# Statistical and Causal Model-Based Approaches to Fault Detection and Isolation

**Seongkyu Yoon and John F. MacGregor**
Dept. of Chemical Engineering, McMaster University, Hamilton, Ont., Canada L8S 4L7

*Both fundamental and practical differences between two common approaches to fault detection and isolation are examined. One approach is based on causal state-variable or parity-relation models developed from theory or identified from plant test data. The faults are then detected and isolated using structured or directional residuals from these models. The multivariate statistical process control approaches are based on noncausal models built from historical process data using multivariate latent variable methods such as PCA and PLS. The faults are then detected by referencing future data against these covariance models, and isolation is attempted through examining contributions to the breakdown of the covariance structure. There are major differences between these approaches arising mainly from the different types of models and data utilized to build them. Each of them has clear, but complementary, strengths and weaknesses. These are discussed using simulated data from a CSTR process.*

## Introduction

Various approaches to fault detection and isolation (FDI) can be classified into three categories: (1) methods based on causal models, (2) methods based on qualitative knowledge, and (3) statistical methods based on correlation models (Figure 1). The causal process models are obtained from theory, or identified empirically from designed experiments. Most common FDI frameworks are based on them. However, obtaining a complete and robust causal model is difficult due to process complexity and dimension. Therefore this approach is generally limited to processes with a small number of variables.

On the other hand, statistical correlation models developed using principal component analysis (PCA) and partial least squares (PLS) can easily handle a large number of variables (hundreds), since they are built from routine operating data in historical databases. These data are readily available and require no plant testing or fundamental knowledge. FDI approaches using these models fall under the category of multivariate statistical process control (MSPC). Although they are very powerful for fault detection, their main limitation lies in their ability to isolate or diagnose faults. It inevitably suggests that combining the approaches may resolve many difficulties inherent in each of them. However, before this can be attempted the fundamental differences between these approaches must be clearly understood.

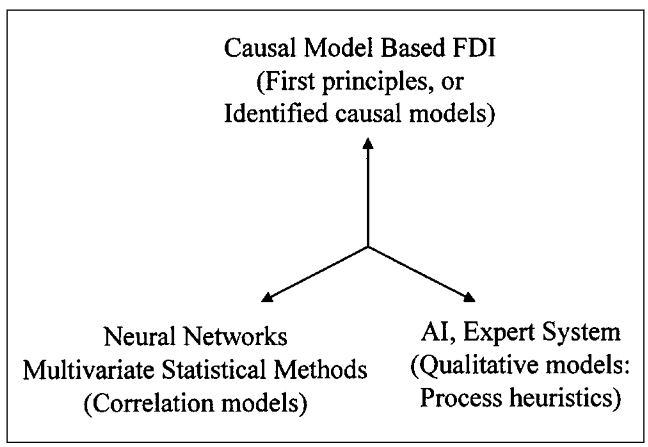


**Figure 1. FDI approaches and resources.**

Correspondence concerning this article should be addressed to J. F. MacGregor.

This article presents the fundamental differences between the causal model-based methods and the multivariate statistical methods to FDI problems, and illustrates them via several simulation studies. The parity-relation approach is selected to represent the causal model-based methods and a PCA-based method to represent the MSPC approaches. By examining the assumptions, model characteristics, and FDI performances to various example problems, the two approaches are overviewed, critical differences are assessed, and their strengths and weaknesses are highlighted. Recognizing the differences between the two approaches should help in the selection of appropriate approaches and provide insight into how the approaches can be combined.

An outline of the article is as follows. In the next section, the parity relation and PCA-based FDI approaches are briefly introduced. The fundamental differences and relationships between the two methods are then addressed. By examining the differences in the models, the data required to build models, the processes to which they are applicable, and the assumptions behind the methods, various fundamental aspects of the FDI system design are compared using two approaches. Practical differences in these approaches are illustrated using simulated data from a continuous stirred-tank reactor process.

## Causal Model-Based Method: Parity-Relation Approach

For details on the causal model-based methods, refer to the following literature (Willsky, 1976; Chow and Willsky, 1984; Isermann, 1984; Basseville, 1988; Frank, 1990; Patton, 1995; Gertler, 1998). In this article, we focus on the parity-relation approach for which design methodologies and techniques have been extensively studied and published by Gertler and coworkers (Gertler, 1998). Parity relations are rearranged and transformed forms of the input−output model equations of processes. They are based on linear process dynamic models in state-space form or in an equivalent input−output form (Gertler and Singer, 1990; Gertler and Monajemy, 1995):

$$y_k^0 = M^0(z)u_k^0, \tag{1}$$

where $y_k^0$ and $u_k^0$ are the true process outputs and inputs acting on the true process, $M^0(z)$. The transfer-function model of a process, $M(z)$ can be obtained either from first principles of a process or from identification using designed experiments satisfying all the identifiability conditions (Ljung, 1999). It can also be derived from the state-space model of the process as $M(z) = [C(zI - A)^{-1}B + D]$, where $A$, $B$, $C$, and $D$ are state-space model matrices. The residuals between the true process outputs, $y_k^0$ and their model estimates, $\hat{y}_k$ can be defined as:

$$r_k = y_k^0 - \hat{y}_k = [I - M(z)]\left[y_k^0 \; u_k^0\right]^T. \tag{2}$$

By introducing faults ($x_k = x_k^0 + \Delta x_k$) on the true process variables, the parity relations are then obtained from the residual equations in terms of the actual measurements ($y_k$ and $u_k$) and the faults as follows:

$$\begin{aligned} O_k &= [I - M(z)][y_k \; u_k]^T \\ &= [I - M(z)][\Delta y_k \; \Delta u_k]^T \\ &= F\Delta x_k, \end{aligned} \tag{3}$$

where the first line in Eq. 3 is called the external parity relation. It is based on the actual measurements and is used for computing all the residuals. The second line in Eq. 3 is called the internal parity relation. It shows how the observed residuals depend upon the unknown faults. The parity relationship becomes nonzero if there is any measurement fault in the variables contained in the model or if there is any process fault that affects the parity relations. Therefore, one can detect faults by checking these primary parity relations. However, fault diagnosis, or isolation tasks are more complicated because nonzero parity relations can result from any faults and so one has to try to isolate the root fault by interrogating all parity relations. To facilitate isolation, parity relations are usually enhanced so that the residual set has one of the following properties (Gertler and Kunwer, 1995):

- *Structured residual sets.* In response to a particular fault, only a fault-specific subset of the residuals becomes nonzero.
- *Fixed-direction residuals.* In response to a particular fault, the residual vector is confined to a fault-specific direction. Isolating a fault amounts to determining to which of the predefined directions the observed residuals lie the closest.

With such residuals, the design of FDI systems using parity relationships involves generating additional parity relations, or rearranging the parity relations such that the FDI system has the properties just listed and fault isolation can be easily done. The directional residuals are usually used when $f = m$ and structural residuals when $f \geq m$, where $f$ and $m$ are the number of specified faults and the number of independent measurements, respectively. To deal with large processes, one can build many FDI systems, one for each individual unit.

When equivalent parity relations are obtained from empirically identified process models using designed plant tests, all identifiability requirements must be satisfied in performing the tests (Ljung, 1999). As with all linear models they are valid only around the operating conditions at which the nonlinear process is linearly approximated. The parity relation approach thus cannot be easily applied to batch or semibatch processes where operating conditions vary continuously. To handle well-defined nonlinear systems, instead of obtaining the structured residuals by algebraic operations on the primary parity equations, the desired nonlinear residual structures can be directly identified (Gertler, 1998).

In this FDI approach, it is usually assumed that there are no process uncertainties such as modeling errors or unmeasured disturbances, and that the explicit model can explain all faults. If any of these assumptions is violated, the performance of the FDI system will be degraded, and the FDI system must be redesigned in a proper manner to account for the assumption violations.

## Statistical Model-Based Method: MSPC Approach

The use of PCA/PLS to build low-dimensional models for the analysis and monitoring of process operations is now well established (Kresta et al., 1991; Wise and Ricker, 1991; Jackson, 1991; MacGregor et al., 1994; Nomikos and MacGregor, 1994; Kourti and MacGregor, 1995), and many industrial applications exist (Kourti and MacGregor, 1996; Kosanovich et al., 1996; Neogi and Schlags, 1998). Provided with historical data ($X$) collected during normal process operation, most of the *common-cause* variations in the process can be expressed in terms of a small number of the principal components (PCs). A cross-validation stopping criterion is usually used to determine the number of principal components ($A$) that capture most of the relevant variations:

$$X = T_A P_A^T + E = \sum_{i=1}^{A} t_i p_i^T + E, \tag{4}$$

where $\hat{X} = T_A P_A^T$ represents the common-cause variations lying in the $A$-dimensional subspace spanned by the first $A$ principal components, and $E$ is the orthogonal space of the residuals. The residual space can also be expressed as

$$E = \tilde{T}\tilde{P}^T = \sum_{i=A+1}^{m} t_i P_i^T,$$

where $\tilde{P}$ is the matrix of loadings associated with the smallest singular values of $X$.

The principal-component loadings ($p_i$) are an orthogonal set of basis vectors for the extracted features called scores ($t_i$). They span the subspace (representational space, $\hat{X} = T_A P_A^T$) of the predictable variations in the training data. The first $A$ principal components define the maximum variance directions in the score space. They give a set of reduced-dimension features that are uncorrelated and account for as much of the total data variance as possible. Since the training set is assumed fault-free, any future score vectors should fall in this reduced space if the process is still fault-free. The remaining principal components, which correspond to the small eigenvalues, span the orthogonal complement ($E$) of the representation subspace. This is referred to as the residual subspace. Note that any true variation is orthogonal to the residual subspace under common-cause variations if there are no errors in model, and thus $\tilde{P}^T x_k^0 = 0$.

If one has both process data $X(n \times m)$, and key quality and productivity data $Y(n \times l)$, then, rather than PCA, one can use PLS to obtain alternative estimates of the latent variable space ($t_1, t_2, \ldots$). PLS will focus more on capturing the high variance directions in the $X$ space that are most correlated with the $Y$ space. The process monitoring approaches would remain unchanged. In the case of dynamic processes each row of $X$ includes lagged values of all measured variables instead of steady-state values. In fact, the monitoring of dynamic batch and semibatch processes using PCA or PLS models developed from the time histories of the variable deviations from their average trajectories has become a well-accepted industrial approach to FDI (Nomikos and MacGregor, 1994, 1995; Kourti et al., 1996; Kosanovich et al., 1996; Neogi and Schlags, 1998).

Having established a PCA model, future behavior can be referenced against this *in-control* model (MacGregor and Kourti, 1995). That is, new observations ($x_{\text{new}}$) are projected onto the plane defined by the loading vectors to obtain their scores ($t_{A,\text{new}} = p_A^T x_{\text{new}}$) and the residuals ($e_{\text{new}} = x_{\text{new}} - \hat{x}_{\text{new}}$, where $\hat{x}_{\text{new}} = P_A t_{A,\text{new}} = P_A P_A^T x_{A,\text{new}}$). Multivariate control charts based on Hotelling's $T^2$ can be plotted based on the first $A$ principal components as follows:

$$T^2 = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2}, \tag{5}$$

where $s_{t_i}^2$ is the estimated variance of $t_i$. An upper control limit on this chart is obtained using the $F$-distribution (Kourti and MacGregor, 1996). This control chart will only detect variation in the plane of the first $A$ principal components that is greater than what can be explained by the common-cause variations. When a new type of special event occurs that was not present in the in-control PCA model, the new observations will move off the plane. This type of event can be detected by computing the squared prediction error (SPE) of the residual for new observations (Kresta et al., 1991). This statistic is also called the $Q$-statistic (Jackson and Mudholkar, 1979), or distance to model (DModX). It is defined as

$$\text{SPE} = \sum_{i=1}^{m} (x_{i,\text{new}} - \hat{x}_{i,\text{new}})^2 = (x_{\text{new}} - \hat{x}_{\text{new}})^T (x_{\text{new}} - \hat{x}_{\text{new}}). \tag{6}$$

When the process is in-control, this SPE statistic represents unstructured residuals that cannot be accounted for by the PCA model. When an unusual event occurs that results in a change in the process mean or covariance structure, it will be detected by a high value of this statistic. The upper control limits for this statistic can be computed based on the reference data (Jackson and Mudholkar, 1979; Nomikos and MacGregor, 1995).

Once a fault is detected by the above multivariate tests, the main approach to fault isolation using PCA/PLS models is the use of contribution plots (MacGregor et al., 1994; Miller et al., 1993; Kourti and MacGregor, 1996). Although the MSPC monitoring charts are very effective for fault detection, fault isolation is more difficult with this approach. When the SPE statistic violates its upper control limit, the contributions of the individual variables ($x_{i,\text{new}} - \hat{x}_{i,\text{new}}$) can be plotted and those variables having large contributions examined to indicate possible causes. Similarly, if the variation in the model space ($T^2$, or $t_1, t_2, \ldots,$) becomes large, then contribution of each variable ($x_i$) to a large value of the $j$th score ($t_j$) is given by

$$\text{Contribution}(x_i) = p_{ij} \Delta x_i, \tag{7}$$

where $p_{ij}$ is the weight of the $i$th variable ($x_i$) in the $j$th latent variable ($p_j$), and $\Delta x_i$ is the change in $x_i$ over the time period in question. More details can be found in the literature (Kourti and MacGregor, 1996). Since these contribution plots come from an underlying correlation model, which does not provide a causal relationship among the variables, they do not provide direct fault isolation. They only show which group of variables is highly correlated with the fault, and it is up to the engineers to use their process insight to provide feasible interpretations. In the case of simple actuator or sensor faults these contribution plots usually can clearly isolate the fault, since it is mainly just one variable whose correlation structure with the other variables has changed (Kourti et al., 1996). In the case of complex or process faults, the isolation is often not clear. However, these plots have been found to be very useful in many applications, because they provide a fault signature that focuses one's attention on a small subset of the large number of process variables, and so greatly restricts the number of fault possibilities that the engineer may have to consider. Given that the only information used in this approach is normal process operating data, this is the best one can expect to do with fault isolation.

To do more, one needs to have additional information. Several approaches using prior fault histories have been proposed (Dunia et al., 1998; Raich and Cinar, 1997; Zhang et al., 1995; Yoon and MacGregor, 2000). These approaches all use some form of fault signatures developed from prior faults for fault isolation. For example, in Yoon and MacGregor, fault signatures consist of the directions of the process movement in both the PCA/PLS model space ($\hat{X} = TP^T$) and the orthogonal residual space ($E$) during the period immediately following its detection of the fault. Once a new fault occurs its signature is compared with those in the fault bank in order to identify the most likely cause.

## Comparisons

Table 1 provides a comparison between the statistical and causal model-based approaches under various headings. The major difference between the approaches is the nature of the models used, and the types of data required to build them. The parity relation/state estimator approaches usually require models that define the causal effects of all inputs on all outputs. For this one needs a first-principles model or an empirical model obtained from identification studies using designed experiments. The data from these identification studies must be persistently excited and satisfy all identifiability conditions (Ljung, 1999; Soderstrom and Stoica, 1989). On

**Table 1. Comparison of FDI Approaches**

| Causal Model-Based Method | Statistical Model-Based Method |
|---|---|
| *Model and Data Requirements*<br>a. First principles model or an empirical model identified from designed experiment<br>b. Model provides causal effects for all inputs on all outputs | a. A correlation model built from undesigned normal operating data<br>b. Model provides no causal effects. It only models the covariance structure among all measured variables under normal operating |
| *Technologies*<br>a. State estimator, Kalman filter, observer, or input/output parity equations<br>b. Parameter estimation | a. Latent variable methods<br>—PCA<br>—PLS |
| *Tools*<br>a. Structured residuals<br>b. Directional residuals | a. Hotelling's $T^2$, score, SPE plots<br>b. Contribution plots<br>c. Directional residuals of prior faults |
| *Size of Problem and Typical Applications*<br>a. Small number of variables (10–20)<br>b. Incorporates only variables for which one has causal models<br>c. Well-defined processes where causal models are available, such as electrical and mechanical processes | a. Large number of variables (hundreds)<br>b. Incorporates all measured variables<br>c. Less-well-defined processes, such as petrochemical, resources (steel, pulp and paper, and semiconductor industries) |
| *Detection*<br>a. Easily done<br>b. Approach: fault breaks causal parity relationship | a. Easily done<br>b. Approach: fault breaks variable covaraince structure existing under normal operations |
| *Isolation*<br>a. Causal model allows for direct isolations of simple and well-modeled faults<br>b. Not possible for unmodeled faults<br>c. Can handle multiple faults with special design | a. Contribution plots provide for easy isolation of simple faults, but ambiguous for complex and multiple faults<br>b. Need additional information on causal effects or on past fault histories to isolate complex faults |
| *Remarks*<br>a. Detection and isolation done together<br>b. Assumes predefined faults and known disturbances | a. Detection and isolation performed sequentially<br>b. Needs representative normal operating data containing all sources of *common-cause* variations<br>c. Often uses only steady-state models, but readily applied to dynamic models |

the other hand, for the statistical approaches a causal model is not even desirable. They require a model for the covariance structure among all measured variables when only common-cause variations are present. Data for building such models are readily available in databases from periods where the plant was operating in a normal and well-behaved manner.

The major strengths and weaknesses of the two different approaches then arise from this difference. The causal model approaches are generally limited to well-defined systems with a small number of variables, while the statistical approaches can easily handle very large and ill-defined processes. On the other hand, the parity equation approach allows for much more direct isolation of known faults through knowledge of the causal structure, while in the statistical approaches the isolation is much more indirect because of the absence of causal information. The statistical approaches must be supplemented with some causal information or prior fault knowledge to provide less ambiguous isolation.

Recently Gertler et al. (1999) proposed an "isolation enhanced PCA" approach. It was shown that, with data from designed experiments, one could use the last principal-component relationships directly as parity relations ($\epsilon_k = \tilde{P}^T x_k = \tilde{P}^T \Delta x_k$, since $\tilde{P}^T x_k^0 = 0$) and apply the same transformation procedures to get isolation properties. Alternatively, one could compute the explicit causal model in Eq. 3 from them ($\tilde{P}^T = WF$, where $W$ is a full-rank square matrix) and apply the parity equation methods directly. However, this approach is clearly still that of the causal model-based approach. Their models are built from designed experiments and provide causal relationships among all the inputs and outputs. The only difference is that PCA is used as the identification method for the causal models rather than the more traditional prediction error methods (Ljung, 1999; Soderstrom and Stoica, 1989). This use of PCA to identify linear causal dependences among the variables from the eigenvectors associated with the smallest eigenvalues is based on the literature on the last principal-component methods for identification (Ku et al., 1995; Negiz and Cinar, 1997) and the literature on total least squares. With no noise, the principal components corresponding to zero eigenvalues would define the space of exact linear relationships among the variables. This approach to identification works well in deterministic situations, but generally is poor compared to prediction error approaches when noise and unmeasured disturbances are present (Negiz and Cinar, 1997).

In the statistical approaches to FDI, latent variable methods such as PCA and PLS are employed in a very different manner. They are used to obtain information on the dominant directions of variations (eigenvectors associated with the largest eigenvalues) that are present in normal operating data. Under these conditions, it is usual to have only a few (2–5) dominant directions that explain most of the process variations arising from these common-causes. The statistical approaches to FDI then rely upon referencing future behavior against this normal behavior defined by the low-dimensional PCA model of the hyperplane. Since the MSPC approach incorporates the effect of all normal disturbances into the PCA model, it will also automatically account for these disturbances in the detection and isolation steps. Clearly, the two FDI approaches are completely different even though PCA is used (albeit for different purposes) in both.

## Simulation Study

In this section, simulation studies on a nonisothermal continuous stirred-tank reactor (CSTR) model (Marlin, 1995, p. 90) are performed to illustrate the two FDI approaches and point out critical differences. The reaction is first order ($A \rightarrow B$), and the reactor system involves heat transfer through cooling coils to remove the heat of reaction. It is assumed that the tank is well mixed and the physical properties are constant. The process has two feed streams: the solvent and the reactant $A$, one product stream, and a cooling water flow to the coils (Figure 2). The reactant ($F_A$) and the cooling water ($F_C$) flows control the reactor outlet concentration ($C_A$) and temperature ($T$), respectively. In this application, however, only the temperature controller is active. Measured process disturbances are the inlet concentrations ($C_{AS}$, $C_{AA}$), the inlet temperature ($T_O$), the solvent flow ($F_S$), and the cooling water temperature ($T_C$). All of these disturbances are simulated to have first-order autoregressive variations under both normal and fault conditions. In addition, it is assumed that unmeasured stochastic disturbances arise from variations in reactive impurities (these will affect the rate of reaction) and from fouling of the cooling coils. These latter two disturbances are also simulated as first-order autoregressive behaviors in the reaction rate ($k$) and heat-transfer ($UA$) constants, respectively. These autocorrelated process disturbances were introduced to add realistic variations into the simulations. (If only white noise is added to the measurements and no autocorrelated disturbances are present, then there would be no need for a controller, and any FDI system would be straightforward.)

In all studies, the true process model was used to obtain the linearized parity relations for the reactor outlet concentration ($C_A$) and the temperature ($T$). The PCA model for the MSPC approach was obtained with a set of observations collected for 200 min from the process under routine operation when no faults were present. (In real processes, much more data would be required to capture all sources of common-cause variations.)

Before investigating fault detection and isolation by the two approaches, we examine the fault characteristics of the process. Based on Eq. 3, the parity relations at any sampling interval are as follows;

$$
\begin{aligned}
O &= F[C_A \; T \; C_{AS} \; C_{AA} \; F_S \; F_A \; T_O \; T_C \; F_C]^T \\
&= F[\Delta C_A \; \Delta T \; \Delta C_{AS} \; \Delta C_{AA} \; \Delta F_S \; \Delta F_A \; \Delta T_O \; \Delta T_C \; \Delta F_C]^T,
\end{aligned}
$$

where $O = [O_{C_A} \quad O_T]^T$, and the parity relation matrix, $F$, obtained from the mechanistic model of the CSTR process is as follows:

$$
F = \begin{bmatrix} 1 & 0 & -0.3254 & -0.0362 & 0.2618 & -6.6072 & 0.0049 & 0.0264 & -0.0034 \\ 0 & 1 & -1.1779 & -0.1309 & 0.6224 & -24.2450 & -0.1678 & -0.8965 & 0.1143 \end{bmatrix}.
$$

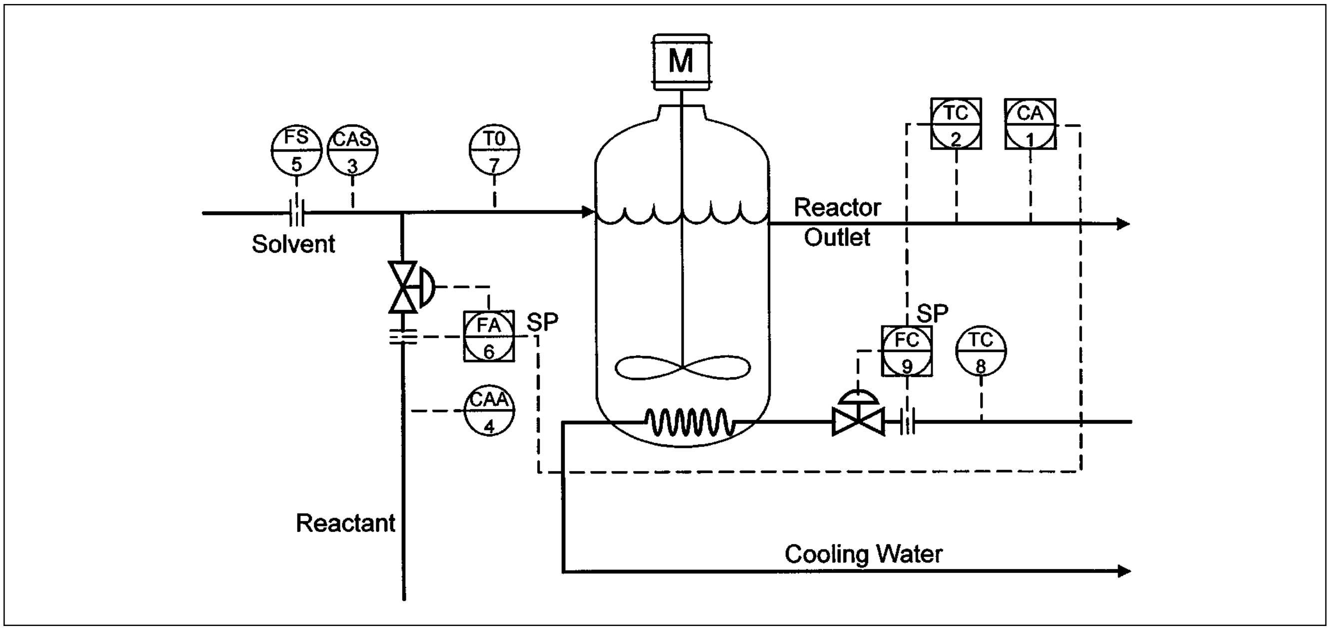


**Figure 2. Process flow diagram of CSTR system.**

As seen in $F$, the parity residual for $C_A$ ($O_{C_A}$) is expressed in terms of all process variables except for $T$ since $F_{12}=0$. Similarly, the primary parity residual for $T$ ($O_T$) is expressed in terms of all process variables except for $C_A$ since $\mathrm{F}_{21}=0$. Thus, in principle, one can isolate only the faults of $C_A$ and $T$ measurements with the primary parity relations. To isolate faults of the other measurements one needs to generate secondary parity relations by taking linear combinations of the primary parity relations.

The $i$th column of parity-relation matrix $F$ defines the fault direction of the $i$th sensor fault $\Delta x_i$ in Eq. 3. That is, the columns of $F$ correspond to the effects of faults in the measurements of $C_A$, $T$, $C_{AS}$, $C_{AA}$, $F_S$, $F_A$, $T_O$, $T_C$, and $F_C$ on the primary residuals $O_{C_A,k}$ and $O_{T,k}$, respectively. Fault isolability is largely dependent on the fault directions. The larger the difference in fault directions among the various faults, the better is the ability to isolate them.

Figure 3 shows the normalized fault directions ($O_i^0=F_i^0$, assuming $\Delta x_j=1$ if $j=i$, otherwise 0) are obtained by scaling all columns of $F$ to unit length. From the columns of the normalized $F^0$ and their plots in Figure 3, one can see that the fault directions of $F_C$, $T$, $T_O$, and $T_C$ are either perfectly collinear or nearly collinear, and hence cannot be isolated from one another. The fault directions of $C_{AS}$, $C_{AA}$, and $F_A$ are also perfectly collinear and cannot be isolated from one another. Furthermore, the angles among the fault directions of the ($F_C$, $T$, $T_O$, $T_C$) group, or of the ($C_{AS}$, $C_{AA}$, $F_A$) group, or that of $F_S$ are small, implying that they could be difficult to isolate from one another in the presence of noises and disturbances. Only a fault in $C_A$ is clearly isolatable from the rest. These exact or near linear dependences between columns of $F$ cause ill-conditioning in the residual equations used in the fault isolation and make the corresponding faults impossible or difficult to distinguish. For this reason, only the residuals in $C_A$ and $T$ obtained from the primary parity equations are used in the following FDI analysis. Deriving residuals for all the other variables (that are collinear or nearly collinear with $T$) through algebraic transformation of these two primary parity equations would yield few useful results in this case. In order to enhance the fault isolability, one needs additional measurements that will break the correlation between the sensor fault directions.

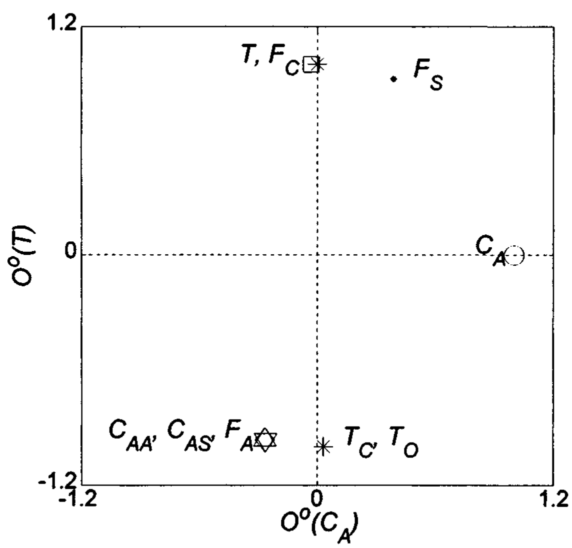


**Figure 3. Normalized fault directions ($O^0$) of CSTR system.**

($C_A$:○; $T$:∗; $C_{AS}$: Δ; $C_{AA}$:◇; $F_S$:●; $F_A$:▽; $T_o$:+, $T_C$:×; $F_C$:□; $O_i^o=F_i^0$)

$$F^0=\begin{bmatrix} 1 & 0 & -0.2662 & -0.2662 & 0.3876 & -0.2629 & 0.0295 & 0.0295 & -0.0295 \\ 0 & 1 & -0.9639 & -0.9639 & 0.9218 & -0.9648 & -0.9996 & -0.9996 & 0.9996 \end{bmatrix}.$$

*FDI of simple faults*

This case study is given to show how the two FDI approaches work for the detection and isolation of simple faults occurring at the normal operating condition. In the first simulation a reactor outlet concentration sensor bias ($\Delta C_A = 0.2$) occurs at 51 min. The fault in $C_A$ is representative of a *simple output sensor fault*. Its effect is not propagated into other process variables (Yoon and MacGregor, 2000). The fault detection and isolation should be relatively easy compared to those of a complex fault.

Figure 4a shows the trends of the primary parity residuals. They clearly indicate that a fault has occurred in the $C_A$ measurement at 51 min. Figure 4b shows two moritoring plots for the MSPC approach. Both the SPE and $T^2$ plots clearly indicate that there is an unusual event around 51 min. The SPE contribution plot in Figure 4c reveals that the measurements contributing most to the abnormal event at 51 min are $C_A$ and the coolant flow rate ($F_C$). In this particular example, faults in these two sources are completely confounded in routine operating data, and so they are difficult to isolate using contribution plots. One can interpret the increase in SPE as a breakdown of the common-cause correlation that is usually present between $C_A$, $F_C$, and the remaining variables. On the other hand, the $T^2$ contribution plot of Figure 4d more clearly isolates the fault as being the $C_A$ sensor because of the large increase in the contribution of $C_A$ starting at 51 min. As a result a simple bias fault has been detected and isolated through a breakdown of the normal correlation among the variables (SPE plot) and a larger variation than normal magnitude of contribution of variation in a variable ($T^2$ plot). In real industrial processes where many additional correlated variables are measured, contribution plots usually can clearly isolate simple sensor faults (such as Kourti et al., 1996).

As a second example, a reactant flow-rate bias ($\Delta F_A = -0.015$) is simulated to occur at 51 min. This is also a simple sensor fault, but one in an input sensor. The fault should be detectable using the two primary parity residual equations ($O_{C_A}$ and $O_T$), since these equations will no longer hold with the fault in $F_A$ being present. The residual plots in Figure 5a show that there is a significant shift in $O_{C_A}$, but not in $O_T$, around 51 min. Thus the fault is detected, but cannot be successfully isolated using only the primary parity residuals. One might conclude that the fault is on the $C_A$ measurement. Generating a secondary parity relation (by a transformation on the primary relations) to specifically isolate the $F_A$ fault also does not help in the isolation in this case. This is because the near collinearity of the fault directions (Figure 3) leads to insufficient sensitivity and a magnification of the noise and disturbance components.

The SPE/T$^2$ plots, based on the PCA model developed from normal plant operating data, are shown in Figure 5b. Both the SPE and $T^2$ plots clearly detect a fault occurring around 51 min. (Note that it is only required that one of the SPE or $T^2$ plots alarms in order to detect a statistically significant event.) The SPE contribution plot in Figure 5c provides no isolation information, since many of the variables show increased contributions after 51 min. However, the $T^2$ contribution plot of Figure 5d clearly isolates the fault as being the $F_A$ sensor because of the large increase in the contribution of $F_A$ starting at 51 min and negligible contributions

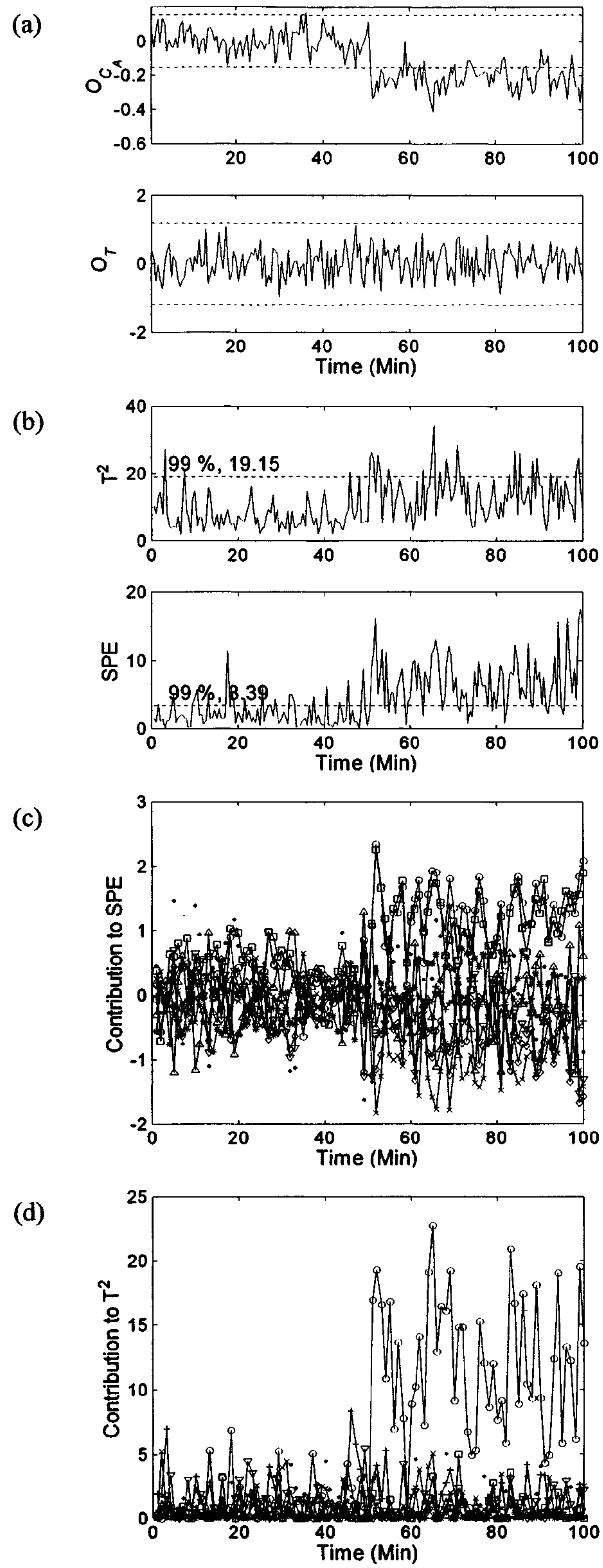


**Figure 4. FDI simple bias fault ($\Delta C_A = 0.2$ at 51 min).**

(a) Parity relation plots; (b) SPE/$T^2$ plots; (c) SPE contribution plot; (d) $T^2$ contribution plot. $C_A$:○; $T$:∗; $C_{AS}$:Δ; $C_{AA}$:◇; $F_S$:●; $F_A$:▽; $T_O$:+, $T_C$:×; $F_C$:□.

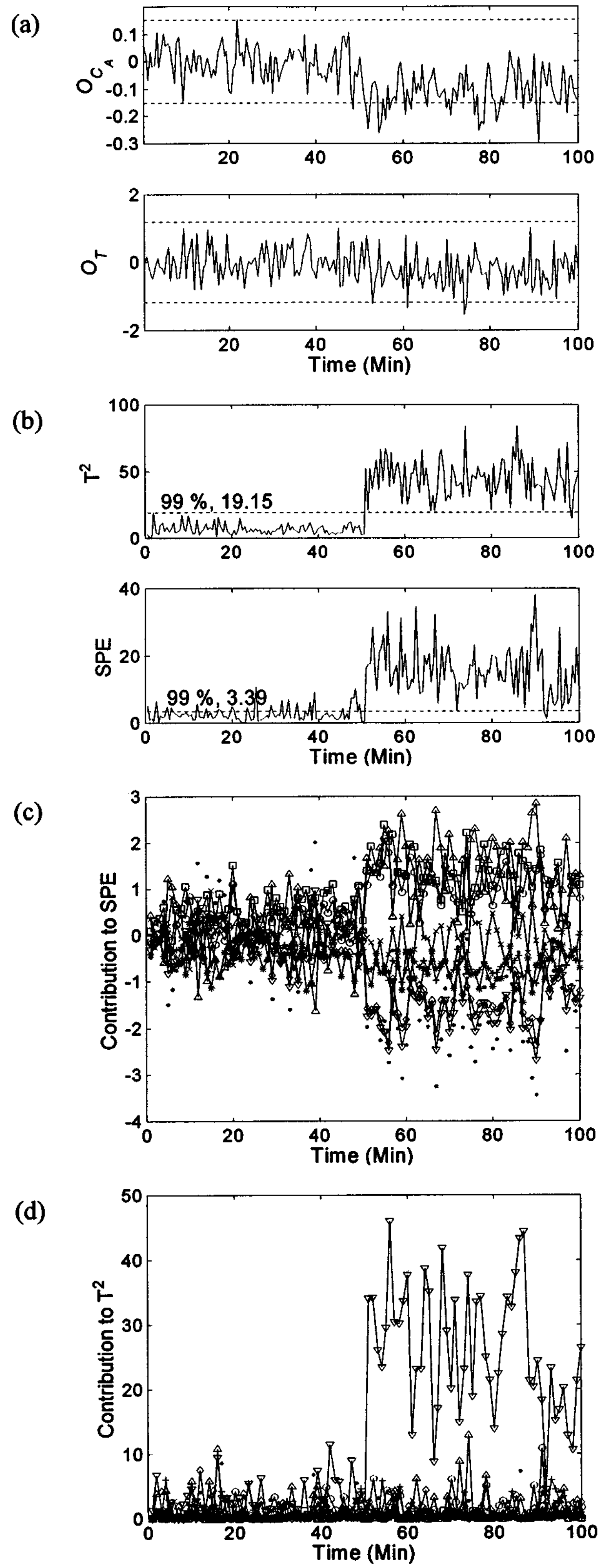


**Figure 5. FDI of simple bias fault ($\Delta F_A = -0.015$ at 51 min).**

(a) Parity relation plots; (b) SPE/$T^2$ plots; (c) SPE contribution plot; (d) $T^2$ contribution plot. $C_A$:○; $T$:∗; $C_{AS}$:Δ; $C_{AA}$◇; $F_S$:●; $F_A$:▽; $T_0$:+, $T_C$:×; $F_C$:□.

from the other variables. The MSPC approach thus clearly detects and isolates this fault. This results from the fact that the fault in $F_A$ significantly breaks the common-cause correlation structure between $F_A$ and the other variables. Another contributing factor is that the effects of the natural disturbances in the system are accounted for more effectively in the MSPC scheme. Their effect on the covariance structure of all the variables is automatically contained in the PCA model and hence in the SPE and $T^2$ tests.

### *FDI of complex sensor faults*

Many faults in a process are complex faults. Isolation of the complex faults usually needs the causal relationship between the root source of the fault and the affected variables. Since MSPC is not based on a causal model, it will not give a clear and unambiguous isolation of the fault. On the other hand, the causal model-based method can promptly isolate faults as long as they are modeled. The object of this example is to show the differences between two FDI methods for complex sensor faults.

A fault is assumed to occur at 51 min in the sensor of the reactor outlet temperature ($T$), which is under closed-loop control. Since the parity relations are based on causal relationships between variables that do not change with the addition of feedback loops, the parity residuals will not be affected by the presence of the feedback controller on $T$. This is illustrated in Figure 6a by showing that the complex sensor fault on $T$ is clearly detected and isolated by the parity residuals. Figure 6b shows the MSPC monitoring plots for the same fault. The abnormality is detected more slowly since it takes time for the feedback controller to propagate the effect of the fault in $T$ into other variables. The corresponding contribution plots, Figure 6c and 6d, indicate that the coolant flow rate ($F_C$) and the reactor outlet concentration ($C_A$) contribute most to the detected fault. This result occurs because the sensor bias in the measured value of the controlled variable ($T$) is eventually eliminated by the PI control action using the manipulated variable, namely the coolant flow rate ($F_C$). The change of the coolant flow rate then lowers the rate of reaction, resulting in an increase in the reactor outlet concentration ($C_A$).

This example illustrates the difficulty in isolating complex faults with the MSPC approach. The contribution plots only show how the covariance structure that existed under normal operation has been broken when a fault occurs. In this case, they reveal that the relationship among nearly all variables has changed somewhat, but for the reasons explained earlier, the relationships of the other variables with $C_A$ and $F_C$ have changed the most. To more clearly isolate the fault using the MSPC approach, causal knowledge or past fault information on the effect of a fault in $T$ could be used (Yoon and MacGregor, 2000).

### *FDI of process faults*

Process faults look like complex sensor faults because they affect many variables and make them difficult to diagnose. The ability to handle process faults is an important property of FDI methods. In this example, it is shown how the two

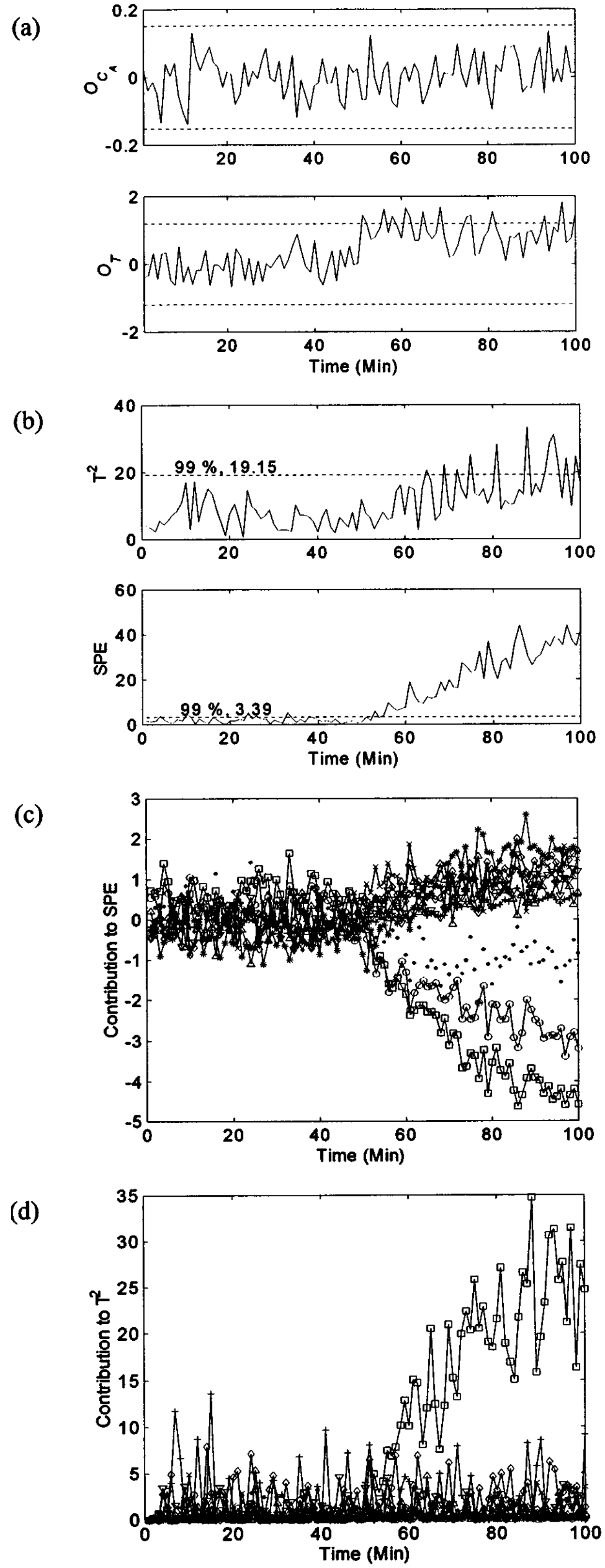


**Figure 6. FDI of complex fault ($\Delta T = -1$°C at 51 min).**

(a) Parity relation plots; (b) SPE/$T^2$ plots; (c) SPE contribution plot; (d) $T^2$ contribution plot. $C_A$:○; $T$:∗; $C_{AS}$: Δ; $C_{AA}$◇; $F_S$:●; $F_A$:▽; $T_O$:+, $T_C$:×; $F_C$:□.

methods work for a process fault and what additional steps are required.

In this study, a sudden degradation of the heat-exchanger performance is simulated as a decrease in the heat-transfer coefficient at 51 min. However, the causal model does not include the equation that relates the heat exchanger fouling to the process measurements, or the other parameters in the model. Figure 7a shows how the causal model-based FDI works for this unmodeled process fault. The causal model-based method does detect that a fault has occurred. The $O_{C_A}$ residual increases, but not enough to violate the limit. However, the temperature residual ($O_T$) drops and violates its lower limit at many points. This detection results from the fact that, with the change in the heat-transfer coefficient, the parity equations no longer hold exactly. However, without a model for the effect of heat-exchanger fouling on the system, isolation is not possible. In order to provide the diagnostic capability on a parametric fault, one would have to generate additional parity relations to incorporate terms for all suspected process faults using special design procedures (Gertler and Kunwer, 1995). Although any number of process faults can be incorporated into the model, the number of independent fault directions (columns in $F$) and independent measurements will limit the number of faults that can be isolated.

Figure 7b shows the SPE/$T^2$ plots for the same process fault. The corresponding contribution plots, Figure 7c and 7d, show that the main contributor to the $T^2$ and the SPE deviations is the coolant flow rate ($F_C$). This results from the fact that a drop in the heat-transfer coefficient causes the temperature controller to increase $F_C$ in order to keep the reactor outlet temperature ($T$) at its setpoint. Most of the remaining variables are unaffected, and hence the contribution plots show that it is mainly $F_C$ whose covariance structure differs from what existed during normal operation. Obviously, the isolation of the process fault is ambiguous since this same pattern could easily have resulted from a simple actuator or sensor fault on $F_C$. However, the MSPC approach clearly detected the unexpected process fault and the contribution plots clearly reduced the number of fault sources that would have to be investigated. One way to enhance the isolability of process faults is to use fault signatures based on past data, as discussed earlier.

### *Effects of data and identification method on model-based FDI performance*

The purpose of this example is to illustrate the effects that the model identification approach used to obtain the parity relations can have on the behavior of the causal model-based FDI scheme. In one case, an equation error identification method (Ljung, 1999) is used to obtain the explicit parity model. This simply corresponds to using a least-squares method on the residuals of the parity equations to estimate the causal model parameters. The residual equations are then obtained from these. In the other case, the last-principal-component approach (Gertler et al., 1999) is used to directly obtain parity relations for residual generation. In all comparisons, the same data are used and generated using random binary sequences (RBS) in all inputs. To simplify the study only two of the variables ($F_C$ and $F_A$) were used as inputs.

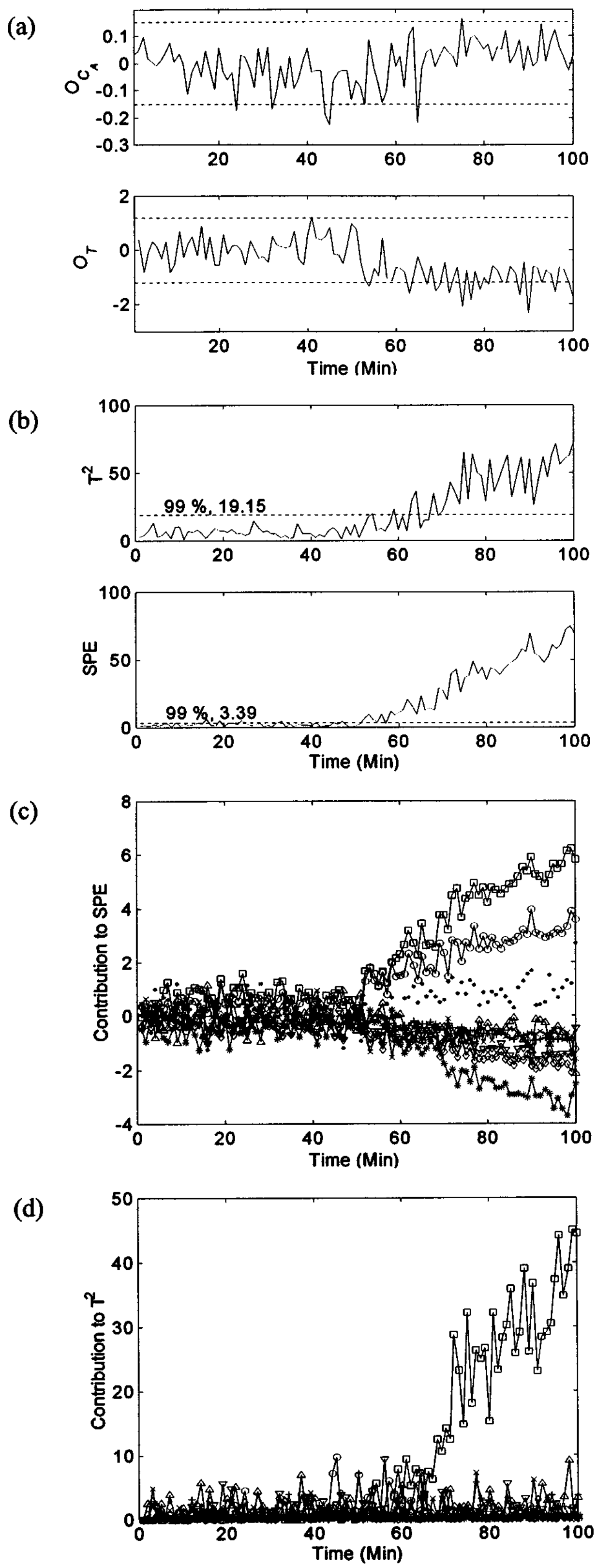


**Figure 7. FDI of process fault (increase in heat exchanger fouling, $\Delta UA = -0.3$ at 51 min).**

(a) Parity relation plots; (b) SPE/$T^2$ plots; (c) SPE contribution plot; (d) $T^2$ contribution plot. $C_A$:○; $T$:∗; $C_{AS}$: Δ; $C_{AA}$ ◇; $F_S$:●; $F_A$:▽; $T_o$:+, $T_C$:×; $F_C$:□.

The steady-state gains between these inputs and outputs are only estimated after data are mean-centered and auto-scaled. Then the real gains are obtained by unscaling the identified parameters and compared against those using an exact linearization of the true fundamental model equations. The performance of the FDI scheme whose parity relations are formulated with the identified gains will be compared with one another and with that of the MSPC approach.

Figure 8 compares the process gains estimated by the two approaches for a wide range of magnitudes used for the RBSs. The RBS magnitudes range between 0.2 and $4.0\sigma$, where $\sigma$ is the standard deviation of the inputs during normal operation period. Each set of gains is estimated with 400 measurements. For large-magnitude RBSs the signal-to-noise ratio is large enough that both methods identify the correct gains (shown by the solid lines), but at lower signal-to-noise ratios the last-principal-component approach provides very poor model compared to the least-squares method. This confirms the results shown by Negiz and Cinar (1997).

To illustrate the impact on the FDI scheme, Figure 9a and 9b contrast the residual plots and thresholds for the parity relations based on the models identified using least squares on the parity equations and the last-principal-component method for an RBS of magnitude $0.4\sigma$. The models identified by the least squares provided almost identical behavior to that of the theoretical model shown in Figure 4a in clearly detecting and isolating the $C_A$ sensor fault. The parity equations based on the last principal components failed to even detect the fault. The poorly identified parity relations resulting from the latter approach also leads to much larger residuals when no fault is present, and hence the need for much wider threshold limits in the FDI scheme. Table 2 compares the FDI thresholds (99%) required for the $C_A$ parity relation, $O_{C_A}$, identified by the different approaches as a function of the RBS magnitude used in generating the data. Clearly, these limits get very large for the models from the last-principal-component method when the signal-to-noise ratio in the data is low, while they remain almost unchanged and close to those for the theoretical model when the models from the least-squares identification are used. This example illustrates that the last-principal-component method of identifying residual relationship must be used with great caution, especially when there is measurement noise and disturbances present in the data used for identification.

The FDI results for the same fault in $C_A$ ($\Delta C_A = 0.2$) using the MSPC approach were previously shown in Figure 4b, where the fault was also clearly detected at 51 min in the SPE plot and isolated by using the SPE/$T^2$ contribution plots.

## Conclusions

The fundamental differences between the causal model-based FDI approaches (as represented by the parity relation approach) and the multivariate statistical process control approaches based on noncausal models developed from historical data using PCA/PLS have been examined. The differences in the nature of the models used (arising from the nature of the data used to build them) are shown to be responsible for the very different approaches to fault detection and isolation employed by the two methods. This model differ-

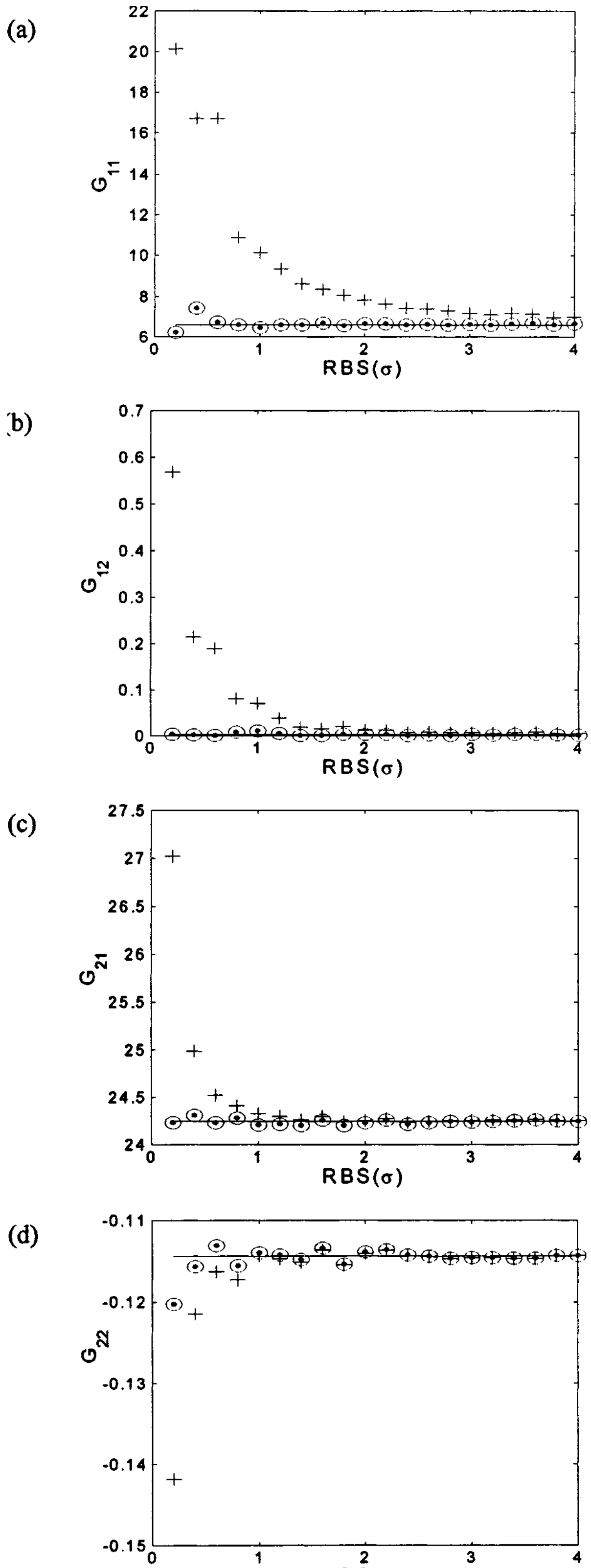


**Figure 8. Process gain estimations by least-squares method and last-principal-component method with full rank RBS as inputs but noisy outputs.**

(a) $G_{11}$; (b) $G_{12}$; (c) $G_{21}$; (d) $G_{22}$. Least-squares method:⊙; last-principal-components method: +; solid lines represent the true gains.

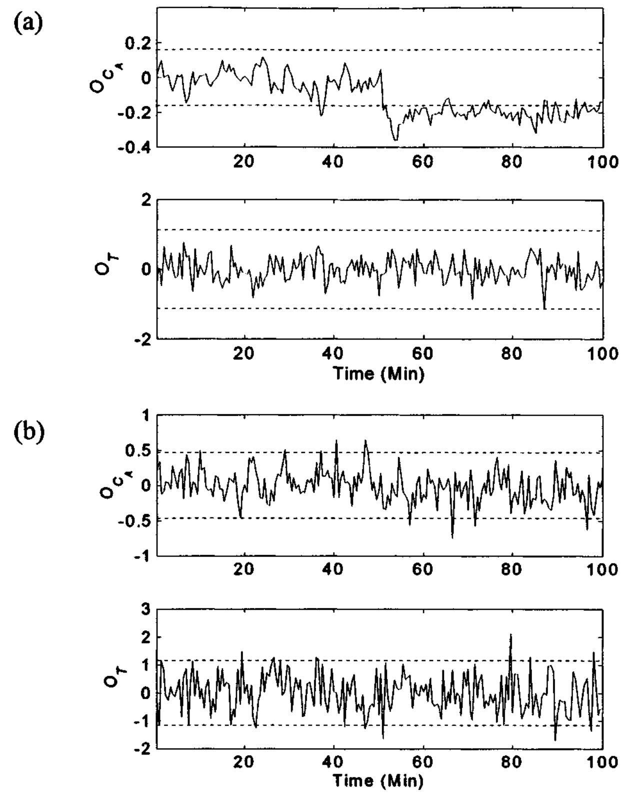


**Figure 9. FDI performance of parity relation approach with mechanistic models and identified models for a bias fault of $C_A$ measurement ($\Delta C_A = 0.2$).**

(a) FDI with the identified model under RBS magnitude of $0.4\sigma$ and least-square regression; (b) FDI with the identified model under RBS magnitude of $0.4\sigma$ and by using last principal components associated with the smallest two eigenvalues ($\sigma$: standard deviation of input signal during normal operation).

ence is also shown to be responsible for their major strengths and weaknesses.

The multivariate statistical approaches are mush easier to develop because of the ready availability of routine operating data, and they can handle a very large number of measured variables. They are capable of detecting almost any type of fault. Their major weakness lies in the ambiguous isolation of faults arising from the noncausal nature of the models. On the other hand, the causal model-based FDI methods can

**Table 2. 99% Thresholds for $C_A$ Parity Relation, $O_{C_A}$ from Models Identified by Least Squares, and Last-Principal-Component Method for Different RBS Magnitudes Used During the Identification**

| RBS Magnitude ($\sigma$) | Least Squares | Last PCs |
|---|---|---|
| 0.2 | 0.1612 | 4.0925 |
| 0.4 | 0.1483 | 0.4629 |
| 0.6 | 0.1557 | 0.3262 |
| 0.8 | 0.1498 | 0.2190 |
| 1.0 | 0.1659 | 0.2121 |

*Note:* Threshold value based on the mechanistic model is 0.156.

both detect and isolate faults as long as causal models are developed for all the variables and the fault structures identified *a priori*. However, this need for theoretical or identified causal models makes the approach more difficult to develop and usually limits it to smaller systems. A more complete comparison was made under a number of different headings (Table 1). Some of these differences and the complementary strengths and weaknesses of the methods were illustrated using a simulated CSTR process.

Because the approaches have such different, but complementary strengths, a potentially fruitful area of research involves ways of combining them in a manner that utilizes the strengths of both. We hope that by pointing out major differences between these approaches and discussing their strengths and weaknesses, this article will serve to foster this research.

## Literature Cited


Basseville, M., "Detecting Changes in Signals and Systems—A Survey," *Automatica*, **24**, 309 (1988).

Chow, E. Y., and A. S. Willsky, "Analytical Redundancy and the Design of Robust Failure Detection Systems," *IEEE Trans. Automat. Contr.*, **AC-29**, 603 (1984).

Dunia, R., and S. J. Qin, "Subspace Approach to Multidimensional Fault Identification and Reconstruction," *AIChE J.*, **44**, 1813 (1998).

Frank, P. M., "Fault Diagnosis in Dynamic System Using Analytical and Knowledge Based Redundancy—A Survey and Some New Results," *Automatica*, **26**, 459 (1990).

Gertler, J., and D. Singer, "A New Structural Framework for Parity Equation Based Failure Detection and Isolation," *Automatica*, **26**, 381 (1990).

Gertler, J., and M. K. Kunwer, "Optimal Residual Decoupling for Robust Fault Diagnosis," *Int. J. Control*, **61**, 395 (1995).

Gertler, J., and R. Monajemy, "Generating Directional Residuals with Dynamic Parity Relations," *Automatica*, **31**, 627 (1995).

Gertler, J., *Fault Detection and Diagnosis in Engineering Systems*, Dekker, New York (1998).

Gertler, J., W. Li, Y. Huang, and T. J. McAvoy, "Isolation Enhanced Principal Component Analysis," *AIChE J.*, **45**, 323 (1999).

Isermann, R., "Process Fault Detection Based on Modeling Estimation Methods—A Survey," *Automatica*, **20**, 387 (1984).

Jackson, J. E., and G. S. Mudholkar, "Control Procedure for Residuals Associated with Principal Component Analysis," *Technometrics*, **21**, 341 (1979).

Jackson, J. E., *A User's Guide to Principal Components*, Wiley, New York (1991).

Kosanovich, K. A., K. S. Dahl, and M. J. Piovoso, "Improved Process Understanding Using Multiway Principal Component Analysis," *Ind. Eng. Chem. Res.*, **35**, 138 (1996).

Kourti, T., and J. F. MacGregor, "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods," *Chemometrics Intelligent Lab. Syst.*, **28**, 3 (1995).

Kourti, T., J. Lee, and J. F. MacGregor, "Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control," *Comput. Chem. Eng.*, **20**, S747 (1996).

Kourti, T., and J. F. MacGregor, "Multivariate SPC Methods for Process and Product Monitoring," *J. Qual. Technol.*, **28**, 409 (1996).

Kresta, J., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can.. J. Chem. Eng.*, **69**, 35 (1991).

Ku, W., R. H. Storer, and C. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Component Analysis," *Chemometrics Intelligent Lab. Syst.*, **30**, 179 (1995).

Ljung, L., *System Identification: Theory for the User*, Prentice Hall, Saddle River, NJ (1999).

MacGregor, J. F., C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multiblock PLS Methods," *AIChE J.*, **40**, 826 (1994).

MacGregor, J. F., and T. Kourti, "Statistical Process Control of Multivariable Processes," *Control Eng. Pract.*, **3**, 403 (1995).

Marlin, T. E., *Process Control: Designing Processes and Control Systems for Dynamic Performances*, McGraw-Hill, New York (1995).

Miller, P., R. E. Swanson, and C. E. Heckler, "Contribution Plots: The Missing Link in Multivariate Quality Control," *Conf. ASQCC and ASA*, Milwaukee, WI (1993).

Negiz, A., and A. Cinar, "Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models," *AIChE J.*, **43**, 2002 (1997).

Neogi, D., and C. E. Schlags, "Multivariate Statistical Analysis of an Emulsion Batch Process," *Ind. Eng. Chem. Res.*, **37**, 3971 (1998).

Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361 (1994).

Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995).

Patton, R., "Robustness in Model Based Fault Diagnosis: The 1995 Situation," *Proc. IFAC Workshop on Online Fault Detection and Supervision in the Chemical Process Industries*, Newcastle, England, p. 55 (1995).

Raich, A., and A. Cinar, "Diagnosis of Process Disturbances by Statistical Distance and Angle Measures," *Comput. Chem. Eng.*, **21**, 661 (1997).

Soderstrom, T., and P. Stoica, *System Identification*, Prentice Hall, Cambridge, UK (1989).

Willksy, A. S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, **12**, 601 (1976).

Wise, B. M., and N. L. Ricker, "Recent Advances in Multivariate Statistical Process Control: Improving Robustness and Sensitivity," *IFAC Int. Symp. ADCHEM*, K. Najim and J. P. Parbary, eds., Toulouse, France, p. 125 (1991).

Yoon, S., and J. F. MacGregor, "Fault Diagnosis with Multivariate Statistical Models: I. Using Steady State Fault Signature," *J. Process Control*, in press (2000).

Zhang, J., E. Martin, and A. J. Morris, "Fault Detection and Classification Through Multivariate Statistical Techniques," *Proc. ACC*, Washington, DC (1995).


*Manuscript received Oct. 29, 1999, and revision received Apr. 4, 2000.*