To the Editor:

The interesting article titled "Relationships Between Statistical and Causal Model-Based Approaches to Fault Detection and Isolation" (this issue) by Yoon and MacGregor was apparently triggered by our recently published ideas on "isolation enhanced" principal component analysis (Gertler et al., 1999). While arguing with the interpretation of some of our results, Yoon and MacGregor are touching upon a number of the most fundamental issues in model-based diagnosis. We believe that a constructive discussion of these issues is extremely useful for the diagnostic community, and for the potential users of diagnostic techniques, and put forth our comments of disagreement in this spirit.

(1) The basic premise of Yoon and MacGregor is that PCA-based modeling and least-squares-based modeling are fundamentally different, because while the former uses passively collected (normal) operating data, the latter requires designed experiments. We do not agree with this statement; in our view, normal operating data may be equally used for either kind of modeling, at the discretion of the model builder.

(2) To support the above premise, Yoon and MacGregor state that LS-type model identification requires persistently exciting inputs, with the implied statement that PCA-based modeling does not. We agree that, in the strict sense, the plant is not identifiable if the inputs do not satisfy certain independence requirements. However, it is always possible to identify models by the assignment of undefined parameters, which lead to valid consistency relations —as long as the input-relations existing in the identification data do not change. These are not "correct" models in the conventional sense, but they allow for the detection of faults and also for the isolation of sensor faults. The isolation scheme, however, breaks down if actuator faults are concerned. This mechanism is described in detail in Chapter 12 of Gertler (1998). With PCA modeling, the presence of relationships concerning the inputs does not lead to outright nonidentifiability; instead, the dimension of the representation subspace will be reduced in the model. This does not create problems in detection, nor in the isolation of sensor faults, but any enhanced isolation scheme breaks down when actuator faults are concerned (this mechanism is described in our isolation-enhanced PCA article). So there is no fundamental difference here. To make actuator faults isolable, the unwanted relationships among variables have to be removed (additional excitation has to be introduced)—and then both approaches will work.

(3) Another point of nondifference between the two approaches (as recognized also by Yoon and MacGregor) is their handling of additive plant faults. Such faults are not represented in normal operating data, so the nominal models (whether PCA or LS) do not know about them. If a plant fault occurs, it causes a deviation from the nominal conditions and is detected by either method. For isolation, the fault-effect needs to be modeled (in either approach); this requires controlled experiments with the fault present.

(4) Perhaps, the most significant difference between the two approaches, in our view, is that while explicit modeling considers the plant separated from its environment, PCA looks at the variables which represent the plant and the environment together. From a diagnostic point of view, explicit modeling produces consistency relations ("if these are the inputs, then these should be the outputs"), while PCA also describes what the inputs (and, consequently, the outputs) should normally be. Thus, the PCA model is richer, however, the additional information it carries may be a disadvantage in certain diagnostic situations, because there is no obvious distinction in the model between internal (plant) and external relationships.

(5) It follows from the previous paragraph that a change in the value of a measured or manipulated input is not a disturbance (or fault) for an explicit model-based algorithm (such as a parity relation), since it does not contradict any consistency relation. So, in Figure 4, it is natural that the parity relations will not capture the jump in the reactant flow rate. It may be considered a disturbance in the PCA framework, as a variable leaving its normal range. However, the data should stay in their nominal hyperplane even with this disturbance present, provided this input had any notable variation while the normal operating data was collected (which apparently was not the case; see the SPE plot in Figure 4b).

(6) The behavior of the two modeling approaches in the presence of noise is a very important issue. For static models with additive output noise, independent of the inputs, the least-squares method is known to provide unbiased estimates. Besides, the parity relation residuals are identical with the prediction errors; thus, an LS model minimizes the variance of the parity relation residuals. PCA modeling appears to lead to biased parameter estimates even in this simple case. Yoon and MacGregor provide some interesting experimental results which support these observations. Also, as they point out, their results are consistent with the findings of Negiz and Cinar (1997)—though the latter addressed the somewhat more complex problem of dynamic model identification. Obviously, this is an area where much more research is needed, in order to gain a clearer theoretical understanding of the reasons behind this behavior, and of the effect it has on diagnostic performance in the various schemes.

Literature cited

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

Gertler, J., W. Li, Y. Huang, and T. J. McAvoy, "Isolation Enhanced Principal Component Analysis," *AIChE J.*, **45**(2), 323 (1999).

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

Janos Gertler
George Mason University
Fairfax, VA 22030

Reply:

We appreciate Prof. Gertler's comments on our article, and are pleased to have the opportunity to respond to them. In principle, we generally agree with his points 3 and 4, but differ with him on points 1 and 2.

Points 1 and 2. In the article, we did not claim that PCA based-modeling and least-squares (LS)-based modeling are fundamentally different because they use different data. PCA, PLS, LS, and so on are estimation methods, any of which can be applied to any data set. If they are applied to data from designed experiments in which the data matrices X (inputs) and Y (outputs) are full rank (in a statistical sense), then all of them can provide estimates of the causal relationships among the inputs and outputs. This type of model is the one desired for the model-based FDI approach. Indeed, Prof. Gertler (Gertler et al., 1999) used PCA as an estimation method applied to data from designed experiments to arrive at a model for the residual equations of the model-based approach. This use of PCA extracts linear relationships among the variables from the principal components associated with the *smallest* eigenvalues of the information matrix. As stated in their article, the residual equations could also have been obtained from the same full rank data by using LS to identify the primary parity equations and then performing algebraic transformations on these.

However, in the MSPC approach one does not try to capture these causal relationships. One wants to capture the natural covariance structure of normal operating data when only *common-cause* variation is present. PCA and PLS are very efficient estimation methods for obtaining a few dominant latent variables that define this space of *high common-cause* variation. This space is defined by the principal components associated with the *largest* eigenvalues of the information matrix. These MSPC models therefore lie in a space orthogonal to that defined by the residual equations of Prof. Gertler's approach. Hence, the information contained in the latent variable model of the MSPC approach and that contained in the residuals of the parity equation approach are mutually exclusive. The parity equation approach essentially models the null space of the data information matrix and provides tests to check when and how the linear relations defining this space change. The MSPC approach models the space of high variance under normal operation and then provides tests for when and how this space changes. This is what we mean when we said that the methods are based on totally different models and, hence, use different tools for FDI.

The other issue is that the normal operating data used to build the MSPC models will almost never contain the information necessary to extract causal models because of feedback controllers, operating constraints, and so on. Although causal models are clearly desirable in order to achieve the full power of the model-based FDI approach, Professor Gertler points out that valid consistency relations can still be obtained from operating data with linear dependencies among the inputs by removing the linearly-dependent inputs (assignment of undefined parameters). These identified consistency relations on a reduced set of variables can then be used for fault detection and for isolation of sensor faults, although many of the sensor faults will now be confounded with (indistinguishable from) faults in variables that have been eliminated from the model through the linear dependencies. Since these consistency relations will be valid only as long as the relationships among the inputs remains the same as they were in the model identification data, this condition needs to be monitored constantly, presumably by using a PCA model of the input space. However, we believe that in this situation where one is using normal operating data and where a large number of such dependencies exist, the MSPC approach is a much more natural one and is easier to implement. It does not require the assignment of parameters, it allows for all the measured variables to be retained, it naturally reveals the confounding of fault information among the variables, it allows for testing that the linear dependencies among the inputs and other variables remain unchanged, and it naturally accounts for the *common-cause* disturbances that are always present.

Point 3. We agree to some extent, but would like to stress that, in spite of this lack of fault information, the contribution plot analysis of the MSPC models often greatly simplifies the search for assignable causes by highlighting a small group of variables that are highly correlated with the fault. There have been many industrial successes in isolating assignable causes for unmodeled faults using this approach (such as Kourti et al., 1996; MacGregor and Kourti, 1998). In our article, the simple examples in Figures 4, 5, 6, and 7 illustrate that the isolation of unmodeled faults is often possible or is greatly simplified by using contribution plots. However, we agree that to unambiguously isolate faults, information on any previously observed faults or controlled experiments with the faults present are usually needed.

Point 4. We agree on this. The MSPC approach considers all measured variables around the plant, not just a selected subset of inputs and outputs. This does provide a richer model in the sense that by using all these additional intermediate (or environmental) measurements and relationships involving them, one is bringing a lot of additional information into a detection and isolation problem. In our experience, the inclusion of all these intermediate measurements greatly enhances one's ability to detect and isolate unknown faults. It also enables the method to be insensitive to missing data (that is, prior sensor fault not yet corrected). This discussion also highlights a difference between the type of faults that the two approaches can easily detect. The explicit model approach only models the effect of a selected subset of inputs (X) on a set of outputs (Y), that is, the internal (plant) relationship. As pointed out by Prof. Gertler, only a fault which breaks these internal input/output relationships will be detected. Any changing correlation structure among the other variables (that is, "external relationships" among the inputs, among the intermediate variables or between the intermediate variables and the inputs or the outputs) is deemed acceptable as long as the input/output relationship between X and Y is unchanged. However, the latent variable models, obtained by using PCA or PLS, model the *common-cause* correlation structure within and between all the groups of variables (internal and external relationships). Any significant breakdown in any part of this much larger correlation structure will be detected as an unusual event or "fault" by the MSPC approach. In this way the latent variable models built from normal operation data are in a sense "richer" and better able to detect and to possibly isolate complex process faults in any part of the process. On the other hand, the parity equations, as long as they define causal relationships, can be more powerful at isolating sensor and actuator faults that occur in the well modeled or "internal" part of the process.

Point 5. We think that this point has arisen from a simple misunderstanding of our simulation that led to Figure 4 (Now, Figure 5). If there had been just a true change in the magnitude of F_A, then Prof. Gerlter's comment is indeed correct. However, we simulated a *bias fault* in the F_A sensor and both methods might be expected to pick it up as indeed they do even though their FDI performances are different, as shown in Figures 4 and 5.

Point 6. This point concerns the estimation algorithms used for identifying causal parity equations or residual models in the explicit model approach. We simply reinforced what other authors have also pointed out, that using PCA in this way as an identification approach for causal models is not a statistically efficient approach. However, this is a side issue related only to the explicit model approach, and has nothing to do with the MSPC approach or with differences between the two approaches.

In summary, we reaffirm that the difference between the two approaches is that one is modeling two very different sets of relationships using two very different data sets. The complementary strengths and differences of the two approaches stem from this difference.

We thank Prof. Gertler for his insightful comments on our article and we hope that his comments together with our response to them will help both the research and the user communities to better appreciate the differences between these two important approaches to FDI.

Literature cited

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

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

MacGregor, J. F., and T. Kourti, "Multivariate Statistical Treatment of Historical Data for Productivity and Quality Improvements," *Foundations of Computer Aided Process Operations*, J. Pekny and G. Blau, eds., Brice-Carnahan, University of Michigan, Ann Arbor Press, in print (1998).

Seongkyu Yoon and John F. MacGregor
Dept. of Chemical Engineering
McMaster University
Hamilton, Ontario, Canada L8S 4L