameter points $\mathbf{\theta}$, namely, the estimates of the unknown parameters $a_i$ and $b_j$. In this work, a feasible guess $\mathbf{\theta}_i$ was carefully chosen from within the full parameter set, while the second guess $\mathbf{\theta}_2$ is a local optimum parameter point obtained by solving the estimation problem eq 25 with one batch of experiment data as a prior. The ranks and condition numbers of the scaled local sensitivity matrixes evaluated at $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ are provided in Table 5. The condition number of matrix $Z^T Z$ is defined as $\text{cond}(Z^T Z) = \|Z^T Z\| \cdot \|(Z^T Z)^{-1}\|$. If the condition number is higher than $1 \times 10^{10}$, then the relative matrix is seen as ill-conditioned.\textsuperscript{36} In the ideal case, it is assumed that the concentration profiles of all components can be measured during the reaction. The ideal case can identify the model quality, i.e., whether the proposed model is over-parameterized or not. On the other hand, in real cases, only the final composition of MBT in the crude product can be obtained, representing the actual estimability under the full unknown parameter set.

Table 5. Ranks and Condition Numbers of the Local Sensitivity Matrix under a Full Parameter Set

| parameter point $\mathbf{\theta}_1$ | parameter point $\mathbf{\theta}_2$ | parameter point $\mathbf{\theta}_1$ | parameter point $\mathbf{\theta}_2$ |
|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| rank($Z$)                           | 10                                  | 10                                  | 10                                  |
| cond($Z^T Z$)                       | $10^{15}$                           | $10^{15}$                           | $3.2 \times 10^1$                   |

The full column ranks of the sensitivity matrixes show that all the unknown kinetic parameters are linearly independent. Hence, $a_i$ and $b_j$ can be uniquely determined in theory. Moreover, in the ideal case, the condition number of $Z^T Z$ is less than $1 \times 10^{10}$, demonstrating that the original parameter estimation problem is well-conditioned, which proves that the proposed model is not over-parameterized when sufficient concentration data are available. Nevertheless, in the real case, we only have the final weight fraction of MBT. Thereby, the condition number of $Z^T Z$ increases drastically to $1.1 \times 10^{15}$. In the real case, it is not possible to determine all unknown $a_i$ and $b_j$. As a consequence, the full unknown parameter set should be reduced to effectively exploit the available data and regain a well-conditioned optimization problem.

The parameter rankings and individual variance contributions given by the estimability analysis are presented in Table 6. The parameter rankings were produced by iterative regression, while the individual variances were calculated by QR decomposition. The parameter ranking is consistent with the ascending order of individual variance contributions because a higher individual variance contribution implies lower estimability.\textsuperscript{36}

According to Table 6, the re-parameterized pre-exponential factors $a_i$ are more estimable than the re-parameterized activation energies $b_j$. More importantly, the individual variance contributions of $b_5$, $b_7$, and $b_8$ are significantly larger than for the other activation energies. This is because $b_5$, $b_7$, and $b_8$ are the re-parameterized activation energies of reaction (iii), side reaction (iv), and the lump tar reaction (v), which have less influence on the output data than the other unknown parameters. Thus, $b_5$, $b_7$, and $b_8$ were marked as the least-estimable parameters and excluded from the full parameter set (i.e., they will not be determined in the estimation problem). Based on the values of parameter point $\mathbf{\theta}_b$, they were fixed as $b_5 = 45$, $b_7 = 54$, and $b_8 = 68$, equivalent to apparent activation energies of 100, 120, and 155 kJ/mol, respectively.

After removing $b_5$, $b_7$, and $b_8$, the full parameter set becomes the estimable parameter set. The parameter points $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ are produced by dropping $b_5$, $b_7$, and $b_8$ from $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$. The ranks and condition numbers of the local sensitivity matrixes $Z'$ evaluated at $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ in the real case are given in Table 7.

Table 7. Ranks and Condition Numbers of the Local Sensitivity Matrix under an Estimable Parameter Set

| parameter point $\mathbf{\theta}_1$ | parameter point $\mathbf{\theta}_2$ |
|-------------------------------------|-------------------------------------|
| rank($Z'$)                          | 7                                   |
| cond($Z'^T Z'$)                     | $4.9 \times 10^6$                   |

4.3. Local Sensitivity Analysis for the MBT Batch Production Process. Local sensitivity analysis can disclose some important characteristics of the system. The sensitivities of the output data with respect to the estimable parameters at $\mathbf{\theta}_1$ are given in Table 6. The sensitivity matrixes show that the estimable parameter set. The parameter points $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ are produced by dropping $b_5$, $b_7$, and $b_8$ from $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$. The ranks and condition numbers of the local sensitivity matrixes $Z'$ evaluated at $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ in the real case are given in Table 7.

Table 6. Estimability Analysis with Local Sensitivity under the Full Parameter Set

| parameter | value | variance contribution | parameter ranking |
|-----------|-------|-----------------------|-------------------|
| $a_1$     | -40   | $1.6 \times 10^{-4}$  | 1                 |
| $a_2$     | -25   | $2.5 \times 10^{1}$   | 5                 |
| $a_3$     | -25   | $8.0 \times 10^{-1}$  | 4                 |
| $a_4$     | -25   | $2.1 \times 10^{-3}$  | 2                 |
| $a_5$     | -50   | $4.0 \times 10^{-1}$  | 3                 |
| $b_1$     | 70    | $6.8 \times 10^{1}$   | 6                 |
| $b_2$     | 45    | $2.7 \times 10^{1}$   | 7                 |
| $b_3$     | 45    | $7.9 \times 10^{1}$   | 8                 |
| $b_4$     | 45    | $4.6 \times 10^{1}$   | 9                 |
| $b_5$     | 70    | $1.2 \times 10^{11}$  | 10                |

| parameter | value | variance contribution | parameter ranking |
|-----------|-------|-----------------------|-------------------|
| $a_1$     | -32.6 | $3.2 \times 10^{-4}$  | 1                 |
| $a_2$     | -28.8 | $5.9 \times 10^{-1}$  | 4                 |
| $a_3$     | -25.8 | $1.5 \times 10^{-3}$  | 2                 |
| $a_4$     | -31.2 | $5.4 \times 10^{-1}$  | 3                 |
| $a_5$     | -43.2 | $9.8 \times 10^{-1}$  | 5                 |
| $b_1$     | 57.3  | $6.9 \times 10^{1}$   | 6                 |
| $b_2$     | 51.9  | $7.8 \times 10^{1}$   | 7                 |
| $b_3$     | 44.9  | $2.3 \times 10^{1}$   | 8                 |
| $b_4$     | 53.6  | $7.6 \times 10^{1}$   | 9                 |
| $b_5$     | 68.1  | $1.1 \times 10^{11}$  | 10                |
Table 8. Scaled Local Sensitivity Matrix of $x_{MBT}(t_f)$ and $m_{p}(t_f)$ with Respect to Parameters at $\theta^*_1$

| parameter | $x_{MBT}(t_f)$ | $m_{p}(t_f)$ |
|-----------|----------------|--------------|
| $a_1$     | $2.0 \times 10^{-1}$ | $2.0 \times 10^{-1}$ |
| $a_2$     | $1.7 \times 10^{-1}$ | $3.5 \times 10^{-2}$ |
| $a_3$     | $1.4 \times 10^{0}$  | $4.7 \times 10^{-2}$ |
| $a_4$     | $-1.3 \times 10^{0}$ | $6.8 \times 10^{-3}$ |
| $a_5$     | $-8.0 \times 10^{-1}$| $1.3 \times 10^{-1}$ |
| $b_1$     | $1.7 \times 10^{-1}$ | $1.7 \times 10^{-1}$ |
| $b_2$     | $1.5 \times 10^{-1}$ | $3.0 \times 10^{-2}$ |

listed in Table 8. Note that higher values of $a_i$ and $b_i$ indicate a higher reaction rate. Therefore, the values of \( \frac{\partial x_{MBT}(t_f)}{\partial a_i} \), \( \frac{\partial x_{MBT}(t_f)}{\partial b_i} \), and \( \frac{\partial m_{p}(t_f)}{\partial b_i} \) are positive because higher reaction rates in reactions (i) and (ii) lead to more ABT and produce more MBT. Similarly, \( \frac{\partial x_{MBT}(t_f)}{\partial a_i} \) > 0 and is significantly larger than other derivatives because MBT is directly produced in reaction (iii). In contrast, \( \frac{\partial x_{MBT}(t_f)}{\partial a_5} \) < 0 because reaction (iv) is competing against reaction (iii), and increasing the rate of reaction (iv) leads to lower MBT production. Analogously, \( \frac{\partial x_{MBT}(t_f)}{\partial a_5} \) < 0 as tar reaction (v) consumes MBT. However, the sensitivity of $m_{p}(t_f)$ with respect to all the kinetic parameters is positive because the mass of crude product does not include the mass of CS$_2$, which is gradually consumed during the reaction.

4.4. Estimation Results. The problem scale and computational performance are listed in Table 9. The WLS approach was applied for parameter estimation with 18 batches of experimental data. The weights of objective functions are set as $w_1 = 20$, $w_2 = 10$, and $w_3 = 1$. There were seven estimable parameters to be solved. A total of 200 random starting points were used to achieve better global optimality, which are generated under uniform distribution within the predefined bounds. The maximum computation time for the optimization on each starting point was set to be $3.6 \times 10^3$ s. The average solving time for each starting point that successfully converged to a local extremum was around 800 s. Because the DAE system is highly nonlinear, some starting points may not return a solution within the maximum computation time or may even converge to an infeasible solution. In detail, 85 points converged to local optima, while the other 115 points resulted in non-convergence or infeasible solutions. Among the 85 converged solutions, the smallest objective value was $4.47 \times 10^3$, which is regarded as the approximated global optimum.

The local extremes and global minimum given by the 85 converged solutions are portrayed in Figure 7. The blue points indicate local optima, while the red dashed line represents the value of the global optimum. Most of the feasible local solutions guarantee weighted residuals of less than 50, which is close to the global optimum. Furthermore, the objective values of several local optimums are close to the global optimum as shown in

Figure 6. Scaled sensitivities $\left[ \frac{\partial}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right) \right]$ at parameter point $\theta^*_2$.

Figure 7. Local optima and global optimum given by the MSP approach.
Figure 7, which implies that the MSP strategy can indeed find the global optimum of this problem. The estimated results for the re-parameterized kinetic parameters are provided in Table 10. The reliability factor

\[ \beta_n = \frac{\Delta \theta_n}{2\hat{\theta}_n} \]

can be computed by the length of the CI and the estimates. Small reliability factors indicate narrow CIs, which further illustrate that the estimated result is close to the true value under a given confidence level of 90%. The activation energies and pre-exponential factors can be obtained via the value under a given confidence interval. In Figure 6, it is indicated that the reaction (i) is the limiting step, which is consistent with the sensitivity analysis shown in Figure 6. Note that \( \alpha_i \) and \( \beta_i \) are fixed.

The average relative error between the model prediction and the data was used to evaluate the overall fitting performance of the proposed model. This average error can be defined as

\[ \Delta y_f = \frac{1}{n_f} \sum_{m \in I} \left| \frac{y_{m,f} - y_{m}^*}{y_{m}^*} \right| \]

where \( \Delta y_f \) denotes the average relative mismatch between the \( f \)th fitted output \( y_{m,f} \) and the \( f \)th measured output \( y_{m}^* \) recorded in the \( m \)th experiment in dataset \( I \), which contains \( n_f \) batches of data.

Analogously, the largest mismatch is defined as

\[ (\Delta y_f)_{\text{max}} = \max_{m \in I} \left| \frac{y_{m,f} - y_{m}^*}{y_{m}^*} \right| \]

where \( (\Delta y_f)_{\text{max}} \) is the maximum relative gap between the fitted and measured values of \( f \)th output \( y_f \).

The overall data fitting performance is summarized in Table 11. The fitting of \( x_{MBT}(t_f) \) is very accurate, with an average mismatch of just 0.38% and the largest relative error of only 1.28%. The prediction of \( m_{xc}(t_f) \) and the reactor pressure is also satisfactory, with average biases of 0.95% and 1.66% and maximum deviations of 2.50% and 3.23%, respectively. It is clear that the proposed model is highly accurate and the estimated parameters are in good agreement with the known training data.

The fitting results of all experiments in the training set are plotted in Figures 8 and 9. In Figure 8, the blue dotted line denotes the fitted values of \( x_{MBT}(t_f) \), while the red dashed line represents the experimental data. The gray area represents the associated confidence region of \( x_{MBT}(t_f) \) under the parameter’s confidence interval. Obviously, the actual measurements are very close to the predicted \( x_{MBT}(t_f) \) within the corresponding confidence interval. In Figure 9, the blue dashed line with squares indicates the predictions of \( m_{xc}(t_f) \), and the red line denotes the actual values. Notwithstanding that the mismatches between the predictions and measurements of \( m_{xc}(t_f) \) are slightly higher than those for \( x_{MBT}(t_f) \), the fitting performance is still very encouraging since all the measured results are within the 90% confidence interval.

The pressure profiles of the fitted and experimental data in the training set are plotted in Figure 10. The red line denotes the pressure data and the blue line represents the pressure predictions. Note that the reaction duration varies in the different experiments. As shown in Figure 10, the model predictions match the real data in most experiments. Even in the batch with the largest deviation, the gap between the prediction and the measured value is still less than 0.3 MPa. Therefore, the proposed model for MBT production accurately predicts the output measurements.

### 4.5. Model Validation with Experimental Data.

Although our model has proven its ability to fit the data in the training dataset, it is essential to assess the prediction quality...
The average and largest mismatches of $x_{\text{MBT}}(t_f)$ between the model prediction and experimental data in the independent test set are just 0.46% and 0.87%, respectively. This result is very satisfying because the prediction mismatches are only slightly higher than with the training data. Indeed, the maximum prediction errors of $m_p(t_f)$, $x_{\text{MBT}}(t_f)$, and the reactor pressure in the test dataset are slightly smaller than with the training dataset. It is obvious that the overall results using the test dataset are consistent with the results using the training dataset, confirming the generality of the established model.

Detailed information about the model predictions and measurements of $x_{\text{MBT}}(t_f)$ is presented in Figure 11. The measurements (red dotted line) are highly consistent with the predictions (blue dashed line) and within the confidence interval (gray area). The estimation quality of the total weight of the crude product can be seen in Figure 12. The predicted mass is very close to the experimental data in most cases within the relative 90% confidence interval.

Furthermore, the predicted pressure curves are compared with the experimental data in Figure 13. The predicted pressures are highly coherent with the experimental data in most batches. Even in the case with the largest pressure deviation, namely, experiment no. 6, the gap between the prediction and the measurement is less than 0.33 MPa.

**4.6. Model Validation with Industrial Data.** The model established in this study was also applied to data from a real industrial batch process, and the gap between the predictions and measurements was examined. Note that the mass of crude product was not recorded in the industrial data.

The overall results are listed in Table 13. Detailed comparisons of the model predictions and industrial data are displayed in Figures 14 and 15. The average prediction errors of $x_{\text{MBT}}(t_f)$ and the reactor pressure are slightly higher than with the experimental data, which might be caused by the effect of scale-up and process uncertainties. For example, the real composition of S/CS$_2$ might derive from the nominal value as the S/CS$_2$ solution is mixed in an upstream mixing tank without an auto control system. Therefore, such mismatches are still acceptable in real production. Moreover, even in the industrial batch that produced the maximum prediction error, the difference in the predicted value of $x_{\text{MBT}}(t_f)$ does not exceed 1.58%, and the mismatch in the pressure is within 3.49%. Therefore, it can be concluded that the proposed model with
estimated parameters shows high prediction accuracy for both experimental and industrial MBT batch production processes.

5. CONCLUSIONS

In this work, a first-principles model of the MBT batch production process has been established based on the existing mechanism of the main reaction pathway and the lump treatment of minor side products. Because of the lack of information regarding the concentration profile, the estimable parameter set was selected through estimability analysis using the local sensitivity matrix. The MSP technique was applied to the estimation problem to improve the global performance and ensure precise model predictions. Moreover, independent unseen data from experimental and industrial batches were used to confirm the prediction performance of the proposed model. In general, the established model illustrated that the reaction (i) is the rate-determining step. In other words, more aniline in the discharged reactants can lead to a higher yield ratio of MBT. Such a conclusion is consistent with a real industrial process.

Currently, it is extremely important to increase profits and improve safety by automatically controlling the chemical production process based on a reliable model. The proposed model will be applied for operation optimization and closed-loop control in real industrial plants. Although the established model can describe the MBT production process with high accuracy, uncertainty cannot be avoided in process modeling and model reduction. Furthermore, other disturbances and measurement errors might also affect the real processes. Therefore, future work will focus on the dynamic optimization of the MBT production process under uncertainties. In addition, it is widely accepted in the chemical industry that MBT synthesis should be transformed from a batch process to a continuous process to enhance production efficiency and achieve better dynamic performance. Hence, formulating the continuous process model is an important task for future work. Optimization of the start-up manipulation and auto-control of...
the continuous process are other essential issues that should be considered in the near future.

■ AUTHOR INFORMATION

Corresponding Author
Zhihong Yuan — State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China; orcid:0000-0002-4680-8203; Email: zhihongyuan@mail.tsinghua.edu.cn

Authors
Enzhi Liang — State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
Song Zhang — China Sunsine Chemical Holdings Ltd., Shanxian City, Shandong 274300, China
Bin Liu — China Sunsine Chemical Holdings Ltd., Shanxian City, Shandong 274300, China
Bujin Qi — China Sunsine Chemical Holdings Ltd., Shanxian City, Shandong 274300, China
Yanpei Nie — China Sunsine Chemical Holdings Ltd., Shanxian City, Shandong 274300, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c06646

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Key Research and Development Program of Shandong Province, China, under grant no. 2020CXGC011205 and from China Sunsine Chemical Holdings Ltd.

■ REFERENCES

(1) Zhao, J.; Cheng, X.; Wang, L.; Wang, J.; Song, C. Hydrogenation of tar residue derived from the synthesis of rubber vulcanization accelerator 2-mercaptobenzothiazole. Asia-Pac. J. Chem. Eng. 2017, 12, 400–405.

(2) Van Der Horst, M.; Hendrikse, K.; Woolard, C. The role of thiol intermediates in 2-mercaptobenzothiazole accelerated sulfur vulcanization of rubber model compounds. J. Appl. Polym. Sci. 2003, 89, 47–54.

(3) Hansson, C.; Agrup, G. Stability of the mercaptobenzothiazole compounds. Contact Dermatitis 1993, 28, 29–34.

(4) Azam, M. A.; Suresh, B. Biological Activities of 2-Mercaptobenzothiazole Derivatives: A Review. Sci. Pharm. 2012, 80, 789–823.

(5) Wu, F.-L.; Hussein, W. M.; Ross, B. P.; McGearry, R. P. 2-Mercaptobenzothiazole and its derivatives: syntheses, reactions and applications. Curr. Org. Chem. 2012, 16, 1555–1580.

(6) Blakemore, P. R. The modified Julia olefination: alkene synthesis via the condensation of metallated heteroarylalkylsulfones with carbonyl compounds. J. Chem. Soc., Perkin Trans. 1 2002, 2563–2585.

(7) Borazjani, N.; Jarrahpour, A.; Rad, J. A.; Mohkam, M.; Behzadi, M.; Ghassemi, Y.; Mirzaeinia, S.; Karbalaei-Heidari, H. R.; Ghanbari, M. M.; Batta, G.; Turos, E. Design, synthesis and biological evaluation of some novel diastereoselective 1,l-β-lactams bearing 2-mercaptobenzothiazole and benzoquinoline. Med. Chem. Res. 2019, 28, 329–339.

(8) Srinivasan, B.; Palanki, S.; Bonvin, D. Dynamic optimization of batch processes: I. Characterization of the nominal solution. Comput. Chem. Eng. 2003, 27, 1–26.

(9) Wu, J. Production Technology of Accelerator M. Rubber Sci. Technol. 2013, 8, 13–17.

(10) Kelly, W. J. Process of manufacturing mercapto-benzo-thiazole. U.S. Patent No. 1631871, 1927, 06.

(11) Verma, A. K., Process Modelling and Simulation in Chemical, Biochemical and Environmental Engineering; CRC Press: 2014.

(12) Jana, A. K. Chemical Process Modelling and Computer Simulation; 2nd Edition PHI Learning Private Limited: 2011.

(13) Banga, J. R.; Balsa-Canto, E. Parameter estimation and optimal experimental design. Essays Biochem. 2008, 45, 195–210.

(14) Safdarnejad, S. M.; Gallacher, J. R.; Hedengren, J. D. Dynamic parameter estimation and optimization for batch distillation. Comput. Chem. Eng. 2016, 86, 18–32.

(15) Reynolds, M. P.; Malz, R. E., Jr. Method for the production of 2-mercaptobenzothiazole. US 6222041, 2001, 04.

(16) Englezos, P.; Kologerakis, N., Applied Parameter Estimation for Chemical Engineers; Marcel Dekker, New York and Basel, 2001.

(17) Seber, G. A. F.; Wild, G. J., Nonlinear Regression; Wiley: 1989.

(18) Almqvist, J.; Cvijovic, M.; Hatzimanikatis, V.; Nielsen, J.; Jirstrand, M. Kinetic models in industrial biotechnology - Improving cell factory performance. Metab. Eng. 2014, 24, 38–60.

(19) Bard, Y. Nonlinear Parameter Estimation; Academic press: New York and London, 1974.

(20) Espie, D. M.; Macchietto, S. Nonlinear Transformations for Parameter Estimation. Ind. Eng. Chem. Res. 1988, 27, 2175–2179.

(21) Jiménez-Villota, D. S.; Acosta-Pavas, J. C.; Betancur-Ramírez, K. J.; Ruiz-Corado, A. A. Modeling and Kinetic Parameter Estimation of the Enzymatic Hydrolysis Process of Lignocellulosic Materials for Glucose Production. Ind. Eng. Chem. Res. 2020, 59, 16851–16867.

(22) Jelemský, M.; Pakšiová, D.; Paulen, R.; Latifi, A.; Fikar, M. Combined Estimation and Optimal Control of Batch Membrane Processes. Processes 2016, 4, 43.

(23) Galirin, M.; Rojas, C.; Hjalmarsson, H. A weighted least-squares method for parameter estimation in structured models. In 53rd IEEE Conference on Decision and Control; IEEE: 2014, pp. 3322–3327.

(24) Smith, L. H.; Mccarty, P. L.; Kitanidis, P. K. Spreadsheet Method for Evaluation of biochemical Reaction Rate Coefficients and Their Uncertainties by Weighted Nonlinear Least-Squares Analysis of the Integrated Monod Equation. Appl. Environ. Microbiol. 1998, 64, 2044–2050.

(25) Fang, F.; Ni, B.-J.; Yu, H.-Q. Estimating the kinetic parameters of activated sludge storage using weighted non-linear least-squares and accelerating genetic algorithm. Water Res. 2009, 43, 2595–2604.

(26) Wang, Y.; Biegler, L. T.; Patel, M.; Wassick, J. Parameters estimation and model discrimination for solid-liquid reactions in batch processes. Chem. Eng. Sci. 2012, 187, 455–469.

(27) Kim, I.-W.; Liebman, M. J.; Edgar, T. F. Robust Error-in-Variables Estimation Using Nonlinear Programming Techniques. AIChE J. 1990, 36, 985–993.

(28) Kim, I.-W.; Liebman, M. J.; Edgar, T. F. Amoco Chemical Company, A sequential error-in-variables method for nonlinear dynamic systems. Chem. Eng. Sci. 1991, 46, 663–670.

(29) Zavala, V. M. Computational Strategies for the Optimal Operation of Large-Scale Chemical Processes; Carnegie Mellon University: 2008.

(30) Hendrix, E. M. T.; G-Tóth, B. a. Introduction to Nonlinear and Global Optimization; Springer: 2010.

(31) Bao, X.; Khajavirad, A.; Saindiniis, N. V.; Tawarmalani, M. Global optimization of nonconvex problems with nonlinear intermediates. Math. Prog. Comput. 2014, 7, 1–37.

(32) Yao, K. Z.; Shaw, B. M.; Kou, B.; McAuley, K. B.; Bacon, D. W. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. Polym. React. Eng. 2003, 11, 563–588.

(33) Kravaris, C.; Hahn, J.; Chu, Y. Advances and selected recent developments in state and parameter estimation. Comput. Chem. Eng. 2013, 51, 111–123.

(34) Box, G. E.; Draper, N. R., Response surfaces, mixtures, and ridge analyses; John Wiley & Sons: 2007, Vol. 649.

(35) Abdi, H.; Williams, L. J. Principal component analysis. WIRES Comput. Stat. 2010, 2, 433–459.

(36) Zhang, Y.; Edgar, T. F. PCA Combined Model-Based Design of Experiments (DOE) Criteria for Differential and Algebraic System Parameter Estimation. Ind. Eng. Chem. Res. 2008, 47, 7772–7783.
(37) Chu, Y.; Hahn, J. Quantitative Optimal Experimental Design Using Global Sensitivity Analysis via Quasi-Linearization. *Ind. Eng. Chem. Res.* 2010, 49, 7782–7794.

(38) Lund, B. F.; Foss, B. A. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* 2008, 44, 278–281.

(39) Bergfeld, M. J.; Gutlein, N.; Wohlfahrt, K.; Aust, E. Process for the preparation of 2-mercaptobenzothiazole. 1994, 11, 5367082.

(40) Buzzi-Ferraris, G.; Manenti, F. Kinetic models analysis. *Chem. Eng. Sci.* 2009, 64, 1061–1074.

(41) Shao, Y.; Zhang, H.; Yan, Y. Least squares estimation of kinetic parameters in batch adsorption of phenol with confidence interval analysis. *Korean J. Chem. Eng.* 2013, 30, 1544–1551.

(42) Beck, J. V.; Arnold, K. J. Parameter estimation in engineering and science; James Beck: 1977.

(43) Diehl, M.; Bock, H. G.; Kostina, E. An approximation technique for robust nonlinear optimization. *Math. Prog.* 2005, 107, 213–230.

(44) Jayasankar, B. R.; Ben-Zvi, A.; Huang, B. Identifiability and estimability study for a dynamic solid oxide fuel cell model. *Comput. Chem. Eng.* 2009, 33, 484–492.

(45) Nagy, Z. K.; Braatz, R. D. Robust nonlinear model predictive control of batch processes. *AIChE J.* 2003, 49, 1776–1786.

(46) Liu, J.; Gnanasekar, A.; Zhang, Y.; Bo, S.; Liu, J.; Hu, J.; Zou, T. Simultaneous State and Parameter Estimation: The Role of Sensitivity Analysis. *Ind. Eng. Chem. Res.* 2021, 60, 2971–2982.

(47) Sargent, R. W. H.; Sullivan, G. R. Development of feed changeover policies for refinery distillation units. *Ind. Eng. Chem. Process Des. Dev.* 1979, 18, 113–124.

(48) Wächter, A.; Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Prog. Study* 2006, 106, 25–57.

(49) Zhigljavsky, A.; Zilinskas, A., *Stochastic Global Optimization*; Springer: 2008.

(50) Villaverde, A. F.; Frohlich, F.; Weindl, D.; Hasenauer, J.; Banga, J. R. Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics* 2019, 35, 830–838.

(51) Nauman, E. B. *Chemical reactor design, optimization, and scaleup*; John Wiley & Sons: 2008.