

Dynamic response surface methodology using Lasso regression for organic pharmaceutical synthesis

Yachao Dong^{1,2}, Christos Georgakis (✉)², Jacob Santos-Marques², Jian Du¹

¹ Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, China

² Department of Chemical and Biological Engineering and Systems Research Institute, Tufts University, Medford, MA 02155, USA

© Higher Education Press 2021

Abstract To study the dynamic behavior of a process, time-resolved data are collected at different time instants during each of a series of experiments, which are usually designed with the design of experiments or the design of dynamic experiments methodologies. For utilizing such time-resolved data to model the dynamic behavior, dynamic response surface methodology (DRSM), a data-driven modeling method, has been proposed. Two approaches can be adopted in the estimation of the model parameters: stepwise regression, used in several of previous publications, and Lasso regression, which is newly incorporated in this paper for the estimation of DRSM models. Here, we show that both approaches yield similarly accurate models, while the computational time of Lasso is on average two magnitude smaller. Two case studies are performed to show the advantages of the proposed method. In the first case study, where the concentrations of different species are modeled directly, DRSM method provides more accurate models compared to the models in the literature. The second case study, where the reaction extents are modeled instead of the species concentrations, illustrates the versatility of the DRSM methodology. Therefore, DRSM with Lasso regression can provide faster and more accurate data-driven models for a variety of organic synthesis datasets.

Keywords data-driven modeling, pharmaceutical organic synthesis, Lasso regression, dynamic response surface methodology

1 Introduction

Automation and computation have contributed substan-

tially to different aspects of chemical sciences, and will continue to make a great impact [1–4]. In the pharmaceutical industry, the availability of high throughput automatic experimentation platforms have enabled multiple experiments to be run in parallel under different conditions [5]. The data-rich experimental environment is crucial for the improvement of quality, safety, speed, and cost efficiency in organic synthesis [6,7]. This brings many advantages in conducting multi-factorial experiments for process characterization and optimization in process development. There are different multivariate tools for pharmaceutical process development and innovation [8,9]. One of the most important tools is the design of experiments (DoE) [10], a method that systematically decides the operating conditions of a set of experiments and maximizes the amount of information that is derived. One of the major advantages of such experimental platforms is the robotic collection of reaction samples at different time instants during each experiment. Detailed analysis of such samples avails a host of time resolved concentration data or chromatographic areas of not only the reactants and products but also the intermediates and byproducts, some of which might be of unknown chemical composition. Before the dynamic response surface methodology (DRSM) modeling approach was postulated, a series of response surface methodology (RSM) models would have to be estimated, one for each of the time instants in which composition measurements were collected. Instead of developing a RSM model for each time instant that data were collected, we have proposed the DRSM that can model all time resolved measurements in all experiments [11–13]. To overcome the same challenge, Domagalski et al. used a strategy that involved the filtering of the data [14], the use of the logarithmic transformed time as a DoE factor, and a genetic algorithm in building the model. Wang et al. proposed a semiparametric model for an ester hydrolysis reaction [15], and compared the mechanistic models and statistical models.

The main theme of the present paper is to examine the use of Lasso regression as a replacement of the stepwise regression (SWR) for eliminating the insignificant parameters of the DRSM model, in terms of both modeling accuracy and computational cost. The value of Lasso regression is well known in the machine learning community, as it induces sparsity in an identified model by using a L_1 norm regularization term [16]. Compared to SWR, the computational time to estimate an accurate modeling using Lasso regression is much faster and thus more attractive, as we will show below. This is especially true when the Lasso algorithm uses the alternating direction method of multipliers (ADMM) [17].

Two organic synthesis problems in pharmaceutical process development will be studied, using the datasets from literature. The overall workflow of the experiment conduction and the model calculation follows four steps: (1) factor identification, (2) DoE, (3) experimentation, and (4) DRSM modeling and further analysis (including optimization, next round of DoE, etc.). We will first study an organic synthesis problem presented by Domagalski et al. in 2015 [14]. We will show how Lasso can provide DRSM models of the time-resolved concentrations of different species with small computational time and great accuracy. We will use the DRSM models of several species to optimize the process operation in order to minimize the usage of catalyst, while the achieved product satisfies the minimum yield constraint. Different types of uncertainty, including deviations from the exact optimal process conditions as well as the uncertainties inherent in the DRSM models, quantified by the associated prediction intervals, have been considered in the optimization. The former type of uncertainty, operational sensitivity, is very important in the pharmaceutical sector for an effective implementation of the optimal conditions [18]. For the latter type of uncertainty, we use the nominal predictions of the DRSM model as well as the related prediction intervals to avoid failing to achieve the desired product quality by model uncertainties, however small they might be. Another important utility of the DRSM models is in identification of the active stoichiometry in the reaction mixture and the subsequent estimation of kinetic models. Most of these tasks have been presented in previous publications [19,20], and are omitted in this article due to space limitation.

The second case study examined here is based on the paper by Dong et al. [20]. Instead of modeling the concentration profiles, as done in the first case study, we will model the extents of reaction to demonstrate the flexibility of the DRSM modeling approach. The extent data are obtained by transforming the concentration data through the assumed stoichiometry of the present reactions. The advantage here is that one needs to estimate a smaller number of DRSM models than those for the measured species, because there are fewer reactions than

the number of measured species. Here again, we observe that Lasso provides similar computational efficiency compared to SWR. This will be very valuable when the chemistry is complex with the exact stoichiometry unknown, and one needs to examine a large set of candidate stoichiometries.

The main contribution of this paper is that DRSM model is newly integrated with Lasso regression, which provides fast and accuracy modeling for the dynamic response data that appears in process development of organic synthesis. The low computational cost of DRSM with Lasso and its accuracy will be critical for multiple usages of the algorithm in modeling the dynamic data of the organic synthesis in pharmaceutical industry, such as automatic stoichiometry identification, process optimization and control.

2 Method and in silico experiment

In this section, we first present a brief description of the DRSM methodology. Next, we summarize the Lasso regression used in estimating the DRSM models. Finally, we present the settings of two case studies, respectively for modeling concentrations and reaction extents.

2.1 DRSM

RSM has been widely used to estimate data-driven models for DoE. If we take its quadratic form as an example, the RSM model is:

$$y(\mathbf{x}) = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \beta_{ij} x_i x_j + \sum_{i=1}^n \beta_{ii} x_i^2, \quad (1)$$

where y , the output variable, is modeled by a polynomial interpolative equation which consists of a constant term, β_0 , plus the linear terms $\beta_i x_i$, the two-factor interaction (2FI) terms $\beta_{ij} x_i x_j$ and the quadratic terms $\beta_{ii} x_i^2$. If the data set is not numerous enough, a model with fewer terms is estimated. All these terms are functions of the input factors $\mathbf{x} = (x_1, \dots, x_n)$, where n denotes the number of input factors. In Eq. (1), the β_q ($q = 0, i, ij, ii$) are model parameters that need to be estimated from the data. The DRSM model is a generalization of the above RSM to handle time-resolved output data, $y(t_k)$. The DRSM generalization of Eq. (1) requires each parameters β_q to become a parametric function of time $\beta_q(t)$. For mathematical convenience one uses a normalized time, τ , instead of t , defined with respect to the batch duration, t_b ; $\tau = t/t_b$.

The initial version of DRSM [11], now denoted by DRSM-1, uses this normalized time, τ , as the independent variable in describing the $\beta_q(\tau)$ parametric functions:

$$\begin{aligned}
y(\mathbf{x}, \tau) = & \beta_0(\tau) + \sum_{i=1}^n \beta_i(\tau)x_i \\
& + \sum_{i=1}^n \sum_{j=i+1}^n \beta_{ij}(\tau)x_i x_j \\
& + \sum_{i=1}^n \beta_{ii}(\tau)x_i^2,
\end{aligned} \quad (2)$$

where both output y and each $\beta_q(\tau)$ are function of the normalized time. Each $\beta(\tau)$ function is parameterized by a linear combination of shifted Legendre polynomials:

$$\begin{aligned}
\beta_q(\tau) = & \sum_{r=0}^R \gamma_{qr} P_r(\tau), \quad q = 0, i, ij, ii \\
\text{with } & i = 1, 2, \dots, n; j = i + 1, \dots, n,
\end{aligned} \quad (3)$$

in which $P_r(\delta)$ denotes the shifted Legendre polynomial of order r , and the γ_{qr} coefficients are the DRSM parameters whose values are estimated from the data. The first three shifted Legendre polynomials are as follows:

$$P_0(\tau) = 1; P_1(\tau) = -1 + 2\tau; P_2(\tau) = 1 - 6\tau + 6\tau^2. \quad (4)$$

A weakness of this model is that as R increases, the estimated model exhibits an undesirable oscillatory behavior at large times. To eliminate this behavior and to achieve great model accuracy with small values of the global parameter R , a second version of DRSM model has been developed [12], denoted as DRSM-2. In DRSM-2, θ , an exponential transformation of time, is used as the temporal independent variable:

$$\theta = 1 - \exp(-t/t_c), \quad (5)$$

where t_c is a time constant representing the slowest component of the apparent dynamics in the process. As a result of this transformation, the DRSM-2 model is described by the following two equations:

$$\begin{aligned}
y(\mathbf{x}, \theta) = & \beta_0(\theta) + \sum_{i=1}^n \beta_i(\theta)x_i \\
& + \sum_{i=1}^n \sum_{j=i+1}^n \beta_{ij}(\theta)x_i x_j \\
& + \sum_{i=1}^n \beta_{ii}(\theta)x_i^2,
\end{aligned} \quad (6)$$

$$\beta_q(\theta) = \sum_{r=0}^R \gamma_{qr} P_r(\theta), \quad q = 0, i, ij, ii$$

$$\text{with } i = 1, 2, \dots, n; j = i + 1, \dots, n, \quad (7)$$

where the γ_{qr} coefficients are called the *local* parameters and appear linearly in the model. On the other hand, the DRSM-2 model has two *global* parameters: the maximum order R , taking only integer values, and the time constant t_c , which has a non-linear effect on the model.

The above discussion concerns the modeling of single response. If multiple responses are measured, e.g., concentrations of n_s species, one develops n_s DRSM-2 models. Furthermore, a constrained variation of the

DRSM-2 algorithm has been introduced [13], where the initial concentration at $t = 0$ are fixed to their known values and all model predictions are forced to be non-negative for problems where concentrations data are modeled. Several values of R , the first global parameter, are explored along with a set of values for t_c , the second global parameter, and the γ_{qr} values are estimated through linear regression. For each R value, the best t_c value is refined through nonlinear regression. Once t_c , for the current R , is fixed, SWR or Lasso regression is employed to eliminate the insignificant γ_{qr} parameters and the corresponding Bayesian information criterion (BIC) value is recorded. The R value, and the corresponding model, with the smallest BIC values is then selected. This model is further refined by the imposition of the constraints on the known initial concentration values and the requirement that all model predictions, related to species concentrations, should be non-negative. For the data where the unit is in percentage (which might be so in reality), we have also considered the constraint to enforce the summation of multiple components to be one, as the independently calculated responses cannot reach exactly one hundred percent. This constraint is important, when (1) all the species could be measured and (2) the data is presented in unit of percentage. However, we do not include such a constraint here because very often there are impurities whose concentrations cannot be measured.

For the estimation of the t_c parameter, we initially try different values, followed by a local non-linear regression for fine-tuning, because the regression of the non-convex model would have encountered challenges on global optima. The elimination of the insignificant γ_{qr} parameters has been achieved through SWR in our previous publications [11–13]. In the present paper, we also use Lasso regression, and we will demonstrate that it reduces the computational effort significantly. Algorithm 1 summarizes the steps employed in the development of the final constrained DRSM-2 model. The previous publication [13] provides more details on the steps involving time constant t_c and constraints (steps 2, 3, 5), while the steps involving Lasso regression and BIC values (steps 4, 6) are presented in the next section of this article. The order of the polynomials is constrained by the number of samples each experiment has; while the form of the DRSM model (linear, 2FI, or quadratic, etc.), relating input to output variables, are constrained by the degree of freedom the DoE offers [13]. However, no specific type of experimental design is required for the DRSM algorithm.

2.2 Lasso regression

We use the bold letter $\boldsymbol{\gamma}$ to denote the vector of all local γ_{qr} parameters with different indices q and r . The simplest way to estimate these local parameters is through the following least square linear regression problem:

Algorithm 1. Procedures in calculating DRSM models

1. For a set of polynomial orders, $R = 1 : R_{\max}$
2. For a set of t_c values, perform linear regression for γ_{qr} and select the t_c that results in the smallest sum of squared errors.
3. Fine-tune t_c via nonlinear regression and fix.
4. Eliminate the insignificant local parameters γ_{qr} , by SWR or Lasso.
5. Re-estimate the significant γ_{qr} parameters by imposing the regression constraints.
6. Tabulate the BIC values.
7. Select R , and the corresponding model, with the smallest BIC value.

$$\min_{\gamma} \left\{ \sum_{i=1}^{n_e} \sum_{k=1}^{m_i} (y_{ik} - \hat{y}_{ik}(\gamma))^2 \right\}, \quad (8)$$

where n_e denotes the number of experiments and m_i denotes the number of measurements in experiment i . In the squared summation terms, y_{ik} denotes the k^{th} measurement in i^{th} experiment, and \hat{y}_{ik} denotes the corresponding DRSM model's prediction, which is a function of the γ parameters. Because linear regression might lead to overfitting, especially for the higher values of R , we have used SWR to eliminate the insignificant parameters in the past. Here we are examining the possible advantages that Lasso regression [21], as the preferred regularization approach, can offer to avoid overfitting. With any regularization approach, a penalty term, $P(\gamma)$, is added to the objective function, and the following minimization is used instead of that in Eq. (8):

$$\min_{\gamma} \left\{ \sum_{i=1}^{n_e} \sum_{k=1}^{M_i} (y_{ik} - \hat{y}_{ik}(\gamma))^2 + \lambda P(\gamma) \right\}, \quad (9)$$

where, λ is the regularization weight parameter. In Lasso regression, the penalty term is the L_1 norm of the vector with all the parameter values, $P(\gamma) = \|\gamma\|_1$. Lasso regression has a well-known sparsity property pushing many of the less significant parameters to a zero value. The alternative L_2 norm penalty, known as Ridge regression, does not provide the sparsity that the L_1 does. Lasso regression can be solved through different optimization methods, and it will be solved using ADMM, a computationally-efficient dynamic programming algorithm based on dual decomposition and augmented Lagrangian [17]. As we will show later, the main attraction of Lasso compared to SWR in the DRSM model estimation is its computational cost. We use the BIC to choose the regularization parameter λ , which is defined as follows.

$$BIC = -2\ln Pr(\mathbf{y}|\hat{\mathbf{y}}) + R_S \ln K_T. \quad (10)$$

In this equation, both \mathbf{y} and $\hat{\mathbf{y}}$ are vectors, and $Pr(\mathbf{y}|\hat{\mathbf{y}})$ is the likelihood function, denoting the probability of observing measurement set \mathbf{y} given that the DRSM model parameter set is estimated to be $\hat{\mathbf{y}}$. K_T is the number of measurements, while R_S is the number of significant $\hat{\mathbf{y}}$ parameters. The log likelihood value in the first term is calculated by the

equation below:

$$\ln Pr(\mathbf{y}|\hat{\mathbf{y}}) = -\frac{1}{2}(\mathbf{y} - \hat{\mathbf{y}})^T \frac{1}{\hat{\sigma}^2} (\mathbf{y} - \hat{\mathbf{y}}) - \frac{K_T}{2} \ln(\hat{\sigma}^2) - \frac{K_T}{2} \ln(2\pi). \quad (11)$$

Here, $\hat{\mathbf{y}}$ denotes the DRSM-estimated values of the output, while $\hat{\sigma}^2$ denotes the estimated variance. The use of Lasso and BIC together helps to build a model with accuracy and sparsity. The Akaike information criterion (AIC) as well as the generalized cross validation (GCV) approach have been tested too, but they do not provide as consistent results as the BIC approach does, which will be discussed in more detail later in Section 3.1.

After Lasso regression completes its task, we fine-tune the resulting model by an additional step applied to the γ parameters that Lasso has retained. Analysis of variance (ANOVA) calculation is performed to test if some of the retained γ parameters are insignificant [22]. Here we use a threshold p -value of 0.05. If such insignificant parameters are still present, they are removed.

2.3 Case study A: modeling of concentrations

First, we examine the ability of the DRSM model to accurately model the datasets related to an organic synthesis reaction, examined by Domagalski et al. [14], which was the first paper, to the best of our knowledge, to model dynamic composition data but used a rather complex modeling approach. In the present paper, we test DRSM modeling methodology against the data of Domagalski et al. [14]. We also compare Lasso and SWR with respect to two criteria: speed of calculations and accuracy of the obtained models. Because Domagalski et al. presented three rounds of DoE, the Lasso vs. SWR comparison is performed with all such data sets. In addition, we use our estimated DRSM models to solve an optimization task defined by Domagalski et al. Furthermore, we demonstrated how some robust optimization versions of the problem can be easily tackled to take uncertainty into consideration.

The case study here relates to a reaction network involving ten species in total, and therefore, there are ten

responses: two starting materials, one intermediate, one product, one catalyst, four impurities generated by the reactions, and water. The initial amount of the starting material 1 is fixed over different experiments, and all of the units of the output responses are provided in equivalent (equiv) mol of the starting material 1. The data are obtained from a simulated model consisting of a set of ordinary differential equations, representing five irreversible and one reversible reaction. Some of the reaction rates are expressed in the Michaelis-Menten form. There are six factors: initial amount of starting material 2, catalyst amount, solvent volume, temperature, trace amounts of water, and mixing speed. Three rounds of DoE are implemented, and each round includes 20 experiments, following a 1/4 fractional factorial design with four replicates at the center. More details can be found in the paper by Domagalski et al. [14].

2.4 Case study B: modeling of reaction extents

Another case study is performed to assess the DRSM obtained with Lasso regression but from a different point of view. This dataset is from Dong et al. 2019 [13], a second organic synthesis problem. The previous publications have shown that this concentration dataset can be well modeled through DRSM with SWR. Similar to case study A, this concentration dataset can be modeled through DRSM with Lasso regression too, which will be omitted here. In order to illustrate that the DRSM methodology is versatile, we model the reaction extents, instead of species concentrations for this case study.

This case study is based on the simulation of a complex reaction network, involving ten species and eight reactions (six reversible and two irreversible). There are three factors involved in the DoE: temperature, initial concentration of species B, and initial concentration of species D. The DoE follows a central composite design with three replicates at the center. Previously, it has been shown that seven, out of eight, reactions can be successfully identified, and the one reaction that was missing has very small reaction extent, leading to concentration change on the same level as the noise in the system [20]. Based on the identified stoichiometries, one can transform the concentration data in experiment i to the reaction extent, as the following formulation shows:

$$\mathbf{z}_i = \Delta \mathbf{c}_i \mathbf{N}_{\text{ID}}^+ \quad (12)$$

In this equation, $\Delta \mathbf{c}_i$ is a $n_T \times n_S$ matrix, and n_T and n_S represent the number of composition samples and measured species, respectively. The j -th row and k -th column of $\Delta \mathbf{c}_i$ is the concentration change of species k at time point j compared to its initial concentration. Matrix \mathbf{N}_{ID} is of dimensions $n_R \times n_S$, representing the identified reaction stoichiometries in the network. Here n_R is the number of independent reactions; while \mathbf{N}_{ID}^+ denotes

the pseudo-inverse of \mathbf{N}_{ID} . Note that \mathbf{N}_{ID} denote the estimated stoichiometries which do not necessarily have to be the true ones in the reaction network. Matrix \mathbf{z}_i is of dimensions $n_T \times n_R$. This transformation will project the time-resolved concentration data of n_S species into a reaction extent space of n_R reactions.

3 Results and discussion

In this section, we first present the results of case study A in detail. This includes the obtained DRSM models for the concentration data, the computational times and model accuracies obtained using Lasso or SWR for different datasets, and the results of process optimization using the DRSM models. Finally, we present the results of case study B, where the reaction extents are modeled.

3.1 Case study A

3.1.1 Results of using Lasso and its comparison with SWR

The Domagalski data relate to 60 experiments. The sixth factor, mixing speed, was shown to be insignificant and is ignored. Because of the plethora of experiments, one can easily estimate a quadratic type of DRSM model. Ten such models are developed, one for each species. When using Lasso, the value of the regularization weight λ is selected so that the BIC value is minimized. Two other approaches were used besides BIC. The AIC is often considered as an alternative to BIC and the GCV is very often used with Lasso and other regularization techniques [23]. As seen in Fig. 1, the BIC values clearly show a minimum versus the λ values, and this helps us to choose the most desirable λ value. No clear minimum is observed in the AIC and GCV plots against the λ values, shown in the two left subplots. The minimum BIC value model of -1344 is obtained at $\lambda = 4 \times 10^{-4}$ and the resultant model has 44 local γ parameters. This exemplary result shown in Fig. 1 relates to the product species. There are consistent minimum values for the BIC criterion in the corresponding plots for other species, which is not the case with the other two criteria: AIC and GCV.

Continuing to focus on the product species, we compare Lasso against SWR in Fig. 2. The red solid line shows the nominal predictions of the SWR, and the green dot-dashed line corresponds to the Lasso model predictions. We show only half of the 60 experiments, just the odd numbered ones; experiment 45 is enlarged as a magnified example. The difference between the two models is clearly seen to be insignificant. The dashed line shows the prediction intervals [13] from Lasso, which are similar to those from SWR. Moreover, we look at the parameters retained by both techniques. The global parameters were the same for both approaches, $R = 2$, $t_c = 2.18$ h. The local

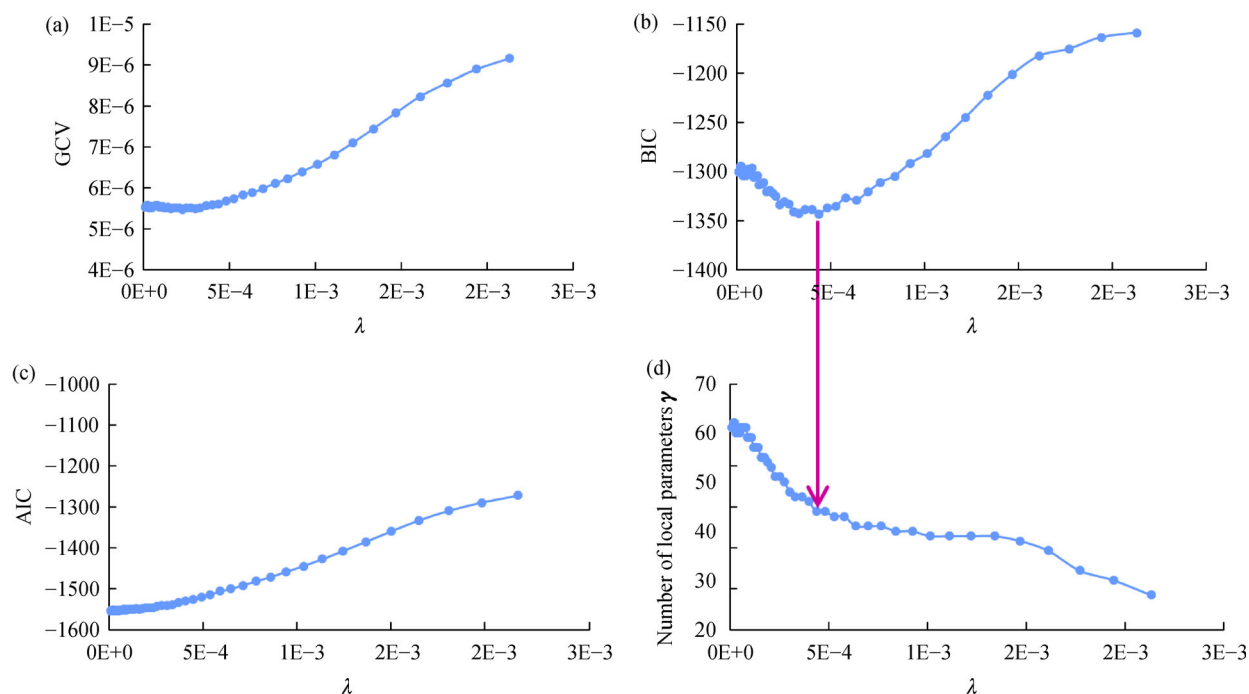


Fig. 1 The dependence of three criteria: (a) GCV, (b) BIC and (c) AIC, on the Lasso weight λ , and (d) the number of local parameters. The BIC plot shows a clear minimum while the other two criteria fail to show such a minimum. For the BIC case, the number of local parameters γ vs. λ is given in the lower right subplot.

parameters, γ , are summarized in Tables 1 and 2, where a blank cell means that the corresponding γ_{qr} is zero. The five factors, x_1 to x_5 , correspond to temperature, solvent volume, catalyst amount, starting material 2 amount, and trace amount of water, in their coded form (varying from -1 to 1). Both Lasso and SWR include 36 local parameters. For most of the terms, both techniques select the same significant orders. For example, considering all the constants, linear and quadratic terms, only x_2 and x_3 effects have different orders in θ using Lasso and SWR. For the cross terms, there are more terms with different orders, but the maximum difference of the coefficient is 0.032 (for order 1 of term x_2x_4), which is only 3% of the maximum coefficient 0.950 (for order 0 of the constant term). Therefore, from both the fitting results and the detailed analysis of model parameters, Lasso and SWR result in very similar DRSM models.

Moreover, we compare the fitting results of all species using Lasso vs. using SWR. The results are summarized in Table 3. The species from S1 to S10 respectively represent (1) catalyst, (2) starting material 1, (3) the intermediate reversibly formed by catalyst and starting material, (4) starting material 2, (5) product, (6) impurity 1, (7) water, (8) impurity 2, (9) impurity 3, and (10) impurity 4. Both techniques lead to the same global parameters over all ten species. Number of significant local parameters, γ , are similar, with the maximum difference of 4. The root mean-square error (RMSE) of the fitted data are also given, and the difference of using Lasso and SWR is less than 10% at

most. The species S5 (product), discussed in the previous paragraph, is a representative species with a RMSE difference of 6%. The table also shows the ratios of the RMSE of Lasso fitting and maximum value of the output, which is between 1% and 7%; the species with large concentrations, such as S2, S4 and S5, also have large absolute values of RMSE.

Therefore, we conclude that the SWR and Lasso do not have significant difference in terms of the calculated DRSM models. However, Lasso has the advantage of much faster solution time. Considering the time consumed in the step of determining the significant local parameters, SWR takes 3.3 s on average, while Lasso takes 0.029 s. Such improvement is valuable, because local parameters need to be determined not only for different responses, but also for different R values. We use both forward and backward SWR [23], and ADMM for Lasso [17].

3.1.2 Models from different datasets

The above models were estimated using all the available data in the paper by Domagalski et al. [14]. This will be denoted as the “Full” data set. However, the data set of 60 experiments were obtained from three rounds of DoE. We now examine the DRSM models that could have been estimated using the different data subsets. We also present a comparison among these models. In each round of DoE, the upper and lower bounds in the factor space were

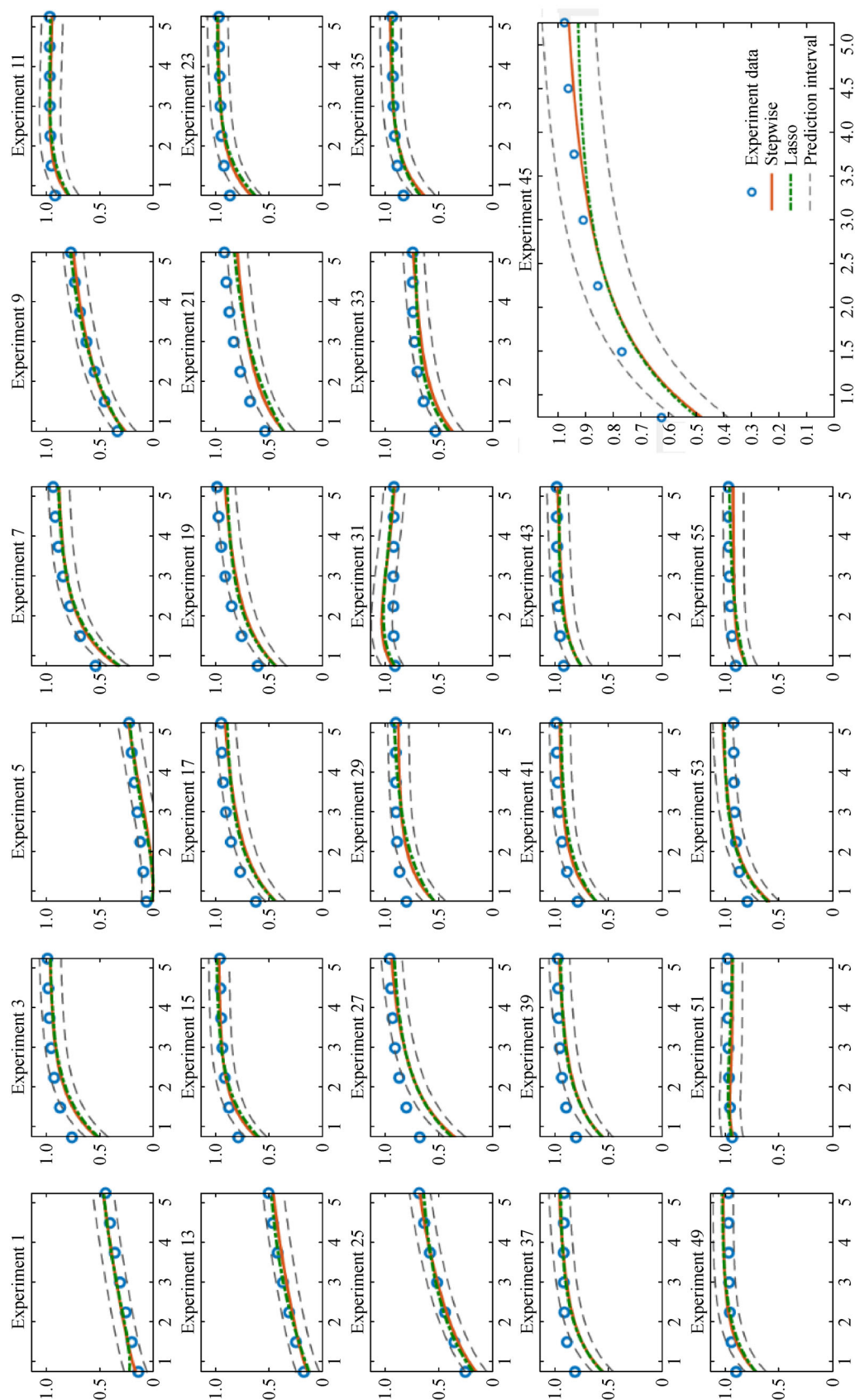


Fig. 2 Comparison of fitting results of using Lasso and SWR for the product species. Time is shown on x-axis with unit of hour, while concentration is on y-axis with unit of equiv.

Table 1 Comparison of γ values for intercept, linear, and quadratic terms estimated through SWR and Lasso for the product species ^{a)}

| Item | Order | $\beta_0(\theta)$ | Linear term | | | | | Quadratic term | | | | |
|---------------------|-------|-------------------|-------------|--------|--------|-------|--------|----------------|---------|---------|---------|---------|
| | | | x_1 | x_2 | x_3 | x_4 | x_5 | x_1^2 | x_2^2 | x_3^2 | x_4^2 | x_5^2 |
| Lasso | 0 | 0.950 | 0.100 | −0.084 | 0.222 | 0.028 | −0.031 | −0.044 | −0.045 | −0.220 | −0.034 | −0.059 |
| | 1 | 0.162 | −0.137 | 0.047 | −0.118 | | | −0.075 | | | | 0.026 |
| | 2 | −0.087 | | | −0.018 | | | | | 0.070 | | |
| Stepwise regression | 0 | 0.945 | 0.105 | −0.091 | 0.216 | 0.041 | −0.035 | −0.036 | −0.047 | −0.222 | −0.035 | −0.052 |
| | 1 | 0.180 | −0.141 | 0.045 | −0.109 | | | −0.072 | | | | 0.033 |
| | 2 | −0.067 | | 0.014 | | | | | | 0.073 | | |

a) A blank cell implies parameter is zero (Lasso) or insignificant (SWR).

Table 2 Comparison of γ values for 2FI terms estimated through SWR and Lasso for the product species ^{a)}

| Item | Order | x_1x_2 | x_1x_3 | x_2x_3 | x_1x_4 | x_2x_4 | x_3x_4 | x_1x_5 | x_2x_5 | x_3x_5 | x_4x_5 |
|---------------------|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Lasso | 0 | 0.020 | −0.071 | 0.011 | −0.017 | 0.041 | 0.042 | −0.021 | | −0.028 | 0.045 |
| | 1 | | −0.131 | 0.037 | | | | −0.021 | −0.039 | 0.020 | 0.043 |
| | 2 | | 0.041 | | | | | | | | |
| Stepwise regression | 0 | 0.018 | −0.069 | | | 0.059 | 0.064 | −0.021 | −0.019 | −0.043 | 0.054 |
| | 1 | | −0.125 | 0.027 | | −0.032 | | −0.023 | −0.027 | 0.039 | 0.060 |
| | 2 | | 0.049 | | | | | | | | |

a) A blank cell implies parameter is zero (Lasso) or insignificant (SWR).

changed. The first round consisted of 20 experiments, 16 for a fractional factorial design and 4 replicates at the center point. The data set will be denoted by “Rd1”. In this case, there are not enough distinct experiments for a quadratic model. Instead, a 2FI model is estimated, considering all possible 2FI terms. In the second DoE round, 20 more experiments were added, bringing the number of available experiments at that time to 40. The quadratic DRSM model using these data will be denoted by “Rd12”.

The estimated values for the two global parameters, R , t_c , for the three datasets are very similar. In fact, the maximum polynomial order, R , for the additional models, “Rd1” and “Rd12” is the same as the one shown in Table 3,

for each of the species. The best t_c value varies slightly between data sets for the same species, but the differences are less than 30%. The comparison of the number of parameters and the RMSE are summarized in Table 4; all of these models were obtained by using Lasso. Models based on the “full” and “Rds12” data sets have very similar values for the number of local parameters γ , while the model based on the “Rd1” data set, has fewer γ parameters because it lacks the five quadratic terms. All three models have comparable values for RMSE when this measure is calculated from the data set through which the model is estimated. This is reported as RMSE_a in Table 4. We also calculated the RMSE for the Rd1 and Rd12 models against the data for all 60 experiments. This is denoted as RMSE_b,

Table 3 DRSM model summary of SWR and Lasso for different species

| Species | R | t_c | Number of variables γ | | RMSE (in unit of equiv) | | | RMSE/max (Lasso) |
|---------|-------------|-------|------------------------------|-------|-------------------------|--------|----------------|------------------|
| | SWR & Lasso | | SWR | Lasso | SWR (ref.) | Lasso | Δ Lasso | |
| S1 | 2 | 1.77 | 37 | 37 | 0.0059 | 0.0062 | 3.8% | 3.8% |
| S2 | 2 | 2.18 | 28 | 31 | 0.0639 | 0.0702 | 9.9% | 6.5% |
| S3 | 2 | 1.77 | 38 | 42 | 0.0061 | 0.0065 | 7.2% | 6.1% |
| S4 | 2 | 2.18 | 28 | 30 | 0.0530 | 0.0570 | 7.4% | 4.4% |
| S5 | 2 | 2.18 | 36 | 36 | 0.0463 | 0.0493 | 6.4% | 4.5% |
| S6 | 1 | 4.24 | 28 | 27 | 0.0035 | 0.0035 | −0.1% | 5.5% |
| S7 | 1 | 2.59 | 22 | 24 | 0.0088 | 0.0089 | 1.4% | 1.7% |
| S8 | 2 | 2.59 | 22 | 20 | 0.0084 | 0.0085 | 1.7% | 6.1% |
| S9 | 2 | 4.24 | 28 | 24 | 0.0005 | 0.0005 | −0.7% | 6.2% |
| S10 | 2 | 4.24 | 26 | 28 | 0.0002 | 0.0003 | 3.4% | 3.2% |

Table 4 Comparison of models estimated from different datasets in case study A

| Species | Number of γ | | | RMSE _a (10^{-2} equiv) | | | RMSE _b (10^{-2} equiv) | | | |
|---------|--------------------|-----|-------|--------------------------------------|------|-------|--------------------------------------|--------------|-------|---------------|
| | Full | Rd1 | Rds12 | Full | Rd1 | Rds12 | Rd1 | Δ Rd1 | Rd12 | Δ Rd12 |
| S1 | 27 | 26 | 31 | 0.62 | 0.54 | 0.41 | 1.25 | 102% | 1.24 | 101% |
| S2 | 28 | 12 | 29 | 7.02 | 9.17 | 6.46 | 17.96 | 156% | 14.38 | 105% |
| S3 | 28 | 24 | 24 | 0.65 | 0.50 | 0.49 | 1.30 | 99% | 1.22 | 87% |
| S4 | 28 | 19 | 31 | 5.70 | 7.42 | 4.53 | 15.48 | 172% | 14.71 | 158% |
| S5 | 36 | 19 | 36 | 4.93 | 7.35 | 4.35 | 22.48 | 356% | 16.23 | 229% |
| S6 | 28 | 24 | 27 | 0.35 | 0.13 | 0.43 | 0.77 | 123% | 0.40 | 14% |
| S7 | 22 | 20 | 24 | 0.89 | 0.87 | 0.99 | 1.97 | 122% | 0.97 | 9% |
| S8 | 22 | 17 | 19 | 0.85 | 0.84 | 0.93 | 1.89 | 121% | 1.03 | 20% |
| S9 | 28 | 28 | 24 | 0.05 | 0.03 | 0.06 | 0.08 | 55% | 0.05 | 10% |
| S10 | 26 | 27 | 26 | 0.03 | 0.02 | 0.03 | 0.06 | 152% | 0.03 | 17% |

in Table 4. These numbers are naturally larger than the RMSE_a values for the same model. In the two shaded columns of Table 4, we report two percentage differences in RMSE values, Δ Rd1 and Δ Rd12, when the predictions of different models are tested against the full set of data from the 60 experiments. Δ Rd1 compares the Full and Rd1 models and Δ Rd12 compares the Full and Rd12 models. Naturally all differences are positive and range between by 55% to 356% in Δ Rd1 and between 9% and 158% for Δ Rd12. For each species the Δ Rd12 values are smaller than the Δ Rd1 values.

Finally, we also compared the DRSM method with those methods used in the paper by Domagalski et al. [14]. To make a fair comparison, we filtered and transformed the dataset according to what was used by Domagalski et al.: (1) we only considered the experiments with yield greater than or equal to 70%; (2) 37 experiments were used for calibration, and the remaining used for verification; (3) the logarithm of the product concentration was modeled as the output. Using the DRSM method, the RMSE values are 0.026 for the calibration data and 0.052 for the verification data. Both these RMSE values are better than most of the models in the paper by Domagalski et al., and are comparable to the best model there, which is using a genetic algorithm, leading to RMSE equal to 0.032 and 0.044 respectively for the calibration and verification data. In addition to the modeling accuracy, DRSM with Lasso is shown to be computationally efficient, while performing a deep heuristic search in genetic algorithm can be much more computationally expensive.

3.1.3 Optimization of process conditions

The use of the RSM model in process optimization is well established in the literature [24–26]. DRSM models, a generalization of the RSM models, can also be used to perform more detailed optimization studies. Without using the duration of the experiment as a factor in the DoE,

which reduces the number of experiments, they enable us to calculate the optimal batch duration to achieve an optimal operation. Moreover, they also enable the constraint with an indefinite time instance, for example, the maximum concentration constraint of an intermediate at any time between the start and the end of the batch. Like any other optimization task, one needs to define the objective function to be maximized or minimized and the required constraints. In a recent publication [27], the DRSM model has been used in the calculation of the optimal operating window, during which all the impurities are below a concentration upper limit.

Here, we solve the optimization problem that Domagalski et al. defined, to minimize the amount of catalyst, while constraining the product concentration to be greater than 90% at the cycle time, which is 6 h. We present three variations of this problem and their respective solutions, mostly to examine some important robustness issues. M1: This is the basic optimization formulation. We calculate the values of the factors that minimize catalyst usage while respecting the product concentration and batch duration constraints. The nominal predictions of the DRSM model are used here. M2: This is a robust variation of the M1 problem above. It accounts for the fact that the implemented factor values might be 5% different from the optimal values calculated from the nominal model. This will be called the implementation robustness. M3: This is further variation of the M2 problem, accounting for the prediction uncertainties of the DRSM model (with 95% confidence interval). Therefore, this model considers both the implementation robustness as well as the prediction robustness.

The solutions to the above three variations to the optimization problem are given in Table 5. The optimal factor levels are given in coded values (varying between -1 and 1) and in the physical units. With M1, the minimum catalyst amount is 0.015 equiv (of starting material 1). However, this will violate the yield constraint when there is

Table 5 Factors in the design space and optimal conditions from different formulations

| Item | | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|-----------------------------|-------|--------------------|-------------------------------------------|---------------------------|-------------------------------|-----------------------|
| | | Temperature /°C | Solvent volume /(L·mol ⁻¹) | Catalyst amount /equiv | Starting material 2 /equiv | Water amount /wt-% |
| Range | Lower | 30 | 0.7 | 0.01 | 1.04 | 0.05 |
| | Upper | 65 | 1.2 | 0.15 | 1.20 | 0.50 |
| Solutions in coded values | M1 | 0.76 | -1.00 | -0.93 | -1.00 | -0.55 |
| | M2 | 0.73 | -0.96 | -0.81 | -1.00 | -0.48 |
| | M3 | 0.40 | -0.84 | -0.37 | -1.00 | -0.60 |
| Solutions in original units | M1 | 60.84 | 0.70 | 0.015 | 1.04 | 0.152 |
| | M2 | 60.31 | 0.71 | 0.024 | 1.04 | 0.167 |
| | M3 | 54.52 | 0.74 | 0.054 | 1.04 | 0.139 |

uncertainty. The M2 solution requires that the yield constraint is respected when the implemented factor vary by 5% from the calculated optimal values. Naturally it asks for a higher amount of catalyst, 0.024 equiv. Finally, if one also considers the DRSM model's uncertainty, more catalyst (0.054 equiv) should be used. The factor levels of the catalyst amount are given in bold because they also represent the objective function of the optimization. The optimal values of the temperature factor are affected by the uncertainties considered especially on the M3 case, while the corresponding values of the other three factors are less sensitive; their coded values change by less than 15%. The optimization results should be further validated through experiments, especially when the DoE is very frugal. This has been studied and discussed in detail previously [27].

With the use of the DRSM model, we can also easily

draw contour plots, which present how the objective function or constraint values change. Such an example is shown in Fig. 3 where the *x*-axis is the catalyst amount (factor 3), which is also the objective function, while factor 1 is on the *y*-axis; other factors are fixed according to the solution of formulation M1. The contour lines show the DRSM nominal estimation of product concentration at 6 h. The blue square is where the optimal solution lies, as it corresponds to the smallest catalyst amount of *x*-axis while satisfying that the product yield is greater than 0.9. We can also generate a series of contour plots by varying multiple factor levels as well as time, as shown in Fig. 4. Across different subplots, we can observe the trend clearly that the increase of the concentration with time will be accelerated by a higher temperature (factor 1), a lower solvent volume (factor 2).

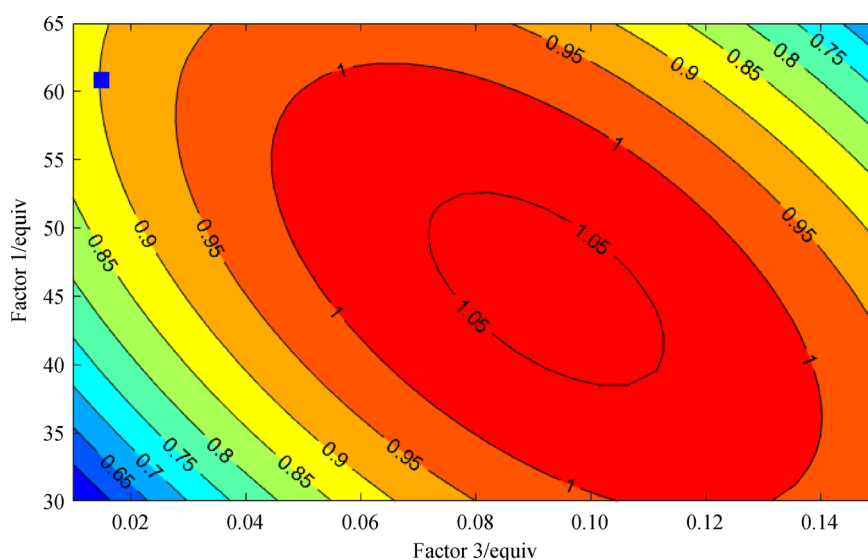


Fig. 3 Contour plot of the concentration for product (in unit of equiv), with factors 1 and 3 varying at different values. Factor 2 is fixed at 0.7 L·mol⁻¹, factor 4 is fixed at 1.04 equiv, and factor 5 is fixed at 0.152 wt-%, while time is fixed at 6 h. Blue square denotes the optimal condition of M1.

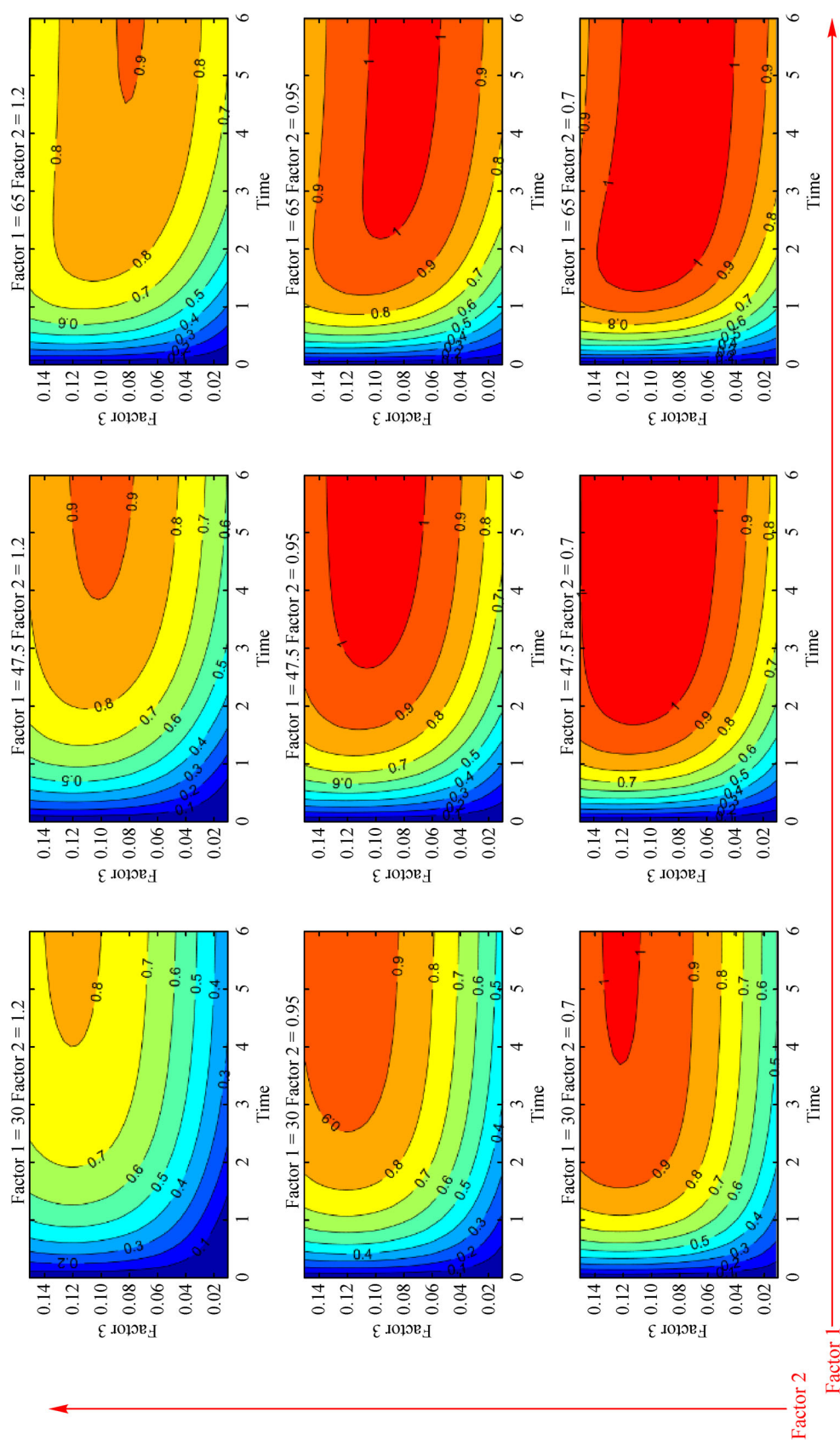


Fig. 4 Contour plots of the concentration for product (in unit of equiv), with factors 1, 2, 3 and time varying at different values with units of °C, L·mol⁻¹, equiv and hour, respectively. Factor 4 is fixed at 1.04 equiv, while factor 5 is fixed at 0.152 wt-%.

3.2 Case study B

Next, we show the results of case study B. Using the reaction extent data, we develop DRSM models for the seven identified reactions (i.e., seven responses), using SWR or Lasso. The results are summarized in Table 6. Both SWR and Lasso lead to the same global parameters R and t_c . For reactions R3 to R7, the difference of Lasso and SWR is quite minimal, in terms of the number of variables γ and the RMSE values.

For reactions R1 and R2, Lasso leads to smaller number of variables, compared to stepwise reactions. Therefore, we look at these two DRSM models in detail. We focus on reaction R2 here, since reaction R2 has a larger RMSE difference. The γ values are given in Table 7. Besides removing some high-order terms, Lasso helps to remove the 2FI term x_2x_3 and the quadratic term x_2^2 entirely. The weight of these removed terms is less than 2% of the maximum weight in the models. Even though Lasso has a total number of variables γ one third less than SWR, the difference in the modeled output is extremely small, as one can barely tell the difference of the SWR and Lasso fitting lines shown in Fig. 5. At the beginning of each experiment, the reaction extent is fixed to be zero, which is a feature that constrained DRSM allows. This figure also includes

the true reaction extent obtained from the ODE system from the simulation. We can observe that both DRSM models can describe the true reaction extents very well.

The differences between the modeled reaction extents and the true values are small for all seven identified reactions. Among them, reaction R6 has the greatest difference, whose fitting results are shown in Fig. 6. For both SWR and Lasso, the obtained model of R6 is the same, as presented in Table 8. The reaction rate of R6 is small, and therefore, the noise in the data, after projection from species concentration space to reaction extent space, leads to more prominent relative errors, as one can observe from the blue circles in the figure. Many other reactions, for example R2, have even more negligible difference, and the profiles of modeled reaction extents and true values overlap each other, as shown previously in Fig. 5.

Finally, the computational time will be briefly discussed. Same as what was observed in case study A, Lasso takes much less time, 0.047 s on average, compared to SWR, 3.8 s on average. This improvement will be very valuable, if one has many candidate stoichiometric groups and wants to develop DRSM models for the reaction extents/rates for a large number of candidate groups. This challenge will appear if the chemistry is unknown and complex, and the continuous-time DRSM models describing each reaction

Table 6 DRSM model summary of SWR and Lasso for different reactions

| Reaction | R | t_c | Number of variables γ | | RMSE (in unit of equiv) | | | RMSE/max (Lasso) |
|----------|-------------|-------|------------------------------|-------|-------------------------|--------|----------------|------------------|
| | SWR & Lasso | | SWR | Lasso | SWR (ref.) | Lasso | Δ Lasso | |
| R1 | 4 | 2.81 | 31 | 21 | 0.0032 | 0.0041 | 27% | 0.4% |
| R2 | 4 | 2.81 | 30 | 20 | 0.0109 | 0.0150 | 38% | 1.2% |
| R3 | 3 | 1.96 | 33 | 32 | 0.0135 | 0.0151 | 11% | 1.5% |
| R4 | 2 | 2.82 | 26 | 26 | 0.0059 | 0.0059 | 0% | 1.1% |
| R5 | 2 | 8.80 | 25 | 25 | 0.0045 | 0.0045 | 0% | 2.9% |
| R6 | 2 | 8.80 | 19 | 19 | 0.0040 | 0.0040 | 0% | 5.5% |
| R7 | 2 | 6.24 | 14 | 12 | 0.0027 | 0.0030 | 10% | 6.6% |

Table 7 Comparison of γ values estimated through SWR and Lasso for reaction R2 ^{a)}

| Item | Order | $\beta_0(\theta)$ | Linear term | | | 2FI term | | | Quadratic term | | |
|---------------------|-------|-------------------|-------------|---------|---------|----------|----------|----------|----------------|---------|---------|
| | | | x_1 | x_2 | x_3 | x_1x_2 | x_1x_3 | x_2x_3 | x_1^2 | x_2^2 | x_3^3 |
| Lasso | 0 | 0.3555 | 0.0584 | 0.0527 | −0.0303 | 0.0064 | −0.0035 | | | | −0.0017 |
| | 1 | 0.3698 | 0.0146 | 0.0456 | −0.0423 | | | | | | −0.0017 |
| | 2 | | −0.0437 | −0.0071 | −0.0120 | −0.0064 | 0.0059 | | | | |
| | 3 | −0.0329 | | | | | 0.0023 | | | | |
| | 4 | −0.0186 | | | | | | | | | |
| Stepwise regression | 0 | 0.3592 | 0.0572 | 0.0511 | −0.0298 | 0.0065 | −0.0036 | −0.0019 | −0.0047 | | 0.0028 |
| | 1 | 0.3880 | 0.0174 | 0.0500 | −0.0436 | | | | −0.0047 | | 0.0028 |
| | 2 | −0.0316 | −0.0489 | −0.0088 | −0.0121 | −0.0043 | 0.0059 | | | | |
| | 3 | −0.0460 | | −0.0048 | 0.0018 | 0.0021 | 0.0023 | | | | |
| | 4 | 0.0144 | 0.0091 | 0.0029 | | | | 0.0019 | | | |

a) A blank cell implies parameter is zero (Lasso) or insignificant (SWR)

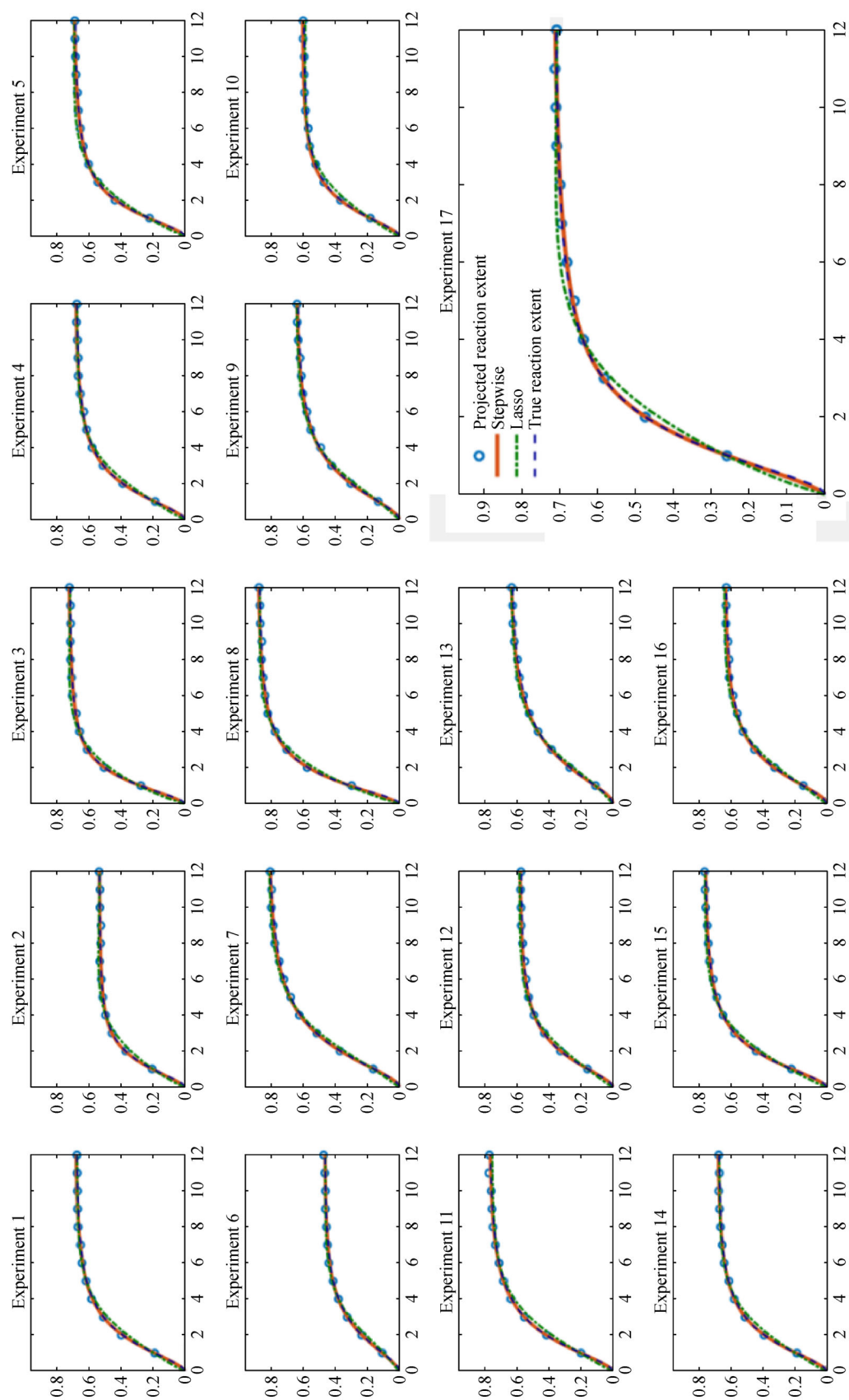


Fig. 5 Fitting results of using SWR and Lasso for reaction R2. Time is shown on x-axis with unit of hour, while reaction extent is on y-axis with unit of equiv. Experiment 17 is enlarged.

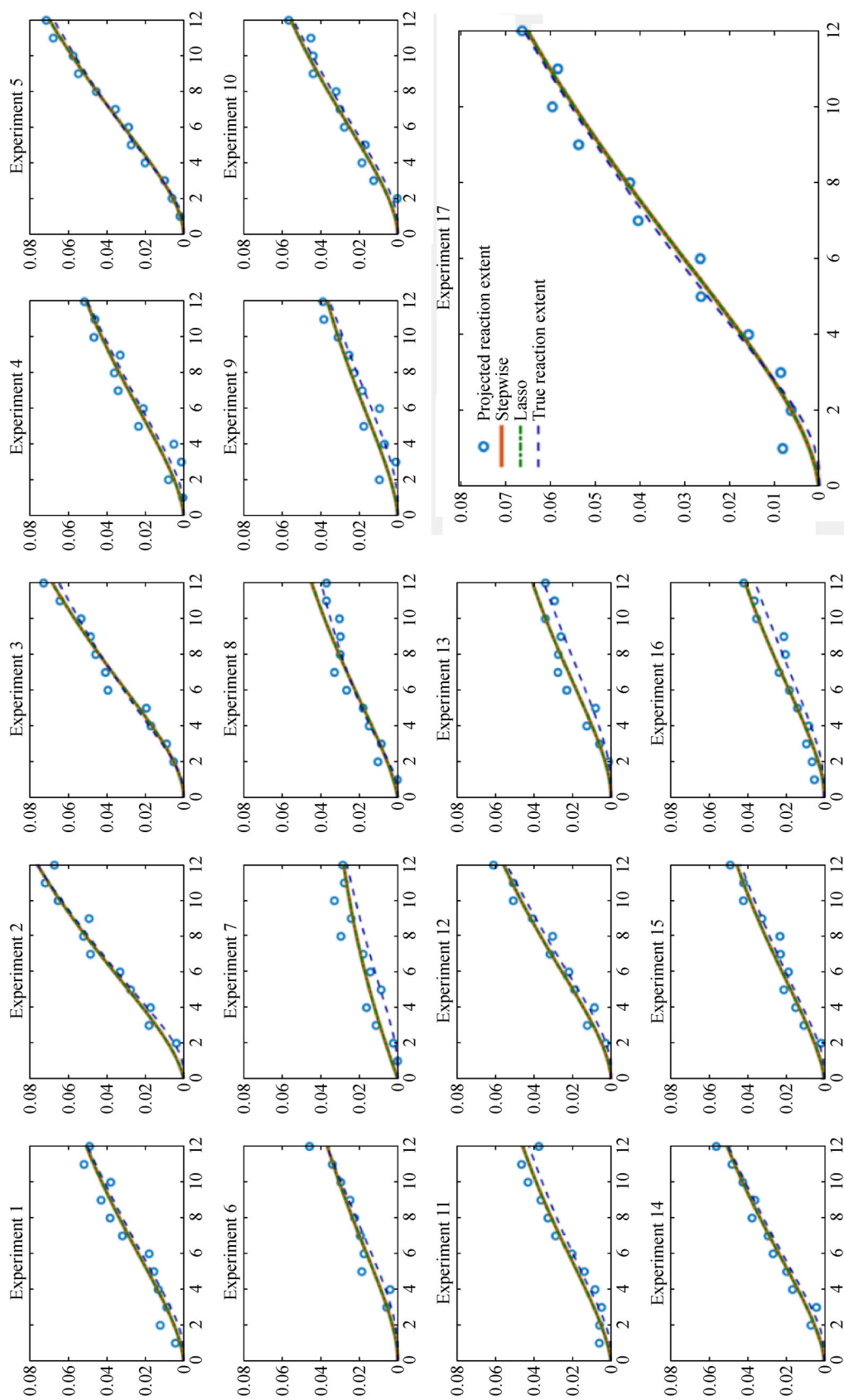


Fig. 6 Fitting results of using SWR and Lasso for reaction R6. Time is shown on x-axis with unit of hour, while reaction extent is on y-axis with unit of equiv. Experiment 17 is enlarged.

Table 8 Comparison of γ values estimated through SWR and Lasso for reaction R6 ^{a)}

| Item | Order | $\beta_0(\theta)$ | Linear term | | | 2FI term | | | Quadratic term | | |
|-----------|-------|-------------------|-------------|---------|--------|----------|----------|----------|----------------|---------|---------|
| | | | x_1 | x_2 | x_3 | x_1x_2 | x_1x_3 | x_2x_3 | x_1^2 | x_2^2 | x_3^3 |
| Lasso/SWR | 0 | 0.0311 | 0.0085 | -0.0029 | 0.0027 | -0.0021 | 0.0018 | 0.0022 | | | |
| | 1 | 0.0446 | 0.0126 | -0.0051 | 0.0050 | -0.0021 | 0.0018 | 0.0051 | | | |
| | 2 | 0.0136 | 0.0041 | -0.0022 | 0.0023 | | | 0.0029 | | | |

a) A blank cell implies parameter is zero (Lasso) or insignificant (SWR).

will be useful in determining whether a specific stoichiometric group is likely to happen or not.

4 Conclusions

In this paper, we introduced an improvement of the DRSM, a data-driven modeling tool for the dynamic response data, by utilizing Lasso regression instead of previously used SWR. This makes the solution process much speedier while maintaining the solution quality. Other algorithmic enhancements added in the article include (1) a non-linear regression to refine the value of the global parameter t_c and (2) the use of an ANOVA step after Lasso regression to further eliminate insignificant model parameters. We tested its ability to model different types of datasets. The computationally efficient DRSM algorithm, with Lasso regression incorporated, can be used for accurately modeling different dynamic datasets, enabling data-driven process development in multiple aspects, such as stoichiometry identification, process optimization and control.

For the case study, we first focus on modeling the concentration data from an organic synthesis problem described by Domagalski et al. in 2015 [14]. There are three advantages in using the systematical data-driven method DRSM: (1) DRSM model is more accurate than most models and as accurate as the best model obtained from genetic algorithm by Domagalski et al. (2) The solution time of DRSM with Lasso is fast. (3) Last but not the least, it does not need the data filter used in the genetic algorithm, and thus enables the modeling of data in a broader range. We further studied its applications in process optimization and how to add robust DRSM constraints to address uncertainty. In the second case study, we demonstrated how DRSM can accurately model the reaction extents, instead of species concentrations, for another example in pharmaceutical industry. This shows the versatility of DRSM approach in modeling different types of data.

Moving forward, it will be interesting to study how to use DRSM in automatic stoichiometry identification and kinetic estimation. Using target factor analysis together with DRSM, we can distinguish the true reactions from the false ones in the underlying network. However, this is based on an expert-generated list of candidate stoichiometries, and false positives might exist. The DRSM algorithm

with Lasso described here, with low computational cost, enables the screening of a large number of stoichiometric groups, which can facilitate the automatic stoichiometry identification.

Acknowledgements Yachao Dong is grateful for the financial support of Fundamental Research Funds for the Central Universities (Grant No. DUT20RC(3)070).

References

1. Coley C W, Eyke N S, Jensen K F. Autonomous discovery in the chemical sciences part I: progress. *Angewandte Chemie International Edition*, 2020, 59: 2–38
2. Van de Vijver R, Vandewiele N M, Bhoorasingh P L, Slakman B L, Khanshan F S, Carstensen H H, Reyniers M F, Marin G B, West R H, Van Geem K M. Automatic mechanism and kinetic model generation for gas- and solution-phase processes: a perspective on best practices, recent advances, and future challenges. *International Journal of Chemical Kinetics*, 2015, 47(4): 199–231
3. Qian F, Tao L, Sun W, Du W. Development of a free radical kinetic model for industrial oxidation of *p*-xylene based on artificial neural network and adaptive immune genetic algorithm. *Industrial & Engineering Chemistry Research*, 2012, 51(8): 3229–3237
4. Shi H, Zhou T. Computational design of heterogeneous catalysts and gas separation materials for advanced chemical processing. *Frontiers of Chemical Science and Engineering*, 2021, 15(1): 49–59
5. Selekmán J A, Qiu J, Tran K, Stevens J, Rosso V, Simmons E, Xiao Y, Janey J. High-throughput automation in chemical process development. *Annual Review of Chemical and Biomolecular Engineering*, 2017, 8(1): 525–547
6. Caron S, Thomson N M. Pharmaceutical process chemistry: evolution of a contemporary data-rich laboratory environment. *Journal of Organic Chemistry*, 2015, 80(6): 2943–2958
7. Ulrich J, Froberg P. Problems, potentials and future of industrial crystallization. *Frontiers of Chemical Science and Engineering*, 2013, 7(1): 1–8
8. Gernaey K V, Cervera-Padrell A E, Woodley J M. A perspective on PSE in pharmaceutical process development and innovation. *Computers & Chemical Engineering*, 2012, 42: 15–29
9. Yue W, Chen X, Gui W, Xie Y, Zhang H. A knowledge reasoning fuzzy-Bayesian network for root cause analysis of abnormal aluminum electrolysis cell condition. *Frontiers of Chemical Science and Engineering*, 2017, 11(3): 414–428
10. Montgomery D C. *Design and Analysis of Experiments*. 8th edition. Hoboken: John Wiley & Sons, 2008

11. Klebanov N, Georgakis C. Dynamic response surface models: a data-driven approach for the analysis of time-varying process outputs. *Industrial & Engineering Chemistry Research*, 2016, 55 (14): 4022–4034
12. Wang Z, Georgakis C. New dynamic response surface methodology for modeling nonlinear processes over semi-infinite time horizons. *Industrial & Engineering Chemistry Research*, 2017, 56(38): 10770–10782
13. Dong Y, Georgakis C, Mustakis J, Hawkins J M, Han L, Wang K, McMullen J P, Grosser S T, Stone K. Constrained version of the dynamic response surface methodology for modeling pharmaceutical reactions. *Industrial & Engineering Chemistry Research*, 2019, 58(30): 13611–13621
14. Domagalski N R, Mack B C, Tabora J E. Analysis of design of experiments with dynamic responses. *Organic Process Research & Development*, 2015, 19(11): 1667–1682
15. Wang K, Han L, Mustakis J, Li B, Magano J, Damon D B, Dion A, Maloney M T, Post R, Li R. Kinetic and data-driven reaction analysis for pharmaceutical process development. *Industrial & Engineering Chemistry Research*, 2020, 59(6): 2409–2421
16. Alpaydin E. *Introduction to Machine Learning*. 3rd edition. Cambridge: MIT Press, 2014
17. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 2011, 3(1): 1–122
18. García-Muñoz S, Dolph S, Ward H W II. Handling uncertainty in the establishment of a design space for the manufacture of a pharmaceutical product. *Computers & Chemical Engineering*, 2010, 34(7): 1098–1107
19. Santos-Marques J, Georgakis C, Mustakis J, Hawkins J M. From DRSM models to the identification of the reaction stoichiometry in a complex pharmaceutical case study. *AIChE Journal*. American Institute of Chemical Engineers, 2019, 65(4): 1173–1185
20. Dong Y, Georgakis C, Mustakis J, Hawkins J M, Han L, Wang K, McMullen J P, Grosser S T, Stone K. Stoichiometry identification of pharmaceutical reactions using the constrained dynamic response surface methodology. *AIChE Journal*. American Institute of Chemical Engineers, 2019, 65(11): e16726
21. Huri N, Feder M. In selecting the Lasso regularization parameter via Bayesian principles, 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE), 2016, 1–5
22. Montgomery D C, Peck E A, Vining G G. *Introduction to Linear Regression Analysis*. 5th edition. London: Wiley, 2012
23. Golub G H, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 1979, 21(2): 215–223
24. Hanrahan G, Lu K. Application of factorial and response surface methodology in modern experimental design and optimization. *Critical Reviews in Analytical Chemistry*, 2006, 36(3–4): 141–151
25. Singh G, Pai R S, Devi V K. Response surface methodology and process optimization of sustained release pellets using Taguchi orthogonal array design and central composite design. *Journal of Advanced Pharmaceutical Technology & Research*, 2012, 3(1): 30–40
26. Bezerra M A, Santelli R E, Oliveira E P, Villar L S, Escaleira L A. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, 2008, 76(5): 965–977
27. Dong Y, Georgakis C, Mustakis J, Lu H, McMullen J P. Optimization of pharmaceutical reactions using the dynamic response surface methodology. *Computers & Chemical Engineering*, 2020, 135: 106778