Anton Friedl, Jiří J. Klemeš, Stefan Radl, Petar S. Varbanov, Thomas Wallek (Eds.)
Proceedings of the 28th European Symposium on Computer Aided Process Engineering
June 10th to 13th, 2018, Graz, Austria.

Data-driven degradation model for batch processes: a case study on heat exchanger fouling

Ouyang Wu[a,b,*], Ala E.F. Bouaswaig[a], Stefan M. Schneider[a], Fernando Moreno Leira[a], Lars Imsland[b] and Matthias Roth[a]

[a]*Automation Technology, BASF SE, 67056 Ludwigshafen, Germany*
[b]*Department of Engineering Cybernetics, NTNU, 7491 Trondheim, Norway*
ouyang.wu@ntnu.no

Abstract

In this paper, we present an industrial case study of a batch process; the degradation effect is fouling in heat exchangers which impedes heat transfer and fluid flow during batch production. The degradation increases from batch to batch until the batch reactor is shut down for a scheduled maintenance. That is, there is a periodic pattern in the batch-to-batch data as a result of the degradation. We propose a novel data-driven modeling technique for predicting the evolution of a batch degradation key performance indicator (KPI). This method extends existing algorithms (based on partial least squares, PLS) for the treatment of within-batch data to the present batch-to-batch degradation problem. An existing missing data algorithm for PLS models is employed to predict degradation in future batches. The proposed approach is then applied to the case study, and predicts the evolution of the fouling KPI with respect to the production planning under predefined process operations, which is a basis for optimal scheduling of batch production and maintenance.

Keywords: fouling, batch process, degradation evolution, unfolding approach, multiway PLS

1. Introduction

The performance and efficiency degradation of industrial equipment and assets is inevitable in their life cycles. The assessment of process condition provides information of degradation, and the following optimization of process operations is meant to mitigate the influence of degradation. In some cases, it is necessary to shut down the plant and carry out the maintenance operation for restoration when the degradation reaches a certain degree. Human experience-based maintenance scheduling is not necessarily optimal on account of degradation in production along with other physical and logistical constraints. However, optimal scheduling asks for information about future degradation evolution. Therefore, the development of a predictive degradation model plays a significant role in plant performance optimization in consideration of degradation effects.

A batch reactor is a multi-purpose unit operation that is widely used in process industry to manufacture diverse products. The flexibility of batch processes is the major reason for its popularity in the chemical industry. By using batch production plants industry can easily adapt to market fluctuations and extend their product portfolio to react to customer needs. A multi-product batch process runs in a finite time period. It includes several stages: the raw materials are charged and the initial conditions are set according to the recipe; then the reaction starts and lasts for some time; the product is transferred into a storage tank and the product final quality variables are recorded at the end of this batch. During the batch running period, the process variables such as temperature, pressure, etc. are measured in real-time, and their profiles are called trajectories. These trajectory data are presented in three-way form with time $t = 1,2,..,T$, variable $k = 1,2,...,K$

and batch $n = 1,2,...,N$, while the initial condition variables and the final quality variables are in two-way form with batch $n = 1,2,...,N$ and variable $k = 1,2,...,K$ (Nomikos and MacGregor, 1994). In comparison with continuous processes, monitoring, control and optimization of batch processes are more challenging. These challenges arise from the problems that are associated with batch processes: absence of steady state (wide operating range) and presence of constraints, irreversible behavior, repetitive nature (Bonvin et al., 2006). In the presented case study, the reaction heat is removed by pumping the reactant through the heat exchangers, and the fouling in the heat exchangers impedes the fluid flow and leads to the increase of the pressure drop across the heat exchanger during the reaction period. As a result, the pressure drop is chosen as a fouling indicator for the heat exchanger (see figure 1). However, the contribution of the fouling towards the real-time pressure drop measurement is hidden and the increasing trend of fouling is masked within the individual batch duration, which results from the variant batch operation conditions. To deal with this problem, the pressure drop measured under selected operation conditions (at the start of the reaction) is considered as the fouling KPI for each individual batch as figure 1 illustrates. The fouling evolution from batch to batch is then indicated by the batch fouling KPIs.

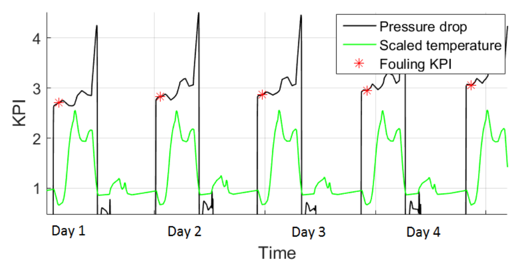

Figure 1: Fouling indication by the pressure drop across the heat exchanger

As for the degradation prediction in continuous process, a variety of degradation examples using different data-driven modeling approaches (such as support vector machine, neural network, etc.) are found in literature (Sun et al., 2008; Aminian and Shahhosseini, 2008; Riverol and Napolitano, 2005). Due to the discrete characteristic of the degradation KPI in batch processes, these approaches do not deal well with the multi-dimension batch data structure. On the other hand, the multiway approach was proposed to solve batch data structure problems by transforming three-dimension data into batch-wise unfolded data, and the multiway PLS approach is then employed in the within-batch modeling to predict the final quality of the individual batches (Nomikos and MacGregor, 1994; García-Muñoz et al., 2003; Wold et al., 2010). As to degradation in batch production, unlike final quality variables barely affected by previous batches, degradation grows from batch to batch. Moreover, degradation prediction usually means to predict degradation evolution in next N batches, and the time horizon for prediction can be longer than one month, while the application of the multiway PLS approach using the within-batch model is only available to predict the final degradation in a single batch.

In this paper, a novel data structure called “campaign” is proposed for the analysis of degradation evolution from batch to batch. A data-driven model is developed based on the unfolded campaign data. Finally, an application of the PLS method and its missing data estimation algorithm enable the prediction of degradation evolution. The paper structure is presented as follows: the proposed approach details are presented in Section 2; the corresponding application on the case study and its results are showing in Section 3; finally, Section 4 draws the conclusions.

2. Method

2.1. The campaign concept in batch processes

In some cases, the batch degradation KPI presents periodic nature as maintenance is carried out to restore its degradation state regularly (see Figure 2). The figure shows the evolution of the fouling KPI in the heat exchanger. Each point represents the fouling degree in an individual batch, and the color and symbol together denote the batch product type from P1 to P9. The product order is planned ahead of time and flexible to modify as necessary. The period between two

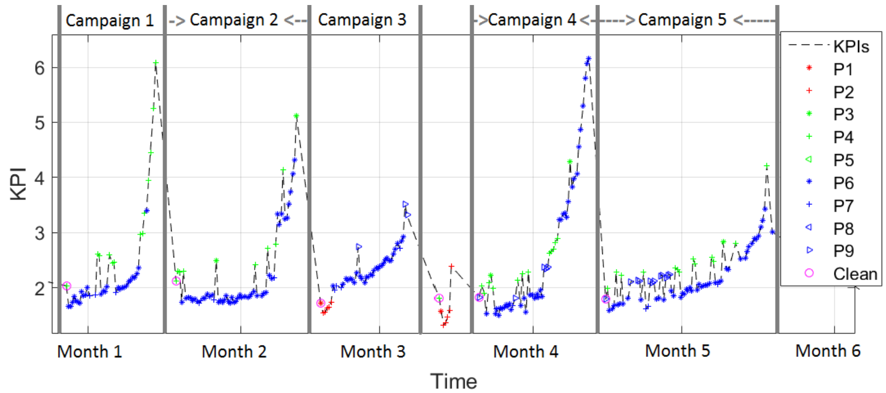

Figure 2: Heat exchanger fouling evolution in batch process and campaigns

maintenance operations is named campaign, which contains a series of batches. The idea of the campaign concept is to capture the dynamics of the degradation evolution. To make effective use of campaign data, further data structure analysis is necessary. The structure of campaign data includes individual batch data firstly, where batch trajectory data have three dimensions: batch, time and variable. Secondly, campaign is a "batch" consisting of a series of batches connected by the degradation evolution.

2.2. Campaign PLS approach and campaign data unfolding method

As one campaign consists of a series of batches, the campaign data structure has four dimensions (campaign, batch, variable and time). The modeling approaches for continuous processes are not applicable to the multi-dimension campaign process, and it requires an unfolding method to transform the multi-dimension campaign data into two-dimension data . In the multiway analysis, the batch-wise unfolding method is developed to transform the three-dimension batch data into two-dimension unfolded data: batch data are extracted horizontally in a time-wise fashion and each batch becomes a single row of data (Nomikos and MacGregor, 1994; García-Muñoz et al., 2003; Wold et al., 2010):

$$X_{batch}^{m} = [xi_1, xi_2, ..., xi_{ki}, xt_1^1, xt_2^1, ..., xt_T^1, xt_1^2, xt_2^2, ..., xt_T^2, ..., xt_1^{kt}, xt_2^{kt}, ..., xt_T^{kt}] \tag{1}$$

where, X_{batch}^{m} is the unfolded batch data of the mth batch, which includes initial conditions and unfolded trajectories: the initial condition variables xi_k, $k = 1, 2, ...ki$ is the index of initial condition variables; trajectory data xt_t^k, $k = 1, 2, ..., kt$ is the index of trajectory variables, $t = 1, 2, ..., T$ is the time index in a single batch.

In a similar methodology, a campaign unfolding method is proposed to obtain 2-dimension unfolded campaign data. The variable and time dimensions are unfolded firstly to obtain the unfolded batch data X_{batch}^{m} using the batch-wise unfolding method. Then the three-dimension campaign data is unfolded as Figure 3 illustrates, where the campaign data are extracted horizontally in a batch-wise fashion and each campaign becomes a single row of different unfolded batch data. For simplicity, unfolded batch data are presented with the index $k = 1, 2, ..., K$. Further, the unfolded campaign data are employed for the degradation modeling. To build a degradation model based on campaign data, the outputs are the evolution of degradation KPIs, and the inputs are the unfolded campaign data which contain control and process information of a single campaign.

In this section, PLS is employed for the campaign-based degradation modeling due to its advantage in dealing with a large number of correlated predictor variables and avoiding over-fitting problem. In PLS, the highly correlated data in the input and output space are modeled separately with orthogonal principal components based on multivariate statistical projection, and linear regression

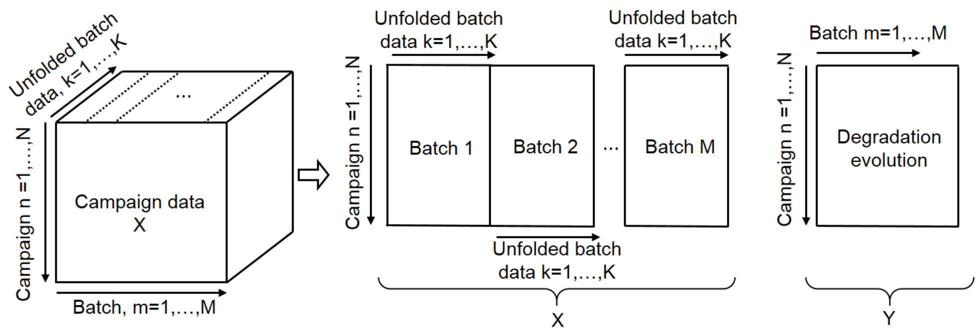

Figure 3: Campaign data structure and unfolding method

is employed on those dimension-reduced principal components (Geladi and Kowalski, 1986; Abdi, 2010). The campaign-based PLS model is:

$$\begin{aligned} X_n &= T_n P^T + E_n, \quad X_n = [X_1^n, X_2^n, ..., X_M^n],\ X_m^n = [x_1^m, x_2^m, .., .x_{k1}^m, z_1^m, z_2^m, ..., z_{k2}^m] \\ Y_n &= T_n C^T + F_n, \quad Y_n = [y_1^n, y_2^n, ..., y_M^n] \end{aligned} \tag{2}$$

where, $X_n \in \mathbb{R}^{1\times M(k_1+k_2)}$ is the nth unfolded campaign; the mth unfolded batch data from the nth campaign $X_m^n \in \mathbb{R}^{1\times(k_1+k_2)}$ are divided into manipulated data z_k^m, $k = 1,2,...,k_2$ and process data $x_k^m, k = 1,2,...,k_1$, while k is the corresponding index; the output $Y_n \in \mathbb{R}^{1\times M}$ is a series of degradation KPI in a campaign, and the element y_m^n is the degradation KPI of mth batch in nth campaign; $T_n \in \mathbb{R}^{1\times L}$ is the score of the nth campaign projecting on the principal component, while L is the dimension; $P \in \mathbb{R}^{M(k_1+k_2)\times L}$ and $C \in \mathbb{R}^{M\times L}$ are the model parameters, called loading matrix, which can be estimated through the NIPALS algorithm (Geladi and Kowalski, 1986).

The campaign-based PLS model can be used to estimate the evolution of batch degradation from the known input observations (historical process data and the manipulated data). In the batch fouling example, the manipulated data are exclusively the production planning, that is, the planned production schedule shows which products are going to be produced and the production order. Given the current batch K, the known data X_n^* are $\{X_{1:K}^n, z_{1:k2}^{K+1:M}\}$, and missing data $X_n^\#$ are $\{x_{1:k_1}^{K+1:M}\}$. The future batch degradation indicators of interest are $\{y_{K+1:M}^n\}$. The missing data estimation approaches for PLS models such as the trimmed sore regression (TSR) are meant to provide estimation of final score $\hat{T}_n$ with the known samples X_n^* (the information update to K), and then output $\hat{Y}_n$ is calculated from the estimated score $\hat{T}_n$. TSR algorithm is proved to be effective in multiway PLS applications (Arteaga and Ferrer, 2002; Nelson et al., 1996; Keivan Rahimi-Adli, 2016).

3. Case study

3.1. Case study description

The case study presents an example of heat exchanger fouling in a polymerization batch process (see the process schematic in Figure 4). This batch process produces multiple water and paper treatment chemicals. The polymerization starts when initiators are added into monomer emulsion in the reactor. The recirculation system works during the reaction, and two parallel heat exchangers help to cool down the reactor. The emulsion flow results in polymer residues depositing in the reactor and the heat exchangers. The fouling in heat exchanger impedes the flow and can even lead to a blockage, and it also decreases the cooling efficiency and prolongs batch duration with less production capacity. Therefore, the cleaning is required to remove the residues in the heat exchanger after running certain number of batches. We focus on the fouling in heat exchangers as it is the short-board of the process.

3.2. Multi-product effects on fouling KPIs

Multiple products in batch process add more complexity to degradation prediction. Different products mean different batch recipes, such as ratios of raw materials, additives and operation conditions, which results in disturbances to fouling KPIs. In figure 2, the fouling KPIs from batch to batch show an increasing trend between two cleanings, especially for the same product type. KPIs with different product type show some biases among them, for example, the green KPIs are relatively larger than the blue KPIs. This is not due to fouling evolution, and it brings more disturbances in degradation modeling. To avoid this, the product biases are calculated by comparing KPIs to some fitted curves of the main product KPIs (P6), and the bias correction is carried out by adding the product-based biases to the KPIs (see the smoothed KPIs in Figure 5).

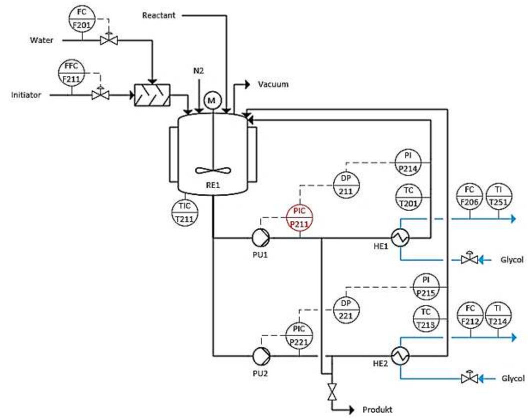

Figure 4: Case study process schematic

3.3. Results

The factors that contribute to fouling evolution in heat exchangers include the product recipes and trajectory process variables like temperature, pressure, flowrate, etc. The pressure drop across the heat exchanger is employed as a batch fouling KPI, and the multi-product biases are excluded to obtain smooth increasing fouling KPIs with less disturbances. To model the fouling KPI evolution in the heat exchanger, the unfolded campaign data is obtained using the proposed method. An equally long campaign is a perquisite for the PLS model, hence the last 30 batches of one cleaning period are collected as one campaign, which includes the dynamics of the fouling evolution. The collected historical data for modeling includes 20 campaigns. 23 input variables are selected from the unfolded batch data, which are production planning variables, landmarks of trajectory data and batch quality data (Wold et al., 2010). The output vector for each campaign includes 30 batch fouling KPIs. The unfolded campaign data X has a row of 460 input variables. The heat exchanger fouling campaign PLS model is developed based on the historical data (20 campaigns), where several campaigns are excluded for cross validation purpose. Based on this model, we can predict fouling evolution in upcoming campaigns of batches by using the missing data estimation method TSR as explained above. We illustrate this using two campaign examples in Figure 6, where three predictions are given using the data up to different stages of the campaign, and compared to the actual fouling KPIs. The first case is K=1, which means given observation data up to Batch 1 and the production planning for the whole campaign, one needs to predict the fouling evolution in the remaining 29 batches. Similar for K=10 and K=20. The fouling KPI predictions in Figure 6 follow the actual KPIs with a smooth increasing curve.

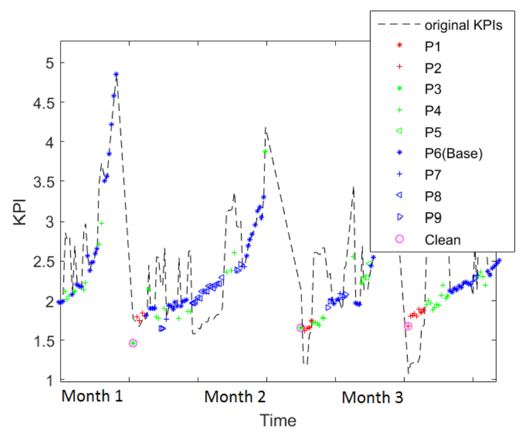

Figure 5: Bias corrected fouling indicators

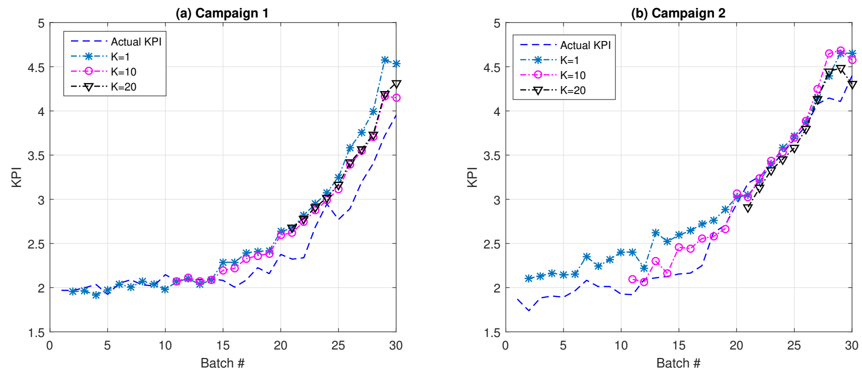

Figure 6: Two fouling KPIs campaign predictions: given information up to Batch $K = 1, 10, 20$

4. Conclusions

This paper focuses on degradation prediction for batch processes with heat exchanger fouling as the case study. To solve the degradation prediction problem, a new "campaign" concept is proposed to structure the batch degradation evolution. Furthermore, an unfolding method is proposed to transform the four-dimension data into two-dimension unfolded campaign data for PLS modeling. The campaign-based PLS approach is then applied in the case study, and the missing data estimation method TSR is employed for the degradation prediction. The validation result illustrates the effectiveness of the proposed campaign PLS approach in predicting heat exchanger fouling evolution, and this fouling predictions can provide further information for better production scheduling and maintenance planning.

Acknowledgments: Financial support is gratefully acknowledged from the Marie Skodowska Curie Horizon 2020 EID-ITN project "PROcess NeTwork Optimization for efficient and sustainable operation of Europe's process industries taking machinery condition and process performance into account PRONTO", Grant agreement No 675215.

References

H. Abdi, 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdisciplinary Reviews: Computational Statistics 2 (1), 97–106.

J. Aminian, S. Shahhosseini, may 2008. Evaluation of ANN modeling for prediction of crude oil fouling behavior. Applied Thermal Engineering 28 (7), 668–674.

F. Arteaga, A. Ferrer, 2002. Dealing with missing data in MSPC: Several methods, different interpretations, some examples. Journal of Chemometrics 16 (8-10), 408–418.

D. Bonvin, B. Srinivasan, D. Hunkeler, 2006. Control and optimization of batch processes: Improvement of process operation in the production of specialty chemicals. IEEE Control Systems 26 (6), 34–45.

S. García-Muñoz, T. Kourti, J. F. MacGregor, A. G. Mateos, G. Murphy, 2003. Troubleshooting of an industrial batch process using multivariate methods. Industrial & engineering chemistry research 42 (15), 3592–3601.

P. Geladi, B. R. Kowalski, 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185 (C), 1–17.

Keivan Rahimi-Adli, 2016. Grey-box modeling, state estimation and optimization of a semi-batch reactor. Master's thesis, TU Dortmund University, Dortmund, Germany.

P. R. C. Nelson, P. A. Taylor, J. F. MacGregor, 1996. Missing data methods in PCA and PLS: score calculations with incomplete observations. Chemometrics and Intelligent Laboratory Systems 35, 45–65.

P. Nomikos, J. F. MacGregor, 1994. Monitoring batch processes using multiway principal component analysis. AIChE Journal 40 (8), 1361–1375.

C. Riverol, V. Napolitano, may 2005. Estimation of fouling in a plate heat exchanger through the application of neural networks. Journal of Chemical Technology & Biotechnology 80 (5), 594–600.

L. Sun, Y. Zhang, X. Zheng, S. Yang, Y. Qin, 2008. Research on the fouling prediction of heat exchanger based on support vector machine. In: Intelligent Computation Technology and Automation (ICICTA), 2008 International Conference on. Vol. 1. IEEE, pp. 240–244.

S. Wold, N. Kettaneh-Wold, J. MacGregor, K. Dunn, 01 2010. Batch process modeling and mspc. In: Comprehensive Chemometrics. Vol. 2. Elsevier, pp. 163–197.