Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

A Comparative Study on Improved DPLS Soft Sensor Models Applied to a Crude Distillation Unit ⋆

Chao Shang * **Xinqing Gao** * **Fan Yang** * **Wenxiang Lyu** * **Dexian Huang** *

* *Department of Automation, Tsinghua University, TNList, Beijing 100084, P.R. China (e-mail: shangc07@mails.tsinghua.edu.cn, gxq08_hor@163.com, yangfan@tsinghua.edu.cn, lvwx79@tsinghua.edu.cn, huangdx@tsinghua.edu.cn)*

Abstract: Soft sensors based on dynamic PLS (DPLS) have been widely used in industrial applications for predicting hard-to-measure quality variables. However, DPLS is prone to over-fitting due to an increasing number of model inputs. A plethora of approaches have been proposed to improve DPLS-based soft sensors, among which variable selection has been a prevailing one. Recently, a new method termed as DPLS-TS has been proposed to penalize dynamic parameters in DPLS using a temporal smoothness regularization, which helps reduce model complexity and deliver smooth predictions for quality variables. In this work we present a comparative study of temporal smoothness regularization and variable selection in terms of their improvements in prediction performance when a large number of lagged time series data are involved. Comparisons are performed through a simulated case of crude distillation unit.

Keywords: Partial least squares, quality prediction, variable selection, temporal smoothness regularization, soft sensor.

1. INTRODUCTION

Partial least squares (PLS) has been a basic tool in developing data-driven soft sensors due to significant correlations existing in process data (Dayal et al. (1997); Qin (1998); Sharmin et al. (2006)). A low-dimensional latent subspace can be described by selecting fewer dominant components that capture most information in input and output data. Because of the dynamics that characterizes industrial processes and the sparsity of quality measurements, dynamic extensions of PLS, i.e., DPLS, have been proposed and successfully implemented by including lagged inputs (Kano et al. (2000); Lin et al. (2007); Facco et al. (2009); Galicia et al. (2011)).

A major concern with DPLS is that an increasing number of model inputs are utilized while the number of available training samples remains unchanged, resulting in the over-fitting problem. Variable selection is an effective method that helps improve soft sensors when tremendous process variables are used as inputs for soft sensor modeling (Mehmood et al. (2012)). The inputs of a soft sensor are usually selected in advance by process engineers according to first-principle experiences. It is hence likely that some less relevant variables would be involved. Depending upon the statistical properties of available data, a variable selection approach is able to identify vital process variables and remove unimportant ones without referring to first-principle knowledge. Therefore the model complexity can be greatly reduced and the prediction performance get enhanced. Wang et al. (2014) demonstrated the efficacy of variable selection by a comparative study of different variable selection methods for static PLS-based soft sensor development. In Liu (2014), variable selection is applied to DPLS in order to prune out redundant process variables.

In addition to the variable selection approach, regularization technique is also an effective approach in the statistical learning theory with the aim to alleviate over-fitting in the face of limited data (Suykens et al. (2002)). Inspired by this idea, a regularized DPLS with temporal smoothness (DPLS-TS) has been proposed in Shang et al. (2015). This formulation incorporates prior information about process dynamics into the latent subspace by adding an extra term to temporal parameters with a smooth regularization constant. In this way, the model enjoys clearer physical interpretations. The related optimization scheme involves a typical eigen-decomposition task that provides computational convenience in practice.

The variable selection and regularization approaches share some similarities because both of them essentially aim at reducing the model complexity, only with differences in their standpoints. In this paper, we make a comparative study of different versions and combinations of both approaches. For variable selection purposes, the PLS with variable importance in projection (PLS-VIP) has been shown to outperform other variable selection methods in terms of prediction accuracy and robustness (Chong and

⋆ This work was supported by the National High-Tech Program of China (2013AA040702), the National Natural Science Foundation of China (21276137), and Tsinghua University Initiative Scientific Research Program. Corresponding author should be addressed to D. Huang at huangdx@tsinghua.edu.cn.

Copyright © 2015 IFAC

Jun (2005); Liu (2014)). Hence PLS-VIP is taken into considerations in this study. An alternative to VIP called group VIP (GVIP) is also proposed especially for group variable selection in DPLS models. In addition, we propose a new DPLS model to simultaneously adopt regularization and variable selection, and investigate performances thereof to gain some further insight into both approaches. A quality prediction task in a crude distillation unit (CDU) is used to test performance of different models in practical scenarios. The influence of tuning parameters on prediction performance is also studied and discussed in detail through the case study.

The organization of this paper is given as follows: Section 2 reviews basics of the DPLS-TS algorithm and variable selection approaches in brief. The concept of GVIP is also inspired. A new DPLS model utilizing both approaches is also proposed. Section 3 compares the methods using an example of a crude distillation unit. Section 4 concludes the paper based on the case study.

2. DYNAMIC PLS WITH TEMPORAL SMOOTHNESS AND VARIABLE SELECTION

In this section, we first review the DPLS model with temporal smoothness presented in Shang et al. (2015). Then we introduce the variable selection method and combine it with DPLS-TS.

Assume that there are N quality samples $\{y(t_1), \cdots, y(t_N)\}$ and m process variables $\{x_1(t), \cdots, x_m(t)\}$. In order to estimate the quality, process variables are used as the inputs of the soft sensor model. At each time t_i, the kth process variable $(1 \leq k \leq m)$ can be augmented into a historical input vector by including d lagged samples, which is described as:

$$\begin{aligned}\mathbf{x}_k(t_i) = &[x_k(t_i), x_k(t_i - \Delta t), \ldots, x_k(t_i - d\Delta t)]^{\mathrm{T}} \\ &\in \mathbb{R}^{d+1},\ 1 \leq i \leq N\end{aligned} \tag{1}$$

where t_i is the sampling time of the ith quality measurement and Δt is the measurement interval for process variables. By stacking historical vector of m process variables into a column, we derive the input vector for DPLS model:

$$\mathbf{x}(i) = \begin{bmatrix} \mathbf{x}_1(t_i) \\ \mathbf{x}_2(t_i) \\ \vdots \\ \mathbf{x}_m(t_i) \end{bmatrix} \in \mathbb{R}^{m(d+1)},\ 1 \leq i \leq N \tag{2}$$

For simplicity, the time t_i is replaced by the index i in the following to enumerate the process samples $\{\mathbf{x}(i),\ 1 \leq i \leq N\}$.

2.1 DPLS with Temporal Smoothness

In this study, the case of univariate output, i.e., PLS1 (Boulesteix et al. (2007)), is considered. Given an input matrix

$$\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)]^{\mathrm{T}} \in \mathbb{R}^{N \times (md+m)} \tag{3}$$

and an output matrix

$$\mathbf{y} = [y(t_1), y(t_2), \ldots, y(t_N)]^{\mathrm{T}} \in \mathbb{R}^N, \tag{4}$$

PLS projects input and output onto a low-dimensional subspace spread by A latent variables (LVs) $\{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_A\}$ $(A \ll md+m)$. Mathematically, the latent variable model is formed as:

$$\begin{aligned}\mathbf{X} &= \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \\ \mathbf{y} &= \mathbf{T}\mathbf{q} + \mathbf{F}\end{aligned} \tag{5}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_A] \in \mathbb{R}^{N \times A}$ denotes the score matrix, and $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_A] \in \mathbb{R}^{(md+m) \times A}, \mathbf{q} = [q_1, q_2, \cdots, q_A]^{\mathrm{T}} \in \mathbb{R}^A$ are the loading matrices for $\mathbf{X}$ and $\mathbf{y}$. Matrices $\mathbf{E}$ and $\mathbf{F}$ represent modeling residuals of $\mathbf{X}$ and $\mathbf{y}$. The objective of the classical PLS1 algorithm is to sequentially solve the following problem:

$$\begin{aligned}&\max_{\mathbf{w}_j}\ \mathbf{w}_j^{\mathrm{T}}\mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j\mathbf{y}_j^{\mathrm{T}}\mathbf{X}_j\mathbf{w}_j \\ &\text{s.t. } \mathbf{w}_j^{\mathrm{T}}\mathbf{w}_j = 1,\ j = 1, \cdots, A\end{aligned} \tag{6}$$

where $\mathbf{w}_j$ is the weight vector for the jth latent variable, computed as the eigenvector of $\mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j\mathbf{y}_j^{\mathrm{T}}\mathbf{X}_j$ corresponding to the largest eigenvalue. The score vector is then derived as $\mathbf{t}_j = \mathbf{X}_j\mathbf{w}_j$. The loading vectors for $\mathbf{X}$ and $\mathbf{Y}$ are calculated as $\mathbf{p}_j = \mathbf{X}_j^{\mathrm{T}}\mathbf{t}_j/\mathbf{t}_j^{\mathrm{T}}\mathbf{t}_j$ and $q_j = \mathbf{y}_j^{\mathrm{T}}\mathbf{t}_j/\mathbf{t}_j^{\mathrm{T}}\mathbf{t}_j$. Then the jth latent variable is removed and the input and output matrices for the $(j+1)$th latent variable are derived as $\mathbf{X}_{j+1} = \mathbf{X}_j - \mathbf{t}_j\mathbf{p}_j^{\mathrm{T}}$ and $\mathbf{y}_{j+1} = \mathbf{y}_j - \mathbf{t}_j q_j$. With these basics, we further analyze the influences of temporal inputs on the model structure by decomposing the elements in LV $\mathbf{t}_j$:

$$\begin{aligned}t_j(i) = \mathbf{x}(i)^{\mathrm{T}}\mathbf{w}_j &= \sum_{k=1}^{m}\mathbf{x}_k(t_i)^{\mathrm{T}}\mathbf{w}_{j,k} \\ &= \sum_{k=1}^{m}\sum_{l=0}^{d} x_k(t_i - l\Delta t)w_{j,k}(l+1).\end{aligned} \tag{7}$$

where $\mathbf{w}_{j,k}$ comes from the decomposition of $\mathbf{w}_j$:

$$\mathbf{w}_j = \begin{bmatrix} \mathbf{w}_{j,1} \\ \mathbf{w}_{j,2} \\ \vdots \\ \mathbf{w}_{j,m} \end{bmatrix} \in \mathbb{R}^{m(d+1)}, 1 \leq j \leq A. \tag{8}$$

Intuitively, each historical input vector $\mathbf{x}_k(t_i)$ contributes to the LV $\mathbf{t}_j$ by a convolution with a coefficient vector $\mathbf{w}_{j,k}$, and each lagged sample corresponds to its own coefficient $w_{j,k}(l+1)$ in (7). The coefficients of proximal historical samples $x_k(t_i - l\Delta t)$ and $x_k(t_i - (l-1)\Delta t)$ should be temporally similar because dynamic data have smoothly varying impacts on inherent features $\{\mathbf{t}_j\}$. In order to encourage temporal smoothness in weighed coefficients, a common choice is to use the L_2 regularization:

Copyright © 2015 IFAC

$$\sum_{l=1}^{d}\left[w_{j,k}(l)-w_{j,k}(l-1)\right]^2 \tag{9}$$

which is denoted as the *temporal smoothness regularization*. Then minimization of (9) can be neatly re-written as:

$$\min_{\mathbf{w}_{j,k}} \mathbf{w}_{j,k}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}}\mathbf{J}\mathbf{w}_{j,k} \tag{10}$$

where

$$\mathbf{J}=\begin{bmatrix}1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1\end{bmatrix}\in\mathbb{R}^{d\times(d+1)}. \tag{11}$$

Taking into considerations coefficient vectors $\mathbf{w}_j = \left[\mathbf{w}_{j,1}^{\mathrm{T}},\mathbf{w}_{j,2}^{\mathrm{T}},\ldots,\mathbf{w}_{j,m}^{\mathrm{T}}\right]^{\mathrm{T}}$ of all m process variables, the smoothness penalty for $\mathbf{w}_j$ is derived as

$$\min_{\mathbf{w}_j} \mathbf{w}_j^{\mathrm{T}}(\mathbf{I}_m\otimes\mathbf{J}^{\mathrm{T}}\mathbf{J})\mathbf{w}_j \tag{12}$$

where $\otimes$ denotes the Kronecker product. Because $\mathbf{w}_j$ is calculated by solving an eigenvector problem, we further modify the optimization problem in (6) by adding the L_2 penalty term, expressed as follows:

$$\begin{aligned}\max_{\mathbf{w}_j}\ & \mathbf{w}_j^{\mathrm{T}}\left[\mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j\mathbf{y}_j^{\mathrm{T}}\mathbf{X}_j-\alpha\|\mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j\|^2(\mathbf{I}_m\otimes\mathbf{J}^{\mathrm{T}}\mathbf{J})\right]\mathbf{w}_j\\ \text{s.t. }& \mathbf{w}_j^{\mathrm{T}}\mathbf{w}_j=1\end{aligned} \tag{13}$$

where $\alpha \geq 0$ denotes the regularization parameter. The first term in (13) maximizes the covariance between $\mathbf{X}_j\mathbf{w}_j$ and $\mathbf{y}_j$, whereas the second term serves to enhance smoothness of coefficients. It is worth noticing that the additional regularization constant $\|\mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j\|^2$ in the second term renders the solution $\mathbf{w}_j$ invariant to the linear scaling of input and output matrices. The optimization problem in (13) simply takes the form of eigen-decomposition of merit, which necessitates a low computational cost in practice.

With A latent variables obtained, the final regression equation can be written as:

$$\mathbf{y}=\mathbf{X}\mathbf{W}\left(\mathbf{P}^{\mathrm{T}}\mathbf{W}\right)^{-1}\mathbf{q}+\mathbf{F} \tag{14}$$

where $\mathbf{W} = [\mathbf{w}_1,\mathbf{w}_2,\cdots,\mathbf{w}_A]$. Given an out-of-sample data point $\mathbf{x}_{\text{new}}$, the prediction model is expressed as:

$$\hat{y}=\mathbf{b}^{\mathrm{T}}\mathbf{x}_{\text{new}} \tag{15}$$

where $\mathbf{b}=\mathbf{W}\left(\mathbf{P}^{\mathrm{T}}\mathbf{W}\right)^{-1}\mathbf{q}$.

2.2 Variable Selection based on the VIP Score

The VIP score is used to evaluate the importance of each process variable (Wold et al. (2001)). When applied in dynamic cases, the VIP score of the l-th lagged variable of the k-th process variable is defined as:

$$\mathrm{VIP}_{l,k}=\sqrt{m(d+1)\sum_{j=1}^{A}\left(q_j^2\mathbf{t}_j^T\mathbf{t}_j\cdot\frac{w_{j,k}^2(l)}{\|\mathbf{w}_j\|^2}\right)\Big/\sum_{j=1}^{A}q_j^2\mathbf{t}_j^T\mathbf{t}_j},\quad 0\leq l\leq d, 1\leq k\leq m. \tag{16}$$

For VIP scores that are larger than a threshold, i.e., $\mathrm{VIP}_{l,k} \geq s_{\mathrm{VIP}}$, the corresponding process variables are considered as having significant impacts on the latent subspace. One can easily verify that the average of squared VIP scores equals 1. For this reason, variables of which VIP scores are greater than one are often selected as dominant input variables (Chong and Jun (2005)).

2.3 DPLS-TS with the Group VIP score

Both temporal smoothness regularization and variable selection aim at reducing model complexity and further alleviate the over-fitting problem. One would be interested in the simultaneous utilization of these two approaches for better prediction performances. In this context, however, the classical VIP score cannot be directly used along with the temporal smoothness regularization. A portion of lagged data are removed and thus the definition of temporal smoothness regularization becomes invalid, as shown in Fig. 1(a). We suggest entirely removing the process variables that are of less significance, rather than some individual lagged measurements, as shown in Fig. 1(b). Hence the group effect on LV subspaces of all lagged measurements $\mathbf{x}_k(t)$ of a certain process variable should be considered instead of the effect of individual lagged measurements. The proposed group VIP (GVIP) score of the k-th process variable in DPLS modeling is therefore derived as:

$$\mathrm{GVIP}_{k}=\sqrt{m\sum_{j=1}^{A}\left(q_j^2\mathbf{t}_j^T\mathbf{t}_j\cdot\frac{\|\mathbf{w}_{j,k}\|^2}{\|\mathbf{w}_j\|^2}\right)\Big/\sum_{j=1}^{A}q_j^2\mathbf{t}_j^T\mathbf{t}_j},\quad 1\leq k\leq m. \tag{17}$$

3. SIMULATED CASE STUDY

3.1 Experimental Design

In this study, HYSYS, a well-known simulation software for chemical processes, is used to generate experimental datasets. The atmospheric column of the crude distillation unit (CDU) is simulated as an object. Fig. 2 provides a systematic sketch of an atmospheric column unit, which arises from Wang et al. (2010). The 100% cut point (ASTM D86) of the top product, naphtha, is set as the quality variable. There are 13 process variables in total selected as the inputs of soft sensors, which are listed in Table 1. All flow rate measurements are contaminated by random noises.

The sampling interval of input variables is set as 2 min. To mimic the multi-rate characteristics in real industrial

Copyright © 2015 IFAC

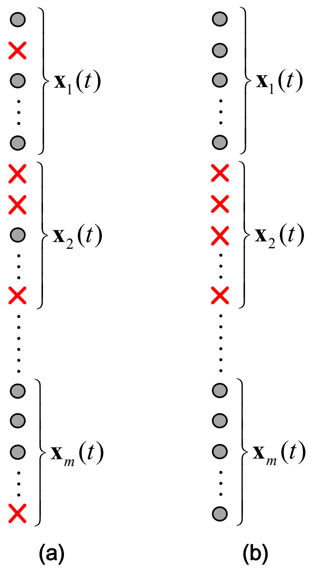

Fig. 1. Classical variable selection and group variable selection in DPLS. (a) Classical variable selection results: Separate lagged measurements are deleted, leading to an invalid definition of temporal smoothness regularization. (b) Group variable selection results: If a certain process variables is determined as irrelevant, all lagged measurements of this variable should be deleted.

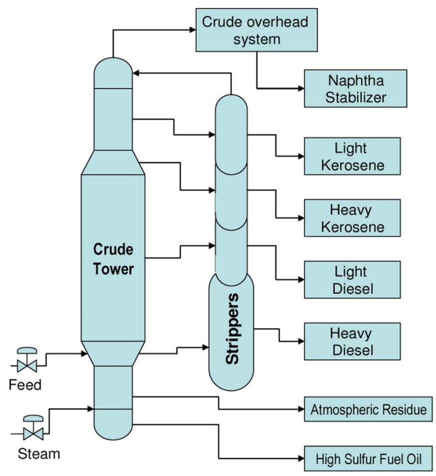

Fig. 2. A systematic sketch of an atmospheric column unit.

processes, the product quality is sampled with an interval varying around 66 min. The dataset consists of 150 samples for training and 150 samples for test. Five versions of DPLS models are introduced for comparison:

(a) the classical DPLS;

(b) DPLS with VIP scores (DPLS-VIP);

(c) DPLS with group VIP scores (DPLS-GVIP);

(d) DPLS-TS;

(e) DPLS-TS with group VIP scores (DPLS-TS-GVIP).

In this study, all optimal tuning parameters for DPLS models, including the number of LVs, and the regularization parameter α for DPLS-TS, are selected using 5-fold cross-validation, and the greater-than-one rule is used as a criterion for adopting VIP and GVIP scores. Before model training, all the inputs and outputs have been zero-mean normalized. To evaluate the performance of different models, the root mean square error (RMSE) is used as the error criterion:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\hat{y}(i) - y(i)]^2} \tag{18}$$

where $\hat{y}(i)$ denotes the predicted output.

3.2 Results and Discussions

In simulations, various models are established, of which the length of lagged variables, d, is chosen consecutively from 1 to 33. The RMSE curves of five approaches in terms of d are presented in Fig. 3, and detailed RMSE values and numbers of LVs are shown in Table 2, where the bold figures and the shaded figures denote the minimal RMSE value in the located row and column, respectively.

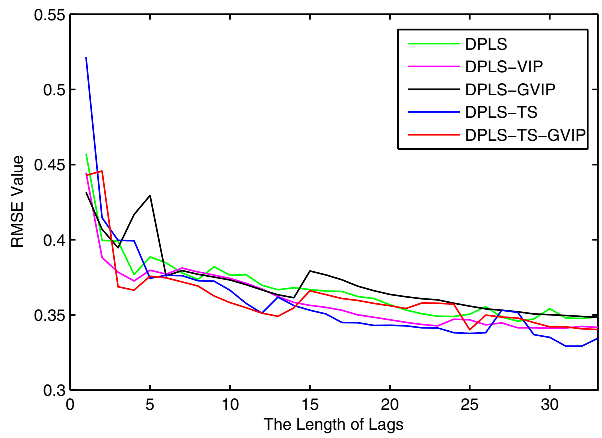

Fig. 3. RMSE curves of different approaches with different lengths of lagged data.

A. Discussions on prediction accuracies

First, we can observe from Table 2 and Fig. 3 that, in most cases, the best prediction performance is achieved by two approaches with temporal smoothness, namely, DPLS-TS and DPLS-TS-GVIP. In addition, DPLS-TS generally has

Table 1. Process Variables in the Crude Distillation Unit

No.	Description
1	Top temperature
2	10# tray temperature
3	18# tray temperature
4	23# tray temperature
5	Feed temperature
6	Top pressure
7	Reflux flow rate ratio
8	Side-drawn 1# flow rate ratio
9	Side-drawn 2# flow rate ratio
10	Side-drawn 3# flow rate ratio
10	Steam flow rate ratio
11	Top recycle flow rate ratio
12	Middle recycle 1# flow rate ratio
13	Middle recycle 2# flow rate ratio

Copyright © 2015 IFAC

Table 2. Results of Different DPLS Models under Different Lengths of Lagged Data

	DPLS		DPLS-VIP		DPLS-GVIP		DPLS-TS		DPLS-TS-GVIP	
d	RMSE	#LVs	RMSE	#LVs	RMSE	#LVs	RMSE	#LVs	RMSE	#LVs
1	0.4573	15	0.4446	9	**0.4316**	9	0.5214	16	0.4428	9
2	0.3996	15	**0.3882**	7	0.4070	8	0.4146	17	0.4457	11
3	0.3992	8	0.3785	7	0.3947	9	0.3996	13	**0.3687**	8
4	0.3769	10	0.3727	7	0.4168	11	0.3994	12	**0.3666**	9
5	0.3886	9	0.3797	7	0.4294	11	**0.3744**	16	0.3757	7
6	0.3846	10	0.3772	7	0.3761	7	0.3764	16	**0.3747**	7
7	0.3780	10	0.3812	7	0.3795	5	0.3761	10	**0.3720**	7
8	0.3736	11	0.3786	5	0.3771	5	0.3727	13	**0.3692**	7
9	0.3821	9	0.3765	5	0.3752	5	0.3725	10	**0.3627**	6
10	0.3764	9	0.3743	5	0.3733	5	0.3664	10	**0.3582**	6
11	0.3768	7	0.3710	5	0.3702	5	0.3577	10	**0.3548**	6
12	0.3698	6	0.3673	5	0.3666	5	**0.3512**	10	**0.3512**	6
13	0.3668	6	0.3623	5	0.3634	5	0.3619	11	**0.3491**	6
14	0.3681	7	0.3584	4	0.3615	5	0.3563	11	**0.3547**	6
15	0.3669	6	0.3564	4	0.3793	5	**0.3532**	12	0.3662	5
16	0.3658	6	0.3551	4	0.3766	5	**0.3506**	11	0.3636	5
17	0.3656	6	0.3531	4	0.3734	5	**0.3449**	10	0.3610	6
18	0.3623	6	0.3501	5	0.3691	4	**0.3449**	10	0.3596	6
19	0.3609	5	0.3485	5	0.3661	4	**0.3431**	9	0.3578	6
20	0.3568	5	0.3468	5	0.3637	4	**0.3431**	9	0.3562	6
21	0.3532	6	0.3451	5	0.3620	4	**0.3428**	9	0.3543	6
22	0.3508	6	0.3437	5	0.3608	4	**0.3415**	9	0.3579	11
23	0.3491	6	0.3428	5	0.3600	4	**0.3414**	10	0.3579	4
24	0.3491	6	0.3472	6	0.3578	4	**0.3382**	8	0.3572	4
25	0.3506	6	0.3468	6	0.3559	4	**0.3377**	8	0.3402	11
26	0.3555	7	0.3434	5	0.3541	4	**0.3383**	8	0.3499	5
27	**0.3486**	4	0.3447	5	0.3531	5	0.3532	12	0.3487	5
28	0.3459	4	**0.3415**	5	0.3521	5	0.3517	11	0.3479	11
29	0.3473	6	0.3415	5	0.3507	4	**0.3369**	6	0.3450	14
30	0.3542	4	0.3412	5	0.3502	4	**0.3352**	6	0.3422	9
31	0.3480	4	0.3414	5	0.3496	4	**0.3292**	7	0.3421	14
32	0.3477	4	0.3424	5	0.3490	4	**0.3293**	7	0.3408	13
33	0.3487	4	0.3417	5	**0.3484**	4	**0.3345**	6	0.3403	14

higher accuracy than DPLS, and the same trend can be found from the comparison between DPLS-TS-GVIP and DPLS-GVIP. It indicates that the temporal smoothness regularization effectively helps utilize dynamic information and thus yields improved prediction results, especially when a large number of lagged data are involved. It is worth mentioning that from the first two rows in Table 2, classical DPLS without temporal smoothness is preferable when d is small. This is because there is few dynamic information incorporated in data, thereby restricting the effect of the regularization term.

Second, we can see that DPLS-VIP outperforms classical DPLS in most cases, indicating the power of variable selection. In addition, when d is large ($d > 26$), more LVs can be achieved by DPLS-VIP in spite of fewer process variables that are used. This demonstrates that some irrelevant process variables are harmful to the latent structure of PLS and it is necessary to perform variable selection (Chong and Jun (2005)). On the other hand, it can be observed that DPLS-GVIP has no evident advantages to DPLS. This is primarily due to the fact that some variables are entirely deleted, leading to a greater information loss. In contrast, DPLS-VIP might only delete some irrelevant lagged measurements more accurately.

Finally we make a comparison between temporal smoothness regularization and variable selection. By comparing the last two columns of Table 2, we can find that when d is small, DPLS-TS-GVIP is superior to DPLS-TS. When d is large, DPLS-TS gives more accurate results. In our opinion, when massive lagged data are involved, DPLS-TS is still able to utilize dynamic information within those less important variables that are simply abandoned by DPLS-TS-GVIP, thereby enjoying better generalizations than DPLS-TS-GVIP. As pointed in Qin (2014), a major direction in the era of big data is the utilization of all available data for more powerful models, including some imperfect data. In this sense, the temporal smoothness regularization is more in line with the spirit of big data herein, thereby being a promising approach to massive process data. In contrast, variable selection might inevitably entail information loss to some extent.

B. Discussions on dimensions of LV subspaces

An interesting fact is that DPLS commonly finds fewer LVs when more lags are used. In this case, more dominant components ought to be extracted instead because more meaningful dynamic information are fused. It implies that DPLS has some fundamental limitations in utilizing dynamic information in that the maximization of the covariance between input and output scores in (6) might be over greedy so that the dynamic structure is unfortunately undermined. Once model dynamics is violated in the first LV, all subsequent LVs will be affected since they are based on the residuals of first LV, leading to a totally inappropriate description of model dynamics. In contrast, DPLS-TS generally extracts more LVs than DPLS, which is as well observed from the comparison between DPLS-

Copyright © 2015 IFAC

TS-GVIP and DPLS-GVIP. This is because the additional regularization term in (13) is able to integrate proper dynamic information into LVs and thus more meaningful LVs could be attained. Therefore, the model dynamics is desirably preserved in the low-dimensional LV subspaces. It demonstrates that temporal smoothness is able to utilize time series lagged data in a more efficient way, yielding a more complicated but reasonable structure.

C. Discussions on the optimal length of lagged data

When more lags are used, over-fitting would become a major concern because the model complexity increases but the number of available data samples remains unchanged. For example, the optimal d^* for DPLS is 28, as indicated by shaded figures in Table 2. When d becomes larger than 28, the prediction performance thereof tends to degrade, as shown in Table 2. Notice that DPLS-VIP, DPLS-GVIP and DPLS-TS have larger d^* than that of classical DPLS. In one sense, they can use more lags because their model complexity is reduced. In addition, the results accords with the fact that the settling time of this CDU process is about one hour. However, DPLS-TS-GVIP has a relatively smaller $d^* = 25$ than the other four methods, mainly because a combination of regularization and variable selection might excessively reduce model complexity. This is also in accordance with the sub-optimality of DPLS-TS-GVIP in terms of prediction accuracy in presence of a large d.

4. CONCLUSION

In this work, several improvements for DPLS soft sensor development based on temporal smoothness regularization and variable selection have been investigated with an industrial case study. We first examined how DPLS model is improved by the temporal smoothness regularization and variable selection approaches. It was found that the prediction RMSE decreases when temporal smoothness or variable selection is adopted to reduce model complexity. The optimal length of lags increases for improved models, indicating that over-fitting problem is better mitigated compared to classical DPLS. In general, temporal smoothness outperforms other methods because it not only gives the minimal prediction error but also helps find more physically meaningful LVs, yielding a reasonable structure. To combine the advantages of both strategies, we propose a simultaneous utilization of temporal smoothness and variable selection called DPLS-TS-GVIP. We find that when d is large the accuracy of DPLS-TS-GVIP is even lower than that of DPLS-TS, implying that some negligible but still useful information is inevitably overlooked by variable selection. It shows the effectiveness of temporal smoothness in that it can appropriately extract information from unimportant variables without being prone to over-fitting. Therefore, the temporal smoothness conforms better to the spirit of big data and would be a promising approach to massive dynamic process data in future.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Yuri A.W. Shardt at the University of Duisburg-Essen for his constructive suggestions.

REFERENCES

Boulesteix, A.L. and Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data *Briefings in Bioinformatics*, 8(1), 32–44.

Chong, I.G. and Jun, C.H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1), 103–112.

Dayal, B. and MacGregor, J.F. (1997). Improved PLS algorithms. *Journal of Chemometrics*, 11(1), 73–85.

Facco, P., Doplicher, F., Bezzo, F., and Barolo, M. (2009). Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. *Journal of Process Control*, 19(3), 520–529.

Galicia, H.J., He, Q.P., and Wang, J. (2011). A reduced order soft sensor approach and its application to a continuous digester. *Journal of Process Control*, 21(4), 489–500.

Kano, M., Miyazaki, K., Hasebe, S., and Hashimoto, I. (2000). Inferential control system of distillation compositions using dynamic partial least squares regression. *Journal of Process Control*, 10(2), 157–166.

Lin, B., Recke, B., Knudsen, J.K., and Jørgensen, S.B. (2007). A systematic approach for soft sensor development. *Computers & Chemical Engineering*, 31(5), 419–425.

Liu, J. (2014). Developing a soft sensor based on sparse partial least squares with variable selection. *Journal of Process Control*, 24(7), 1046–1056.

Mehmood, T., Liland, K.H., Snipen, L., and Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69.

Qin, S.J. (1998). Recursive PLS algorithms for adaptive data modeling. *Computers & Chemical Engineering*, 22(4), 503–514.

Qin, S.J. (2014). Process data analytics in the era of big data. *AIChE Journal*, 60(9), 3092–3100.

Shang, C., Huang, X., Suykens, J.A.K., and Huang, D. (2015). Enhancing dynamic soft sensors based on DPLS: A temporal smoothness regularization approach. *Journal of Process Control*, 28, 17–26.

Sharmin, R., Sundararaj, U., and Shah, S.L. (2006). Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant. *Chemical Engineering Science*, 61(19), 6372–6384.

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. (2002). *Least Squares Support Vector Machines*, volume 4. World Scientific.

Wang, D., Liu, J., and Srinivasan, R. (2010). Data-driven soft sensor approach for quality prediction in a refining process. *IEEE Transactions on Industrial Informatics*, 6(1), 11–17.

Wang, Z., He, Q., and Wang, J. (2014). Comparison of different variable selection methods for partial least squares soft sensor development. In *American Control Conference (ACC), 2014*, 3116–3121. IEEE.

Wold, S., Sjstrm, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

Copyright © 2015 IFAC