AIChE

Input Reconstruction for Statistical-Based Fault Detection and Isolation

Udo Schubert
Dept. of Process Dynamics and Operation, Berlin Institute of Technology, Berlin D-10623, Germany

Uwe Kruger
Chemical Engineering Program, The Petroleum Institute, P.O. Box 2533, Abu Dhabi, United Arab Emirates

Günter Wozny and Harvey Arellano-Garcia
Dept. of Process Dynamics and Operation, Berlin Institute of Technology, Berlin D-10623, Germany

DOI 10.1002/aic.12693
Published online July 18, 2011 in Wiley Online Library (wileyonlinelibrary.com).

In this work, an input reconstruction scheme for detecting and isolating sensor, actuator, and process faults is proposed. The scheme uses model-based and statistical-based FDI methods, which yields an improved analysis of abnormal operation conditions in chemical processes. The main advantage of the proposed approach over existing works lies in the reconstruction of system inputs and the subsequent estimation of fault signatures. This advantage is demonstrated through simulation examples and the analysis of recorded process data from a reactive batch distillation column. © 2011 American Institute of Chemical Engineers *AIChE J,* 58: 1513–1523, 2012
Keywords: unknown input observer, input reconstruction, fault detection, fault isolation, simultaneous fault detection

Introduction

Several factors render the monitoring of complex processes in the chemical industry a difficult task. From the process perspective, it is the ever increasing complexity caused by increasing levels of automation, process intensification, and optimal operation at process constraints[1] for example. From the perspective of plant operators, difficulties arise from the out-of-the-loop operation of such systems, which implies that operators are less engaged in interactions to control the process.[2] Over the past few decades, an additional but not less important factor relates to the introduction of stringent legislation on emissions and process safety that require more advanced levels of process automation and monitoring.[3]

Meeting the increasing demand for more rigorous monitoring of such complex systems requires, therefore, the development of suitable and efficient methods that support both: (i) the monitoring of dynamic operation by plant personnel and (ii) the ability of effective fault detection and isolation (FDI). Significant progress has been achieved over the past few decades by introducing approaches that use causal models for model-based fault detection and identification (MBFDI).[4,5] The research literature has proposed an alternative research direction that involves statistical-based techniques, which are collectively referred to as multivariate statistical process control (MSPC).[6,7] The main advantage of MSPC is that it can handle the typically large variable sets that are routinely measured at steady-state conditions but may not work well if the process exhibits a significant dynamic behavior.[8] In contrast, MBFDI relies on causal models that accurately describe complex dynamic behavior. However, the development of such models is difficult, time consuming, and consequently, expensive for large-scale systems.[5]

To overcome the limitations of individual approaches, several contributions motivate a combination of MBFDI and MSPC, e.g., Ref. 9. For MSPC-based fault diagnosis, Yoon & MacGregor[8] as well as Gertler and Cao[10] studied the incorporation of the fault isolation capability of MBFDI. The general lack of an accurate description of the dynamic process behavior using MSPC remains with this approach, which has the potential to lead to incorrect and misleading results. Sotomayor and Odload[11] and Ding et al.[12] recently examined the incorporation of subspace model identification (SMI) into MBFDI in order to handle the modeling task for large-scale systems, which presents a framework for monitoring the typically large number of highly correlated variables, as described in Kourti,[6] using MBFDI methods. Ding et al.[12] suggested the design of a residual generator in the parity space, where standard MBFDI methods are not applicable. Sotomayor & Odload[11] use a state space model for MBFDI while addressing only simple sensor and actuator faults.

However, a scheme that efficiently uses individual techniques from MBFDI and MSPC has not yet been proposed in literature. This work introduces an input reconstruction scheme for sensor, actuator, and process FDI to (i) directly address dynamic process behavior by incorporating causal models into the MSPC framework and (ii) the utilization of statistical inference that is based on multivariate statistics by the MBFDI approach. The utility of the proposed scheme is demonstrated first through (i) an introductory example that exhibits a simple actuator fault and (ii) a mechanistic model of a continuous reactor that is prone to multiple fault

Correspondence concerning this article should be addressed to U. Schubert at usc@zmms.tu-berlin.de or U. Kruger at ukruger@pi.ac.ae.

conditions that involve a sensor bias and a simultaneously occurring fault describing a complex parametric process fault condition. Finally, the article presents the analysis of recorded data from a reactive batch distillation unit for which a complex process fault condition was recorded.

Motivation

This section motivates the rationale behind the proposed scheme which uses a combination of MBFDI and MSPC techniques. The next section describes the concept of input reconstruction and how to use it for detecting and diagnosing sensor, actuator and complex process faults. In the context of the proposed scheme, the advantage of MBFDI relates to the use of causal models, which are identified state space models. This, consequently, addresses the problem of providing detailed mechanistic models for large-scale systems directly.[8] However, MBFDI requires a priori knowledge of the dynamic properties of the specific fault conditions. Moreover, such mechanistic models are difficult to obtain for large-scale systems.[5] On the other hand, MSPC-based process monitoring is robust in detecting deviations from a steady state operation, but conventional fault diagnosis using contribution plots[13] or variable reconstruction based contributions[14] may yield ambiguous and misleading results for process fault conditions. This stems from the associated complex interactions between recorded process variables, which may suggest that a faulty variable affects several other output variables at the same time.[8,14]

The proposed scheme is based on reconstructed process inputs using measured outputs which can describe different fault scenarios, such as sensor, actuator, and process faults. The problem of existing first-principle models is circumvented by using subspace model identification (MSPC-based) in order to create suitable process models. State observers (MBFDI based) together with records of input/output variables use these models to track the process behavior with respect to common cause variation, model uncertainty, as well as fault conditions. The generated model residuals are further analyzed to detect an abnormal behavior using a few key performance indicators (KPI), i.e., by multivariate fault detection indices (MSPC based). The next section describes details of the observer-based input reconstruction (MBFDI based) and associated monitoring statistics (MSPC based).

FDI Scheme Details

The scheme is based on stable discrete time-invariant causal models in state-space form:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{E}\mathbf{w}(k) \tag{1}$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \mathbf{F}\mathbf{v}(k), \tag{2}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state vector, $\mathbf{y}(k) \in \mathbb{R}^l$ represents the output vector, $\mathbf{u}(k) \in \mathbb{R}^m$ denotes the input vector, $\mathbf{v}(k) \in \mathbb{R}^r$ describes output uncertainties (e.g., measurement noise, or sensor faults), and $\mathbf{w}(k) \in \mathbb{R}^s$ represents uncertainty of states (e.g., due to common cause process noise, or process and actuator faults) for the kth sample.

Practical experience has shown that industrial processes can be approximated with sufficient accuracy for specific operation regions by Eqs. 1 and 2.[15] Using recorded input/output data, subspace model identification (SMI) methods estimate simultaneously the appropriate system order, calculate system matrices **A**, **B**, **C**, and **D**, and also extract statistical parameters describing the disturbance distribution during normal operation.[16] The accuracy of the identified model is a crucial factor for the performance of any FDI-scheme, particularly for preventing high false alarm rates. More precisely, the identification of sufficiently accurate models requires recorded datasets that stem, in the ideal case, from designed experiments and should provide an adequate level of excitation in the considered input channels.

Model-based FDI approaches use Eq. 1 together with state observers to create an estimate $\widehat{\mathbf{x}}(k)$ of the state variables and evaluate the prediction error of the measured outputs $\mathbf{e}(k) = \mathbf{y}(k) - \widehat{\mathbf{y}}(k)$.[5] Any significant increase in this error is indicative of anomalous behavior. To increase the robustness of the state estimation, the literature advocates the use of unknown input observers (UIOs) in case of unmeasured disturbances with known distributions described in **E**.[17] For FDI applications, input reconstruction or the estimation of system inputs have also been proposed for the detection of sensor and actuator faults.[18–21]

For a reconstruction of fault signatures, information about the fault characteristic has to be supplied through the matrix **E** for process and actuator faults and matrix **F** for sensor faults. Moreover, in the presence of faulty sensor measurements, the system has to be augmented with sensor fault characteristics to achieve a robust state estimation. This gives rise to define the input vector $\mathbf{v}(k)$ that describes a sensor fault. More precisely, the sensor fault is described dynamically by a matrix $\mathbf{A}_v$. Equation 3 shows that a sensor fault is based on a state-space representation and includes an unknown input vector $\boldsymbol{\xi}(k) \in \mathbb{R}^r$ with $r \leq l$. The dynamics governed by $\mathbf{A}_v$ determine how fast the additional states $\mathbf{v}(k)$ can follow a sensor fault $\boldsymbol{\xi}(k)$ (e.g., step-wise, drifting, or random-walk charakteristic). Following this argument, the eigenvalues of $\mathbf{A}_v$ should allow a reasonably fast system response, while, at the same time, being significantly slower than the sensor noise.

$$\mathbf{v}(k+1) = \mathbf{A}_v\mathbf{v}(k) + \boldsymbol{\xi}(k). \tag{3}$$

During normal operation, the fault vector $\xi(k)$ can be considered to follow a zero mean Gaussian distribution. Augmenting the state vector by the fault vector $\boldsymbol{\xi}$, the original system described in Eqs. 1 and 2 becomes:

$$\underbrace{\begin{pmatrix}\mathbf{x}(k+1)\\ \mathbf{v}(k+1)\end{pmatrix}}_{\mathbf{x}_a(k+1)} = \underbrace{\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_v\end{bmatrix}}_{\mathbf{A}_a}\underbrace{\begin{pmatrix}\mathbf{x}(k)\\ \mathbf{v}(k)\end{pmatrix}}_{\mathbf{x}_a(k)} + \underbrace{\begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}}_{\mathbf{B}_a}\mathbf{u}(k) + \underbrace{\begin{bmatrix}\mathbf{E} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}}_{\mathbf{E}_a}\underbrace{\begin{pmatrix}\mathbf{w}(k)\\ \boldsymbol{\xi}(k)\end{pmatrix}}_{\mathbf{w}_a(k)}. \tag{4}$$

To describe the impact of additive sensor faults $\xi(k)$ upon the outputs, **F** is set to the identity matrix since each sensor fault affects only a single output variable. Consequently, a prediction error that results from a sensor fault, will be compensated through the observer by nonzero elements for $\widehat{\mathbf{v}}(k)$ in the estimate $\widehat{\mathbf{x}}_a$ of the augmented state vector in Eq. 4.

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \mathbf{F}\mathbf{v}(k) = \begin{bmatrix}\mathbf{C} & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{x}(k)\\ \mathbf{v}(k)\end{pmatrix} + \mathbf{D}\mathbf{u}(k) \tag{5}$$

$$\mathbf{y}(k) = \mathbf{C}_a\mathbf{x}_a(k) + \mathbf{D}\mathbf{u}(k). \tag{6}$$

Following Simani et al.,[22] a full order UIO has the following form

$$\mathbf{z}(k+1) = \mathbf{G}\mathbf{z}(k) + \mathbf{T}\mathbf{B}\mathbf{u}(k) + \mathbf{K}\mathbf{y}(k) \tag{7}$$

$$\hat{\mathbf{x}}(k) = \mathbf{z}(k) + \mathbf{H}\mathbf{y}(k) \tag{8}$$

with $\mathbf{z}(k) \in \mathbb{R}^n$ is the state vector of the UIO, $\hat{\mathbf{x}}(k)$ is the estimate of the state vector $\mathbf{x}(k)$ and $\mathbf{F}$, $\mathbf{T}$, $\mathbf{H}$ and $\mathbf{K}$ are the design matrices of the observer that achieve the de-coupling property. Necessary and sufficient conditions for the existence of the observer are $rank(\mathbf{CE}) = rank(\mathbf{E})$ and detectability of ($\mathbf{A}_1$, $\mathbf{C}$), with $\mathbf{A}_1 = \mathbf{A} - \mathbf{E}(\mathbf{CE})^{\dagger}\mathbf{CA}$.[22] Following the rank condition, the number of unknown inputs must not be greater than the number of independent measurements. By choosing $\mathbf{H} = \mathbf{E}(\mathbf{CE})^{\dagger}$,[23] the matrix $\mathbf{G}$ derives from $\mathbf{G} = \mathbf{I} - \mathbf{HC}$. The observer gain $\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2$ is calculated by placing poles to achieve a stable matrix $\mathbf{G} = \mathbf{A} - \mathbf{HCA} - \mathbf{K}_1\mathbf{C}$ and subsequently solving $\mathbf{K}_2 = \mathbf{FH}$.

If the existence conditions of the chosen design approach are met, the UIO generates state estimates that are robust against sensor faults with the dynamic properties described in $\mathbf{A}_v$ and the distribution $\mathbf{F}$ among the outputs, as well as process or actuator faults with the corresponding characteristic $\mathbf{E}$.[17,22] While the reconstructed sensor fault signature $\widehat{\mathbf{v}}(k)$ is immediately available as part of the estimate of the augmented state vector $\widehat{\mathbf{x}}_a$, the reconstruction of the propagation of process or actuator faults $\widehat{\mathbf{w}}(k)$ is discussed next.

Apart from process noise, the vector $\mathbf{w}(k)$ in Eq. 1 can also represent modeling errors, input uncertainty or process faults.[19,24] This can be addressed by reconstructing $\mathbf{w}(k)$ through inversion of Eq. 1 using robust state estimates:[11,19]

$$\widehat{\mathbf{w}}(k) = \mathbf{E}^{\dagger}(\widehat{\mathbf{x}}(k+1) - \mathbf{A}\widehat{\mathbf{x}}(k) - \mathbf{B}\mathbf{u}(k)) = \mathbf{E}^{\dagger}(\widehat{\mathbf{x}}(k+1) - \widehat{\mathbf{x}}(k+1|k)) \tag{9}$$

where $\mathbf{E}^{\dagger} = [\mathbf{E}^T\mathbf{E}]^{-1}\mathbf{E}^T$ represents the generalized inverse of $\mathbf{E}$, which is designed for specific fault scenarios, and $\widehat{\mathbf{x}}(k+1|k)$ denotes the deterministic model prediction of the state vector for the sample $k+1$ using the state estimate $\widehat{\mathbf{x}}(k)$ and the control input $\mathbf{u}(k)$ from the previous step. In case of unmodeled behavior, this will lead to a state prediction error $\mathbf{e}(k+1|k) = \widehat{\mathbf{x}}(k+1) - \widehat{\mathbf{x}}(k+1|k)$ that is being weighted using matrix $\mathbf{E}$.

According to (1) and (2), both $\mathbf{v}(k)$ and $\mathbf{w}(k)$ are assumed to be described by a zero mean multivariate Gaussian distribution under normal operating conditions. Any significant change in the first or second order statistics of either input vectors is therefore indicative of the presence of a sensor or process fault. Consequently, we can calculate a Hotelling's T^2 statistic, as defined in (10), which is sensitive to "process and actuator faults", as they affect the state-to-input relationship. A second Hotelling's T^2 statistic can be constructed from estimates of $\mathbf{v}$ to detect "sensor faults". The covariance matrices for $\mathbf{v}$ and $\mathbf{w}$, $\widehat{\mathbf{S}}_{vv}$ and $\widehat{\mathbf{S}}_{ww}$, can be estimated from normal operation condition (NOC) data. The corresponding control limits can be determined from a F-distribution based on the number of degrees of freedom, the number of reference samples K and a given significance level α (e.g., 1 or 5%). These statistics yield $T^2(k) \leq {T_\alpha}^2$ during NOC and $T^2(k) > {T_\alpha}^2$ for the existence of fault conditions.

$$T^2(k) = \widehat{\mathbf{w}}^T(k)\widehat{\mathbf{S}}_{ww}^{-1}\widehat{\mathbf{w}}(k) \tag{10}$$

The following sections present case studies to demonstrate the capability of the proposed scheme to detect various fault conditions and the usefulness of the scheme in analyzing the reconstructed sequences of $\widehat{\mathbf{v}}$ and $\widehat{\mathbf{w}}$ for fault diagnosis.

Illustrative Example

A simulation example is discussed first so as to illustrate the calculation of T^2 statistics using input reconstruction for an existing linear state-space model. The simulation model is the discrete-time version of a fourth-order process[19] and defined through the following matrices:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -32.624 & -0.148 & 0 & 11.654 \\ 0 & 0 & 0 & 1 \\ 0.982 & 0.279 & 0 & -3.123 \end{pmatrix} \quad B = \begin{pmatrix} 0 \\ 12.354 \\ 0 \\ -9.115 \end{pmatrix}.$$
$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad D = 0; \tag{11}$$

From this continuous time version, the corresponding discrete-time model matrices have been obtained using a zero order hold with $dt = 0.01$ sec. From the output variables, only the third state variable cannot be measured. This process has a single input and the corresponding actuator is considered to be faulty here, which can be described by a superposition of the nominal input $\mathbf{u}_0$ according to Eq. 12.

$$\mathbf{u}(k) = \mathbf{u}_0 + \Delta\mathbf{u}(k) \tag{12}$$

A period of 800 samples representing NOC was simulated for a sinusodial input $k \cdot \sin(\omega t)$ with $\omega = 2$ and $k = 0.2$ superimposed by gaussian noise with standard deviation $\sigma = 0.05$ The output measurements were also corrupted by Gaussian distributed noise with standard deviation $\sigma = 0.01$ Here, an UIO has been designed to achieve a stable state estimation with eigenvalues of matrix $\mathbf{G}$ (see Eq. 7) placed at $eig(\mathbf{G}) = [0.9564, 0.9569, 0.9568, 0.9439]^T$, which is then robust against deviations of the true input from the measured nominal input $\mathbf{u}_0$.[22] The results are listed in Appendix A and the estimation results are illustrated in the top left and right plot in Figure 1. This gives rise to set $\mathbf{E} = \mathbf{B}$ and $\mathbf{w}(k) = \Delta\mathbf{u}(k)$ in Eq. 1, since the unknown portion of the true system input $\mathbf{u}(k)$ that superimposes the measured value is unknown but has the same dynamic characteristic $\mathbf{B}$.

The lower left plot in Figure 1 shows a violation of the T^2-control limit ${T_{1,899}}^2 = 6.6466$ for a significance of $\alpha = 99\%$ within a test dataset covering 200 samples, in which an actuator fault has been introduced after 100 intervals of preceding NOC. Using the input residual (9), the true system input $\mathbf{u}(k)$ could be reconstructed leading to the description of the fault signature $\Delta\mathbf{u}(k)$ through the reconstruction $\widehat{\mathbf{w}}(k)$.

Simultaneous Sensor and Process Faults

In this example, we use a more complex simulation example of a jacketed continuous reactor with a coolant recycle and a first order chemical reaction $A \rightarrow B$ to demonstrate the ability to detect the occurrence of a simultaneous sensor and actuator fault condition. The reactor, as depicted in Figure 2, is affected by product accumulation on the reactor wall, which depends on various operational parameters.

A 4th order linear reactor model is derived through linearization of a nonlinear differential equation system at a stable steady state condition. A discrete version of the model ist obtained through transformation using a sample rate of $dt = 1$ sec. The state vector consists of the current jacket temperature $T_{j,1}$, the concentration c_A of substance A, the reactor

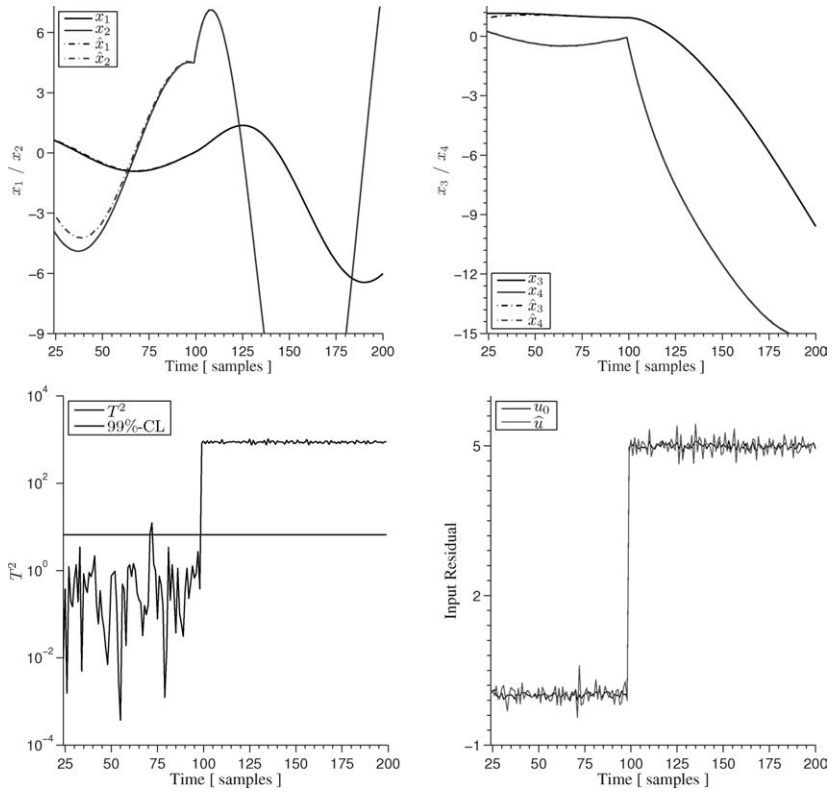

Figure 1. State estimation results for state variables (top left and right), T^2 statistic (lower left) for the detection of the actuator fault and corresponding fault signature (lower right).

temperature T_r, and the temperature $T_{j,2}$ of the coolant when it returns into the jacket.

$$\mathbf{x}^T(k) = [\,x_1(k) \quad x_2(k) \quad x_3(k) \quad x_4(k)\,] = [\,T_{j,1}(k) \quad c_A(k) \quad T_r(k) \quad T_{j,2}(k)\,] \tag{13}$$

The system matrices **A** and **C** are defined as follows:

$$\mathbf{A} = \begin{pmatrix} -\frac{Q_c}{V_j} - \frac{UA}{V_j\rho_c c_{P,c}} & 0 & \frac{UA}{V_j\rho_c c_{P,c}} & \frac{Q_c}{V_j} \\ 0 & -\frac{Q_f}{V_r} - k_r^0 & -c_A^0 k_r^0 \frac{dE}{RT_r^{0^2}} & 0 \\ \frac{UA}{V_r\rho_A c_{P,A}} & -\frac{\Delta H}{\rho_A c_{P,A}} k_r^0 & -\frac{Q_f}{V_r} + \frac{\Delta H}{\rho_A c_{P,A}} k_r^0 \frac{dE}{RT_r^{0^2}} - \frac{UA}{V_j\rho_A c_{P,A}} & 0 \\ \frac{1}{\tau_{\mathrm{he}}}(1-\lambda_{\mathrm{he}}) & 0 & 0 & -\frac{1}{\tau_{he}} \end{pmatrix} \tag{14}$$

$$\mathbf{B}^T = [\,0 \quad 0 \quad 0 \quad b_4\,] \tag{15}$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16}$$

Therein, the subscript c denotes the coolant liquid, A the concentration of component A, j the reactor jacket, f the feed flow and r the reactor. The superscript 0 denotes state variables at the operation point, which has been chosen for the linearization of the system. The reaction rate is defined as $k_r^0 = k_0 \exp(-dE/RT_r^0)$. The coefficient b_4 in Eq. 15 describes the impact of the external cooling liquid valve, which is the control input, on the jacket return temperature $T_{j,2}$. Furthermore, τ_{he} denotes the heat exchanger time constant and λ_{he} is the heat transfer efficiency. Equation 16 highlights that the only unmeasured state variable is the concentration. It should be noted that the description of complex process faults through the matrix **E** is not a trivial task, since process faults may exhibit time-varying or nonlinear behavior. In this case study, the affected parameter is the heat transfer coefficient U in Eq. 14 because of the undesired product accumulation. Equation 17 illustrates this in terms of the heat transfer coefficients $\alpha_{1/2}$, which describe the impact of the different liquids on both sides of

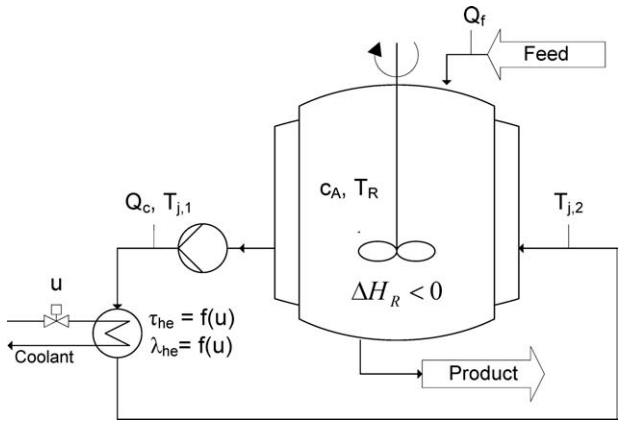

Figure 2. Schematic diagram of the continuous reactor.

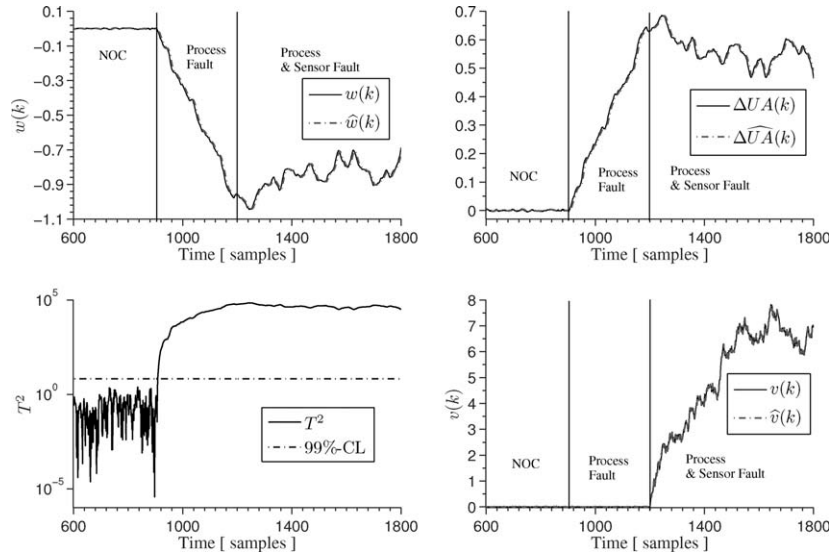

Figure 3. Inknown input $\hat{\mathbf{w}}(k)$ reconstruction (top left), corresponding reconstruction of the changing heat exchange coefficient (top right), T^2-statistic for process fault detection (bottom left) and estimated sensor fault signature $\hat{\mathbf{v}}(k)$ (bottom right).

the reactor wall (1 - jacket side, 2 - inside reactor), on the heat transfer, and the contribution of the reactor wall material is included through wall thickness δ and heat conductivity λ. When the product accumulates on the reactor wall, it forms a layer with increasing thickness δ_1, and thus, the heat conductivity λ_1, and the parameter U degrades according to Eq. 17. Consequently, the elements a_{11}, a_{13}, a_{31}, and a_{33} of the system dynamic matrix in Eq. 14 vary, when the layer thickness on the inner side of the reactor wall changes.

$$\frac{1}{U} = \frac{1}{\alpha_1} + \frac{\delta}{\lambda} + \frac{\delta_1}{\lambda_1} + \frac{1}{\alpha_2} \tag{17}$$

Next, the representation of this parametric fault is derived, which is suitable for the integration into the system description of Eq. 4. Furthermore, in order to capture this changing parameter, $U = U_0 + \Delta U$ is defined, which describes that the nominal value is being superimposed with the unknown portion ΔU that relates to the variation of the accumulated layer described by (δ_1, λ_1). Using this definition, $\mathbf{A}$ can be rewritten as $\mathbf{A} = \mathbf{A}_0 + \Delta\mathbf{A}$ so as to describe the resulting impact of this changing parameter upon the original system according to Eq. 19. Inserting this form into Eq. 14 gives rise to:

$$\mathbf{x}(k+1) = (\mathbf{A}_0 + \Delta\mathbf{A})\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \tag{18}$$

$$\mathbf{x}(k+1) = \mathbf{A}_0\mathbf{x}(k) + \Delta\mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \tag{19}$$

with

$$\Delta\mathbf{A} = \begin{pmatrix} -\frac{\Delta UA}{V_j\rho_c c_{P,c}} & 0 & \frac{\Delta UA}{V_j\rho_c c_{P,c}} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\Delta UA}{V_r\rho_A c_{P,A}} & 0 & -\frac{\Delta UA}{V_r\rho_A c_{P,A}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{20}$$

By multiplying the state vector with Eq. 20 yields:

$$\Delta\mathbf{A}\mathbf{x}(k) = \begin{pmatrix} -\frac{\Delta UA}{V_j\rho_c c_{P,c}}(x_1(k) - x_3(k)) \\ 0 \\ \frac{\Delta UA}{V_r\rho_A c_{P,A}}(x_1(k) - x_3(k)) \\ 0 \end{pmatrix}, \tag{21}$$

which, in turn, leads to the definition of the fault distribution matrix $\mathbf{E}$ and the corresponding unknown input $w(k)$ according to the following decomposition:

$$\Delta\mathbf{A}\mathbf{x}(k) = \underbrace{\begin{bmatrix} -\frac{1}{V_j\rho_c c_{P,c}} \\ 0 \\ \frac{1}{V_r\rho_A c_{P,A}} \\ 0 \end{bmatrix}}_{\mathbf{E}} \cdot \underbrace{\Delta UA(x_1(k) - x_3(k))}_{w(k)}. \tag{22}$$

Thereby, a UIO can be designed to be robust against parametric uncertainty w.r.t. the heat transfer coefficient of the reactor wall. Additionally, a fault on the third temperature sensor is introduced. This sensor measures the temperature of the coolant liquid before it returns into the jacket and is thus important for the reactor operation because this measurement is crucial for achieving a tight reactor temperature control. With the number of unknown sensor faults $r = 1$, the matrix $\mathbf{A}_v$ in Eq. 3 reduces to a scalar and is chosen to be $a_v = .5^{25}$ (which corresponds to a time constant of approximately one second), while the sensor fault distribution matrix is set to $\mathbf{F}^T = [\,0 \quad 0 \quad 1\,]$.

With the complete definition of the system being affected by process and sensor faults according to Eqs. 1 and 2, the augmented form of Eq. 4 is derived and a stable UIO is designed by placing the eigenvalues of matrix $\mathbf{G}$ (see Eq. 7) at $\text{eig}(\mathbf{G}) = [-0.0987, -0.1005, -0.0963, -0.0996, -0.0993]^T$ to be robust against both unknown inputs gathered in $\mathbf{w}_a(k) = (w(k), \xi(k))^T$. The resulting observer can be found in Appendix B.

For fault detection, a training dataset covering 800 samples of NOC by adding gaussian noise with standard deviation $\sigma = 0.01$ to represent common cause process variations $\mathbf{w}(k)$ has been simulated. From this dataset, the covariance

matrix $\mathbf{S}_{ww}$ is determined using the reconstructed process fault propagation $\hat{w}(k)$.

Another set of 900 samples for non-NOC has been simulated, in which a stochastic trend in the heat exchange coefficient is introduced through $w(k)$ and a simultaneous sensor drift that begins after the process fault evolved for some time. The time trends of the true fault propagation $w(k)$ and the reconstructed values as well as the heat exchange coefficient are illustrated in Figure 3 (top left and top right). It can be seen that the departure from NOC is detected timely through the the T^2-control limit ${T_{1,900}}^2 = 6.6424$ for a significance of $\alpha = 99\%$ (bottom left) and that the reconstruction $\hat{w}(k)$ is robust against the simultaneously occurring sensor fault. Furthermore, the estimated signature $\hat{v}(k)$ for the sensor inaccuracy is shown in Figure 3 (bottom right). It should be noted, that the estimated fault signatures accurately represent the true fault signatures, since no further uncertainty has been introduced into the system.

Process Fault Detection in a Reactive Batch Distillation Unit

The section describes the application of the proposed monitoring scheme to a pilot-scale reactive batch distillation column. This process provides the integrated reaction and separation of methyl acetate. Because of this integration and operation in a semi-batch mode, the dynamic behavior is regarded complex.[1,26] It is consequently desirable to prevent any major disturbance, which results from an abnormal operation. Moreover, a timely detection and diagnosis of fault conditions is essential for an appropriate operator intervention. The case studied in this section describes a fault inside the cooling system, which is a critical component for safe process operation. The reaction carried out by this process is:

$$\mathrm{HAc} + \mathrm{MeOH} \rightleftharpoons \mathrm{MeAc} + \mathrm{H_2O}, \tag{23}$$

where methyl acetate (MeAc) is produced by the reaction of acetic acid (HAc) and methanol (MeOH) supported by a catalyst. Different fault scenarios for this process have been studied in Lieftucht et al.[14] using a regression-based MSPC approach. From the defined scenarios there, one specific process fault is analyzed here so as to present a benchmark for contrasting the performance of the proposed FDI scheme with the MSPC approach in Lieftucht et al.[14]

Plant operation

The plant operates in a semi-batch mode and consist of a column with two catalytic packings in the lower part and an integrated separation packing in the upper part. A schematic flowsheet of this process is shown in Figure 4. The required amount of Methanol is provided through an initial holdup in the reboiler, whereas the acetic acid is fed to the column between reaction and separation section. It follows from Eq. 23 that water accumulates in the reboiler as a heavy fraction from the reaction, which leads to a permanently drifting boiling point of the mixture in the reboiler. Consequently, the process does not operate in a steady state mode. However, the composition profiles along the column height show only moderate variations despite this nonstationarity and can, thus, be omitted from the analyzed operation conditions.[27]

This process operates in an open-loop fashion with respect to the product composition. For the conducted experiments, the reboiler duty, the reflux preheating and the condenser are under feedback control using regulatory PI controller and switching control in order to maintain a constant heat flow, a complete condensation or a constant reflux temperature level. Multiple steady state operating points may arise at this plant, from which only a few are economically feasible. Therefore, monitoring the dynamic state of the system is pivotal and of significant interest for detecting a deviation from NOC.[28]

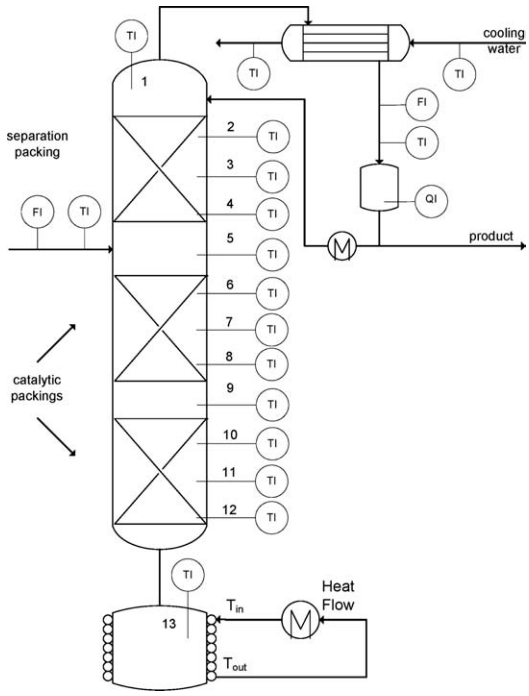

Figure 4. Schematic diagram of the reactive distillation column with instrumentation.

Process fault description

The condenser at the top of the column was affected several times by an actuator fault. Process data of one of such incident is analyzed here. During the column operation, the condenser is typically used to control the pressure inside the column and represents therefore a safety relevant part. The objective is to use an appropriate amount of coolant such that total condensation of the passing steam can be achieved. This coolant flow, as shown by this example, may fail abruptly. Since this flow is not measured, this behavior may remain unnoticed until the fault propagates through the column. Consequently, this may lead to a deviation from the desired operation point.

The effects of this fault were locally restricted to the column head, as shown by the recorded variables in Figure 5. The coolant inlet temperature remains constant for the whole period of the fault, but the coolant and condensate outlet temperatures increase both about $20°C$. Since the reflux temperature is controlled, it remains on the specified level despite the increasing temperature in the reflux splitter. Because less steam would condensate with higher temperatures in the condensator, heavier fractions accumulate at the top of the column, which causes a gradual increase in temperatures in the splitting section. Moreover, the methanol concentration measurement in the condensate increased. This, however, is highly undesired, since it negatively affects the process yield. Furthermore, the situation could become critical once the temperature of the remaining coolant holdup in the condenser becomes not sufficient for the complete

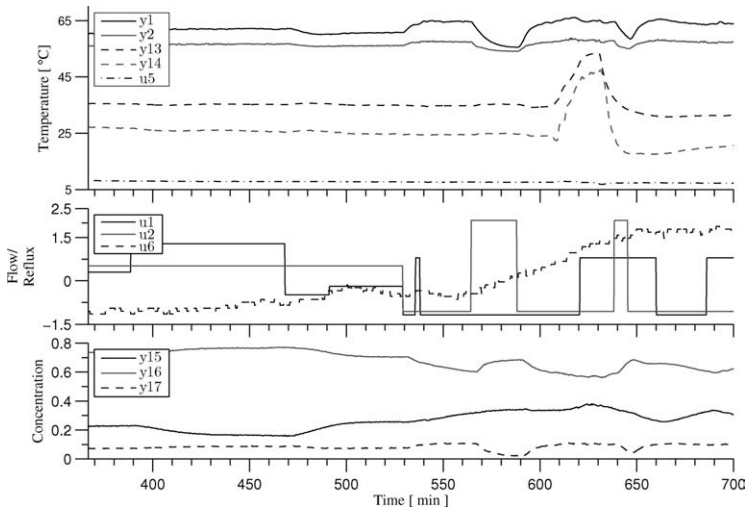

Figure 5. Disturbance propagation affecting the column top.

condensation of teh rising steam. Consequently, the pressure inside the column would start to rise rapidly and a plant shutdown may be needed.

Model identification and fault detection

A dataset from normal operation, covering 3.9 h (1400 samples, with $dt = 10$ sec) of recorded process variables, has been used to identify a state space model. This data set described the process in a dynamically excited mode and was generated for an advanced controller design.[27] The identified SMI model was therefore expected to provide a good approximation of causal relationships between the input and output variables.[10]

Using a slightly reduced set of variables in comparison to the complete set of recorded variables listed in Table 1, the application of SMI suggests the inclusion of $n = 6$ state variables so as to model the causal relationships between the $m = 6$ input and $l = 15$ output variables. Because of the high redundancy of temperature measurements in the reaction section, two measurements (y_7 and y_10) were removed. The corresponding model matrices are listed in Appendix C, together with information about the accuracy of the model for the identification dataset.

The chosen assignment of input and output variables differs partially from the mapping shown in Lieftucht et al.,[14] which is explained in detail. The temperature of the boiling holdup in the reboiler has been included as an input variable since it represents the changing composition due to the accumulation of water. The flow of condensate is also used as an input variable, given that it is proportional to the steam flow through the column.[26] In addition to feed flow and reflux ratio, the feed temperature, and, more important, the coolant inlet temperature is also included as an input variable. The coolant flow itself, which is of importance for the analyzed fault, could not be included as it was not measured.

The FDI scheme has been designed by applying the following procedure:

(1) Generation of a linear state space model using SMI.

(2) Design of a stable Luenberger-type state observer.

(3) Reconstruction of input variables for the NOC-dataset and determination of input residuals.

(4) Covariance estimation of input residuals.

(5) Choice of a confidence level and control limits for the T^2-statistic.

After the scheme has been set up, process data has been analyzed including the variable recordings during the fault event. The T^2-statistic has been evaluated using the estimated covariance information from the NOC, together with the input residuals, which were calculated using the measured inputs and the corresponding reconstructions. In this example no specific unknown input has been included to illustrate the generic fault detection and diagnosis capability of the FDI-scheme. Therefore, the UIO reduces to a conventional Luenberger observer.[22] As outlined in the left plot in Figure 6, the T^2 statistic violates its control limit about 610 min into the data set, which is almost immediately after the coolant flow broke down. Insignificant violations during NOC arise around 540, 580, and 590 min.

Fault diagnosis

The analysis of the same data in Lieftucht et al.[14] shows that a large number of variables contributes to this event. The 6 main variables identified with a decreasing contribution were: condensate temperature, y_{13}, separation section temperatures y_1, y_2, y_3, to a lesser extent y_{12}, and the feed temperature u_2. In comparison to these reported results, the right plot in Figure 6 highlights that two input residuals shows a significant contribution to this event. It should be noted that the input residuals show the deviation of the reconstructed inputs from the measured values after they have been scaled to unit variance (with respect to NOC-variations).

The two residuals which have the most relative contribution are (i) the coolant inlet temperature and (ii) the condensate flow. Whereas the measured coolant temperature is considered too low, the steam flow is regarded to be too high,

Table 1. Recorded Process Variables of the Reactive Batch Distillation Column

Type	No.	Notation	Unit
Output	$y_1 \cdots y_{12}$	Temperatures 1···12	°C
	y_{13}	Temperature condensate	°C
	y_{14}	Temperature coolant out	°C
	y_{15}	Concentration *MeOH*	mol/mol
	y_{16}	Concentration *MeAc*	mol/mol
	y_{17}	Concentration H_2O	mol/mol
Input	u_1	Feed flow acetic acid *HAc*	mol/s
	u_4	Temperature reboiler	°C
	u_2	Temperature feed	°C
	u_3	Temperature coolant in	°C
	u_5	Reflux ratio	%
	u_6	Condensate flow	mol/s

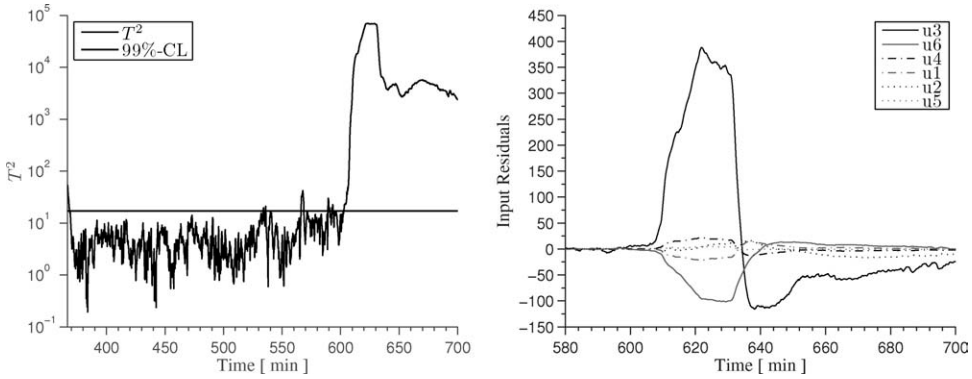

Figure 6. T^2 statistic during faulty operation condition (left) and fault diagnosis of the coolant fault using scaled input residuals (right).

in order to explain the observed behavior of output variables using the dynamic model.

A more comprehensive picture for fault diagnosis can be generated using the spiderweb-illustration for key performance indicators.[3] For the period of fault evolution, Figure 7 shows the corresponding snapshots. Each spiderweb shows the deviation of the individual input residuals from their nominal mean-free value. The last snapshot marks the situation right before the fault has been removed by reactivating the coolant flow.

From this representation, a characteristic picture is generated, which can be interpreted directly. These illustrations imply that the measured output variables could be reconstructed, if

- the coolant inlet temperature (u_5) would be much higher (measured u_5 is too low) and
- the uprising steam flow (represented by the condensate flow u_6) would be lower (measured u_6 is too high).

This analysis already restricts the root cause of this event to the overhead condenser, since both variables pose inputs of this subsystem. Moreover, in contrast to the MSPC-based analysis, these inputs follow strict causal relations in a conventional condenser:

- If the coolant temperature increases, both coolant outlet and condensate outlet temperatures increase.
- If the steam flow decreases, the condensate temperature approaches the coolant inlet temperature closer than before.

The described effects give rise to the possibility of a failing coolant inlet temperature measurement (scenario A) or a decreased coolant flow (scenario B). Subsequently, the root cause is determined, once the operator varies the coolant flow, which will have no effect on both the coolant outlet and condensate temperature. If a failing sensor would have been the root cause, both outlet temperatures would have responded to this action. An experienced operator can therefore use the diagnostic information depicted in the spiderweb-illustrations in Figure 7 to identify: (i) that the condenser subsystem is responsible for the violation of the T^2 statistic and (ii) confirm that the failing coolant flow is the root cause of this event by appropriate intervention. The results show, therefore, that the root cause of this event can be traced down before a critical condition arises (due to insufficient cooling). The corresponding period of fault recovery is illustrated through Fig. 8. Therein, the shape of the polytopes clearly show that the process comes back to normal operation condition.

Conclusions

This article has proposed an input reconstruction scheme on the basis of MBFDI and MSPC techniques for detecting and diagnosing sensor, actuator, and process faults. The utility of the proposed scheme has been demonstrated using two examples so as to simulate a sensor bias and multiple fault conditions. In addition to the simulation examples, relying

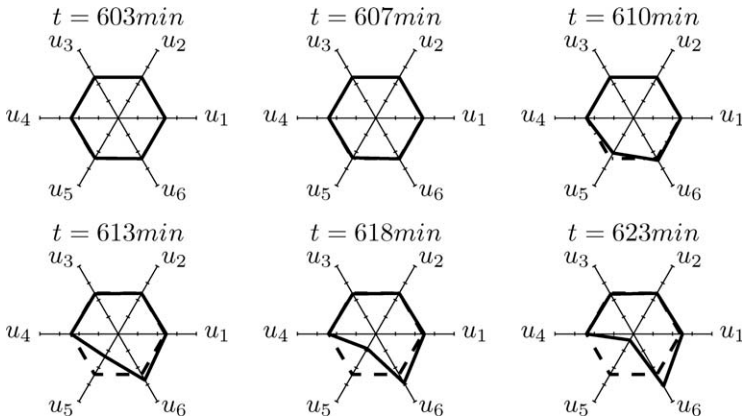

Figure 7. Spiderweb illustration of input residuals for process-monitoring during fault propagation.

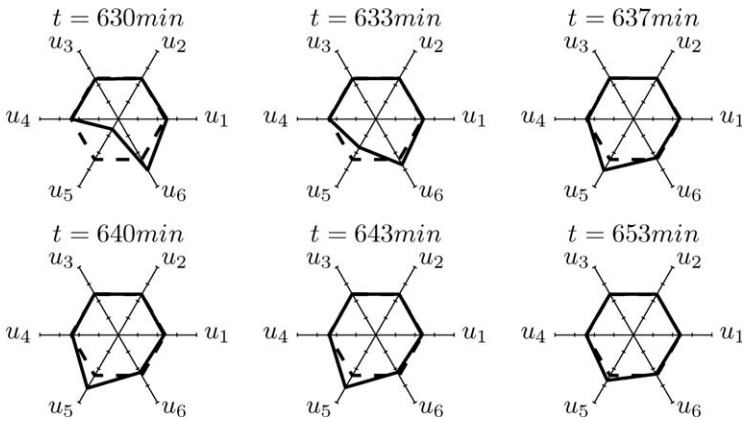

Figure 8. Spiderweb illustration of input residuals for process-monitoring during fault recovery.

on known system matrices, the proposed scheme has also been applied to recorded data from an reactive batch distillation column. For this experimental application study, a process model has been identified using subspace model identification.

The crucial factor affecting the performance of the proposed approach is the sufficient accuracy of the identified model, which depends on the availability of recorded reference data and a priori knowledge of the process operation. While the fault distribution could be derived for the simulation examples, the fault distribution is difficult to obtain for industrial applications. This issue needs to be investigated further. The application studies in this article, however, have shown that the proposed approach offers a generic reconstruction of measured system inputs, which, in turn, allowed the detection of fault condition and a subsequent isolation of potential root causes.

Each application study has demonstrated that the combination of techniques from both, the MBFDI and the MSPC domain, has the potential to improve the performance of the individual approaches and helps to overcome their respective limitations. The potential for ambiguous or misleading diagnosis results in MSPC has been circumvented through the generation of input residuals. Morover, the use of multivariate statistics simplifies the construction of monitoring statistics. Finally, the generated polytope illustration forms a visual aid to support the operator during the presence of abnormal situations.

Acknowledgments

The authors gratefully acknowledge the financial support of this project from the German research foundation (DFG) and the Petroleum Institute, Abu Dhabi, United Arab Emirates. Also thank Prof. Sebastian Engell of the Technical University Dortmund, Germany, for providing access to the recorded data of the reactive distillation process.

Literature Cited

1. Schmidt-Traub H, Górak A. *Integrated Reaction and Separation Operations: Modelling and Experimental Validation*. Berlin: Springer-Verlag, 2006.
2. Wickens CD, Lee J, Liu YD, Gordon-Becker S. *Introduction to Human Factors Engineering*, 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 2003.
3. Schuler H. Automation in chemical industry. *ATP Automatisierungstechnische Praxis*. 2006;54:363–371.
4. Isermann R. *Fault-Diagnosis Systems—An Introduction from Fault Detection to Fault Tolerance*; XVIII of *Robotics*. Berlin: Springer Verlag GmbH, 2006.
5. Ding SX. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Berlin: Springer, 2008.
6. Kourti T. Application of latent variable methods to process control and multivariate statistical process control in industry. *Int J Adapt Control Signal Process*. 2005;19:213–246.
7. Xie L, Kruger U, Lieftucht D, Littler T, Chen Q, Wang SQ. Statistical monitoring of dynamic multivariate processes part 1: Modeling autocorrelation and cross-correlation. *Ind Eng Chem Res*. 2006;45:1659–1676.
8. Yoon S, MacGregor JF. Fault diagnosis with multivariate statistical models part I: using steady state fault signatures. *J Process Control*. 2001;11:387–400.
9. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Comput Chem Eng*. 2003;27:293–311.
10. Gertler J, Cao J. PCA-based fault diagnosis in the presence of control and dynamics. *AIChE J*. 2004;50:388–402.
11. Sotomayor OAZ, Odloak D. Observer-based fault diagnosis in chemical plants. *Chem Eng J*. 2005;112:93–108.
12. Ding SX, Zhang P, Naik A, Ding EL, Huang B. Subspace method aided data-driven design of fault detection and isolation systems. *J Process Control*. 2009;19:1496–1510.
13. Kourti T, MacGregor JF. Multivariate SPC methods for process and product monitoring. *J Quality Technol*. 1996;28:409–428.
14. Lieftucht D, Völker M, Sonntag C, Kruger U, Irwin GW, Engell S. Improved fault diagnosis in multivariate systems using regression-based reconstruction. *Control Eng Pract*. 2009;17:478–493.
15. Favoreel W, Moor BDe, Overschee P Van. Subspace state space system identification for industrial processes. *J Process Control*. 2000;10:149–155.
16. Qin SJ. An overview of subspace identification. *Comput Chem Eng*. 2006;30:1502–1513.
17. Hui S, Zak SH. Observer design for systems with unknown inputs. *Int J Appl Math Comput Sci*. 2005;15:431–446.
18. Xiong Y, Saif M. Unknown disturbance inputs estimation based on a state functional observer design. *Automatica*. 2003;39:1389–1398.
19. Edwards C. A comparison of sliding mode and unknown input observers for fault reconstruction. In: *Proceedings of the 43rd IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, Vol. 5, 2004:5279–5284.
20. Gao Z, Ding SX. Sensor fault reconstruction and sensor compensation for a class of nonlinear state-space systems via a descriptor system approach. *IET Control Theory Appl*. 2007;1:578–585.
21. Fang H, Shi Y, Yi J. A new algorithm for simultaneous input and state estimation. Presented at the American Control Conference, 2008:2421–2426.
22. Simani S, Patton RJ, Fantuzzi C. *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. New York: Springer-Verlag, 2002.
23. Chen J, Patton RJ. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston: Kluwer Academic Publishers, 1999.
24. Yoon S, MacGregor JF. Statistical and causal model-based approaches to fault detection and isolation. *AIChE J*. 2000;46:1813–1824.

25. Saif M, Guan Y. A new approach to robust fault detection and identification. *Aerospace Electron Syst IEEE Trans*. 1993;29:685–695.
26. Gesthuisen R, Dadhe K, Engell S, Völker M. Systematischer Reglerentwurf für eine semikontinuierliche Reaktivrektifikation. *Chemie Ingenieur Technik*. 2002;74:1433–1438.
27. Völker M, Sonntag C, Engell S. *Controlling reactive distillation*. In: *Integrated Reaction and Separation Operations—Modelling and Experimental Validation*. Schmidt-Traub H, Górak A, editors. Berlin: Springer, 2006:299–338.
28. Forner F. Anfahren von Reaktivrektifikationsprozessen in Kolonnen mit unterschiedlichen Einbauten. Ph.D. thesis, Technische Universität, Berlin, 2008.
29. Overschee P. Subspace identification for linear systems. 2002. Available at http://www.mathworks.com/matlabcentral/.
30. Overschee PV, Moor BD.*Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Dordrecht: Kluwer Academic Publishers, 1996.

Appendix A: UIO First Example

The UIO for the first example are given below. The eigenvalues of the matrix **G**, that determine the dynamic of the estimation error are placed at eig(**G**) = [0.9564, 0.9569, 0.9568, 0.9439]^T.

$$H = \begin{bmatrix} 0.0000 & -0.0000 & -0.0067 \\ 0.0088 & -0.0066 & -1.3188 \\ -0.0000 & 0.0000 & 0.0050 \\ -0.0067 & 0.0050 & 0.9999 \end{bmatrix} \quad \text{(A1)}$$

$$G = \begin{bmatrix} 0.9000 & 0.0100 & -0.0000 & -0.0004 \\ -0.2487 & 1.0006 & -0.0000 & -0.0017 \\ -0.0000 & 0.0000 & 0.9564 & 0.0000 \\ -0.0004 & 0.0001 & 0.0000 & 0.9569 \end{bmatrix} \quad \text{(A2)}$$

$$K1 = \begin{bmatrix} 0.0983 & 0.0000 & 0.0075 \\ -0.0732 & 0.0066 & 1.3948 \\ 0.0000 & 0.0436 & 0.0050 \\ 0.0071 & -0.0050 & -0.9568 \end{bmatrix} \quad \text{(A3)}$$

$$K2 = \begin{bmatrix} 0.0001 & -0.0001 & -0.0196 \\ 0.0088 & -0.0066 & -1.3197 \\ -0.0000 & 0.0000 & 0.0048 \\ -0.0064 & 0.0048 & 0.9567 \end{bmatrix} \quad \text{(A4)}$$

$$K = \begin{bmatrix} 0.0985 & -0.0001 & -0.0122 \\ -0.0644 & 0.0000 & 0.0751 \\ 0.0000 & 0.0436 & 0.0098 \\ 0.0007 & -0.0002 & -0.0001 \end{bmatrix} \quad \text{(A5)}$$

$$T = \begin{bmatrix} 1.0000 & 0 & 0.0000 & 0.0067 \\ -0.0088 & 1.0000 & 0.0066 & 1.3188 \\ 0.0000 & 0 & 1.0000 & -0.0050 \\ 0.0067 & 0 & -0.0050 & 0.0001 \end{bmatrix} \quad \text{(A6)}$$

Appendix B: UIO Second Example

The UIO for the simultatneous fault condition and analytically derived fault distribution in the reactor simulation are supplied below. The eigenvalues of the matrix **G**, that determine the dynamic of the estimation error are placed at eig(**G**) = [−0.0987, −0.1005, −0.0963, −0.0996, −0.0993]^T.

$$H = \begin{bmatrix} 0.3079 & -0.4616 & 0.0000 \\ 0.0001 & -0.0001 & 0.0000 \\ -0.4616 & 0.6921 & -0.0000 \\ 0.0192 & -0.0288 & 0.0000 \\ -0.0192 & 0.0288 & 1.0000 \end{bmatrix} \quad \text{(B1)}$$

$$G = 10^3 \cdot \begin{bmatrix} -0.0000 & 0.0002 & -0.0031 & 0.0000 & -0.0000 \\ -0.0007 & 0.0010 & -0.0681 & -0.0000 & 0.0000 \\ 0.0001 & 0.0001 & -0.0022 & 0.0000 & -0.0000 \\ 0.0759 & 0.0000 & 3.3687 & 0.0004 & -0.0005 \\ -0.0759 & -0.0000 & -3.3687 & -0.0005 & 0.0004 \end{bmatrix} \quad \text{(B2)}$$

$$K1 = 10^3 \cdot \begin{bmatrix} 0.0007 & 0.0036 & 0.0000 \\ 0.0007 & 0.0681 & -0.0000 \\ 0.0004 & 0.0025 & 0.0000 \\ -0.0758 & -3.3687 & 0.0005 \\ 0.0758 & 3.3687 & -0.0004 \end{bmatrix} \quad \text{(B3)}$$

$$K2 = 10^3 \cdot \begin{bmatrix} 0.0014 & -0.0022 & -0.0000 \\ 0.0312 & -0.0468 & 0.0000 \\ 0.0010 & -0.0015 & -0.0000 \\ -1.5317 & 2.2966 & -0.0005 \\ 1.5317 & -2.2966 & 0.0004 \end{bmatrix} \quad \text{(B4)}$$

$$K = 10^3 \cdot \begin{bmatrix} 0.0022 & 0.0014 & 0.0000 \\ 0.0319 & 0.0213 & 0.0000 \\ 0.0014 & 0.0010 & 0.0000 \\ -1.6074 & -1.0721 & -0.0000 \\ 1.6074 & 1.0721 & 0.0000 \end{bmatrix} \quad \text{(B5)}$$

$$T = \begin{bmatrix} 0.6921 & 0 & 0.4616 & -0.0000 & -0.0000 \\ -0.0001 & 1 & 0.0001 & -0.0000 & -0.0000 \\ 0.4616 & 0 & 0.3079 & 0.0000 & 0.0000 \\ -0.0192 & 0 & 0.0288 & 1.0000 & -0.0000 \\ 0.0192 & 0 & -0.0288 & -1.0000 & 0.0000 \end{bmatrix} \quad \text{(B6)}$$

Table C1. Simulation and Prediction Errors in Percentage for the Identification Data Set

Variable	Simulation error %	Prediction error %
y_1	11.8835	3.6281
y_2	13.8250	10.3154
y_3	34.0772	6.9864
y_4	7.4874	4.5371
y_5	6.7583	4.2542
y_6	6.8188	4.3213
y_7	10.3832	3.5716
y_8	7.1165	3.1296
y_9	15.6076	2.9699
y_{10}	6.3454	2.7418
y_{11}	13.5711	5.2352
y_{12}	8.8491	3.7594
y_{13}	8.9292	4.8151
y_{14}	11.0556	8.1772
y_{15}	11.7371	8.8121

Appendix C: SMI Results Third Example

These are the system matrices that stem from subspace model identification on a data set covering approximately 3.9h of normal process operation, using the toolbox from[29] with CVA-weighting.

$$A = \begin{bmatrix} 0.9880 & 0.0006 & -0.0045 & -0.0048 & -0.0060 & 0.0112 \\ 0.0015 & 0.9926 & -0.0298 & 0.0109 & -0.0221 & 0.0097 \\ 0.0029 & 0.0149 & 0.9998 & -0.0300 & 0.0142 & -0.0427 \\ 0.0101 & 0.0107 & 0.0224 & 0.9570 & 0.0211 & -0.0255 \\ 0.0049 & 0.0028 & -0.0051 & 0.0166 & 0.9999 & -0.0034 \\ 0.0076 & 0.0087 & 0.0468 & -0.0289 & 0.0487 & 0.9382 \end{bmatrix} \tag{C1}$$

$$B = \begin{bmatrix} 0.0835 & -0.0586 & -0.0763 & -0.0204 & 0.0112 & 0.0267 \\ -0.0159 & -0.0782 & 0.1645 & -0.0121 & -0.0306 & -0.0808 \\ 0.0507 & 0.2527 & 0.3427 & -0.0032 & -0.0231 & -0.0394 \\ 0.1560 & 0.2380 & 0.1476 & -0.0358 & 0.0114 & 0.0378 \\ -0.1247 & -0.0703 & -0.0208 & -0.0091 & 0.0123 & 0.0349 \\ -0.1168 & 0.3603 & 0.1189 & 0.0020 & 0.0106 & 0.0567 \end{bmatrix} \tag{C2}$$

$$C = \begin{bmatrix} 0.1241 & -0.1057 & 0.2007 & -0.0632 & 0.1870 & -0.1278 \\ 0.0827 & -0.0754 & 0.1885 & -0.0689 & 0.1580 & -0.1420 \\ 0.0978 & -0.2828 & -0.0041 & -0.0023 & 0.0872 & -0.2375 \\ -0.0299 & -0.0116 & 0.0588 & 0.0653 & 0.0681 & -0.0918 \\ -0.0357 & -0.0387 & 0.0469 & 0.0752 & 0.0535 & -0.0949 \\ -0.0349 & -0.0625 & 0.0539 & 0.0720 & 0.0506 & -0.0911 \\ -0.1184 & 0.0151 & -0.0723 & 0.1538 & 0.0314 & -0.0600 \\ -0.1408 & 0.0071 & -0.0554 & 0.1423 & 0.0206 & -0.0810 \\ -0.0660 & 0.1230 & -0.1224 & 0.2030 & 0.0450 & -0.0254 \\ 0.0046 & 0.1382 & -0.1328 & 0.1970 & -0.0053 & -0.0146 \\ -0.5079 & -0.0492 & -0.1698 & 0.0405 & -0.0565 & -0.0042 \\ -0.1949 & 0.0691 & -0.1157 & 0.0208 & 0.0192 & 0.0164 \\ 0.2582 & 0.1590 & -0.0447 & -0.0763 & -0.0319 & 0.1435 \\ -0.3041 & -0.0860 & -0.0634 & 0.0992 & -0.0529 & -0.1132 \\ 0.0603 & -0.2490 & 0.2675 & -0.0677 & 0.1607 & -0.1112 \end{bmatrix} \tag{C3}$$

$$D = \begin{bmatrix} -0.0808 & 0.0558 & 0.2091 & -0.0020 & -0.0184 & 0.0072 \\ -0.0586 & 0.0357 & 0.3557 & -0.0215 & -0.0436 & -0.0880 \\ -0.0267 & 0.0466 & 2.3382 & 0.0287 & -0.0660 & 0.0373 \\ 0.0924 & 0.0239 & -0.0661 & -0.0515 & 0.0249 & -0.0012 \\ 0.0704 & 0.0251 & 0.1791 & -0.0654 & 0.0376 & -0.0134 \\ 0.0596 & 0.0196 & 0.1450 & -0.0228 & 0.0529 & -0.0163 \\ -0.0572 & -0.0143 & -0.4575 & 0.0102 & -0.0506 & 0.0054 \\ -0.0776 & 0.0327 & -0.1650 & 0.0202 & 0.0025 & -0.0448 \\ -0.0847 & -0.0420 & -0.6462 & 0.0039 & -0.0116 & 0.0363 \\ -0.0792 & 0.0014 & 0.8427 & -0.0024 & -0.0594 & -0.0974 \\ 0.1268 & -0.0051 & -1.6705 & -0.1355 & -0.0381 & -0.0173 \\ 0.0753 & 0.0576 & -1.9372 & -0.0643 & 0.1691 & 0.1075 \\ 0.0652 & 0.0139 & -0.0041 & -0.0209 & -0.0321 & -0.0602 \\ -0.0607 & 0.0058 & 0.0762 & 0.0299 & -0.0125 & 0.0692 \\ 0.0067 & -0.0285 & 0.2780 & 0.0389 & 0.0499 & -0.0246 \end{bmatrix} \tag{C4}$$

The accuracy of the model can be described using simulation and prediction errors according to Van Overschee & De Moor.[30] The results for the identification dataset are shown in Table C1.

Manuscript received Mar. 2, 2011, and revision received May 18, 2011.