

Soft Sensor Based on Recursive Kernel Partial Least Squares for 4-carboxybenzaldehyde of an Industrial Terephthalic Acid Hydropurification Process

Zhi Li, Weimin Zhong, Xin Peng, Wenli Du, Feng Qian *

Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai, 200237, China
 fqian@ecust.edu.cn

Terephthalic acid is a raw material for polyester and textile industry. However, the by-product 4-Carboxybenzaldehyde in TA is harmful to the polymer process since it can lower the polymerization rate and the average molecular weight. Thus, a hydropurification process is built to decrease 4-CBA. In this process, 4-CBA in TA is purified by hydrogen in water at 270 °C–290 °C under 7.9 MPa pressure over 0.5 wt. % carbon-coated palladium catalyst in a fixed-bed reactor. The activity of the catalyst will gradually decrease with the process running. The most important quality index for this process is the content of 4-CBA in the product. However, in real plant, the content of 4-CBA is analysed every two hours in a laboratory. It is a very large delay for control system so the operation conditions of the process could not be adjusted in time. These may result in a maximum two hours product failure. In this paper, a first principle model of the process is developed based on Aspen Plus. The accuracy of the model is verified by the model results and the actual plant results. Based on this model, a series of sensitive analysis are performed. Six variables include 4-CBA content in TA, react temperature, react pressure, hydrogen flow rate, catalyst activity and Slurry concentration are the main factors influencing the 4-CBA content in the product. Though the aspen model is accurate, these parameters in aspen model are not easy to adjust quickly. In order to predict 4-CBA content quickly, a soft sensor for 4-CBA is developed using recursive kernel partial least squares considering the nonlinearity and slow time-varying of the process. Results show that the prediction accuracy of this method is very high and it is easy for engineers to handle.

1. Introduction

Pure terephthalic acid (PTA) is one of the most important raw materials for polymer industry (Li et al., 2016). There are two main sections in PTA process. The first is *p*-xylene (PX) oxidation process. PX in acetic acid solvent is oxidized to TA in a continuous stirred tank reactor by air or molecular oxygen (Qian et al., 2012). However, the product TA contains an impurity, namely, 4-carboxybenzaldehyde (4-CBA). TA with high content of 4-CBA is called crude terephthalic acid (CTA). 4-CBA can lower the polymerization rate and the average molecular weight in polymer process, besides, it is not easy to separate using physical method as the molecular weight of 4-CBA is similar to TA. The second section, hydropurification process is implemented to decrease 4-CBA in CTA. In this section, CTA is first mixed with deionized water in a storage tank, and then the slurry is gradually heated to about 273 °C through five continuous heat exchangers. After that, the slurry and hydrogen are injected to the top of a fixed-bed reactor. 4-CBA reacts with hydrogen and converts to *p*-toluic acid in the presence of 0.5 wt. % carbon-coated palladium (Pd/C) catalyst, and then, *p*-toluic acid is easy to separate from TA by crystallization and centrifugation (Azarpour et al., 2012). In this process, the content of 4-CBA is decrease from 3,000 ppm to less than 25 ppm. This can achieve the material requirement of polymer industry (Li et al., 2015).

In real CTA purification process, 4-CBA content in final product is the most important quality index. However, rather than observed directly, it is tested every two hours in a laboratory. This is a very large delay for the whole process, results from the laboratory can not respond to system status in real time. These may result in

a maximum two hours product failure. Although a lot of real process simulators such as Aspen, PRO-II are popular in this field, it is still not easy for engineers and workers to handle. Besides, with the progress of production, the operation conditions need to be adjusted in real time. It is not easy to change all of the conditions in a simulator. For this reason a lot of soft sensors are developed to solve these problems. The common method is data based regression algorithm, such as PCA (Farsang et al., 2015), PLS (Facco et al., 2009) and KPLS (Godoy et al., 2014), etc. CTA hydropurification process is a nonlinear process and the catalyst activity is a slow time-varying variable, it decreases gradually after the process begins. PCA and PLS can only handle linear problem, KPLS is difficult to solve slow time varying problem. Thus, a recursive kernel partial least squares method is developed. The kernel here is used to solve the non-linearity, recursive method here is used to solve the slow time varying problem. The following sections of this paper is organized as follows, in section 2, details of this process are discussed. In section 3, 6 essential variables which influence the final 4-CBA are picked out by doing a series of sensitive analysis. Section 4 describes the proposed recursive kernel partial least squares and the real plant data are applied using this method, also, discussions of the results are given in this section. Finally, conclusions are drawn in section 5.

2. CTA hydropurification process

CTA hydropurification process is of great importance in PTA process. CTA containing 3000 ppm 4-CBA from PX oxidation process and deionized water are mixed together in a storage tank, the mass fraction of the CTA slurry is about 29%, then the slurry is pumped through five continuous heat exchangers. The temperature of the slurry increases from atmosphere to 273 °C. In this pre-heat process, CTA is dissolved in the deionized water. Then the slurry and hydrogen are injected to the top of a fix-bed reactor. The reactor bed is fulfilled with 0.5.wt. % carbon-coated palladium (Pd/C) catalyst. In the reactor, 4-CBA reacts with hydrogen under the pressure of 7.9 MPa and inverts to *p*-toluic acid. The residence time is about 20 min. Next, the product goes out from the bottom of the reactor. After five continuous crystallizers, the temperature of the product decreases gradually from 273 °C to atmosphere, at the same time, PTA crystallized with the temperature goes down. In the final product, the content of 4-CBA is less than 25 ppm. The flowchart is shown in Figure 1.

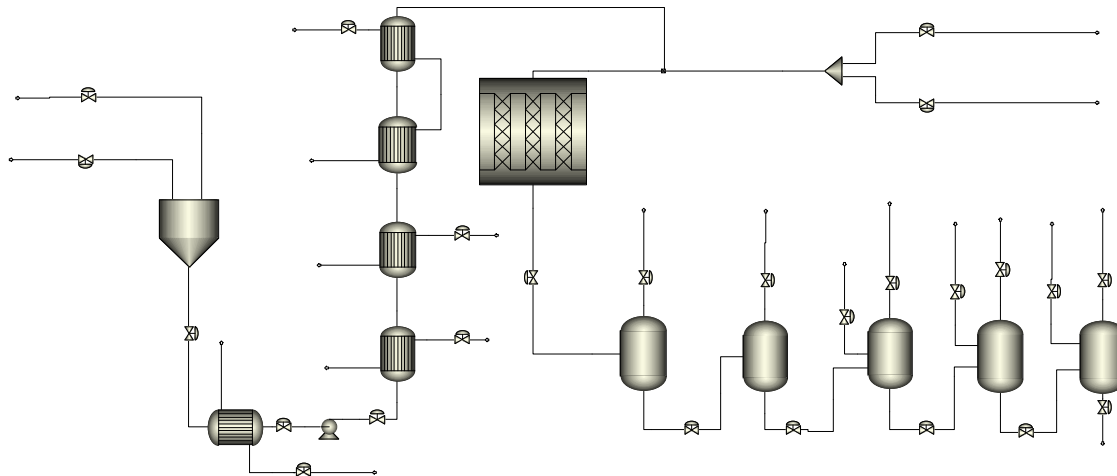


Figure 1: Process flowchart of pure terephthalic acid hydropurification process

CTA hydropurification reaction is a complicated chemical process. The reactions are shown in Eq(1) and (2). Reactions Eq(1) and Eq(2) are the main reactions. The intermediate product is 4-hydroxymethylbenzoic acid, and the final product is *p*-toluic acid. Reaction Eq(3) is a side decarbonylation reaction, and the final product is benzoic acid (Zhang et al., 2008).

The kinetics of the reactions are listed below. Eqs(3 - 7) are main and side reactions(Zhou et al., 2006a). The kinetic data are shown in Table 1 (Zhou et al., 2006b).



$$-\frac{dc_{4-CBA}}{dt} = k_{01}e^{-\frac{E_1}{RT}}C_{4-CBA}^{n_1}C_{H_2}^{n_2} \quad (3)$$

$$\frac{dc_{4-HMBA}}{dt} = k_{01}e^{-\frac{E_1}{RT}}C_{4-CBA}^{n_1}C_{H_2}^{n_2} - k_{02}e^{-\frac{E_2}{RT}}C_{4-HMBA}^{n_3}C_{H_2}^{n_4} \quad (4)$$

$$\frac{dc_{p-toluic\ acid}}{dt} = k_{02}e^{-\frac{E_2}{RT}}C_{4-HMBA}^{n_3}C_{H_2}^{n_4} \quad (5)$$

$$-\frac{dc_{4-CBA}}{dt} = k_{03}e^{-\frac{E_3}{RT}}C_{4-CBA}^{n_5} \quad (6)$$

$$\frac{dc_{BA}}{dt} = k_{03}e^{-\frac{E_3}{RT}}C_{4-CBA}^{n_5} \quad (7)$$

Table 1: Kinetic data of the reactions

Item	Data in reference (Zhang et al., 2008)	Data in this study (Li et al., 2015)
	$k_{01} = 0.67$	$k_{01} = 0.69$
Frequency factor	$k_{02} = 0.2558$	$k_{02} = 0.5971$
	$k_{03} = 69$	$k_{03} = 68$
Reaction order	$n_1 = 0.98, n_2 = 0.26, n_3 = 0.70, n_4 = 0.60, n_5 = 0.30$	
Activation energy	$E_1 = 18.66\text{kJ/mol}, E_2 = 28.04\text{kJ/mol}, E_3 = 724.143\text{kJ/mol}$	

3. Sensitivity analysis and variables selection

Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs (Saltelli, 2002). In the whole process, there are 7 main variables who may influence the final product, namely, 4-CBA content in CTA, mass flow of CTA, mass flow of deionized water, mass flow of hydrogen, reactor temperature, reactor pressure and catalyst activity. Sensitivity analysis here is used to better understanding the relationships between inputs and output of the process and find out the input variables which are most relevant to the output. Based on Aspen plus software, sensitivity analysis of these 7 input variables are done. The variation range of these variables is set to be $\pm 10\%$ (The temperature of reactor cannot be too high since the catalyst activity losses too fast at higher temperature. Therefore, the variation range of temperature is set to be $\pm 5\%$). Figure 2 shows the influence of these 6 variables to final 4-CBA. The normal conditions of these input and output variables are listed in Table 2. Also, variation range, percentage are listed too.

Table 2 and Figure 2 demonstrate clearly that catalyst activity is the most relevant input variable to final 4-CBA, reactor pressure has little influence to final 4-CBA. Therefore, 6 variables are selected to be the input variables.

Table 2: Data of sensitivity analysis of 7 variables

Variables	Normal condition	%	Range	Final 4-CBA (2.059 kg/h in normal condition)	%
4-CBAin	176 kg/h	$\pm 10\%$	158.4 ~ 193.6 kg/h	1.346 ~ 2.694 kg/h	-34.6% ~ 30.8%
CTA	80,000 kg/h	$\pm 10\%$	72,000 ~ 88,000 kg/h	1.591 ~ 2.461 kg/h	-22.7% ~ 19.5%
H ₂ O	141,111 kg/h	$\pm 10\%$	127,000 ~ 155,222 kg/h	0.946 ~ 3.499 kg/h	-54.1% ~ 69.9%
H ₂	30 kg/h	$\pm 10\%$	27 ~ 33 kg/h	2.514 ~ 2.655 kg/h	22.1% ~ -19.6%
T	273 °C	$\pm 5\%$	259.35 ~ 286.65 °C	5.019 ~ 0 kg/h	143.8% ~ -100%
P	7.9 MPa	$\pm 10\%$	7.11 ~ 8.69 MPa	2.059 ~ 2.058 kg/h	0% ~ 0%
CA	0.69	$\pm 10\%$	0.621 ~ 0.759	3.861 ~ 0.765 kg/h	87.5% ~ -62.9%

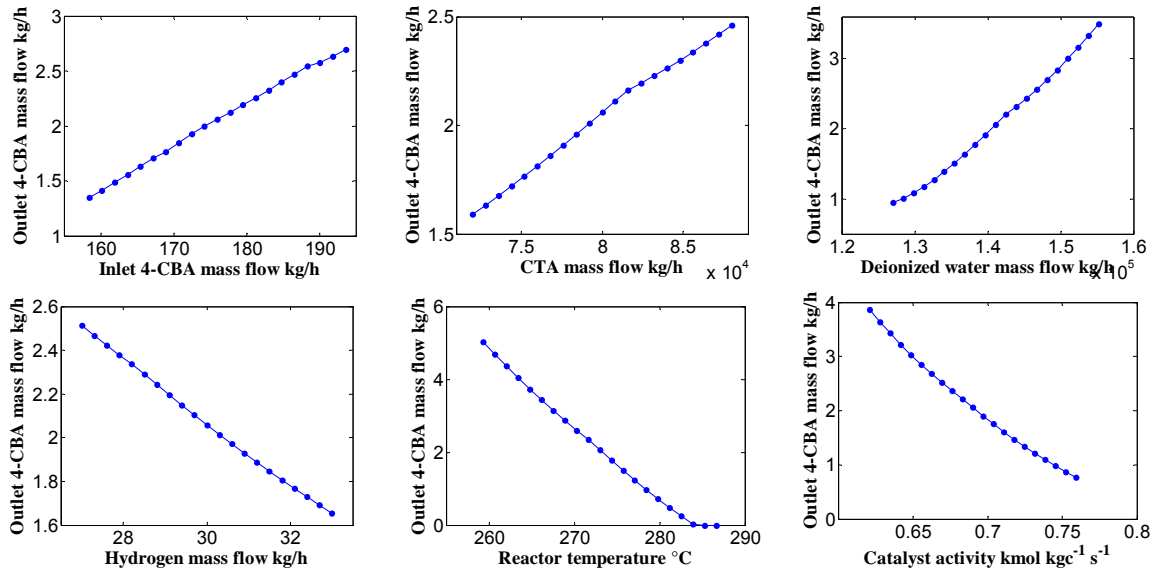


Figure 2: Sensitivity analyses of 6 variables

4. Methodology and application illustration

CTA hydropurification process is a complex nonlinear and time varying process, traditional PLS method can not handle the nonlinear problem. Though KPLS is capable of dealing with nonlinear problem, the time varying problem of this process is still waiting to be solved. Thus, recursive kernel partial least squares method is developed to build the soft sensor model for this process. The procedures of this method are shown in Table 3.

Table 3: RKPLS method

Recursive kernel partial least squares
Step 1. Get data from the real process
Step 2. Data standardization X_0
Step 3. Kernelization of $X_0 \rightarrow K_0$
Step 4. Update when a new X_{new} comes to get X_1
Step 5. Update the new kernel of $X_1 \rightarrow K_1$
Step 6. PLS method
Step 7. Return to step 4
Step 8. Calculate Y_{pred}

Gaussian kernel is used to establish kernelization in step 3 to calculate K_0 . The update of X is as follows,

$$X_1 = \lambda X_0 + \gamma X_{new}, Y_1 = \lambda Y_0 + \gamma Y_{new} \quad (8)$$

where λ is forgetting factor, $0 \leq \lambda \leq 1$, $\gamma = 1 - \lambda$. In this study, λ is set to be 0.8. It can be seen from Figure 3 that the new X_1 contains information both in X_0 and X_{new} . The new kernel K_1 is calculated as follows,

$$K_1 = K_0 - t_0 t_0^T K_0 - K_0 t_0 t_0^T + t_0 t_0^T K_0 t_0 t_0^T \quad (9)$$

where $t_0 = K_0 Y_0$, $t_0 \leftarrow t_0 / \|t_0\|$. After step 5, using traditional PLS to calculate the matrix of P and W.

Then the prediction of Y_{pred} is calculated as follows,

$$Y_{pred} = K_t P (W^T K_0 P)^{-1} W^T * Y_t \quad (10)$$

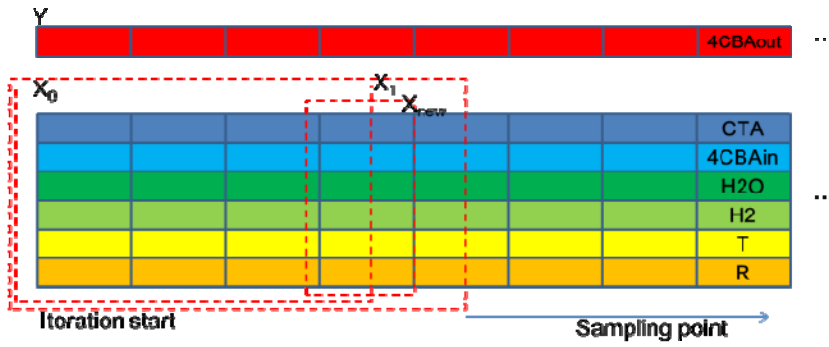


Figure 3: Soft sensor model developments for CTA hydropurification process using RKPLS

It is clear that the computational cost of this method is higher than KPLS and PLS, because of the kernelization step in every step. However, this method can solve both nonlinear and slow time varying problems. As it is not easy to get real plant data directly, 400 sampling points of these 6 input variables and corresponding output 4-CBA results are collected using Aspen Plus. In the simulation, the secondary variables are given random fluctuations to simulate the disturbances in real plant. Prediction results using PLS, KPLS and RKPLS methods are shown in Figure 4. It can be seen that PLS method is not suitable in this complex process. The prediction results of KPLS method is getting worse and worse as the number of sampling points goes on. That is because the catalyst is slowly deactivating, KPLS method can not handle this slow time varying problem. It is clear that the prediction results of RKPLS method is very good from the beginning to the end, this demonstrates that this method is suitable for this nonlinear and slow time varying CTA hydropurification process.

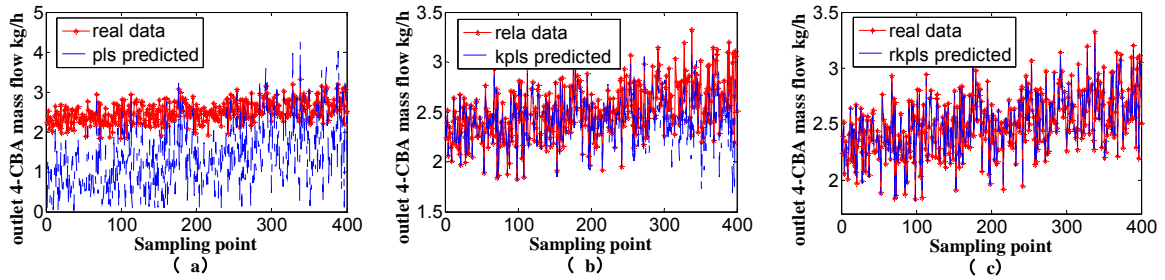


Figure 4: Prediction results using PLS, KPLS and RKPLS

Besides, the root mean square error of calibration (RMSECV), root mean square error of prediction (RMSEP), max error, min error and mean error in this study are chosen as the performance indexes to compare the results using PLS, KPLS and RKPLS method. The lower the values of RMSECV and RMSEP, the better the soft sensor model.

The calculation formula of RMSECV is Eq(11)

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

where, n is the number of sampling points in the calibration set, y_i is the real value of sample i in the calibration set, \hat{y}_i is the prediction result using regression model.

The calculation formula of RMSEP is Eq(12)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{\tilde{y}}_i - \tilde{y}_i)^2}{n}} \quad (12)$$

where, n is the number of sampling points in the prediction set, \tilde{y}_i is the real value of sample i in the prediction set, $\hat{\tilde{y}}_i$ is the prediction result using regression model.

Table 4 shows the error comparisons among these methods. It can be seen that RKPLS is better than the other two methods in CTA hydropurification process soft sensor modelling.

Table 4: Error comparisons among PLS, KPLS and RKPLS

	MinCV	MaxCV	MeanCV	RMSECV	MinPre	MaxPre	MeanPre	RMSEPre
PLS	8.164E-03	2.047E+00	8.182E-01	1.102E+00	6.429E-04	2.086E+00	1.840E+00	1.645E+00
KPLS	3.635E-05	1.883E-02	7.422E-03	8.484E-03	5.040E-05	4.096E-02	9.386E-03	1.129E-02
rKPLS	6.843E-06	5.077E-02	1.375E-02	1.668E-02	4.140E-06	2.426E-02	4.600E-03	6.288E-03

5. Conclusion

In this paper, a new recursive kernel partial least squares method is developed. This RKPLS method is applied to build a soft sensor for a complex nonlinear and slow time-varying process, terephthalic acid hydropurification process. The results demonstrate that the proposed method can efficiently capture the time-varying and nonlinear relationship in process variables. The prediction results using RKPLS are better than PLS and KPLS method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61422303, 21376077) and Fundamental Research Funds for Central Universities.

References

- Azarpour A., Zahedi G., 2012. Performance analysis of crude terephthalic acid hydropurification in an industrial trickle-bed reactor experiencing catalyst deactivation. *Chemical Engineering Journal*, 209, 180-193.
- Facco P., Doplicher F., Bezzo F., Barolo M., 2009. Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. *Journal of Process Control*, 19, 520-529.
- Farsang B., Balogh I., Németh S., Székvölgyi Z., Abonyi J., 2015. PCA based data reconciliation in soft sensor development - Application for melt flow index estimation. *Chemical Engineering Transactions*, 43, 1555-1560.
- Godoy J. L., Zumoffen D. A., Vega J. R., Marchetti J. L., 2014. New contributions to non-linear process monitoring through kernel partial least squares. *Chemometrics & Intelligent Laboratory Systems*, 135, 76-89.
- Li Z., Zhong W., Liu Y., Na L., Feng Q., 2015. Dynamic Modeling and Control of Industrial Crude Terephthalic Acid Hydropurification Process. *Korean Journal of Chemical Engineering*, 32, 597-608.
- Li Z., Zhong W., Wang X., Luo N., Qian F., 2016. Control structure design of an industrial crude terephthalic acid hydropurification process with catalyst deactivation. *Computers & Chemical Engineering*, 88, 1-12.
- Saltelli A., 2002. Sensitivity Analysis for Importance Assessment. *Risk Analysis*, 22, 579-590.
- Zhang S., Zhou J., Yuan W., 2008. Mathematical simulation of hydrorefining reactor for terephthalic acid. *Chemical Reaction Engineering and Technology*, 24.
- Zhou J., Zhang T., Sui Z., 2006a. Hydropurification of Terephthalic Acid over Pd/C I. Thermodynamics and Feature Analysis. *Journal of East China University of Science and Technology (Natural Science Edition)*, 32, 374-380.
- Zhou J., Zhang T., Sui Z., 2006b. Hydropurification of Terephthalic Acid over Pd/C II. Apparent Kinetics of 4-CBA Hydrogenation on Catalysts of Different Sizes. *Journal of East China University of Science and Technology (Natural Science Edition)*, 32, 503-507.