CHEMICAL ENGINEERING TRANSACTIONS

VOL. 39, 2014

A publication of AIDIC
The Italian Association
of Chemical Engineering
www.aidic.it/cet

Guest Editors: Petar Sabev Varbanov, Jiří Jaromír Klemeš, Peng Yen Liew, Jun Yow Yong
Copyright © 2014, AIDIC Servizi S.r.l.,
ISBN 978-88-95608-30-3; **ISSN** 2283-9216

DOI: 10.3303/CET1439121

Synergy between Data Reconciliation and Principal Component Analysis in Energy Monitoring

Barbara Farsang*, Sandor Nemeth, Janos Abonyi

University of Pannonia, Department of Process Engineering, P.O. Box 158, H-8201, Hungary
farsangb@fmt.uni-pannon.hu

Monitoring of energy consumption is central importance for the energy-efficient operation of chemical processes. Fault detection and process monitoring systems can reduce the environmental impact and enhance safety and energy efficiency of chemical processes. These solutions are based on the analysis of process data. Data reconciliation is a model-based technique that checks the consistence of measurements and balance equations. Principal component analysis is a similar multivariate model based technique, but it utilises a data-driven statistical model. We investigate how information can be transferred between these models to get a more sensitive tool for energy monitoring. To illustrate the capability of the proposed method in energy monitoring, we provide a case study for heat balance analysis in the well-known Tennessee Eastman benchmark problem. The results demonstrate how balance equations can improve energy management of complex process technologies.

1. Introduction

Over the last ten years the global energy consumption increased by 30 % (Rühl, 2013). Chemical companies are faced with rising energy and material costs. Increasing energy efficiency and reducing energy usage become the most important factors of competitiveness.

Energy efficiency can be defined in several ways (Xia and Zhang, 2010):

- *Performance efficiency* is characterised by production, cost, energy sources and environmental impact.
- *Operational efficiency* is evaluated by considering the proper coordination of different system components.
- *Equipment efficiency* is indicated by capacity, specifications and standards, constraints, maintenance.
- *Technology efficiency* includes the reduction of life cycle cost and coefficients in the conversing/ processing/transmitting rate, in addition improving of the novelty and optimality of processes.

Energy monitoring systems can be used to improve energy efficiency and reduce the energy consumption. The purpose of energy management is “to enable understanding of energy consumption data; identify underlying factors which impact upon consumption; and set appropriate targets that allow you to review performance” (Carbon Trust, 2010). Monitoring systems improve the energy efficiency in processes, because these systems calculate actual energy use, estimate the needed energy for normal operation and highlight where energy use can be improved. These systems can detect energy wastages caused by human error, equipment malfunction or poor process control. Significant difference between measured and theoretically required values indicates the abnormal behaviour. These abnormal situations can cause a significant impact on the safety and economy of the process industry. When the monitoring system responds fast and supports the control of the unusual situation, the economic loss can be significantly reduced. A detailed overview of these systems is given in Bayindir et al. (2011).

In chemical processes robust methods are required for process monitoring due to the safety, economic operation and production specifications.

Multivariate statistical data-based models techniques are powerful tools for process monitoring. Although multivariate statistical models do not directly reduce operation costs, when financial indicators are

Please cite this article as: Farsang B., Nemeth S., Abonyi J., 2014, Synergy between data reconciliation and principal component analysis in energy monitoring, Chemical Engineering Transactions, 39, 721-726 DOI:10.3303/CET1439121

calculated based on exact and accurate process values a more realistic picture is available for the decision makers.

The most commonly used model is the Principal Component Analysis (PCA). PCA is applied in various areas of chemical engineering, e.g. process monitoring, quality control, disturbance detection, sensor fault diagnosis and process fault diagnosis (Misra, 2002). When a priori model is not available, measured values involve the model, because process variables are linked by a set of constraints, e.g. balance equations. Using PCA the correlation among variables can be found under normal operating conditions and information can be extracted from process data. The main idea of PCA is to replace a large number of interrelated variables by a few uncorrelated variables (Wold, 1987).

The performance of model based process monitoring systems highly depends on the quality of the model. Hence, good PCA based solutions require accurate and validated historical process data with high information content. Measurements are always affected by errors due to imperfect instruments, signal transmission, power fluctuation, improper instrument installation and miscalibration. To minimize random errors pre-processing of data is necessary. Data reconciliation (DR) technique is a useful tool, because this method uses the balance equations and physical-chemical laws so the consistency of data is provided. Jiang et al. (2013) summarized the principle of DR and presented a study to illustrate the capability of data reconciliation for operational data accuracy. Sometimes it is difficult to measure important variables, which influence the energy uses, e.g. steam flow. In this case, DR technique is used to reconcile the measurements and to estimate unmeasured variables. DR and PCA were already combined in some applications. It has been shown that data reconciliation can improve the quality and sensitivity of PCA model by reducing the number of principle components (Amand, 2001).

In this paper we show a stronger relationship between PCA and DR techniques and we propose a multivariate model based energy monitoring system using the synergistic combination of PCA tools, data reconciliation and flowsheeting simulator.

The paper is organised as follows: in Section 2 we describe the synergy between PCA and DR. The application of the proposed fault diagnosis system is illustrated in energy balance of Tennessee Eastman Process. In Subsection 3.1 the analysed process is introduced. The results are presented in Subsection 3.2. Section 4 summarizes the paper with same key results.

2. Similarity of PCA and DR projections

PCA and DR both perform optimal projection of the process data into a (linear) multivariate model. The model of PCA is defined by the covariance matrix of the data, while the model of the DR is defined by material and energy balance equations, usually given in a system of linear equations, $Ax = b$ (A is incidence matrix, x vector contains variables and b is the source vector). The classical data reconciliation is formulated by the following equation:

$$\hat{x} = (I - V_{\tilde{d}} A^T (A V_{\tilde{d}} A^T)^{-1} A) x + V_{\tilde{d}} A^T (A V_{\tilde{d}} A^T)^{-1} b = P_{DR} x + c \tag{1}$$

where x represents the measured variables, I is an identity matrix, $V_{\tilde{d}}$ is the variance matrix of the error, P_{DR} is the projection matrix and c is a constant shift vector.

The Projection matrix of PCA is determined based on covariance matrix of normalized data pairs. The covariance matrix (F) is decomposed three matrices with singular value decomposition: $F = USV^T$, where the columns of U are eigenvectors of covariance matrix, the diagonal elements of S are the eigenvalues, and the columns of V is represent the right singular vectors. According to the number of principal components (*p*), the first *p* columns of eigenvectors are selected, and the projection matrix is formulated based on these *p* vectors as:

$$P_{PCA} = U_p^T \, U_p \tag{2}$$

To compare projection matrices Krzanowski similarity factor is applied. Krzanowski (1979) defined a factor to measure the similarity between matrices by comparing the hyper planes spanned by eigenvectors. This factor characterizes the angle (Θ) between two hyper planes, because Krzanowski similarity factor shows the squared cosine values between all the combinations of the first p principal components from two matrices (X, Y):

$$s_{PCA}(X,Y) = \frac{1}{p} \sum_{\mathrm{i=1}}^{\mathrm{p}} \sum_{\mathrm{j=1}}^{\mathrm{p}} \cos^2 \Theta_{\mathrm{i,j}} = \frac{\mathrm{trace}\left(\mathrm{U}_{\mathrm{X,p}}^{\mathrm{T}} \mathrm{U}_{\mathrm{Y,p}} \mathrm{U}_{\mathrm{Y,p}}^{\mathrm{T}} \mathrm{U}_{\mathrm{X,p}}\right)}{\mathrm{p}} \tag{3}$$

where U_p matrix contains the eigenvectors, *p* is the number of principal component.

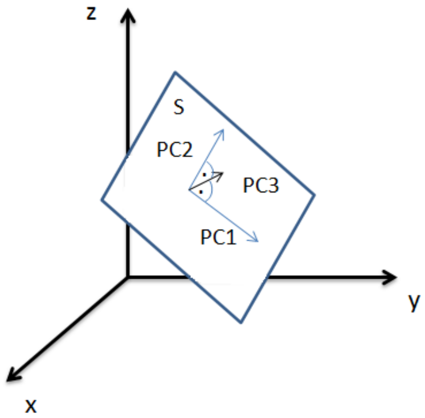

Figure 1: The normal vector of the plain is determined by the neglected principal component (PC3)

The concept is illustrated in three dimensions (see Figure 1). In this example the third principal component is (PC3) perpendicular to the plane determined by the data, so the neglected eigenvector as a normal vector of the plane can define the coefficients in equation of plain, like $ax + by + cz = d$.

This interpretation is useful when a data – driven model is needed, so balance equations should be detected from the correlation between variables. It should be noted that this approach results in the application of total least squares (TLS) technique. TLS is type of errors-in-variables regression, in which observational errors on both dependent and independent variables are taken into account (Ganger, 2008).

With the use of this approach not only the projection matrix of P_{PCA} can be calculated and its similarity to the projection matrix of P_{DR} can be evaluated, but by using TLS coefficients parameters of the balance equations can also be (re)calculated.

These approaches are verified based on a case study detailed in Section 3. Projection matrix of data reconciliation and principal component analysis are compared based on Krzanowski similarity factor. Based on data (which come from the process) we determine the relationship between the variables. The results demonstrate how balance equations can improve energy management of complex process technologies.

3. Results and discussion

In this work we use energy balances of Tennessee Eastman Process to illustrate the synergy between projection matrices of data reconciliation and principal component analysis. In this section we use the nomenclature of Tennessee Eastman Process, so users of Tennessee Eastman model can easily identify variables if they would reproduce our experiments. The numbers after the variable name identify streams in system (Figure 2 helps coupling the streams and numbers).

The operating cost (TS) of the technology is determined by the loss of raw materials (purge stream, byproducts and dissolved reactant in product), compressor work and steam flow:

$$TC = PC \cdot FTM(9) + PrC \cdot FTM(13) + CC \cdot CW + SC \cdot FS \qquad (4)$$

where PC, PrC, CC and SC are the cost of purge, product stream, compressor and steam, FTM(9), FTM(13) are the component flow of purge and product streams, CW is the compressor work, FS is the steam rate.

In this study we do not examine how optimization and decision support techniques rely on examined variables and how these variables influence the manipulated variables. We only deal with the reliability of the measurements and how this uncertainty appears in the estimated operating costs. This equation draws attention to importance of accurate measurement of flow rate. In this paper we analysed the effect of flows rates to total cost and present the role of DR and PCA in energy balances. In Subsection 3.1 we present this process and highlight the analysed parts of technology. Simulation results can be found in Subsection 3.2.

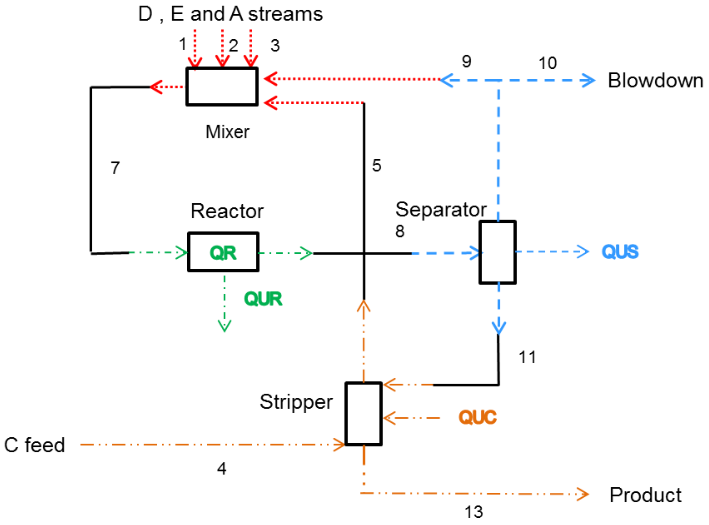

Figure 2: Analysed streams in Tennessee Eastman Process (numbers identify streams in Eq(5))

3.1 Tennessee Eastman Process

Downs and Vogel (1993) prepared a process model of an industrial chemical process to develop, study and evaluate process control technology. This model is often used for evaluate and compare different data-analysing methods.

The system includes five major unit operations: reactor, condenser, gas-liquid separator, compressor and stripper. The gaseous reactants are fed to the reactor. In reactor four reactions take place: all reactions are exothermic and irreversible. The reactor product stream (in gas-phase) passes through a cooler for condensing the products and in the separator the two phases (products and reactant) are separated. The reactants are recirculated; purge stream removes inert components and byproducts from system. The liquid stream of separator contains dissolved reactants which are removed in stripper. The bottom stream of stripper is the product; the overhead stream recycles back to the reactor feed.

In Figure 2 we highlight the analysed streams and heat sources. The energy balances of operation units can be defined easily in a matrix form: $\mathrm{Ax} = \mathrm{b}$.

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & HST(7) & -HST(8) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & HST(8) & -HST(9) & -HST(10) & -HST(11) & 0 \\ 0 & 0 & 0 & HST(4) & -HST(5) & 0 & 0 & 0 & 0 & HST(11) & -HST(13) \\ HST(1) & HST(2) & HST(3) & 0 & HST(5) & -HST(7) & 0 & HST(9) & 0 & 0 & 0 \end{bmatrix} \quad (5)$$

$$\mathbf{b}' = \left[-RH + UAR \cdot (TCR - TWR) \quad UAS \cdot (TST(8) - TWS) \quad UAC \cdot (TST(8) - 100) \quad 0 \right]$$

$$\mathbf{x}' = \left[FTM(1) \quad FTM(2) \quad FTM(3) \quad FTM(4) \quad FTM(5) \quad FTM(7) \quad FTM(8) \quad FTM(9) \quad FTM(10) \quad FTM(11) \quad FTM(13) \right]$$

where FTM is the mole flow of streams, HST is the specific enthalpy of streams (it depends on the composition and temperature of stream), RH is the released reaction heat, UAR and UAS are the transferred heat to water in reactor and separator, UAC is the transferred heat from steam, TWR is the reactor temperature, TWS is the separator temperature and TCC is the stripper temperature.

Measurements of process variables are always affected by errors, so variables do not satisfy energy balances. Data reconciliation can minimize the balance error ($\mathrm{Ax} - \mathrm{b}$) by optimal projection of the project variables to the model equations.

Table 1: Square balance error (sum(mean((Ax-b)^2)) and specific total cost (TC) in case of raw and different ways reconciled values

x vectors	Square balance error	TC ($/h)
Raw measurements	12.0662	55,975
Reconciled values with time-dependent A and b matrices	6.72 10^{-29}	54,271
Reconciled values with time-independent A and b matrices	1.21 10^{-28}	54,924
Projected points with PCA (5 principal components)	0.5899	56,084
Projected points with PCA (7 principal components)	0.5974	55,974

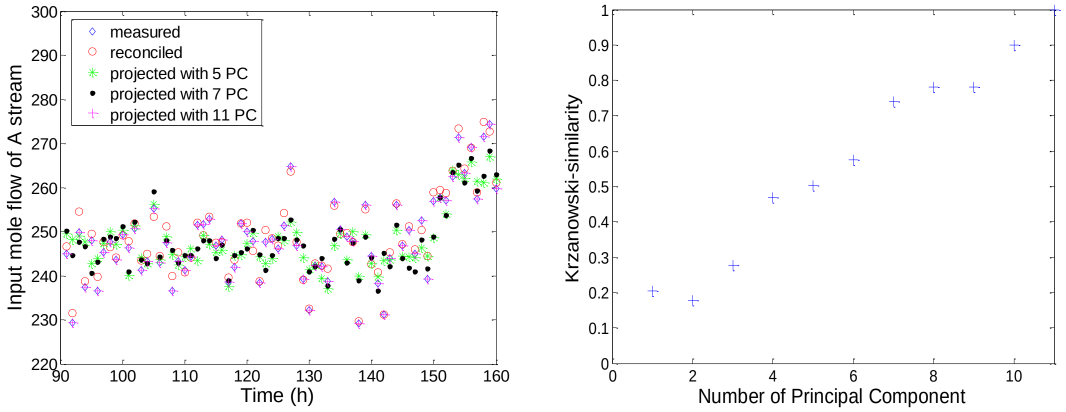

Figure 3: a) Comparison of measured, reconciled with time-dependent and time-independent projection matrix and projected values in case of product stream. b) Similarity of eigenvectors of projection matrix of PCA and DR

These model equations represent the dependency of process variables (e.g. stream 11 is the liquid product of separator and the inlet stream of stripper, so the enthalpy of the outlet and inlet stream should be identical – apart from the heat loss in pipeline). PCA can also detect such relationship, so the combined application of PCA and DR can increase the sensitivity and accuracy of energy balance based energy monitoring.

3.2 Results and discussion

Firstly we compared the projections of DR and PCA and we analysed how they influence the estimation of the total cost of process. Since temperature is measured more accurate (0.1 - 0.5 %) than flow rates (1.5 - 3 %) (Lipták, 2003), we assumed perfect temperature measurements and focus on balancing the flow rates. This assumption allows the application of linear data reconciliation. Since temperature varies in time, therefore elements of incidence and source matrices (which are non-linear function of the temperature) can change over time resulting in a linear parameter varying (LPV). Table 1 shows that projections based on time dependent or time-independent A and b matrices are almost identical, so the process in the studied operating regime is almost linear. Principal component based projection also increases the reliability of the data, so indirectly it also reduces the balance error.

The last column of Table 1 highlights that value of total cost depends on the quality of the variables. It should be noted that neither PCA nor DR reduces the total cost but these techniques give more realistic cost estimation. To illustrate the similar effects of these techniques Figure 3a shows reconciled and projected values of mole flow of A stream (FTM(1)).

We also calculated the Krzanowski similarity factor (see Figure 3b) to compare the eigenvectors of the projection matrices of DR and PCA.

In the third step we examined how we can get information about balance equation from data using TLS based interpretation of the eigenvectors. Data from steady state operation were analyzed because we tried to avoid the unhandled effects of process dynamics. The heat balance of separator (2^{nd} model equation is Eq.5 which contains four variables) is used for the demonstration of the approach. We collected the necessary data from simulator (FTM(8:11)). Eigenvalues of the covariance matrix of the data show that there is a strong connection between variables. The first tree principal components define a hyperplane and the remaining one eigenvector defines the parameters of the equation defining of this plane. TLS can estimate these parameters:

$$n = [5.4498 \quad -4.1473 \quad -0.3231 \quad -1] \qquad (6)$$

Since PCA is based on normalized process values, the extracted equation describes the relationship between the normalized variables. When the effect of the normalization is taken into account it can be shown that the extracted model identical to the balance equation, so the PCA based Eq.6 defines the same hyperplane as balance equation based data reconciliation.

The proposed technique can be used to verify PCA and DR models and detect significant changes in the process affecting energy efficiency.

4. Conclusions

Energy monitoring requires validated data and informative alarms related to abnormal operations. Principal component analysis and data reconciliation are widely used techniques to improve the accuracy and reliability of data. We found strong relationship between these techniques, and we presented how we can infer the coefficient matrix of DR from the projection matrix of PCA.

The whole concept is illustrated based on the well-known Tennessee Eastman case study. In this study we assumed perfect temperature measurement and balanced flow rates. The resulted linear parameter varying model gave almost the same performance as a global linear model, so we showed that in the studied operating regime linear data reconciliation technique can be effectively applied. The operating cost of the technology has been calculated to show the effect of the projections. It has been shown that increasing the reliability of the data highly modifies the estimated cost, so PCA and DR are useful tools when the estimated cost is used in real time control or optimization. In further work we analyse how model equations of DR and PCA can be merged together, how the analogy of the two techniques can be demonstrated in more complex examples, and how temperature measurements can be balanced to further improve the accuracy of cost estimation.

Acknowledgements

The research of Barbara Farsang and Janos Abonyi was realized in the frames of TÁMOP 4.2.4.A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system convergence program”. The infrastructure of research is supported by the frame of the TÁMOP 4.2.4.A/11/1-KONV-2012-0071 project. The projects were subsidized by the European Union and cofinanced by the European Social Fund.

References

Amand T., Heyen G., Kalitventzeff B., 2001, Plant monitoring and fault detection: Synergy between data reconciliation and principal component analysis, Computers & Chemical Engineering, 25, 501-507, DOI: 10.1016/S0098-1354(01)00630-5.

Bayindir R., Irmak E., Colak I., Bektas A., 2011, Development of a real time energy monitoring platform, International Journal of Electrical Power & Energy Systems, 33, 137-146, DOI: 10.1016/j.ijepes.2010.06.018.

Carbon Trust, 2010, Monitoring and Targeting, <www.carbontrust.com/media/31683/ctg008_monitoring_and_targeting.pdf>, Accessed 07.02.2014.

Downs J. J., Vogel E. F., 1993, A plant-wide industrial process control problem, Computers & Chemical Engineering, 17, 3, 245-255, DOI: 10.1016/0098-1354(93)80018-I.

Ganger W., 2008, The Singular Value Decomposition, <www.math.ethz.ch/education/bachelor/lectures/hs2012/other/linalg_INFK/svdneu.pdf>, Accessed 15.03.2014.

Jiang X., Liu P., Li Z., 2012, A data reconciliation based approach to accuracy enhancement of operational data in power plants, Chemical Engineering Transactions, 35, 1213-1218 DOI:10.3303/CET1335202.

Krzanowski W., 1979, Between-groups comparison of principal components, Journal of the American Statistical Society, 74, 367, 703-707, DOI: 10.1080/01621459.1979.10481674.

Lipták B.G., 2003, Instrument Engineers' Handbook, Volume 1, Fourth Edition: Process Measurement and Analysis, CRC PRESS, Boca Raton, Florida, United States of America, ISBN: 0-8493-1082-0 (v. 1)

Misra M., Yue H.H., Qin S.J., Ling C., 2002, Multivariate process monitoring and fault diagnosis by multiscale PCA, Computers & Chemical Engineering, 26, 1281-1293, DOI: 10.1016/S0098-1354(02)00093-5.

Rühl C., 2013, BP Statistical Review of World Energy 2013, <www.bp.com/en/global/corporate/about-bp/energy-economics/statistical-review-of-world-energy-2013.html>, Accessed 07.02.2014.

Wold S., Esbensen K., Geladi P., 1987, Principal Component Analysis, Chemometrics and Intelligent Laboratory Systems, 2, 37-52, DOI: 10.1016/0169-7439(87)80084-9.

Xia, X., Zhang, J., 2010, Energy efficiency and control systems-from a POET perspective, Methodologies and Technology for Energy Efficiency, 1, 255-260, DOI: 10.3182/20100329-3-PT-3006.00047.