Using Structured and Unstructured Estimators for Distillation Units: a Critical Comparison

Fabrizio Bezzo*, Riccardo Muradore, Massimiliano Barolo
DIPIC – Dipartimento di Principi e Impianti di Ingegneria Chimica
Università di Padova, via Marzolo 9, I-35131 Padova PD, Italy

Abstract
The paper discusses the use of several structured and unstructured estimators for the control of a binary distillation column. An extended Kalman filter, three multivariate regression-based estimators, and a novel subspace identification-based estimator are compared in terms of design effort, on-line computational demand and control performance.

Keywords: inferential sensors, subspace identification methods, distillation columns.

1. Introduction
Modern production control methods require the availability of reliable and timely information on product quality. When hardware quality measurement instrumentation cannot be used on-line, a common practice is to make use of the available process knowledge (first-principles models and on-line “secondary” measurements) to estimate the desired “primary” variables. Quality estimators based on first-principles models are usually denoted as structured. If a reliable physical model of the plant is not available (or is too burdensome for on-line use), unstructured estimators are preferred, which use some form of multivariate statistical methods to relate the secondary measurements to the primary ones.

Following the work of Roffel et al. (2003), in this paper a structured estimator and several unstructured estimators are developed and critically compared to assess their effectiveness for the control of a distillation column. An extended Kalman filter, three multivariate regression-based estimators, and a novel subspace identification-based estimator are compared in terms of design effort, on-line computational demand, and performance in closed-loop operation.

2. Case Study
A continuous distillation column model will be used as a case study; this model will be referred to as “the plant” in the following. The 12-tray column separates an ethanol/water mixture. Column design and operating parameters can be found in Bezzo et al. (2004). The column is described by a detailed tray-to-tray model taking into account mass and energy balances, with nonlinear tray hydraulics based on the actual

* Author to whom correspondence should be addressed: fabrizio.bezzo@unipd.it
tray geometry. Vapour-liquid equilibria are described by an NRTL model. The control objective is to keep the distillate and bottoms compositions close to their setpoints in spite of unmeasured disturbances on the feed. The LV control configuration is used. It is assumed that tray temperature measurements are available on-line in order to reconstruct the product composition values that are used as controller inputs. Note that, since the purpose of this study is to evaluate the performance of different composition estimators and not to devise the best control strategy for this column, we are not considering the (simpler) control approach where two pilot tray temperatures are considered as the controlled variables.

3. Estimator Design

The vector of estimated compositions at sampling time $k$ will be indicated with $\hat{z}(k) = \begin{bmatrix} \hat{x}_D(k) & \hat{x}_B(k) \end{bmatrix}^T$, where the $x_D$ and $x_B$ are the ethanol mole fraction in the distillate and bottoms (respectively), and $\hat{\cdot}$ indicates an estimated property. It will be assumed that only two tray temperatures (whose optimal locations are to be determined) can be made available to each estimator. A set of closed-loop data (under conventional temperature control) is also supposed to be available; these data are used to select the optimal temperature locations for each estimator, and to calibrate the estimators.

3.1 Extended Kalman filter (EKF)

The structured estimator considered in this work is the well-known extended Kalman filter (Söderström, 2002), whose formulation requires a state-space model of the plant. In order to keep the on-line calculation time reasonably low, this model must be a simplified representation of the plant. Therefore, several simplifications were made, such as constant vapour flows in each section of the column, piecewise linear vapour-liquid equilibria, linear tray hydraulics. A discrete-time representation of this model was then obtained, so that the EKF can be written as:

$$
\begin{align*}
\dot{\hat{x}}(k+1) &= \left[ F(\hat{x}(k)) - K(k) C_{opt}(\hat{x}(k)) \right] \hat{x}(k) + K(k)y_{opt}(k), \\
\hat{z}(k) &= L \hat{x}(k)
\end{align*}
$$

where $\hat{x}$ is the state vector estimation, $F(\cdot)$ and $C_{opt}(\cdot)$ are the Jacobians of the nonlinear functions of the model, $K$ is the Kalman gain, $L$ is the matrix relating the full state vector to the distillate and bottoms compositions, $y_{opt}$ is the optimal set of temperature measurements. A two-step procedure was devised to select the elements of $y_{opt}$. The first temperature measurement location is selected in such a way that an EKF using this single measurement provides the largest explained variance of the vector of estimated compositions. A similar approach is used to select the second location, with the EKF now employing two measurements (with the first one determined in the previous step).

3.2 Static linear estimator (SLE)

The static linear estimator takes the form:

$$
\hat{z}(k) = A y_{opt}(k) + b
$$


where matrix $A$ and vector $b$ are computed using a least-squares regression approach. Note that if more (than two) temperature measurements were available, an alternative approach could be to rely on all the available measurements and use partial least-squares regression to eliminate the input measurement collinearity and to design the estimator (Mejdell and Skogestad, 1991).

In order to select the optimal temperature measurement locations, a procedure developed by Muradore et al. (2004) was used. Let $Y$ and $Z$ be the matrices containing the temperatures on each tray and the compositions $x_D$ and $x_B$ for a sufficient large time interval (calibration data), respectively. The procedure selects the optimal locations in a sequential manner using the sample correlation matrix between $Y$ and $Z$, and the explained variance of the $Z$-block as a stopping rule.

3.3 Dynamic linear estimator (DLE)

Augmenting the input matrix $Y_{opt}$ with past temperature measurements produces a dynamic linear estimator in the form:

$$
\hat{z}(k) = A[y_{opt}(k) \ y_{opt}(k-1) \ \cdots \ y_{opt}(k-\tau)] + b.
$$

In this case, the $A$ and $b$ matrices are determined via least-squares regression, while the number $\tau$ of past measurements is selected by trial and error. The optimal temperature locations are calculated in the same way as done for the SLE.

3.4 Static nonlinear estimator (SNLE)

The static nonlinear estimator uses a nonlinear mapping between the optimal selected temperatures and the estimated concentrations. A second-order polynomial function is used to relate the input and output data:

$$
\hat{z}(k) = A_1 y_{opt}(k) + A_2 y_{opt}^2(k) + b,
$$

where the polynomial coefficients are determined via least-squares regression. The optimal temperature locations are selected as done for the SLE.

3.5 Subspace identification (SID) based estimator

The last unstructured estimator considered in this work is based on a subspace identification approach (Van Overschee and De Moor, 1993). The advantage of this approach is that the identification algorithm directly provides the dynamic model order, so that the problem related with the determination of $\tau$ in the DLE is circumvented. The first step in SID is to identify a linear state-space dynamic model for concentrations and temperatures. With reference to equation (5), calibration data matrices $Y$ and $Z$ are used to compute matrices $F$, $K$, $L$, $C$. These matrices allow mapping the innovation $[e_z(k) \ e_y(k)]^T$, which is strictly related to the process and measurement noise in the plant, into compositions and temperatures. Therefore, the identified model takes the form:

$$
\begin{align*}
x(k+1) &= F x(k) + K [e_z(k) \ e_y(k)] \\
z(k) &= L x(k) + e_z(k) \\
y(k) &= C x(k) + e_y(k).
\end{align*}
$$

The analysis of the observability Gramian allows to select the temperature locations that are more related to the state vector. Finally, a Kalman filter is designed:
\[
\hat{x}(k+1) = F \hat{x}(k) + G(k) \left[ y_{\text{opt}}(k) - C_{\text{opt}} \hat{x}(k) \right],
\]
\[
\tilde{z}(k) = L \hat{x}(k) + J \left[ y_{\text{opt}}(k) - C_{\text{opt}} \hat{x}(k) \right],
\]

where \( G(k) \) is the Kalman gain and matrix \( J \) takes into account the covariance between \( e_y \) and \( e_z \).

\[8]

**Figure 1.** Product composition responses to a feed flow disturbance for the ideal and real case.

### 4. Simulation Results

A closed-loop test is used to compare the estimator performances when a +10% step feed flow disturbance is delivered to the column. All simulations are compared to an “ideal” case, where it is assumed that the product compositions are available in real time and the quality controllers’ inputs/outputs are continuous. Figure 1 compares the control responses of the ideal case to a “real” case where the composition signal is sampled and delayed to take into account the analyser delay (which is assumed equal to 10 min). The quality loops are tuned according to the relay-feedback procedure suggested by Shen and Yu (1994). Note that the performance of the real case is quite unsatisfactory and advocates the use of on-line composition estimators.

**Table 1.** Selected temperature sensors and accuracy indices for different estimator designs.

<table>
<thead>
<tr>
<th>Selection</th>
<th>( x_D ) RMS ([\times 10^3])</th>
<th>ExVar</th>
<th>( x_B ) RMS ([\times 10^3])</th>
<th>ExVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKF ( T_2, T_7 )</td>
<td>0.436</td>
<td>99.14</td>
<td>1.38</td>
<td>90.10</td>
</tr>
<tr>
<td>SLE ( T_3, T_7 )</td>
<td>1.76</td>
<td>63.13</td>
<td>2.06</td>
<td>77.22</td>
</tr>
<tr>
<td>SNLE ( T_3, T_6 )</td>
<td>1.68</td>
<td>66.18</td>
<td>1.96</td>
<td>79.27</td>
</tr>
<tr>
<td>DLE ( (\tau=1 \text{ min}) ) ( T_3, T_8 )</td>
<td>1.57</td>
<td>70.48</td>
<td>1.53</td>
<td>87.43</td>
</tr>
<tr>
<td>DLE ( (\tau=2 \text{ min}) ) ( T_3, T_8 )</td>
<td>1.42</td>
<td>75.96</td>
<td>1.30</td>
<td>90.95</td>
</tr>
<tr>
<td>SID ( T_3, T_9 )</td>
<td>1.13</td>
<td>84.92</td>
<td>2.06</td>
<td>77.69</td>
</tr>
</tbody>
</table>

Table 1 shows the optimal temperature sensors for each estimator, and compares the estimation accuracy in terms of root mean square (RMS) error and explained variance (ExVar) as obtained in the estimator design stage (the bottom tray is tray number 1).
Even if the ExVar and the RMS are the most sensible indices to assess the optimal sensor location, they appear to be less significant in order to compare the closed-loop performance of different estimators. In fact, Figure 2 shows that, although the EKF explains a considerably larger variance of the design data, all the other estimators exhibit a good performance in the closed-loop simulations (note that the profiles of the actual product compositions are reported on each graph).

![Figure 2. Product composition responses to a feed flow disturbance using different estimators.](image)

The SLE causes a not completely satisfactory control performance on $x_B$. Making the estimator nonlinear or dynamic greatly improves the performance. However, a potential issue arising in the design of SNLE’s and DLE’s is that a systematic method would be desirable to choose either the order of the polynomial function (SNLE) or the sampling delay and the number of past measurements (DLE). The SID-based estimator overcomes the shortcomings of the DLE, although it appears that the control response is
quite noisy in this case study. A way to counteract this problem could be to slightly detune the composition controllers. Anyway, the results achieved with this novel approach are promising, and further investigation is called for. The above estimators exhibit important differences in terms of design effort. The SLE, SNLE, DLE estimators are all very easy to design. The SLE does not need any tuning parameter, whereas the SNLE and DLE require the tuning of the polynomial order and of parameter $\tau$, respectively, as previously commented on. On the other hand, the “standard” EKF design is much more demanding since a sufficiently reliable model must be available, and a time-consuming filter tuning is necessary. The SID-based estimator is a Kalman filter, too, but the identification procedure generates a low-order black-box model that does not require tuning, thus greatly simplifying the design of the estimator. As far as the on-line computation time is concerned, it is quite obvious that SLE, SNLE, DLE are very little demanding. On the contrary, the computational burden of the EKF is quite heavy because of the Riccati equation update and of the fast sampling for robustness. The SID estimator appears to be a good compromise, since a low-order linear model is used.

5. Conclusions

Several structured and unstructured estimators have been compared in terms of design effort, on-line computational demand and effectiveness in a closed-loop test for composition control in a distillation column. The nonlinear and dynamic unstructured estimators proved to be a sensible choice for quality control. A subspace identification-based estimator (where a reasonably simple, easy-to-identify dynamic model is used within a Kalman filter framework) proved to be a promising alternative.

References


Acknowledgements

This research was carried out in the framework of the MIUR-PRIN 2002 project \textit{Operability and controllability of middle-vessel distillation columns} (ref. no. 2002095147:002).