CHEMICAL ENGINEERING TRANSACTIONS

VOL. 61, 2017

A publication of AIDIC

The Italian Association
of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Petar S Varbanov, Rongxin Su, Hon Loong Lam, Xia Liu, Jiří J Klemeš
Copyright © 2017, AIDIC Servizi S.r.l.
ISBN 978-88-95608-51-8; **ISSN** 2283-9216

DOI: 10.3303/CET1761259

Batch Process Scheduling under Uncertainty using Data-Driven Multistage Adaptive Robust Optimization

Chao Ning, Fengqi You*

Cornell University, Ithaca, New York 14853, USA
fengqi.you@cornell.edu

This paper proposes a novel data-driven batch process scheduling approach based on multistage adaptive robust optimization coupled with robust kernel density estimation (RKDE). The kernelized iteratively re-weighted lease squares (KIRWLS) algorithm combined with kernel tricks are adopted to learn the probability density function from outlier-corrupted uncertain processing time data. We then propose a data-driven outlier-resilient uncertainty set for scheduling problem using the extracted distributional information. The proposed framework exhibits robustness to contamination of uncertainty data by integrating robust optimization with robust statistics. The batch process scheduling is then formulated as a data-driven multistage decision-making problem. By introducing affine decision rules for recourse variables, the resulting data-driven multistage adaptive robust optimization problem can be solved efficiently. We apply the proposed data-driven multistage adaptive robust optimization to a multipurpose batch process scheduling problem using a dataset to demonstrate the superiority of the proposed method. Our proposed approach generates $13,851 more profits than those of multistage adaptive robust optimization with box set. Compared with the multistage adaptive robust optimization using kernel density estimation (KDE), the result returned from the proposed method generates $4,064 more profits.

1. Introduction

In the past decades, batch processes are becoming increasingly important in modern manufacturing industries (Hegyháti and Friedler, 2010). Due to their customer-oriented nature, multiproduct batch plants have been widely employed in manufacturing high-value-added products, such as fine chemicals and pharmaceuticals. Therefore, considerable research effort has been made in batch process scheduling (Sun et al., 2016). The uncertain nature of some parameters in scheduling models necessitates the research in batch process scheduling under uncertainty (Shi and You, 2016). Among these methods, robust optimization gains popularity due to its strong ability to hedge against uncertainties and also because of its computational attractiveness (Ben-Tal et al., 2004), and it has been applied in biomass processing network (Gong et al., 2016), biofuel supply chain (Tong et al., 2014) and microalgae production (Gong and You, 2017). Recently, a nested stochastic robust optimization framework was proposed to handle multi-scale uncertainties (Yue and You, 2016). However, most of the existing robust optimization based approaches assume a perfect knowledge on the region of uncertainty parameters or uncertainty set, which might not be realistic in some applications. In practice, what process operators have is the historical data of these uncertain parameters, and this is especially true in the era of big data. Besides, datasets in process industries are often contaminated with outliers. Data outliers arise for various reasons, such as sensor malfunction, human recording error and disturbance in processes (Liu et al., 2004). A main drawback of the existing robust scheduling methods is the lack of a proper mechanism to deal with these outliers. Thus, their uncertainty sets will be distorted in the presence of outliers. Consequently, their robust solutions could be very conservative, meaning a less profitable schedule when using real datasets.

This paper proposes a novel data-driven multistage adaptive robust optimization (ARO) modeling framework and its solution strategy to address batch process scheduling under uncertainty. The kernelized iteratively re-weighted lease squares (KIRWLS) algorithm is adopted to extract the probability density functions of uncertain parameters from corrupted uncertain processing time data (Kim and Scott, 2012). The RKDE can be interpreted as a weighted version of kernel density estimation (KDE), in which weights of outliers are diminished (Kim and Scott, 2012). In this way, the RKDE approach captures a reliable probability distribution in the presence of

Please cite this article as: Ning C., You F., 2017, Batch process scheduling under uncertainty using data-driven multistage adaptive robust optimization, Chemical Engineering Transactions, 61, 1567-1572 DOI:10.3303/CET1761259

outliers in uncertainty data. We then propose a data-driven uncertainty set for multistage ARO using quantile functions. The proposed data-driven multistage ARO exhibits robustness to contamination of uncertainty data by exploiting the idea from robust statistics in the context of robust optimization. To address the computational challenge, we introduce affine decision rules for recourse variables, and reformulate the problem as a deterministic problem. We apply the proposed data-driven multistage ARO to an industrial scale multipurpose batch process scheduling under processing time uncertainty.

2. Robust kernel density estimation

We assume uncertain parameters follow some unknown distribution with a density function *f*, and $\mathbf{u}_1$, …, $\mathbf{u}_N$ are *N* realizations of uncertainties. The KDE is shown as follows.

$$\hat{f}_{KDE}(\mathbf{u}) = \frac{1}{N}\sum_{i=1}^{N} K_h(\mathbf{u},\mathbf{u}_i) \tag{1}$$

where K_h is a kernel function with a positive bandwidth *h*. The Gaussian kernel, which is given below.

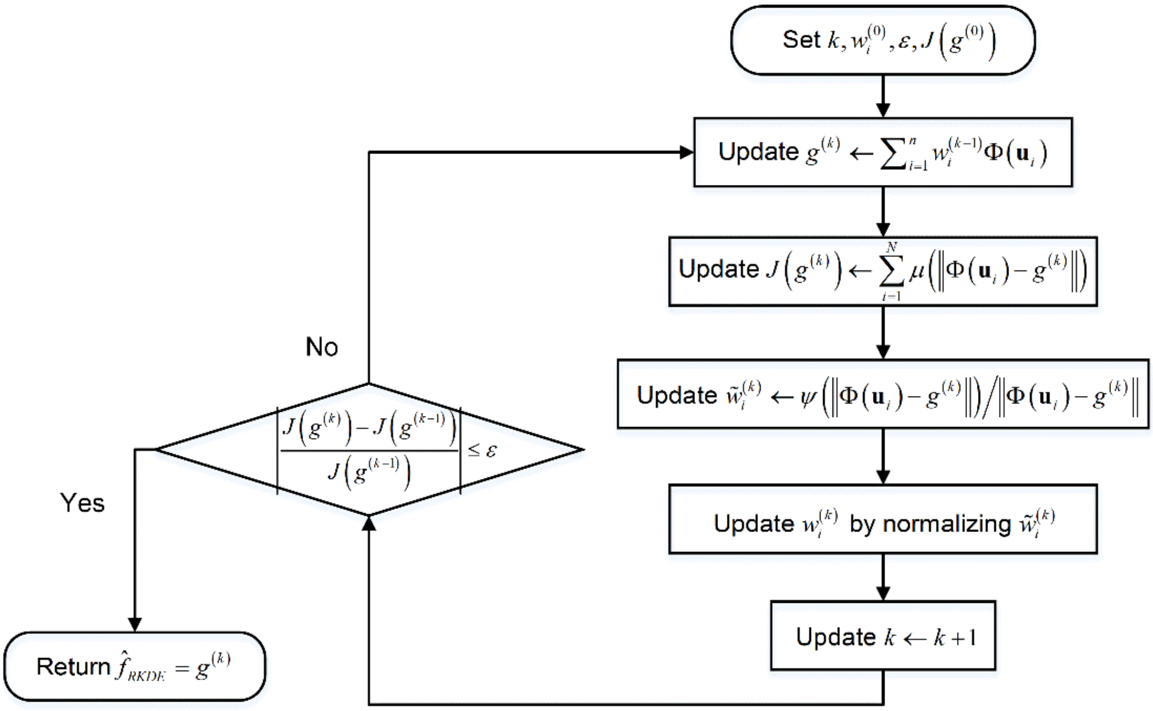
$$K_h(\mathbf{u},\mathbf{u}_i) = \left(\frac{1}{\sqrt{2\pi}h}\right)^d \exp\left(-\frac{\|\mathbf{u}-\mathbf{u}_i\|}{2h^2}\right) \tag{2}$$

Since Gaussian kernel is a positive semi-definite kernel, there exists a mapping Φ from $\boldsymbol{R}^d$ to *H* such that $K_h(\mathbf{u},\mathbf{u}_i) = \langle \Phi(\mathbf{u}), \Phi(\mathbf{u}_i) \rangle$ (Kim and Scott, 2012). *H* is an infinite dimensional Hilbert space of functions, also known as reproducing kernel Hilbert space (RKHS).

We employ RKDE, and it is casted as a minimization problem shown as follows (Kim and Scott, 2012).

$$\min_{g\in H} \sum_{i=1}^{N} \mu\left(\|\Phi(\mathbf{u}_i) - g\|\right) \tag{3}$$

where μ is a robust loss function, and the notation ψ represents the first derivative of robust loss function μ. The objective function is denoted as $J(g)$. The KIRWLS algorithm is given below (Kim and Scott, 2012).



Set $k, w_i^{(0)}, \varepsilon, J\left(g^{(0)}\right)$
Update $g^{(k)} \leftarrow \sum_{i=1}^{n} w_i^{(k-1)} \Phi(\mathbf{u}_i)$
Update $J\left(g^{(k)}\right) \leftarrow \sum_{i=1}^{N} \mu\left(\left\|\Phi(\mathbf{u}_i) - g^{(k)}\right\|\right)$
Update $\tilde{w}_i^{(k)} \leftarrow \psi\left(\left\|\Phi(\mathbf{u}_i) - g^{(k)}\right\|\right) \Big/ \left\|\Phi(\mathbf{u}_i) - g^{(k)}\right\|$
Update $w_i^{(k)}$ by normalizing $\tilde{w}_i^{(k)}$
Update $k \leftarrow k+1$
$\left|\frac{J\left(g^{(k)}\right) - J\left(g^{(k-1)}\right)}{J\left(g^{(k-1)}\right)}\right| \le \varepsilon$
No
Yes
Return $\hat{f}_{RKDE} = g^{(k)}$

Figure 1: Flowchart of the KIRWLS algorithm.

3. Data-driven multistage ARO based batch process scheduling model

In this section, the data-driven multistage ARO based process scheduling model is proposed. In a scheduling problem, the following information is given: the structure of the facility, recipe data, equipment units and their

sizes, storage capacities of intermediates, product prices and processing times (Chu and You, 2015). The data-driven multistage ARO based process scheduling model is presented as follows.
The objective function of the scheduling model is expressed as the revenue from selling products minus the cost of purchasing raw materials, and the expression is given as follows.

$$\max \sum_{s} price_s \left(ST_{sN} - ST_{s0} \right) \tag{4}$$

where the index *s* presents state, *N* is the total number of time points, *prices* is the price of chemical *s* and *STsn* is a continuous decision variable representing the storage amount of chemical *s* at time point *n*.
Constraint (5) specifies that the batch size of task *i* should not exceed its maximum batch size or fall below its minimum batch size when the task *i* is scheduled.

$$b_i^{\min} W_{inn'} \le B_{inn'} \le b_i^{\max} W_{inn'} \quad \forall n \in N, n' \in N_n^+, j \in J \tag{5}$$

where $B_{inn'}$ represents the batch size of task *i* that starts at time point *n* and finished before time point n', $W_{inn'}$ is a binary variable, and is equal to 1 when the task *i* starts at time point *n* and finished before time point n'. The parameters $b_i^{\max}$ and $b_i^{\min}$ are the maximum and minimum batch sizes of task *i*, respectively.
Constraint (6) specifies the balance of equipment units.

$$E_{jn} = E_{j(n-1)} + \sum_{i \in I_j} \left(\sum_{n' \in N_n^+} W_{inn'} - \sum_{n' \in N_n^-} W_{in'n} \right) \quad \forall n \in N, j \in J \tag{6}$$

where E_{jn} indicates whether equipment unit *j* is used at time point *n*, and its upper bound is given by,

$$E_{jn} \le 1 \quad \forall j, n \tag{7}$$

Constraints (8) and (9) specify mass balance and storage capacities.

$$ST_{sn} = ST_{s(n-1)} + \sum_{i \in TO_s} \rho_{is}^{O} \sum_{n' \in N_n^-} B_{in'n} - \sum_{i \in TI_s} \rho_{is}^{I} \sum_{n' \in N_n^+} B_{inn'} \quad \forall s, n \tag{8}$$

$$ST_{sn} \le C_s^{\max} \quad \forall s, n \tag{9}$$

where *TIs* and *TOs* are the sets of tasks that utilize chemical *s* as input and output, respectively. ρ_{is}^{I} and ρ_{is}^{O} are the material balance coefficients for chemical *s* as input and output of task *i*, respectively. $C_s^{\max}$ stands for the maximum storage capacity for chemical *s*.

Constraint (10) enforces that the time duration between two time points is no less than the processing times in each equipment unit.

$$T_{n'}\left(\mathbf{d}^{n'}\right) - T_n\left(\mathbf{d}^{n}\right) \ge \sum_{i \in I_j} \left(d_{in} \cdot W_{inn'} + vd_{in} \cdot B_{inn'} \right) \quad \forall \mathbf{d} \in U, j \tag{10}$$

where $\mathbf{d}^n$ represents the realizations of uncertain processing times available up to time point *n*. d_{in} is the fixed processing time of task *i*, and vd_{in} is the variable processing time of task *i*. Note that $T_n(\mathbf{d}^n)$ is a general function of the past uncertainty realizations. The set *U* is the data-driven uncertainty set defined as follows, which is decision-dependent.

$$U = \left\{ d_{in} \left| \begin{array}{l} \hat{F}_{RKDE}^{(i)\,-1}(\alpha) \le d_{in} \le \hat{F}_{RKDE}^{(i)\,-1}(1-\alpha) \quad \forall i, n \\ \displaystyle\sum_{i \in I_j} \sum_{n \in N} \left(\sum_{n' \in N_n^-} W_{in'n} \right) d_{in} \le \sum_{i \in I_j} \sum_{n \in N} \left(\sum_{n' \in N_n^-} W_{in'n} \right) \left(1 + \xi_i \phi\right) d_i^0 \quad \forall j \end{array} \right. \right\} \tag{11}$$

where set N_n^- represents the set of time points prior to *n*, d_i^0 and ξ_i are defined in (12) and (13), respectively.

$$d_i^0 = \left[\hat{F}_{RKDE}^{(i)\,-1}(\alpha) + \hat{F}_{RKDE}^{(i)\,-1}(1-\alpha) \right] \Big/ 2 \tag{12}$$

$$\xi_i = \left(\hat{F}_{RKDE}^{(i)\,-1}(1-\alpha) - d_i^0 \right) \Big/ d_i^0 \tag{13}$$

Constraints (14)-(17) specify the initial and final conditions.

$$W_{inn'} = 0 \quad \forall i, n, n' \in N \setminus N_n^+ \tag{14}$$

$$B_{inn'} = 0 \quad \forall i, n, n' \in N \setminus N_n^+ \tag{15}$$

$$T_1 = 0 \tag{16}$$

$$E_{jN} = 0 \quad \forall j \tag{17}$$

The nonnegative and integral constraints are shown as follows.

$$B_{inn'}, E_{jn}, ST_{sn}, T_n \geq 0 \quad \forall i, n, n', s, j \tag{18}$$

$$W_{inn'} \in \{0,1\} \quad \forall i, n, n' \tag{19}$$

The multistage ARO with general decision rules is known to be computationally intractable (Ben-Tal et al., 2004). Therefore, we adopt the widely used affine decision rules to address this computational challenge. An affine decision rule for timing decision variables is adopted as follows.

$$T_n\left(\mathbf{d}^n\right) = \tau_0^{(n)} + \sum_i \sum_{n' \leq n} \tau_{in'}^{(n)} \cdot d_{in'} \tag{20}$$

$$\sum_{n'' \in N_{n'}^-} W_{in''n'} = 0 \Rightarrow \tau_{in'}^{(n)} = 0 \tag{21}$$

By plugging the affine decision rule of (20) into (21), we can reformulate it as follows.

$$\sum_{i' \in I} \sum_{n'' \in N} \left(\hat{F}_{RKDE}^{(i)\ -1}\left(1-\alpha\right) \cdot r_{i'n''}^{jnn'} - \hat{F}_{RKDE}^{(i)\ -1}\left(\alpha\right) \cdot q_{i'n''}^{jnn'} \right) + \sum_{j' \in J} \sum_{i' \in I_{j'}} \sum_{n'' \in N} \left(\left(1+\xi\phi\right) \cdot s_{j'i'n''}^{jnn'} \right) \leq \tau_0^{(n')} - \tau_0^{(n)} - \sum_{i \in I_j} vd_i \cdot B_{inn'} \tag{22}$$

We adopt the Glover's linearization technique to handle the bilinear term, which is a product of a continuous variable and a binary variable.

$$\sum_{n''' \in N_{n''}^-} W_{i'n'''n''} = 1 \Rightarrow s_{j'i'n''}^{jnn'} \geq p_{j'}^{jnn'} \tag{23}$$

$$\sum_{n''' \in N_{n''}^-} W_{i'n'''n''} = 0 \Rightarrow s_{j'i'n''}^{jnn'} \leq 0 \tag{24}$$

$$s_{j'i'n''}^{jnn'} \leq p_{j'}^{jnn'} \tag{25}$$

The resulting data-driven multistage adaptive robust optimization based scheduling problem can be solved efficiently using the state-of-the-art branch-and-cut methods implemented in solvers like CPLEX.

4. Case study

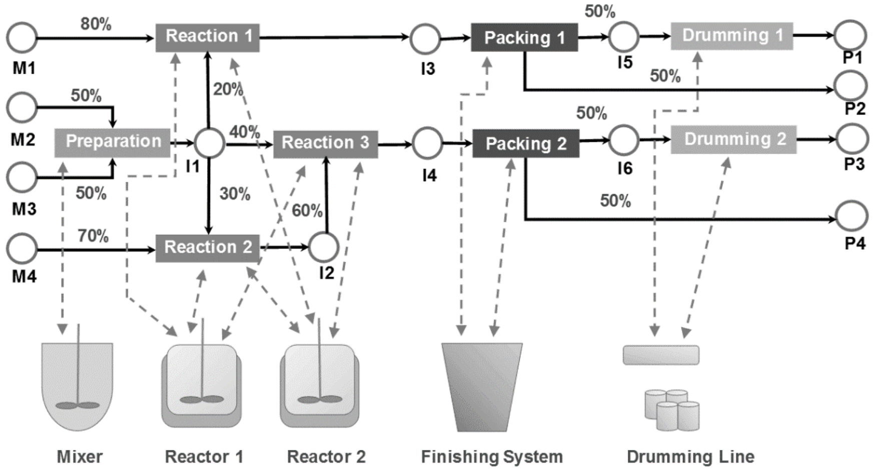


Figure 2: State task network of the multipurpose batch process.

In this section, an industrial multipurpose batch process (Yue et al., 2013) is presented to demonstrate the advantages of the proposed data-driven multistage ARO based approach. Figure 1 depicts the state-task network of this batch process (Chu et al., 2013). This network consists of 14 chemicals and 8 operation tasks (Wassick et al., 2012). In this application, the fixed processing time of all tasks are subject to uncertainty (Chu et al., 2013). The processing time data of Reaction 1 and Reaction 2 are used to construct uncertainty sets. It is worth noting that these processing time data are contaminated by outliers (Chu et al., 2015). The fixed processing times of other tasks are assumed to change up to 20 % of their corresponding nominal values.

In this case study, we demonstrate the superiority of the proposed data-driven multistage ARO based scheduling approach through comparing it with other methods. These methods include multistage ARO with box uncertainty set, multistage ARO based on KDE, multistage ARO based on RKDE using Huber loss function and multistage ARO based on RKDE using Hample loss function. All these optimization methods are modeled in GAMS 24.7.3, and are implemented on a computer with an Intel (R) Core (TM) i7-6700 CPU @ 3.40 GHz and 32 GB RAM. The optimality tolerance is set to be 0.001 % for the solver CPLEX 12.6.3.

In the schedule results of multistage ARO with box uncertainty set and multistage ARO based on KDE, the idle periods are actually reserved to hedge against the outliers in the dataset. It is worth pointing out that these data outliers might be ascribed to human recording error or sensor malfunctions. Therefore, its robust scheduling result is over-conservative and less profitable. Table 1 shows that the multistage ARO based on KDE method generates 22.3 % more profits than the multistage ARO with box set. The reduction of conservatism is due to the probability distribution information leveraged by the multistage ARO based on KDE.

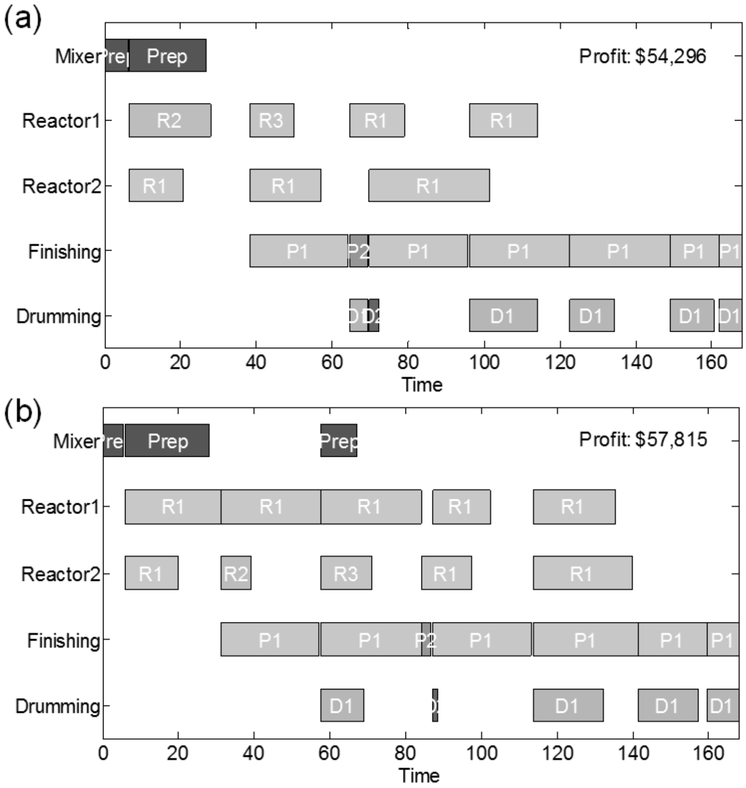


Figure 3: Gantt charts of data-driven multistage ARO using RKDE with different robust loss function: (a) Huber loss function, (b) Hampel loss function.

Table 1: Comparisons of model and solution statistics between different methods in the case study.

	Multistage ARO with box set	Data-driven multistage ARO with KDE	Data-driven multistage ARO with RKDE (Huber loss)	Data-driven multistage ARO with RKDE (Hampel loss)
Binary variables	1,320	1,320	1,320	1,320
Continuous variables	95,401	95,401	95,401	95,401
Constraints	215,457	215,457	215,457	215,457
CPU (s)	1,294.8	363.7	299.0	261.6
Max. Profit ($)	43,964	53,751	54,296	57,815

Figure 3 shows the Gantt charts of the data-driven multistage ARO based on RKDE using two kinds of robust loss functions, including Huber loss function and Hampel loss function (Kim and Scott, 2012). Due to the integration of robust statistics into robust optimization, the robust schedule is able to tolerate the data outliers in processing times. Comparing Figure 3 (a) and (b), we can readily conclude that the Hampel loss function is more suitable than the Huber loss function in the process scheduling problem. The data-driven multistage ARO based on RKDE (Hampel loss function) generates the highest profit, which is 31.5 % and 7.6 % higher than the results of the multistage ARO with box set and the multistage ARO based on KDE, respectively.

5. Conclusions

This paper proposed a novel data-driven multistage ARO coupled with RKDE batch process scheduling model. The proposed framework exhibited robustness to contamination of uncertainty data by integrating robust optimization with robust statistics. The novel contributions of this work include the novel data-driven multistage ARO based batch process scheduling model and its solution strategy, as well as the data-driven outlier-resistant uncertainty set based on RKDE for multistage ARO. We applied the proposed data-driven multistage ARO modelling framework and solution strategy to a multipurpose batch process scheduling to demonstrate the superiority of the proposed method. Future works may consider extensions of the proposed method for nonlinear and/or nonconvex mixed-integer ARO problems, and the comparison of the impacts of different loss functions on the results of data-driving multistage ARO frameworks.

References

Ben-Tal A., Goryashko A., Guslitzer E., Nemirovski A., 2004, Adjustable robust solutions of uncertain linear programs, Mathematical Programming, 99, 351-376.

Chu Y., Wassick J.M., You F., 2013, Efficient scheduling method of complex batch processes with general network structure via agent-based modeling. AIChE Journal, 59, 2884-2906.

Chu Y., You F., 2013, Integrated scheduling and dynamic optimization of sequential batch processes with online implementation. AIChE Journal, 59, 2379-2406.

Chu Y., You F., 2015, Model-based integration of control and operations: Overview, challenges, advances, and opportunities, Computers & Chemical Engineering, 83, 2-20.

Chu Y., You F., Wassick J.M., Agarwal A., 2015, Integrated planning and scheduling under production uncertainties: Bi-level model formulation and hybrid solution method. Computers & Chemical Engineering, 72, 255-272.

Gong J., Garcia D. J., You F., 2016, Unraveling Optimal biomass processing routes from bioconversion product and process networks under uncertainty: An adaptive robust optimization approach, ACS Sustainable Chemistry & Engineering, 4, 3160-3173.

Gong J., You F., 2017, Optimal processing network design under uncertainty for producing fuels and value-added bioproducts from microalgae: Two-stage adaptive robust mixed integer fractional programming model and computationally efficient solution algorithm, AIChE Journal, 63, 582-600.

Hegyháti M., Friedler F., 2010, Overview of industrial batch process scheduling, Chemical Engineering Transactions, 21, 895-900.

Kim J., Scott C., 2012, Robust kernel density estimation, Journal of Machine Learning Research, 13, 2529-2565.

Liu H., Shah S., Jiang, W., 2004, On-line outlier detection and data cleaning, Computers & Chemical Engineering, 28, 1635-1647.

Shi H., You F., 2016, A computational framework and solution algorithms for two-stage adaptive robust scheduling of batch manufacturing processes under uncertainty, AIChE Journal, 62, 687-703.

Sun L., Zou X., Dong H., Wang S., 2016, Simultaneous optimization of short-term scheduling and heat integration schemes for multipurpose batch plants, Chemical Engineering Transactions, 52, 355-360.

Tong K., You F., Rong G., 2014, Robust design and operations of hydrocarbon biofuel supply chain integrating with existing petroleum refineries considering unit cost objective, Computers & Chemical Engineering, 68, 128-139.

Wassick J.M., Agarwal A., Akiya N., Ferrio J., Bury S., You F., 2012, Addressing the operational challenges in the development, manufacture, and supply of advanced materials and performance products. Computers & Chemical Engineering, 47, 157-169.

Yue D., You F., 2013, Sustainable scheduling of batch processes under economic and environmental criteria with MINLP models and algorithms, Computers & Chemical Engineering, 54, 44-59.

Yue D., You F., 2016, Optimal supply chain design and operations under multi-scale uncertainties: Nested stochastic robust optimization modeling framework and solution algorithm. AIChE Journal, 62, 3041-3055.