Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 52-1 (2019) 685–690

Robust State Estimation for Fouling Evolution in Batch Processes using the EM Algorithm ⋆

Ouyang Wu *,** **Lars Imsland** ** **Edmund Brekke** **
Stefan M. Schneider * **Ala E.F. Bouaswaig** * **Matthias Roth** *

* *Automation Technology, BASF SE, 67056 Ludwigshafen, Germany*
** *Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway*
(e-mail: ouyang.wu@ntnu.no; lars.imsland@ntnu.no)

Abstract: Degradation such as fouling has significant negative impact on the efficiency of process manufacturing. Creating transparency about degradation is essential in the operations of the plants. This paper focuses on the state estimation of fouling in a multipurpose batch process, and the problem is formulated using an industrial case study. Due to the lack of a good model of fouling, the state estimation is integrated with parameter estimation using the expectation maximization (EM) algorithm, where the variational approximation method is employed for the intermediate robust state estimation in the *E*-step. The proposed state estimation is applied in the case study, which demonstrates the efficacy of the estimation of a recipe-independent key performance indicator for the batch-to-batch fouling.

Keywords: Batch-to-batch fouling, EM algorithm, T-distribution, Variational approximation

1. INTRODUCTION

In process industries, it is inevitable that the performances of the plants degrade during their lifetime. Fouling, a typical degradation in chemical plants, refers to undesirable accumulation of residues in internal or external surfaces of equipment, which results in significant economic loss (Ibrahim, 2012). Many studies, such as Teruel et al. (2005), Radhakrishnan et al. (2007) and Rodriguez and Smith (2007), applied data analytics and optimization approaches in fouling problems, and most of the published studies dealt with continuous processes. Compared to continuous processes, batch processes present a wide operating range, irreversible behavior and repetitive nature (Bonvin, 2006), which brings challenges to the modeling and optimization considering degradation. A case study of a chemical batch process is considered in this paper, where the batch reactors along with the heat exchangers suffer from fouling. Fouling that evolves from batch to batch leads to an increase in the pressure drop across the exchangers. Eventually, plant shutdowns for cleaning of the units are required to avoid a system failure. Nevertheless, frequent cleaning leads to increasing time of shutdown and therefore a decrease in the capacity of production. Finding a reliable key performance indicator (KPI) for the degree of fouling become requisite for the optimization of any operations that are affected by fouling. One of the solutions is the use of state estimation.

First-principle models are usual when implementing state estimation in industrial applications. In some cases, such a model is not available, and the state estimation problem has to be solved in combination with parameter estimation, see Zia et al. (2008), where a data-driven model is employed. Shumway and Stoffer (1982) and Gibson and Ninness (2005) proposed the application of the expectation maximization (EM) algorithm for system identification using Gaussian linear models, in which both model parameters and hidden states are estimated iteratively. Gaussian distributions present good analytical properties for linear models, but it cannot handle outliers which results in robustness issues such as inaccurate estimates for both model parametes and states. As an alternative, the Student's t-distribution provides an option to handle potential outliers through its heavy-tailed shape and therefore improves the accuracy of estimation (Liu and Rubin, 1995). Agamennoni et al. (2012) proposed a variational approximation solution to a robust linear state estimation problem, where the heavy-tailed measurement noise is modeled using inversed Wishart distributed covariance parameters. Zhu et al. (2013) applied the variational Bayesian method in the robust sensor fusion problem with t-distributed noise terms to estimate both the states and noise parameters.

Compared to the state estimation problems in Agamennoni et al. (2012) and Zhu et al. (2013), this paper focuses on smoothing of fouling indicators in the historical batch data without any known models. A linear state-space-form structure is employed to describe the fouling dynamics, in which a t-distributed noise term is applied to handle outliers in the industrial data. The state estimation of the fouling KPI is combined with the model parameter estimation based on the EM framework, where the variational approximation is taken as the *E*-step of the EM algorithm to calculate the intermediate state estimates. The main contribution of this paper is the integration of the EM algorithm and the variational approximation for robust

⋆ Financial support is gratefully acknowledged from the Marie Skodowska Curie Horizon 2020 EID-ITN project "PRONTO", Grant agreement No 675215.

10.1016/j.ifacol.2019.06.142

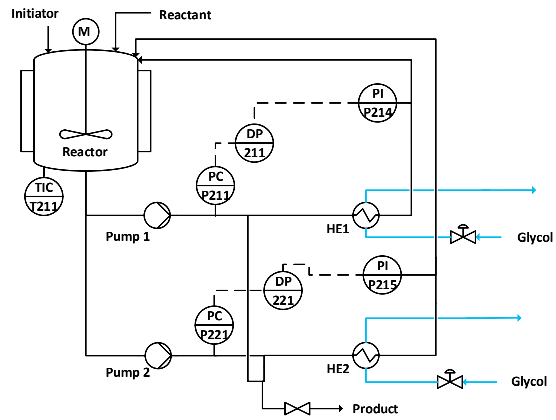

Fig. 1. Batch process schematics: reaction section

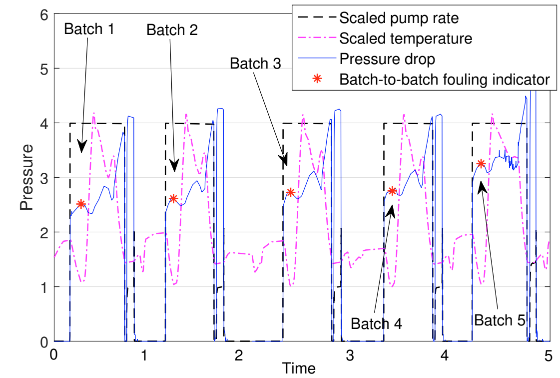

Fig. 2. Pressure drop measurements (5 batches)

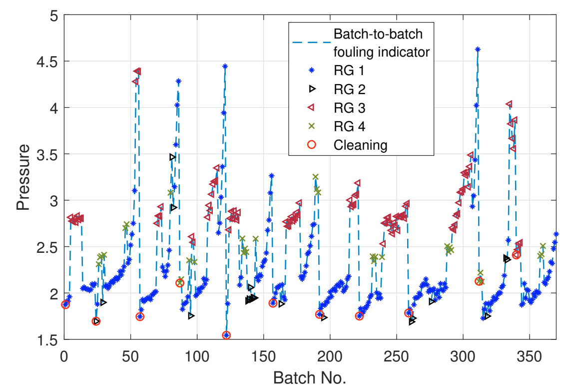

Fig. 3. Batch-to-batch fouling indicator

state estimation of the fouling KPI in an industrial example. The remainder of this paper is structured as follows. The problem formulation for the case study is presented in Section 2. The robust state estimation using the EM framework is included in Section 3. The state estimation results of the case study are illustrated in Section 4 and conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

The multipurpose batch plant in the considered example produces many types of products according to the recipes. In each batch run, raw materials are discharged and blended in a vessel according to the recipe, and the blended fluid is discharged into the reaction section as shown in Fig. 1. The polymerization starts with the inflow of the initiator, and the reactor temperature is controlled during the exothermic reaction using two parallel recirculation loops including pumps and heat exchangers.

The reaction section is highly prone to fouling. During the reaction phase, polymer residuals are accumulated in the reactor and heat exchangers. One of the main impacts of fouling is the reduced heat transfer from the product to the coolant, which results in longer batches by limiting the cooling capacity. Moreover, fouling leads to an increased pressure drop over the heat exchanger resulting in a smaller recirculation flow. As a result, the reaction section needs to be shut down and cleaned once the residuals approach an unacceptable level. As shown in Fig. 2, the pressure drop measurement for overall batch is increasing, which matches the fouling trend from batch to batch; within an individual batch, the pressure drop, which is affected by the varying operating condition such as temperature, fails to indicate the actual fouling trend; therefore, a batch-to-batch fouling indicator is developed by sampling the pressure drop at the starting point of the reaction in each batch (denoted as asterisks in Fig. 2), where the operation conditions keep the same. The fouling indicator is depicted in Fig. 3, which shows a periodic pattern for the evolution of fouling with repeated cleaning operations denoted by the red circle. The concept of *campaign* is the batch series between two consecutive cleanings, in which the evolution of fouling varies and is independent among different campaigns (Wu et al., 2018). As Fig. 3 illustrates with the denotation from Recipe Group (RG) 1 to RG 4, the fouling indicator is also recipe-dependent due to varying physical properties in different recipes. It is necessary to develop a recipe-independent KPI to avoid confusion in the fouling indication.

In this paper, we aim to develop state estimation for a recipe-independent fouling KPI, which can be further used in the modelling of fouling evolution. Wu et al. (2018) employed a static model to fit the fouling indicator and calculated biases of the indicator values due to different recipes. To improve the performance of model fitting, we employ a state space model to describe the batch-to-batch fouling dynamics within campaigns as presented in Eq. 1. Because of the lack of the first-principle model, a linear first-order model structure is employed to describe the evolution of the fouling KPI from batch to batch. The parameters of the state equations are set to be campaign-dependent to fit the varying dynamics of campaigns as illustrated in Fig. 3. Besides, a set of linear function factors is used in the measurement equation to model the recipe effects on the fouling indicator.

$$\begin{aligned} x_{k+1}^j &= A_j x_k^j + B_j + w_k^j \\ y_k^j &= C_{r_k^j} x_k^j + D_{r_k^j} + e_k^j \end{aligned} \tag{1}$$

where, the observed data set consists of the fouling indicator y_k^j and the group identifier of the recipe $r_k^j \in I = \{1, 2, ..., M\}$, and M is the total number of the recipe groups; the hidden state set includes the recipe-independent fouling KPI x_k^j; the subscript and superscript $j \in J = \{1, 2, ..., L\}$ refer to the jth campaign, and L is the number of campaigns in the picked historical data; the subscript k denotes the kth batch in the jth campaign; for instance, x_k^j denotes the fouling KPI at the kth batch of the jth campaign, where $k \in K = \{1, 2, \ldots, N_j\}$ and the length of campaign N_j varies according to the fouling evolution; $\{(A_j,\ B_j)|j \in J\}$ is the parameter set for the campaign-depended state equation, while $\{(C_i,\ D_i)|i \in I\}$ is for the recipe-dependent measurement equation; C_i acts as the slope factor between the state and the observation, while D_i is the corresponding intercept term; the initial state x_1^j is assumed to be identical independent distributed (i.i.d) Gaussian $x_1^j \sim \mathbf{N}(\mu_x, \sigma_x^2)$; the process noise w_k^j is assumed to follow i.i.d Gaussian $w_k^j \sim N(0, \sigma_{wj}^2)$; the measurement noise e_k^j is assumed to follow i.i.d Student's t-distribution $e_k^j \sim t(0, \sigma_e^2, \nu)$ in consideration of measurement outliers. The expression of the t- distribution is presented as follows (Liu and Rubin, 1995; Fang and Jeong, 2008):

$$e_k^j = \frac{\Gamma(v/2 + 1/2)}{\Gamma(v/2)\sqrt{\pi\sigma_e^2}} \cdot (1 + \frac{{e_k^j}^2}{\nu\sigma_e^2})^{-\frac{v+1}{2}} \quad (2)$$

where σ_e^2 refers to the variance, and ν denotes the degree of freedom; Gamma function $\Gamma(z) = \int_0^\infty y^{z-1}e^{-y}dy$; the t-distribution can be further decomposed by introducing variance scaling variables s_k^j: $e_k^j|s_k^j$ follows Gaussian distribution, and s_k^j is i.i.d Gamma distributed:

$$e_k^j|s_k^j \sim \mathbb{N}(0, \sigma_e^2/s_k^j), \quad s_k^j|\nu \sim \Gamma(\nu/2, \nu/2) \quad (3)$$

where s_k^j depends on the degree of freedom ν; as ν goes to infinity, s_k^j equals 1 with probability 1, and e_k^j becomes Gaussian distributed; while as ν gets closer to zero, e_k^j becomes heavy-tailed; the variance scaling variable s_k plays an important role in giving lower weight for outlier data, and therefore improve the performance of both the modeling and state estimation.

3. METHODOLOGY

In this section, the maximum likelihood (ML) estimation approach is employed to estimate both the parameters and the hidden variables in the fouling model (1).

3.1 Revisit of EM algorithm

The log likelihood can be written as a function of the parameter set Θ as follows:

$$L(\Theta) = \ln P(C_{obs}|\Theta) = \ln \int P(C_{obs}, C_{mis}|\Theta)dC_{mis} \quad (4)$$

where, C_{obs} is the observed data set, and C_{mis} is the missing/hidden data or state set; the hidden variables may make maximizing (4) difficult, and the integral over hidden variables can be intractable (Beal et al., 2003). By introducing an auxiliary distribution over missing data $q(C_{mis})$, a tractable lower bound $F(q(C_{mis}), \Theta)$ on $L(\Theta)$ is obtained using Jensen's inequality, given as:

$$\begin{aligned} L(\Theta) \geqslant &\int q(C_{mis}) \ln P(C_{obs}, C_{mis}|\Theta)dC_{mis} - \\ &\int q(C_{mis}) \ln q(C_{mis})dC_{mis} \equiv F(q(C_{mis}), \Theta) \end{aligned} \quad (5)$$

A variant ML approach called the Expectation Maximization (EM) algorithm solves the parameter estimation problem in the presence of hidden variables (Dempster et al., 1977). The EM algorithm maximizes the lower bound $F(q(C_{mis}), \Theta)$ in an iterative framework, which consists of two iterative procedures as the E step and the M step. In the E step, it infers posterior distributions over hidden variables given the previous parameter estimate by maximizing $F(q(C_{mis}), \Theta)$ with respect to $q(C_{mis})$; the part that is a function of Θ in the lower bound (5) is known as Q function $Q(\Theta|\Theta_{(n)})$; in the M step, new parameter set $\Theta_{(n+1)}$ is obtained by maximizing the lower bound or the Q function with respect to Θ.

$$\begin{aligned} &E \text{ step: } q(C_{mis})_{(n)} = \arg\max_{q(C_{mis})} F(q(C_{mis}), \Theta_{(n)}) \\ &\quad Q(\Theta|\Theta_{(n)}) = E_{q(C_{mis})_{(n)}} \ln P(C_{obs}, C_{mis}|\Theta) \\ &M \text{ step: } \Theta_{(n+1)} = \arg\max_{\Theta} Q(\Theta|\Theta_{(n)}) \end{aligned} \quad (6)$$

Here, the subscript (n) denotes the iteration number; starting from an initial parameter set $\Theta_{(0)}$, the iterations repeat until some convergence criteria are met, and $L(\Theta)$ reaches to a local maxima (Moon, 1996). More discussion about initialization and convergence of the EM algorithm can be found in McLachlan and Krishnan (2007).

3.2 ML estimation approach using EM algorithm

In the batch-based fouling model (1), C_{obs} consists of the fouling indicator and the recipe identity throughout the historical batch series $Y = \{y_k^j|k \in K, j \in J\}$, $R = \{r_k^j|k \in K, j \in J\}$, while C_{mis} is the set of the fouling states and the variance scaling variables ($X = \{x_k^j|k \in K, j \in J\}$, $S = \{s_k^j|k \in K, j \in J\}$); the model parameter set Θ includes the state equation parameters $\{(A_j,\ B_j)|j \in J\}$, the measurement equation parameters $\{(C_i,\ D_i)|i \in I\}$, the initial state and the noise term parameters $\{\mu_x, \sigma_x^2, \sigma_e^2, \sigma_{wj}^2, \nu|j \in J\}$. To start with the application of the EM algorithm, the Q function is derived as follows:

$$\begin{aligned} &Q(\Theta|\Theta_{(n)}) = E_{q(S,X)_{(n)}} \ln P(Y, S, X, R|\Theta) \\ =&E_{q(S,X)_{(n)}}\{\sum_{j=1}^{L}\sum_{k=1}^{N_j-1} \ln P(x_{k+1}^j|x_k^j, \Theta) \\ &+ \sum_{j=1}^{L}\sum_{k=1}^{N_j}\sum_{i=1}^{M} P(r_k^j = i) \ln P(y_k^j|x_k^j, s_k^j, r_k^j = i, \Theta) \\ &+ \sum_{j=1}^{L} \ln P(x_1^j|\Theta) + \sum_{j=1}^{L}\sum_{k=1}^{N_j} \ln P(s_k^j|\Theta)\} + C \end{aligned} \quad (7)$$

where C is the constant term $\ln P(R)$. The log likelihood functions are further derived as follows:

$$\ln P(y_k^j|x_k^j, s_k^j, r_k^j = i, \Theta) = -\frac{1}{2}\ln\frac{\sigma_e^2}{s_k^j} - \frac{(y_k^j - C_i x_k^j - D_i)^2}{\sigma_e^2/s_k^j}$$

$$\ln P(x_{k+1}^j|x_k^j, \Theta) = -\frac{1}{2}\ln\sigma_w^2 - \frac{(x_{k+1}^j - A_j x_k^j - B_j)^2}{\sigma_{wj}^2}$$

$$\ln P(x_1^j|\Theta) = -\frac{1}{2}\ln\sigma_x^2 - \frac{(x_1^j - \mu_x)^2}{\sigma_x^2}$$

$$\ln P(s_k^j|\Theta) = -\ln\Gamma(\frac{\nu}{2}) + \frac{1}{2}\nu\ln(\frac{\nu}{2}) + (\frac{\nu}{2} - 1)\ln s_k^j - \frac{\nu}{2}s_k^j$$

The expectation terms in the Q function are computed using the posterior distribution $q(S, X)_{(n)}$, where the derivation of the E step (state estimation) are presented in Section 3.3. In the M step, the Q function is maximized with respect to Θ, and the derived parameter updating expressions are presented in Appendix A. However, it is difficult to find an analytical solution for the parameter ν due to the complex nonlinear structure. Instead, a numerical optimization method is employed to obtain ν from the following nonlinear programming problem:

$$\nu^{(n+1)} = \arg\max_{\nu} \sum_{j=1}^{L} \sum_{k=1}^{N_j} E_{q(s_k^j)_{(n)}} \ln P(s_k^j|\nu) \quad (8)$$

which is solved numerically as a nonlinear programming problem, using the MATLAB fmincon function.

3.3 Robust state estimation with t-distributed noise terms

In this section, an approximate inference to the state estimation with t-distributed noise terms is presented. Given the estimate model parameter $\Theta_{(n)}$, the robust state estimation is carried out as the E step when implementing the EM algorithm. Considering the independence of states in different campaigns, the state estimation is implemented for campaigns separately. The superscript and subscript j denoting the campaign index are omitted for notational simplicity, and the model representation is given as:

$$\begin{aligned} x_{k+1} &= Ax_k + B + w_k \\ y_k &= C_{r_k} x_k + D_{r_k} + e_k \end{aligned} \quad (9)$$

In the E step, maximizing the lower bound of the log likelihood $L(\Theta_{(n)})$, $F(q(X,S), \Theta_{(n)})$, with respect to $q(X,S)$ yields $q(X,S) = P(X,S|Y,R,\Theta_{(n)})$. The hidden variables X and S are coupled and present difficulties in calculating the exact posterior probability analytically. An approximation approach is considered by adopting the variational appropriation (Beal et al., 2003), where the posterior is constrained to be a simpler factorized approximate $q(X,S) \approx q(X)q(S)$. With the absence of the statistical dependencies between state X and scaling variable sequences S, analytical approximations of $q(X)$ and $q(S)$ are available in an iterative framework. The local maxima of the log likelihood lower bound is reached and the state estimates converge after a number of iterations. The new lower bound $F(q(X), q(S), \Theta_{(n)})$ is derived according to the equation (5), given as:

$$\begin{aligned} F(q(X), q(S), \Theta_{(n)}) = \int q(S) \Bigg[\int q(X) \cdot \\ \ln \frac{P(Y,X,R|S,\Theta_{(n)})}{q(X)} dX + \ln \frac{P(S|\Theta_{(n)})}{q(S)} \Bigg] dS \end{aligned} \quad (10)$$

By taking functional derivatives of $F(q(X), q(S), \Theta_{(n)})$ with respect to both $q(X)$ and $q(S)$, the lower bound is optimized iteratively using the following updates:

$$\begin{aligned} q(X)_{(t+1)} &= C_X \exp[E_{q(S)_{(t)}} \ln P(Y,X,R|S,\Theta_{(n)})] \\ q(S)_{(t+1)} &= C_S \exp[E_{q(X)_{(t+1)}} \ln P(Y,S,R|X,\Theta_{(n)})] \end{aligned} \quad (11)$$

where, subscript (t) denotes the iteration number for the variational approximation; C_X and C_S are the normalization constants to force the normalization of $q(X)$ and $q(S)$; $q(X)_{(t+1)}$ is obtained by fixing $q(S)$ as $q(S)_{(t)}$, and $q(S)_{(t+1)}$ is obtained by fixing $q(X)$ as $q(X)_{(t+1)}$. As $q(S)$ can be factorized in this form: $q(S) = \prod_{k=1}^{N} q(s_k)$, $q(s_k)$ is further derived as:

$$\begin{aligned} q(s_k)_{(t+1)} =& C_{s_k} \exp[E_{q(x_k)_{(t)}} \ln P(y_k|x_k, r_k, s_k, \Theta_{(n)}) + \\ & \ln P(s_k|\Theta_{(n)})] \end{aligned} \quad (12)$$

The conjugate prior distribution of s_k is Gamma distribution, and posterior $q(s_k)$ is also a Gamma distribution (Liu and Rubin, 1995; Fang and Jeong, 2008); therefore, the distribution parameter of $q(s_k)$ is derived by rearranging the exponential terms in (12), given as:

$$s_k|_{q(\cdot)} \sim \Gamma\left(\frac{\nu+1}{2}, \frac{\nu+\delta_k^{(t)}}{2}\right)$$

where, $\delta_k^{(t)}$ is the information term including y_k and $q(x_k)_{(t)}$, denoting as $\delta_k^{(t)} = E_{q(x_k)_{(t)}}(y_k - C_{r_k}x_k - D_{r_k})^2/\sigma_e^2$; the expectation terms in $\delta_k^{(t)}$ are calculated and presented in Appendix B. The distribution $q(X)$ cannot be marginalized in a similar way as $q(S)$ due to the state dynamics. Instead, the parameter of the marginalized distribution $q(x_k)$ and $q(x_k, x_{k-1})$ are calculated using Kalman methods (Shumway and Stoffer, 2011). Given the state equation structure and the expectation terms $E_{q(s_k)_{(t)}}(s_k)$, the Kalman filter and Kalman smoother are derived and presented in Appendix B.

The posteriors $q(X)_{n+1}$ and $q(S)_{n+1}$ are obtained once the algorithm converged. More discussion about the convergence criteria can be found in Agamennoni et al. (2012). The corresponding expectation terms required in the Q function are calculated accordingly (Liu and Rubin, 1995; Shumway and Stoffer, 2011), given as

$$\begin{aligned} E(x_k^2) =& x_{k|N}^2 + P_{k|N} \\ E(x_{k+1}x_k) =& x_{k+1|N}x_{k|N} + P_{k+1,k|N} \\ E_{q(s_k)_{t+1}}(s_k) =& (\nu+1)/(\nu+\delta_k) \\ E(\ln s_k) =& \phi\left(\frac{\nu+1}{2}\right) + \ln\left(\frac{\nu+\delta_k}{2}\right) \end{aligned} \quad (13)$$

where, $E(\cdot)$ here denotes expectation $E_{q(X)_{(n+1)}q(S)_{(n+1)}}(\cdot)$ and the distribution parameters are calculated in Appendix B. Due to the factorization of the posteriors, the joint expectation terms are decomposed as follows:

$$\begin{aligned} E(x_k s_k) =& E(x_k)E(s_k) \\ E(x_k^2 s_k) =& E(x_k^2)E(s_k) \end{aligned} \quad (14)$$

The overall procedures of the EM-based state estimation approach is presented in Table 1. The iterations in the E-step considers the relative change in the lower bound $F(q(X), q(S), \Theta_{(n)})$ of two neighbouring iterations as the convergence criteria and terminates if the value reaches a predefined tolerance. Similar for the iterations of the EM algorithm, the relative changes in the parameters or the Q function are monitored as the convergence condition. The robust state estimation algorithm presented here is generic and can adapt to other similar systems.

4. RESULTS

In this section, the proposed fouling KPI estimation method is applied to the historical data set, which consists of 11 campaigns with the total batch number 368 and four recipe groups as illustrated in Fig. 3. In this example, the state X, the recipe-independent fouling KPI, is set equal to the value of the fouling indicator when using recipes from RG 1, where C_1 and D_1 are fixed as one and zero, respectively. The reason for choosing RG 1 is because the batch runs using RG 1 are in a majority and are frequently scattered in each campaign. Considering recipes' intercept effect alone, other scaling parameters $C_{2:4}$ are fixed as one. Then, the physical interpretation of the intercept terms

$D_{2:4}$ is the constant deviation on the fouling indicator y_k of batches that use different recipes from ones in RG 1.

The estimates of the model parameters and the batch series of the fouling KPI are obtained by implementing the algorithm on the case study as table 1 shows. The parameters $\{A_{1:11}, B_{1:11}\}$, which are used to fit the varying batch-to-batch evolution of fouling in different campaigns, have little use for the KPI interpretation, while the intercept terms $D_{1:4} = [0, -0.05, 0.8, 0.42]$ show clear deviations of the indicator because of the variations in batch recipes. Two campaign examples are presented to show the results of the state estimation as Fig. 4 and Fig. 5 illustrate. The dotted line denotes the fouling indicator with recipe-based deviations, and the four symbols denote the recipe type of each batch that the fouling indicator corresponds to. The circle with dashed line denotes the estimated fouling KPI, while the dashed line refers to the smoothed output, the fouling indicator. The batch-series estimates for the recipe-independent fouling KPI present a good fit into the series of the fouling indicator and the resistance to the outliers as Fig. 4 and Fig. 5 illustrate. In additions, the computation for this example takes about five minutes CPU time with over one thousand iterations for the EM algorithm, which is feasible for the off-line use to improve the indication of fouling. The estimated recipe-independent fouling KPI can be further used in the modeling and optimization of the batch process with explicit consideration of fouling effects.

5. CONCLUSION

The problem of robust state estimation with unknown model parameters is discussed in this paper and is illustrated using an industrial example. An ML estimation approach is developed to solve the state estimation problem. The robustness of estimation, which refers to the ability for the resistance to outliers, is handled using t-distributed noise terms. Because of both the unknown parameters and

Table 1. Implementation of EM algorithm on fouling modeling and state estimation

Input: observed data Y, R

Output: estimates of hidden state X and parameters Θ

Initialization: parameter initial guess Θ_0

Repeat: $n = n + 1$
E-step: robust state estimation on Model (1) given $\Theta_{(n)}$.
Repeat: $t = t + 1$
for j = 1,2,...,L
for k = 1,2,...,N_j
Kalman filter (B.1) given $q(s_k^j)_{(t)}$.
end
for k = N_j,$N_j - 1$,...,1
Calculate $q(x_k^j)_{(t+1)}$ and $q(x_k^j, x_{k-1}^j)_{(t+1)}$ using Kalman smoother (B.2).
end
for k = 1,2,...,N
Calculate $q(s_k)_{(t+1)}$ using (B.3) given $q(x_k)_{(t+1)}$.
end
end
Until the state estimation converged
Calculate Q function (7) using expectation terms (13, 14) given $q(S)_{(n)}$, $q(X)_{(n)}$.
M-step: maximize Q function and update parameter set $\Theta_{(n+1)}$ using (A.1-A.8, 8).
Until the EM algorithm converged

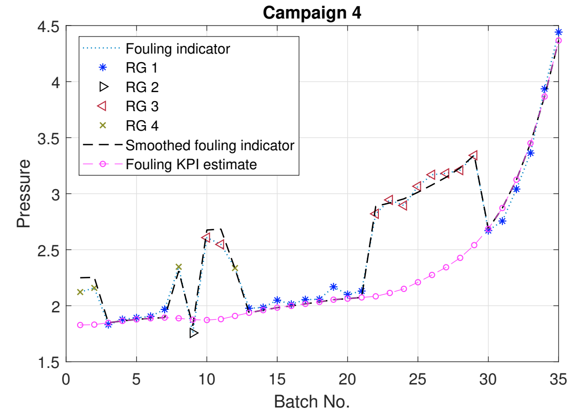

Fig. 4. Estimated KPI series in Campaign 4

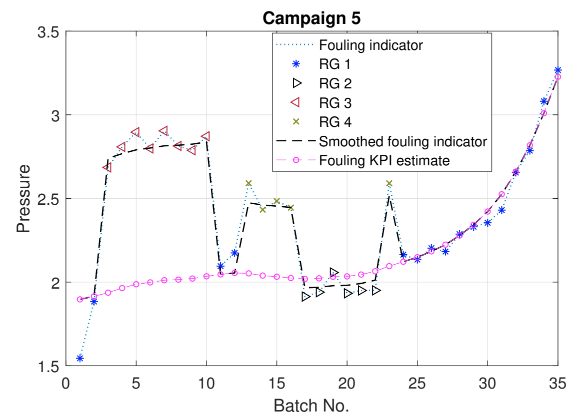

Fig. 5. Estimated KPI series in Campaign 5

states, the EM algorithm is applied to compute the ML estimation of the corresponding states and model parameters iteratively. The variational approximation method is used in the E step to calculate analytical approximation of the states given the t-distributed noise terms. To solve a practical problem of fouling indication, the state estimation approach is tailored to calculate the recipe-independent fouling KPI from the case study example, and the results further show the efficacy of the proposed method.

REFERENCES

Agamennoni, G., Nieto, J.I., and Nebot, E.M. (2012). Approximate inference in state-space models with heavy-tailed noise. *IEEE Transactions on Signal Processing*, 60(10), 5024–5037.

Beal, M.J. et al. (2003). *Variational algorithms for approximate Bayesian inference*. Ph.D. thesis, university of London.

Bonvin, D. (2006). Control and optimization of batch processes. *IEEE control systems*, 26(6), 34–45.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Fang, Y. and Jeong, M.K. (2008). Robust probabilistic multivariate calibration model. *Technometrics*, 50(3), 305–316.

Gibson, S. and Ninness, B. (2005). Robust maximum-likelihood estimation of multivariable dynamic systems.

Automatica, 41(10), 1667–1682.
Ibrahim, H.A.H. (2012). Fouling in heat exchangers. In *MATLAB-A Fundamental Tool for Scientific Computing and Engineering Applications- Volume 3*. InTech.
Liu, C. and Rubin, D.B. (1995). Ml estimation of the t distribution using em and its extensions, ecm and ecme. *Statistica Sinica*, 19–39.
McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.
Moon, T.K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6), 47–60.
Radhakrishnan, V., Ramasamy, M., Zabiri, H., Do Thanh, V., Tahir, N., Mukhtar, H., Hamdi, M., and Ramli, N. (2007). Heat exchanger fouling model and preventive maintenance scheduling tool. *Applied Thermal Engineering*, 27(17-18), 2791–2802.
Rodriguez, C. and Smith, R. (2007). Optimization of operating conditions for mitigating fouling in heat exchanger networks. *Chemical Engineering Research and Design*, 85(6), 839–851.
Shumway, R.H. and Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis*, 3(4), 253–264.
Shumway, R.H. and Stoffer, D.S. (2011). State-space models. In *Time series analysis and its applications*, 319–404. Springer.
Teruel, E., Cortes, C., Diez, L.I., and Arauzo, I. (2005). Monitoring and prediction of fouling in coal-fired utility boilers using neural networks. *Chemical Engineering Science*, 60(18), 5035–5048.
Wu, O., Bouaswaig, A.E., Schneider, S.M., Leira, F.M., Imsland, L., and Roth, M. (2018). Data-driven degradation model for batch processes: a case study on heat exchanger fouling. In *Computer Aided Chemical Engineering*, volume 43, 139–144. Elsevier.
Zhu, H., Leung, H., and He, Z. (2013). A variational bayesian approach to robust sensor fusion based on student-t distribution. *Information Sciences*, 221, 201–214.
Zia, A., Kirubarajan, T., Reilly, J.P., Yee, D., Punithakumar, K., and Shirani, S. (2008). An em algorithm for nonlinear state estimation with model uncertainties. *IEEE Transactions on Signal Processing*, 56(3), 921–936.

Appendix A. M-STEP: UPDATES

The parameter updates in the M step are presented as follows:

$$A_j^{(n+1)} = \frac{\sum_{k=1}^{N_j-1}[E(x_{k+1}^j x_k^j) - B_j^{(n)} E(x_k^j)]}{\sum_{k=1}^{N_j-1} E((x_k^j)^2)} \tag{A.1}$$

$$B_j^{(n+1)} = \frac{\sum_{k=1}^{N_j-1}[E(x_{k+1}^j) - A_j^{(n+1)} E(x_k^j)]}{N_j - 1} \tag{A.2}$$

$$C_i^{(n+1)} = \frac{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i)[y_k^j E(x_k^j s_k^j) - D_i^{(n)} E(x_k^j s_k^j)]}{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i) E((x_k^j)^2 s_k^j)} \tag{A.3}$$

$$D_i^{(n+1)} = \frac{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i)[y_k^j E(s_k^j) - C_i^{(n+1)} E(x_k^j s_k^j)]}{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i) E(s_k^j)} \tag{A.4}$$

$$\sigma_{wj}^{2\ (n+1)} = \frac{\sum_{k=1}^{N_j}\sum_{i=1}^{M} P(r_k^j = i)\Delta_{wj}}{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i)} \tag{A.5}$$

$$\sigma_e^{2(n+1)} = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N_j}\sum_{i=1}^{M} P(r_k^j = i)\Delta_e}{\sum_{j=1}^{L}\sum_{k=1}^{N_j} P(r_k^j = i) E(s_k^j)} \tag{A.6}$$

$$\mu_x^{(n+1)} = \frac{\sum_{j=1}^{L} E(x_1^j)}{L} \tag{A.7}$$

$$\sigma_x^{(n+1)} = \frac{\sum_{j=1}^{L}[E(x_1^j) - \mu_x^{(n+1)}]^2}{L} \tag{A.8}$$

where, $E(\cdot)$ denotes $E_{q(X)_{(n+1)}q(S)_{(n+1)}}(\cdot)$,

$$\Delta_e = \sum_{j=1}^{L}\sum_{k=1}^{N_j}\sum_{i=1}^{M} P(r_k^j = i)\{[(y_k^j)^2 - 2D_i^{(n+1)} y_k^j + (D_i^{(n+1)})^2]\cdot E(s_k^j) + (C_i^{(n+1)})^2 E[(x_k^j)^2 s_k^j] + 2C_i^{(n+1)}(D_i^{(n+1)} - y_k^j) E(x_k^j s_k^j)\}$$

$$\Delta_{wj} = \sum_{k=1}^{N_j-1}\{E[(x_{k+1}^j)^2] - 2B_j^{(n+1)}\cdot E(x_{k+1}^j) + (B_j^{(n+1)})^2 + (A_j^{(n+1)})^2 E[(x_k^j)^2] + 2A_j^{(n+1)}\cdot [B_j^{(n+1)} E(x_k^j) - E(x_{k+1}^j x_k^j)]\}$$

Appendix B. E STEP: ITERATIVE KALMAN UPDATES

The Kalman filter and Kalman smoother as well as lag-one covariance are presented in equations (B.1) and (B.2) (Shumway and Stoffer, 2011).

$$\begin{aligned} x_{k|k-1}^{(t+1)} &= A x_{k-1|k-1}^{(t+1)} + B \\ P_{k|k-1}^{(t+1)} &= A P_{k-1|k-1}^{(t+1)} A^\top + Q_k^{(t)} \\ K_k^{(t+1)} &= P_{k|k-1}^{(t+1)} C_{r_k}^\top (C_{r_k} P_{k|k-1}^{(t+1)} C_{r_k}^\top + R_k)^{-1} \\ x_{k|k}^{(t+1)} &= x_{k|k-1}^{(t+1)} + K_k^{(t+1)}(y_k - C_{r_k} x_{k|k-1}^{(t+1)} - D_{r_k}) \\ P_{k|k}^{(t+1)} &= (I - K_k^{(t+1)} C_{r_k}) P_{k|k-1}^{(t+1)} \end{aligned} \tag{B.1}$$

where, $Q_k = \sigma_w^2$, $R_k^{(t)} = \sigma_e^2 / E_{q(s_k)_t}(s_k)$; the superscript (t) represents the iteration index of the variationl approximation method. Kalman smoother is then calculated based on the result of Kalman filter:

$$\begin{aligned} J_{k-1}^{(t+1)} &= P_{k-1|k-1}^{(t+1)} A (P_{k|k-1}^{(t+1)})^{-1} \\ x_{k-1|N}^{(t+1)} &= x_{k-1|k-1}^{(t+1)} + J_{k-1}^{(t+1)}(x_{k|N}^{(t+1)} - x_{k|k-1}^{(t+1)}) \\ P_{k-1|N}^{(t+1)} &= P_{k-1|k-1}^{(t+1)} + J_{k-1}^{(t+1)}(P_{k|N}^{(t+1)} - P_{k|k-1}^{(t+1)}) J_{k-1}^{(t+1)\top} \\ P_{k-1,k-2|N}^{(t+1)} &= P_{k-1|k-1}^{(t+1)} J_{k-2}^{(t+1)\top} + J_{k-1}^{(t+1)}(P_{k,k-1|N}^{(t+1)} - A P_{k-1|k-1}^{(t+1)}) J_{k-2}^{(t+1)\top} \end{aligned} \tag{B.2}$$

where, the initial lag-one covariance at time N is calculated as $P_{N,N-1|N}^{(t+1)} = (I - K_N^{(t+1)} C_{I_N}) A P_{N-1|N-1}^{(t+1)}$; the calculation of $E_{q(s_k)_{(t)}}(s_k)$ is based on (12):

$$E_{q(s_k)_{(t)}}(s_k) = \frac{\nu + 1}{\nu + \delta_k^{(t)}} \tag{B.3}$$

The expectation terms in $\delta_k^{(t)}$ are calculated as:

$$\begin{aligned} E_{q(x_k)_{(t)}}(x_k) &= x_{k|N}^{(t)} \\ E_{q(x_k)_{(t)}}(x_k^2) &= x_{k|N}^{(t)\ 2} + P_{k|N}^{(t)} \end{aligned} \tag{B.4}$$