Data Processing and Reconciliation in Chemical Process Operations

By J. Romagnoli and M. C. Sánchez, Academic Press, San Diego, 2000, 270 pp., $79.95.

Data reconciliation is a branch of process monitoring, which has its formal roots in a seminal article by Kuehn and Davidson (1961). The technique consists of filtering measurements of a process with the object of determining reliable estimates of process variables. Typically, this is performed by contrasting measurements with an existing mathematical model of the process, determining the existence of biases and leaks. The process is in general rooted on the use of maximum likelihood models.

The technique is possible simply because there is a redundancy of measurements. This redundancy is seldom hardware redundancy (duplicate measurements of the same variable), and is almost always spatial or software redundancy, which basically arises when measurements do not conform to model equations.

In industrial practice, this technique has concentrated on models based on material balances and has found its usage in production accounting and instrumentation maintenance. Several types of software exist commercially that also use models including component and energy balances. Thus, advanced users perform parameter estimation for online optimization. The use of more detailed models of the process has been a matter of some (in my opinion unwarranted) controversy.

The book covers the field well, highlighting its most important aspects with efficiency. After an introductory chapter touching on the estimation problem, the book discusses the issue of variable classification, followed immediately by steady-state data reconciliation, including a very useful chapter on sequential processing. The issue of gross errors is discussed next. The book then turns to dynamic data reconciliation, the parameter estimation problem, and, in the final chapters, an important discussion on new tendencies and case studies.

The objective of classification is on the one hand to eliminate variables of the model, which lead to singularities and uncertainties (these are called unobservables), and, on the other hand, to further decompose the model so that only essential varibles (called redundent measured) are used for the core of the data reconciliation procedure. The authors make a brief review of the different approaches, and they concentrate on the QR decomposition approach in a separate chapter. Although there exist earlier suggestions to use QR for classification (Swartz, 1989), the authors had studied it extensively suggesting its use in a variety of situations (Sánchez and Romagnoli, 1996). The example used in Chapter 3 is rich and illustrates the technique well. The authors ingored illustrating graph oriented methods, a decision that I share. A small section discussing the advantages and disadvantages of all equation-oriented methods, including those not illustrated, like the Madron's Gauss Jordan factorization, is somehow missing. However, I think that this is an easy task to ask, and I am not sure if it is so easy to accomplish without additional research. Thus, this omission does not diminish these excellent chapters. Finally, Q-R factorization is emphasized, because it is easy to perform using software like Matlab, it can be used as part of some implementations of data reconciliation (as shown in Chapter 5), and, in general, we can see it as an elegant alternative to the matrix projection method.

The book illustrates steady-state data reconciliation very well. In particular, the chapter on sequential processing is a very important contribution. We should thank the authors for reminding the data reconciliation field the valuable results discussed by Mikhail (1976).

One of the most important issues in data reconciliation is the identification of gross errors, which consists mainly of biased instruments and leaks. The former do not follow the statistical assumption of normal distribution of errors and therefore invalidate the data reconciliation procedure altogether and the latter constitute a model deviation. Serial elimination of biases is treated using the method developed by Romagnoli and Stephanopoulos (1981) and Romagnoli (1983) in the early time of research in this area. The authors also present a combined procedure that takes advantage of the use of thresholds based on the chi-square distribution. Although these methods are adequate, there are in my view several others that merit more coverage. For example, the serial elimination procedure involving the so-called measurement test, which is in use in commercial software, is not discussed. Also, the use of generalized likelihood ratio (GLR) proposed by Narasimhan and Mah (1987), as well as the unbiased estimators (UBET) proposed by Rollins and Davis (1992) and Rollins and Devanthan (1993), should have been in my opinion included in more detail, especially because the latter seem to be powerful.

After this book was finished, some relevant additional work on the identification and estimation of gross errors was performed. The reader of the book should be aware that this body of work exists and consult it for a more complete understanding of the issue. First, an important article on gross error detection for nonlinear data reconciliation was published by Renganathan and Narasimhan (1999, 2000). In the case of linear systems, the problem of uncertainty in the gross error location is of importance. Some aspects of it were already pointed out by Iordache et al. (1985), problems were reported by Rollins and Davis (1992) and some remedies were offered by Sánchez and Romagnoli (1994). Finally, Jiang and Bagajewicz (1999) showed that this is not a phenomenon inherent to the identification method used, but rather a problem that will surface independently of the method used. The theory was later used to improve the performance and automation of several methods such as SEGE, UBET, and GLR (Bagajewicz et al., 2000). Some advanced serial compensation procedures (Jiang and Bagajewicz, 1999) including the improvement of the SEGE method presented in the book, which was published by Sánchez et al. (1999), as well as the performance measure that takes into account uncertainties (OPFE) developed by Jiang and Bagajewicz (1999) are new findings that will complement well the material offered in the book.

The chapter on dynamic data reconciliation is well written and covers key aspects. Considering that a whole book can be written just on this issue alone, the authors have given just a brief introduction of the most basic aspects and have left some work intentionally out. Good judgment has been exercised in this case, because, as the authors say, the book emphasizes steady-state data reconciliation. The reader can consult the work of Darouach and Zasadzinski (1991), which points out that many times the Kalman filter is singular and suggests the use of the generalized Kalman filter, the fault detection GLR method due to Narasimhan and Mah (1988), as well as the work due to Rollins and Devanathan (1993). The work of Albuquerque and Biegler (1995, 1996) is also of interest, but it is included in the chapter of parameter estimation.

Another great chapter is the one on parameter estimation. This is an increasingly important issue these days, because industry is rapidly shifting towards real time online optimization. This means that accurate models have to be used to optimize the process and suggest changes in production online. Thus, reconciling the data against these models is of paramount importance. The chapter covers extensively the latest and most important work in this area and presents excellent examples. The same can be said about the chapter on variance estimation, which covers the subject in a complete manner.

Finally, the last two chapters on new trends and case studies are informative and well written. The authors are very fond of these chapters and rightfully so. For example, they cover the analysis of the work of Johnston and Kramer, which represents a more general formulation of the problem, of which the classical maximum likelihood model that gives rise to the least-square formulation is a particular case. They also touch in detail on the issue of robust estimation (M-estimators, QQ-plots, and trust functions). Finally, they review the use of principal component analysis in data reconciliation. These are tendencies that the authors point out as new directions for research and development. While others may think of some other issues, there is no doubt that these recent techniques need to be researched further. This has already started to happen in issues regarding the use of PCA. Unfortunately, results on the use of PCA analysis, published after the book went to press, are discouraging (Bagajewicz et al., 1999, 2001).

The chapter on case studies presents several real industrial cases of applications of data reconciliation techniques. In addition, they are valuable because almost all the techniques presented in the book are being used. There is a distillation column example that also shows the importance of data reconciliation for online optimization, especially when a distributed control environment is used.

Finally, as the authors warn in the preface, the book is written for advanced readers and may be used for advanced undergraduates in chemical engineering for teaching purposes.

To sum up, this is an excellent and well-written book. Even though there are other books on the subject, this book contains valuable information that has not been put together before and thoroughly worked-out examples, which reveal the great potential of data reconciliation for good and proper process monitoring.

Literature cited

Albuquerque, J. S., and L. T. Biegler, "Decomposition Algorithms for On-line Estimation with Nonlinear DAE Models," *Comp. & Chem. Eng.*, **19**, 1031 (1995).

Albuquerque, J. S., and L. T. Biegler, "Data Reconciliation and Gross-Error Detection for Dynamic Systems," *AIChE J.*, **42**(10), 2841 (1996).

Bagajewicz, M., and Q. Jiang, "Gross Error Modeling and Detection in Plant Linear Dynamic Reconciliation," *Comput. and Chem. Eng.*, **22**(12), 1789 (1998).

Bagajewicz, M., Q. Jiang, and M. Sánchez, "Performance Evaluation of PCA Test for Multiple Gross Error Identification," *Comput. and Chem. Eng.*, **23**, Supp. S589 (1999).

Bagajewicz, M., Q. Jiang, and M. Sánchez, "Removing Singularities and Assessig Uncertainties in Two Efficient Gross Error Collective Compensation Methods," *Chem. Eng. Commun.*, **178**, 1 (2000a).

Bagajewicz, M., Q. Jiang, and M. Sánchez, "Performance Evaluation of PCA Tests in Serial Elimination Strategies for Gross Error Identification," *Chem. Eng. Commun.*, in press (2001).

Darouach, M., and M. Zasadzinski, "Data Reconciliation in Generalized Linear Dynamic Systems," *AIChE J.*, **37**(2), 193 (1991).

Iordache, C., R. S. H. Mah, and A. C. Tamhane, "Performance Studies of the Measurement Test for Detection of Gross Errors in Process Data," *AIChE J.*, **31**, 1187 (1985).

Jiang, Q., M. Sánchez, and M. Bagajewicz, "On The Performance of Principal Component Analysis in Multiple Gross Error Identification," *Ind. & Eng. Chem. Res.*, **38**(5), 2005 (1999).

Jiang, Q., and M. Bagajewicz, "On a Strategy of Serial Identification with Collective Compensation for Multiple Gross Error Estimation in Linear Data Reconciliation," *Ind. Eng. Chem. Res.*, **38**, 5, 2119 (1999).

Kuehn, D. R., and H. Davidson, "Computer Control. II. Mathematics of Control," *Chem. Eng. Prog.*, **57**, 44 (1961).

Mikhail, E., *Observations and Least Squares*, IEP Series, Harper and Row, New York (1976).

Narasimhan, S., and R. S. H. Mah, "Generalized Likelihood Ratio Method for Gross Error Identification," *AIChE J.*, **33**(9), 1514 (1987).

Narasimhan, S., and R. S. H. Mah, "Generalized Likelihood Ratios for Gross Error Identification in Dynamic Processes," *AIChE J.*, **34**, 1321 (1988).

Rollins, D. K., and J. F. Davis, "Unbiased Estimation of Gross Errors in Process Measurements," *AIChE J.*, **38**(4), 563 (1992).

Rollins, D. K., and S. Devanathan, "Unbiased Estimation in Dynamic Data Reconciliation," *AIChE J.*, **39**(8), 1330 (1993).

Romagnoli, J., "On Data Reconciliation Constraints Processing and Treatment of Bias," *Chem. Eng. Sci.*, **38**, 1107 (1983).

Renganathan, T., and S. Narasimhan, "A Strategy for Detection of Gross Errors in Nonlinear Processes," *Ind. Eng. Chemistry Res.*, **38**, 6, 2391 (1999).

Renganathan, T., and S. Narasimhan, "Addition/Correction: A Strategy for Detection of Gross Errors in Nonlinear Processes," *Ind. & Eng. Chemistry Res.*, **39**(1), 243 (2000).

Romagnoli, J., and G. Stephanopoulos, "Rectification of Process Measurement Data in the Presence of Gross Errors," *Chem. Eng. Sci.*, **36**, 1849 (1981).

Sánchez, M. C., and J. Romagnoli, "Monitoreo de Procesos Continuos: Análisis Comparativo de Técnicas de Identificación y Calculo de Bias en los Sensores," AADECA 94-XIV Simposio Nacional de Control Automático, Argentina (1994).

Sánchez, M., and J. Romagnoli, "Use of Orthogonal Transformations in Classification/Data Reconciliation," *Comp. Chem. Eng.*, **20**, 483 (1996).

Sánchez, M., J. Romagnoli, Q. Jiang, and M. Bagajewicz, "Simultaneous Estimation of Biases and Leaks in Process Plants," *Comput. and Chem. Eng.*, **23**(7), 841 (1999).

Swartz, C. L. E., "Data Reconciliation for Generalized Flowsheet Applications," American Chemical Society of National Meeting, Dallas, TX (1989).

Miguel Bagajewicz
School of Chemical Engineering and
Materials Science
University of Oklahoma
Norman, OK 73019