Softsensor Development and Disturbance Estimation through Two-Stage Subspace Identification

Seunghyun Lee, Manabu Kano, and Shinji Hasebe

Abstract—Softsensor design is one of the key technologies in industry, because important variables such as product quality are not always measured on-line. Therefore, to reduce off-specification products and enhance productivity, the development of an accurate softsensor is crucial. In the present work, two-stage subspace identification (SSID) is proposed to develop highly accurate softsensors that can take into account the influence of unmeasured disturbances on estimated key variables. The procedure of two-stage SSID is as follows: 1) identify a state space model by using measured input and output variables, 2) estimate unmeasured disturbance variables from residual variables, and 3) identify a state space model to estimate key variables from the estimated disturbance variables and the other measured input variables. The proposed two-stage SSID can estimate unmeasured disturbances without assumptions that the conventional Kalman filtering technique must make; thus it can outperform the Kalman filtering technique when innovations are not Gaussian white noises or the properties of disturbances do not stay constant with time. The superiority of the proposed method over conventional methods is demonstrated through their application to an industrial ethylene fractionator.

I. INTRODUCTION

Product quality is not always measured on-line and its estimates are useful for realizing feedback control; thus softsensors play an important role in achieving better industrial productivity. To build softsensors, statistical methods have been widely used. When many process variables are used as input variables of a statistical model, the highly correlated nature of process data must be taken into account. In distillation processes, for example, tray temperatures close to each other change in nearly the same way. Applying a statistical modeling method to such highly correlated data causes a collinearity problem. To solve this problem, composition estimators using partial least squares (PLS) have been widely used [1], [2], [3]. Kano et al. (2000) further investigated PLS-based inferential models, which can estimate product compositions of a multicomponent distillation column from on-line measured process variables [4]. They compared steady-state, static, and dynamic inferential models and found that the estimation accuracy could be greatly improved by using dynamic models. More recently, Kamohara et al. investigated the integration of a softsensor with multivariate statistical process control (MSPC) to check the reliability of both the softsensor and an analyzer [5]. The softsensor and the MSPC system were developed by using dynamic PLS, and the effectiveness of their integration was demonstrated through its application to an industrial distillation process.

In recent years, much research on subspace identification (SSID) has been conducted, and several applications of SSID to softsensor design have been reported [6]. However, the performance of the conventional methods based on the Kalman filtering technique is limited due to the assumption that innovations are Gaussian white noises and the properties of disturbances stay constant with time. In other words, the conventional methods do not use measured variables effectively, while measured variables contain valuable information on a process including unmeasured disturbances that have serious influence on key variables.

In the present work, two-stage SSID is proposed to develop highly accurate softsensors that can take into account the influence of unmeasured disturbances on key variables explicitly. The proposed method can estimate unmeasured disturbances without assumptions that the conventional Kalman filtering technique must make. The usefulness of the proposed method is demonstrated through its application to an industrial ethylene fractionator.

The remainder of this paper is organized as follows: In section 2, SSID is explained in brief. In section 3, the proposed two-stage SSID method is described in detail. Application results are provided in section 4, where dynamic PLS, SSID, and two-stage SSID are compared in prediction performance. Finally, conclusions are given in section 5.

II. CONVENTIONAL SSID-BASED METHOD

In this section, a conventional softsensor design method based on SSID is briefly explained.

A. Subspace Identification

SSID is a method for identifying the following state space model directly from input-output data by using QR decomposition and singular value decomposition (SVD) [7].

\[ x(t+1) = Ax(t) + Bu(t) + Ke(t) \] (1)
\[ y(t) = Cx(t) + Du(t) + e(t) \] (2)

where, \( x \), \( u \), \( y \), and \( e \) denote state, input, output, and innovation vectors respectively, and \( K \) is the Kalman gain.

Although various algorithms for SSID have been proposed and they use different algorithms for estimating coefficient matrices, all algorithms adopt fundamentally the same approach: derive a subspace spanned by state variables from a part of input-output data, and then estimate coefficient matrices using the derived subspace and the other part of input-output data.
B. Conventional SSID-based Method

In the conventional SSID-based softsensor design method, a state space model is identified through SSID, state variables are estimated by Kalman filtering technique, and finally key variables such as quality variables are estimated.

To build a softsensor, state variables \( x \) need to be estimated from measured output variables \( y_m \) and measured input variables \( u \), because key variables \( y_q \) are not measured on-line. The process model is given by:

\[
x(t + 1) = Ax(t) + Bu(t) + Ke_m(t)
\]

\[
\begin{bmatrix}
    y_q(t) \\
    y_m(t)
\end{bmatrix} = 
\begin{bmatrix}
    C_q \\
    C_m
\end{bmatrix} x(t) +
\begin{bmatrix}
    D_q \\
    D_m
\end{bmatrix} u(t) +
\begin{bmatrix}
    e_q(t) \\
    e_m(t)
\end{bmatrix}
\]  

(4)

On the basis of this model, estimates of state variables \( \hat{x} \) are derived through the following filtering equations:

\[
\hat{x}(t+1 | t) = A\hat{x}(t | t) + Bu(t)
\]

\[
\hat{x}(t | t) = \hat{x}(t | t-1) + K (y_m(t) - C_m \hat{x}(t | t-1) - D_m u(t))
\]

(6)

Finally, \( y_q \) is estimated through

\[
y_q(t) = C_q \hat{x}(t | t) + D_q u(t)
\]

(7)

Stochastic effects including disturbances and noises can be taken into account by using the Kalman filter. For the effective functioning of the Kalman filter, disturbances should be generated from Gaussian white noises and dynamics from the white noises to the output variables should stay constant with time. However, in actual processes, such assumptions are not always valid.

III. TWO-STAGE SUBSPACE IDENTIFICATION

In this section, the proposed two-stage SSID is explained in detail. The procedure for two-stage SSID is as follows: 1) identify a state space model by using measured input and output variables, 2) estimate unmeasured disturbance variables from residual variables, and 3) identify a state space model to estimate key variables from the estimated disturbance variables and the other measured input variables.

A. Identification (1)

It is assumed that a process can be described by a linear state space model of the form:

\[
x(t + 1) = Ax(t) + Bu(t) + w(t)
\]

\[
\begin{bmatrix}
    y_q(t) \\
    y_m(t)
\end{bmatrix} = Cx(t) + e(t)
\]

(9)

where \( w \in \mathbb{R}^n \) and \( e \in \mathbb{R}^{q+m} \) are white noise sequences, \( u \in \mathbb{R}^d \) are measured input variables except manipulated variables \( u_{mv} \in \mathbb{R}^q \), \( u_s \in \mathbb{R}^s \) are unmeasured disturbance variables, \( y_q \in \mathbb{R}^q \) are key variables to estimate, and \( y_m \in \mathbb{R}^m \) are other measured output variables.

First, a state space model from measured input variables \( u_{mv} \) and \( u_d \) to output variables \( y_m \) is identified through SSID. At this stage, it is assumed that \( u_{mv} \) and \( u_d \) satisfy the persistently exciting condition.

B. Estimation of disturbance variables

The influence of the measured input variables \( u_{mv} \) and \( u_d \) on the output variables \( y_m \) is modeled in the previous stage. However, there must be residual because \( y_m \) are affected not only by the measured input variables but also by unmeasured factors including \( u_s \). In other words, residual variables \( \Delta y_m \) defined as

\[
\Delta y_m = y_m - y_m
\]

have valuable information about the unmeasured disturbance variables \( u_s \). To estimate \( u_s \) with accuracy, it is worth estimating \( u_s \) from \( \Delta y_m \) and using the estimated disturbances \( \hat{u}_s \) as input variables together with \( u_{mv} \) and \( u_d \).

Input signals have to be persistently exciting so that a statistical model is identifiable. However, it is possible that the residual variables \( \Delta y_m \) are linearly dependent and not persistently exciting. Therefore, it is necessary to derive \( \hat{u}_s \) that satisfy the persistently exciting condition. Although the dimensionality reduction of \( \Delta y_m \) can be achieved through principal component analysis (PCA), principal components cannot be used as \( \hat{u}_s \) because they might not satisfy the persistently exciting condition.

The first step to estimate \( \hat{u}_s \) is to define block Hankel matrices \( U_{1|k,s} \in \mathbb{R}^{k \times N} \) and \( \Delta Y_{1|k,m} \in \mathbb{R}^{km \times N} \):

\[
U_{1|k,s} =
\begin{bmatrix}
    u_s(1) & u_s(2) & \cdots & u_s(N) \\
    u_s(2) & u_s(3) & \cdots & u_s(N+1) \\
    \vdots & \vdots & \ddots & \vdots \\
    u_s(k) & u_s(k+1) & \cdots & u_s(N+k-1)
\end{bmatrix}
\]

(11)

\[
\Delta Y_{1|k,m} =
\begin{bmatrix}
    \Delta y_m(1) & \Delta y_m(2) & \cdots & \Delta y_m(N) \\
    \Delta y_m(2) & \Delta y_m(3) & \cdots & \Delta y_m(N+1) \\
    \vdots & \vdots & \ddots & \vdots \\
    \Delta y_m(k) & \Delta y_m(k+1) & \cdots & \Delta y_m(N+k-1)
\end{bmatrix}
\]

(12)

where \( N \) is large enough in comparison with \( km \). These block Hankel matrices satisfy the following equation derived through LQ decomposition.

\[
\begin{bmatrix}
    U_{1|k,s} \\
    \Delta Y_{1|k,m}
\end{bmatrix} =
\begin{bmatrix}
    L_o & L_b & L_c
\end{bmatrix}
\begin{bmatrix}
    Q_1 \\
    Q_2
\end{bmatrix}^T
\]

(13)

\[
\begin{bmatrix}
    L_o \\
    L_b
\end{bmatrix} Q_1^T
\]

(14)

By using (14), \( u_s \) can be estimated from \( Q_1 \), which is derived from \( \Delta y_m \).

The block Hankel matrix \( \Delta Y_{1|k,m} \) is decomposed by SVD:

\[
\Delta Y_{1|k,m} = USV^T
\]

(15)

\[
S = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_{km})
\]

(16)

\[
V = [v_1 \ v_2 \ \cdots \ \ v_N]
\]

(17)

where the singular values are in descending order. All row vectors of \( \Delta Y_{1|k,m} \) are mean centered and their standard deviations are scaled to be unity.
A matrix $Q ∈ ℝ^{N×kl}$ is generated from the orthogonal matrix $V$ by selecting column vectors corresponding to significant (non-zero) singular values.

$$Q = [q_1, q_2, \ldots, q_{kl}]$$  \hspace{1cm} (18)

$$q_i = [q_{1i}, q_{2i}, \ldots, q_{Ni}]^T$$  \hspace{1cm} (19)

where $l$ is the number of significant singular values, and it is the same as the number of the estimated disturbance variables $\hat{u}_s$.

The estimated disturbance variables $\hat{u}_s$ are derived by solving the following equation:

$$\hat{U}_{1:k,s} = LQ^T$$  \hspace{1cm} (21)

where $L ∈ ℝ^{kl×kl}$ is a lower triangular matrix.

$$L = \begin{bmatrix} L_{11} & L_{21} & L_{22} \\ L_{k1} & L_{k2} & \cdots & L_{kk} \end{bmatrix}$$  \hspace{1cm} (22)

Here, it is assumed that

$$L_{ii} = diag\{σ_{i(i-1)+1}, σ_{i(i-1)+2}, \cdots, σ_{ii}\} .$$  \hspace{1cm} (23)

The estimated disturbance variables $\hat{u}_s$ are persistently exciting of order $k$, because $L$ is a regular matrix and $Q$ has full column rank.

In the following part of this subsection, the procedure for solving (21) is described. As for the first $2l$ rows of (21), the following equations are derived.

$$\hat{u}_s(2) = L_{21} \begin{bmatrix} q_{11} \\ \vdots \\ q_{1l} \end{bmatrix} + L_{22} \begin{bmatrix} q_{1(l+1)} \\ \vdots \\ q_{1(2l)} \end{bmatrix}$$  \hspace{1cm} (24)

$$= L_{11} \begin{bmatrix} q_{21} \\ \vdots \\ q_{2l} \end{bmatrix}$$  \hspace{1cm} (25)

$$\hat{u}_s(3) = L_{21} \begin{bmatrix} q_{21} \\ \vdots \\ q_{2l} \end{bmatrix} + L_{22} \begin{bmatrix} q_{2(l+1)} \\ \vdots \\ q_{2(2l)} \end{bmatrix}$$  \hspace{1cm} (26)

$$= L_{11} \begin{bmatrix} q_{31} \\ \vdots \\ q_{3l} \end{bmatrix}$$  \hspace{1cm} (27)

Similar equations can be derived regarding $\hat{u}_s(i)$ for $i = 2, 3, \cdots, N$, and they result in

$$\begin{bmatrix} L_{21} & L_{22} \end{bmatrix} = L_{11} \begin{bmatrix} q_{21} & q_{22} & \cdots & q_{2(N-1)l} \\ \vdots & \vdots & \cdots & \vdots \\ q_{1(l+1)} & q_{2(l+1)} & \cdots & q_{1(N-l)l} \end{bmatrix} + L_{11} \begin{bmatrix} q_{(l+1)1} & q_{(l+1)2} & \cdots & q_{(l+1)(N-l)1} \\ \vdots & \vdots & \cdots & \vdots \\ q_{1(2l)} & q_{2(2l)} & \cdots & q_{1(N-l)(2l)} \end{bmatrix}$$  \hspace{1cm} (28)

This equation can be rewritten as

$$L_{21} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} - L_{21} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} = L_{22} \begin{bmatrix} q_i^{T+1,1} \\ \vdots \\ q_i^{T+1,2l} \end{bmatrix} + L_{11} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} \Xi_i - L_{11} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix}$$  \hspace{1cm} (29)

where

$$\tilde{q}_{i,j} = \begin{bmatrix} 0 & \cdots & 0 & q_{(N-j+1)i} & \cdots & q_{Ni} \end{bmatrix}^T$$  \hspace{1cm} (30)

$$\Xi_i = \begin{bmatrix} 0_{i×i} \\ I_{N-i} \end{bmatrix}$$  \hspace{1cm} (31)

and $I_j$ is a unit matrix. Equation (29) post-multiplied by $Q_i$ becomes

$$L_{21} \begin{bmatrix} I_i & \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} \end{bmatrix} Q_i =$$

$$\begin{bmatrix} L_{22} \begin{bmatrix} q_i^{T+1,1} \\ \vdots \\ q_i^{T+1,2l} \end{bmatrix} + L_{11} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} \Xi_i - L_{11} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix} \end{bmatrix} \begin{bmatrix} q_i^T \\ \vdots \\ q_i^{T+1} \end{bmatrix}$$  \hspace{1cm} (33)

where

$$Q_i = \begin{bmatrix} q_1 & \cdots & q_l \end{bmatrix} .$$  \hspace{1cm} (34)

In (33), only the matrix $L_{21}$ is unknown. Thus, $L_{21}$ can be determined by solving (33). In the same way, a more general equation can be derived.

$$\begin{bmatrix} L_{i1} & \cdots & L_{i(i-1)} \end{bmatrix} \begin{bmatrix} I_{l(i-1)} - \begin{bmatrix} q_{i1}^T,1 \\ \vdots \\ q_{i1}^{T+1,1} \end{bmatrix} Q_{l(i-1)} \end{bmatrix} =$$

$$\begin{bmatrix} L_{ii} \begin{bmatrix} q_{i(i-1)+1,1}^T \\ \vdots \\ q_{i(i-1)+1,2l}^T \end{bmatrix} + L_{i(i-1)} \begin{bmatrix} q_{i1}^T \\ \vdots \\ q_{i1}^{T+1} \end{bmatrix} \Xi_i - L_{i(i-1)} \begin{bmatrix} q_{i1}^T \\ \vdots \\ q_{i1}^{T+1} \end{bmatrix} \end{bmatrix} Q_{l(i-1)}$$  \hspace{1cm} (35)
By solving (35) in sequence, \( L \) is derived. Finally, the estimated disturbance variables \( \hat{u}_s \) are calculated from (21).

It is possible to derive \( \hat{u}_s \), which is sufficiently excited to build a statistical model through SSID, by using sufficiently large \( k \) in (12). Since all input variables, \( u_{mv}, u_d, \) and \( \hat{u}_s \), must be persistently exciting, the parameter \( k \) needs to be larger than or equal to the settling time of the process. From a practical viewpoint, only the dynamics between influential inputs and outputs need to be taken into account.

In addition, \( \hat{u}_s \) at \( k = 1 \) are persistently exciting when \( \Delta y_m \) are generated through an ARMA (autoregressive moving-average) process, because the ARMA process is persistently exciting of infinite order. In such a case, \( u_s \) can be calculated by PCA, and \( \hat{u}_s \) are the same as principal component scores.

C. Identification (2)

The two-stage SSID-based softsensor is developed by using measured input variables, \( u_{mv} \) and \( u_d \), and the estimated disturbance variables, \( \hat{u}_s \), as inputs in the state space model of the form:

\[
x(t+1) = Ax(t) + B \begin{bmatrix} u_{mv}(t) \\ u_d(t) \\ \hat{u}_s(t) \end{bmatrix}
\]

\[
\hat{y}_q(t) = Cx(t) + D_q \begin{bmatrix} u_{mv}(t) \\ u_d(t) \\ \hat{u}_s(t) \end{bmatrix}.
\]

This state space model can be identified through SSID.

The accuracy of the softsensor does not deteriorate even when unmeasured disturbances are significant, because the two-stage SSID-based softsensor can take into account the influence of unmeasured disturbances on the measured output variables \( y_m \) and the key variables \( y_q \). In addition, the dynamics from the manipulated input variables \( u_{mv} \) and \( u_d \) to the key variables \( y_q \) can be modeled accurately, because the estimated disturbance variables \( \hat{u}_s \) do not correlate to the measured input variables \( u_{mv} \) and \( u_d \).

D. On-line estimation

On-line estimation is executed in the following procedure: 1) \( \hat{y}_m \) is calculated through the first state space model and \( \Delta y_m \) is derived, 2) \( \hat{u}_s \) is calculated by solving (21), in which \( Q \) is determined from \( \Delta y_m \), and 3) \( \hat{y}_q \) is calculated through the second state space model.

IV. CASE STUDY

In this section, the usefulness of the proposed two-stage SSID-based softsensor is demonstrated through its application to an industrial ethylene fractionator [5].

A. Ethylene fractionator

The schematic diagram of the industrial ethylene fractionator is shown in Fig. 1. This ethylene fractionator consists of two columns: the bottom column T431 and the top column T432. The feed stream enters the bottom column, and the product ethylene is drawn from the top column. The main specification is the ethane concentration in the ethylene product (16). This fractionator is controlled by multivariable model predictive control. The numbers of controlled variables, manipulated variables, and disturbance variables are seven, four, and three, respectively. The controlled variables are the ethane concentration and methane concentration in the ethylene product, T431 tray #29 temperature (1), T431 differential pressure, T342 differential pressure, condenser pot level, and reboiler pot level. Manipulated variables are the T431 reboiler flow rate (7), T432 internal reflux flow rate (9), T432 purge flow rate (10), and T432 top pressure (12). The disturbance variables are the T431 feed flow rate, T431 feed ethane concentration (13), and C351 #4 suction pressure (14). Here, C351 is a propylene refrigerant. Its #4 suction pressure affects propylene refrigerant temperature and reboiler heat duty. The numbers in parentheses correspond to those shown in Fig. 1 and Table I.

B. Softsensor Design

Softsensor design is one of the key technologies for reducing off-specification products and producing enhancing productivity when on-line analyzers are not available. In this study, softsensors, which estimate the ethane concentration in the ethylene product, are developed through dynamic PLS, SSID, and two-stage SSID.

For building softsensors, operation data obtained in the period from January 1 to February 20, 2002 were used. All variables used for modeling are listed in Table I. Variables #7, #9, #10, and #12 are classified into manipulated variables \( u_{mv} \), variables #8, #11, #13, and #14 are classified into other measured input variables \( u_d \), variables #1, #2, #3, #4, #5, #6, and #15 are classified into measured output variables \( y_m \), and variable #16 is the key quality variable \( y_q \) to estimate.

1) Dynamic PLS: Kano et al., who investigated steady-state, static, and dynamic PLS-based inferential models, found that the estimation accuracy could be greatly improved by using dynamic PLS (DPLS) models [4]. In addition, Kamohara et al. applied a DPLS-based softsensor to an industrial distillation process [5], which was investigated...
TABLE I

VARIABLES USED FOR MODELING

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T431 tray #29 temperature</td>
</tr>
<tr>
<td>2</td>
<td>T431 bottom temperature</td>
</tr>
<tr>
<td>3</td>
<td>T431 top temperature</td>
</tr>
<tr>
<td>4</td>
<td>T431 tray #37 temperature</td>
</tr>
<tr>
<td>5</td>
<td>T431 tray #129 temperature</td>
</tr>
<tr>
<td>6</td>
<td>Flow rate from T432 to T431</td>
</tr>
<tr>
<td>7</td>
<td>T431 reboiler flow rate</td>
</tr>
<tr>
<td>8</td>
<td>Product ethylene flow rate</td>
</tr>
<tr>
<td>9</td>
<td>T432 internal reflux flow rate</td>
</tr>
<tr>
<td>10</td>
<td>T432 purge flow rate</td>
</tr>
<tr>
<td>11</td>
<td>T432 reflux ratio</td>
</tr>
<tr>
<td>12</td>
<td>T432 top pressure</td>
</tr>
<tr>
<td>13</td>
<td>T431 feed ethane concentration</td>
</tr>
<tr>
<td>14</td>
<td>C351 #4 suction pressure</td>
</tr>
<tr>
<td>15</td>
<td>V359 level (cooling propylene)</td>
</tr>
<tr>
<td>16</td>
<td>Ethane concentration in the ethylene product</td>
</tr>
</tbody>
</table>

in the present work. Therefore, DPLS is used here for developing a softsensor.

The operation data including current measurements and those 5, 10, 15, 20, 25, 30, 35, 40, 45 minutes before were used to build a softsensor. The number of latent variables was adjusted to 20. Here, all input variables and an output variable were mean-centered and their standard deviations were scaled to be unity.

2) Conventional SSID: An advantage of using SSID for softsensor design is that the dynamics of a multivariable process can be easily taken into account, whereas the number of input variables drastically increases in a dynamic PLS approach and input variables should be appropriately selected to achieve the good prediction performance.

In the present work, two types of softsensors were built through conventional SSID to evaluate the effect of using the measured input variables \( u_d \) on the estimation accuracy.

- softsensor C1 : \( u_{mv} \rightarrow [y_q, q_m] \)
- softsensor C2 : \( u_{mv}, u_d \rightarrow [y_q, q_m] \)

Here, \([a] \rightarrow [b]\) represents a state space model from \( a \) to \( b \). To identify state space models, N4SID (numerical algorithms for subspace state space system identification) was used [8]. The numbers of state variables \( x \) in C1 and C2 were determined to 15 on the basis of the estimation results for the validation data. All input and output variables were mean-centered and their standard deviations were scaled to be unity.

3) Two-stage SSID: In the same way as conventional SSID, two types of softsensors were built through two-stage SSID.

- softsensor TS1
  - the first model : \( u_{mv} \rightarrow [y_q] \)
  - the second model : \( u_{mv}, \hat{u}_s \rightarrow [y_q] \)
- softsensor TS2
  - the first model : \( u_{mv}, u_d \rightarrow [y_q] \)
  - the second model : \( u_{mv}, u_d, \hat{u}_s \rightarrow [y_q] \)

To identify state space models, N4SID was used. The results of the past research show that the settling time of this ethylene fractionator is about 50 minutes [5]. Consequently, the value of \( k \) is set to 10. The number of estimated disturbance variables \( \hat{u}_s \) and the number of state variables \( x \) were determined as shown in Table II. These numbers of variables were selected on the basis of the estimation results for the validation data. All input and output variables were mean-centered and their standard deviations were scaled to be unity.

C. Estimation results

The developed softsensors were validated by using operation data obtained from the industrial ethylene fractionator during three different periods of time: (a) December 9 through December 16, 2001, (b) December 21 through December 31, 2001, and (c) May 1 through May 10, 2002.

The estimation results are shown in Fig. 2 and Table III. Here, measurements and estimates of ethane concentration in the ethylene product (Ethane conc.) are scaled. The estimation accuracy of the softsensors was evaluated by both the correlation coefficient \( R \) and RMSE (root mean squared error).

In periods (a) and (b), the softsensors C2 and TS2, both of which use \( u_{td} \) as input variables, function better than the others. In addition, the softsensors TS1 and TS2, both of which use \( \hat{u}_s \) as input variables, function better than the softsensors C1 and C2, respectively. The results show that the use of measured input variables \( u_{td} \) and estimated unmeasured disturbance variables \( \hat{u}_s \) is effective for improving the estimation accuracy of softsensors. In the same period, the DPLS-based softsensor shows better estimation accuracy than the softsensors C1 and TS1, both of which do not use \( u_{td} \) as input variables.

In period (c), the estimation accuracy, RMSE, of all softsensors deteriorates significantly, and large bias is observed in Fig. 2. Such bias is generally caused by changes in operating condition. As shown in Fig. 3, T432 top pressure (12) in period (c) is quite different from that in the other

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>THE NUMBER OF VARIABLES USED IN TWO-STAGE SSID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conventional SSID</td>
</tr>
<tr>
<td></td>
<td>C1</td>
</tr>
<tr>
<td>State variables ( x )</td>
<td>First model</td>
</tr>
<tr>
<td>Second model</td>
<td>6</td>
</tr>
<tr>
<td>Estimated disturbance variables ( \hat{u}_s )</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>COMPARISON OF SOFTSENSORS DEVELOPED THROUGH DPLS, SSID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(C1 and C2), and Two-stage SSID (TS1 and TS2).</td>
</tr>
<tr>
<td>( R )</td>
<td>(a)</td>
</tr>
<tr>
<td>DPLS</td>
<td>0.75</td>
</tr>
<tr>
<td>Conventional SSID</td>
<td>0.13</td>
</tr>
<tr>
<td>C1</td>
<td>0.88</td>
</tr>
<tr>
<td>C2</td>
<td>0.70</td>
</tr>
<tr>
<td>TS1</td>
<td>0.90</td>
</tr>
<tr>
<td>TS2</td>
<td>0.90</td>
</tr>
</tbody>
</table>
periods (a), (b), and (modeling). Several other variables in period (c) are also different. The softsensor C2 achieves as high correlation coefficient $R$ as TS2, but its RMSE is much larger. The combination of high $R$ and large RMSE is caused by bias. The DPLS-based softsensor achieves the smallest RMSE, but its $R$ is lower than those of C2 and TS2. The correlation coefficients $R$ of C2 and TS2 are markedly larger than those of C1 and TS1.

Consequently, both $u_d$ and $\hat{u}_s$ should be used to develop a highly accurate softsensor.

V. CONCLUSIONS

In the present work, two-stage subspace identification (SSID) was proposed to develop an accurate softsensor that could take into account the influence of unmeasured disturbances on estimated key variables. The procedure of two-stage SSID is as follows: 1) identify a state space model by using measured input and output variables, 2) estimate unmeasured disturbance variables from residual variables, and 3) identify a state space model to estimate key variables from the estimated disturbance variables and the other measured input variables. The proposed two-stage SSID can estimate unmeasured disturbances without assumptions that the conventional Kalman filtering technique must make; thus it can outperform the Kalman filtering technique when innovations are not Gaussian white noises or the properties of disturbances do not stay constant with time. The superiority of the proposed method over conventional methods, dynamic PLS and SSID, was demonstrated through their application to the industrial ethylene fractionator. To develop accurate softsensors, it is very effective to use measured input variables including manipulated variables $u_{mv}$ and disturbances $u_d$ and also estimated unmeasured disturbance variables $\hat{u}_s$. The proposed two-stage SSID can cope with these three types of inputs systematically; thus it can realize highly accurate softsensors.

VI. ACKNOWLEDGMENTS

This research was partially supported by the JFE 21st Century Foundation.

REFERENCES