Process performance analysis in large-scale systems integrating different sources of information

Inês M. Cecílio*, Knut Rapp**, Nina F. Thornhill*

* Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: i.cecilio09@imperial.ac.uk).
**ABB Process Automation Division, NO-0603 Oslo, Norway

Abstract: Process auditing using historical data can identify causes for poor performance and reveal opportunities to improve process operation. To date, the data used has been limited to process measurements; however other sources hold complementary information about the process behavior. This paper proposes a new approach to root-cause diagnosis, which also takes advantage of the information in utility, mechanical and electrical data, alarms and diagrams. Its benefit is demonstrated in an industrial case study, by tackling an important challenge in root-cause analysis: large-scale systems. This paper also defines specifications for a semi-automated tool to implement the proposed approach.

Keywords: chemical industry, data fusion, fault diagnosis, large-scale systems, performance analysis, plantwide, operators

1. INTRODUCTION

It is important to analyse the historical records of process operation because they give insights on the process behaviour under normal and abnormal conditions (Qin, 2003), and can also be used for root-cause diagnosis, control performance assessment, operational decision support, as well as maintenance. (Thornhill, 2007; Choudhury et al., 2007)

This is also true in large-scale processes, but it becomes difficult to extract useful information when the number of variables rises to some thousands. Specific challenges are visualisation, computational memory and speed, and the complex dependencies amongst variables due to recycles, energy integration and control. Indeed, survey of literature shows that most contributions on process performance analysis target only single unit operations or reasonable-sized processes, far from the reality of very large processes such as in the oil and gas industry.

An additional challenge in the analysis of process performance is when the necessary data is not available or cannot be found.

The analysis of process historical records is a time-consuming task, normally undertaken by contracted services companies. The experience of one of the co-authors as a consultant in such a company motivated the specification of functional requirements for tools to help the experts with this task. Therefore this paper presents work in progress towards the development of such tools. Its proposal is to address process analysis by taking advantage of the information held in different sources. The value of this approach has been recognized by the ARC Advisory Group in a report for a process analytics client (Fiske, 2009).

Recently, some progress has been done with the integration of different sources of information, with Cecilio et al. (2011) demonstrating the value of integrating data from the mechanical, electrical and utility systems with process data, and Yim et al. (2006) and Maurya et al. (2003) who combined the connectivity information from process schematics with data-driven analysis methods. However, most contributions for process performance analysis are still based on process measurements alone.

The contributions of this paper are thus to:

- demonstrate the value of an integrated analysis approach;
- propose additional sources of information and analyse their dissimilarities so as to outline technical challenges;
- devise a systematic approach to apply the proposed concept in large-scale processes; and
- define the functional specifications to semi-automate the integrated analysis approach.

The paper is structured as follows. Section 2 reviews the contributions to process performance analysis in terms of sources of information exploited and addressing large-scale systems. Section 3 introduces the idea of integrated analysis and discusses what it involves. Then, section 4 uses an industrial case study of a large-scale process to demonstrate the concept. The approach used in this section is manual, thus section 5 defines the functional requirements for its semi-automation. The paper ends with conclusions and comments.

2. BACKGROUND

2.1 Sources of information in process performance analysis

The analysis of process performance can exploit different sources of information. Indeed, in their detailed review series

Venkatasubramanian et al. (2003 a, b, c) classify the methods of analysis according to the type of knowledge used.

The methods with greatest impact are based on operational data from process variables (Thornhill, 2007). The historical data can be used to build data models of normal operation and of various faulty scenarios, which are then used as a reference for the running process behaviour (Qin, 2003). Many of these data models are supported by dimensionality reduction techniques like principal component analysis.

Another approach to using historical process data is that of plant-wide disturbance analysis. Here, no data models are built and, instead, the time histories of the various variables are compared to extract features indicating the direction of the root-cause (Choudhury et al., 2007; Bauer and Thornhill, 2008). To date, research on data-driven approaches has been limited to the use of data from process variables, and the integration of data from other systems in the plant has not been considered.

Besides operational data, other sources of process knowledge are first principles models, which add to the analysis the physical and chemical dependencies between the variables. Besides numerical models, information on the causal relationships between variables can be represented qualitatively, mainly in the form of signed digraphs. This technique was comprehensively reviewed by Maurya et al. (2003), who also showed how to derive systematically these qualitative models from model equations. Recently, some authors have extracted this knowledge from process schematics; Yim et al. (2006) used schematics represented electronically under the CAEX standard, and Jiang et al. (2009) converted the process schematic to a digraph based on the information of the controllers.

The combination of process data and qualitative causal information has shown benefits for the analysis (Chiang and Braatz, 2003). However, until now only a few studies have used the process schematics as the source for that causal information (Thambirajah et al., 2009; Jiang et al., 2009).

This paper defends the integration of schematics and process data, as well as other sources of information like alarms and data from the mechanical, electrical and utility systems.

2.2 Process performance analysis in large-scale systems

Large-scale systems represent a challenge in any domain dealing with data analysis and visualisation. In the field of process performance analysis, various authors have addressed this subject, and the contributions can be roughly classified in two approaches.

One strategy is to reduce the large set of data to a smaller set, and then base the analysis on the latter. The reduction of dimensionality is based on the dependencies amongst variables, which in chemical processes are due to the physical and chemical principles governing the process operation, such as mass and energy balances. Some of the most popular techniques for the purpose are Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Partial Least Squares (PLS) (Russel et al., 2000).

Using this approach, Maurya et al. (2005) reduced by 40% the computational complexity in qualitative trend analysis by exploiting the redundancy on measured data with PCA. Thornhill et al. (2002) used the scores of spectral PCA to extract smaller clusters of measurements having the same disturbance.

A second approach to large-scale systems was introduced by Wold et al. (1996) which consists of dividing the variables into conceptually meaningful blocks before applying multivariate modelling techniques. Some of the algorithms proposed by these authors are consensus PCA, hierarchical PCA, multiblock PLS, and hierarchical PLS. Westerhuis et al. (1998) and Qin et al. (2001) demonstrated that the multiblock models can be calculated directly from the standard PCA and PLS techniques.

The success of this approach depends on a proper variable blocking, which requires knowledge of the process or access to the process diagrams. Qin et al. (2001) divided the data according to process sections that describe a unit or a specific physical or chemical operation.

In this paper the authors use an alternative approach to root-cause analysis in large-scale process. Here, the reduced set of variables under analysis changes at each step, as the focus of analysis moves along the process towards the root-cause.

3. INTEGRATED ANALYSIS

This section examines the types of information available for integrated analysis, and compares their technical characteristics for posterior research specifications.

The idea of integrated analysis is to take advantage of the separate parts of information available in different sources. The proposal is justified because the behaviour of a process is influenced by a combination of factors, namely the process conditions, control tuning and structure, the condition of equipment and utility systems, and the actions of operators.

Information on these factors is held in (i) measurements from process and control variables, usually collected continuously by process control systems, (ii) measurements from the utility systems, machinery and electrical drives, usually collected by condition monitoring systems, and (iii) event logs, recording occurrences such as interventions of operators and alarms.

In addition to operational records, process schematics could also be integrated to provide qualitative topological information, helpful to give meaning to data.

The systematisation and automation of this integrated analysis should help the experts responsible for analysing the plant operation. Specifications for such tools require understanding the unique characteristics of the data to be integrated. These are:

- *format/support*: data from process, utility, mechanical and electrical systems are numerical and continuous if pertaining to physical variables; for on-off variables, like valve openings, data is generally numerical but binary. Alarms are held in text format, and diagrams are most commonly supported in paper or pdf files.

- *dynamics*: process and utility systems are mostly dependent on mass and heat phenomena with time constants in the order of minutes. On the contrary, the dynamics of rotating machinery are determined by the rotational speed, in the order of 1 rotation per second or faster; similarly the duration of events in electrical systems is typically below 1 second.

- *sampling rate*: consistent with the systems dynamics, process and utility measurements are sampled at a slower rate than machinery condition and electrical measurements. These latter might be available at the same sampling rate as process measurements but, in such cases, the dynamic events are usually not captured.

- *databases*: it is common that different domains in the plant have technology supplied by different vendors or different business units in the same company, leading to data being stored in different databases.

- *time synchronisation*: different databases may have different clocks, raising a problem of time synchronisation between different sources of data and alarms.

- *diagrams*: tags from different systems and equipment are generally figured in different diagrams, posing practical problems of localising and handling several diagrams.

4. INDUSTRIAL CASE STUDY

This section demonstrates the concept of integrated analysis with a combination of process data, utility data, alarms and diagrams from an industrial plant. To apply this concept in a large-scale process, a systematic approach was devised based on a consulting study by the industrial co-author.

4.1 Process description and motivations for analysis

The case study is based on industrial data and diagrams from a gas processing plant. The system can be classified as large-scale since more than 7500 measurement tags and around 350 diagrams are available. For confidentiality reasons, dates are omitted, and systems and tags are anonymised.

The motivation for the analysis is an unplanned process shut-down resulting from the propagation of disturbances through a series of systems.
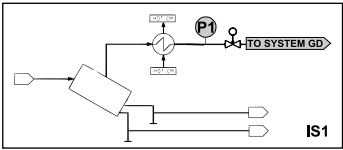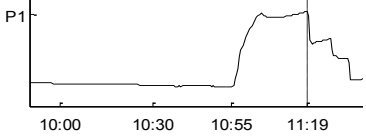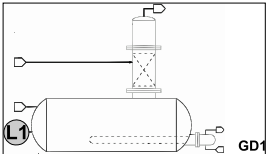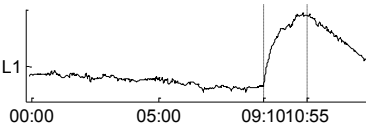
4.2 Integrated analysis

An unplanned process shut-down is the starting point for this analysis. The aim is to find the closest possible root cause.
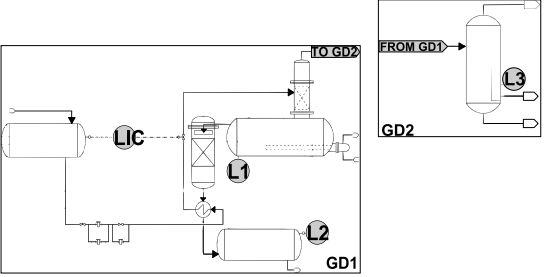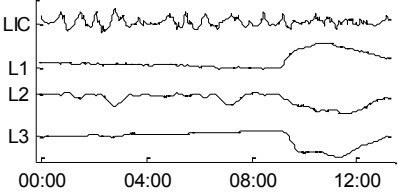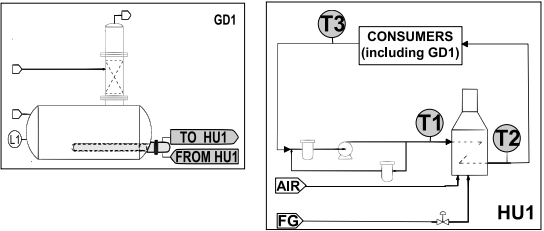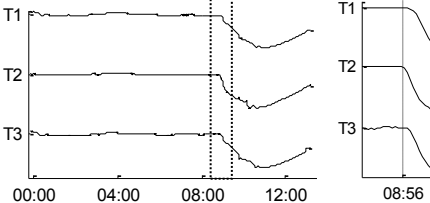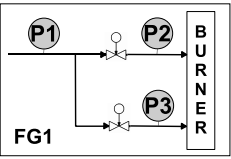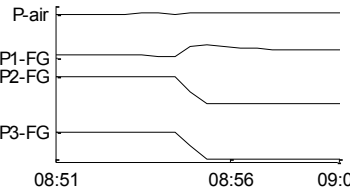
Table 1 is used to describe the process of diagnosis in a structured way. The flow of analysis proceeds from left to right, top to bottom. The rows represent different envelopes of analysis, where envelope refers to a reduced set of variables which is being analysed. Each row the envelope moves towards the root cause. The columns classify the type of information used within each envelope.

As an example, the first row of the table should be read as follows. To start investigating the unplanned shut-down at 11:19, the *alarm* log is searched around that time and a high pressure alarm in system IS is found triggering the shut-down. The safety instrument associated with this alarm is the first envelope of analysis (step 1). Thus, the instrument is localised in a *diagram* to understand its location, and the corresponding *time history* is analysed around that time. The conclusions from this analysis lead to a new envelope, in the second row, and the description of the diagnosis goes forth in this manner.

Table 1. Root cause diagnosis with integrated analysis applied to an industrial case study.

	Alarms	Diagrams	Process/Utility data
Step 1	At 11:19 the alarm log reports a high-pressure shut-down order by instrument P1-IS	P1-IS is localised in drawing IS1; it is situated at the outlet of system IS, immediately before system GD.	The time trend of P1-IS reveals that the pressure increase leading to the alarm trip started at 10:55.
Step 2	At 10:55 the alarm log reports a high-level trip of system GD (L1-GD). As seen in IS1, GD is immediately downstream of the previous envelope; when GD trips the valve at its inlet (outlet of IS1) is closed.	The envelope of analysis is now around L1-GD. A second drawing shows that L1-GD indicates the level of a reboiler.	The time trend of L1-GD reveals that the level increase leading to the alarm trip started at 09:10.

Table 1. (continuation) Root cause diagnosis with integrated analysis applied to an industrial case study.

	Alarms	Diagrams	Process/Utility data
Step 3	No records of alarms before this point because main alarm log is overwritten periodically.	The envelope of analysis is expanded upstream and downstream of the reboiler, along the process streams. A third diagram, GD2, is needed to visualise one of the downstream directions. The level tags in this new envelope are selected.	The level disturbance was also present in the equipment downstream of the reboiler. However, while in the reboiler the level increased (L1-GD), downstream it decreased (L2-GD, L3-GD). This suggests a reduction of the flow out of the reboiler.
Step 4		Diagram GD1 is searched for the elements that can influence the flow between the reboiler and the equipment downstream. The heating utility is identified, and the relevant temperature tags are selected from the corresponding diagram, HU1.	The temperature of the heating utility suffered a decrease throughout, which can explain the reduction of gas flow from the reboiler. The time delays between the start of the disturbance suggest that it entered this system between T1-HU and T2-HU, in the burner.
Step 5		Diagram HU1 (row above) reveals the burner is fed by a fuel gas stream (FG) and by air. Thus, the envelope of analysis is expanded to these two systems. Pressure tags upstream of the burner in both systems are selected from the corresponding diagrams (above, only a part of the fuel gas system is shown).	The pressure of air (P_{air}) shows no disturbance, whereas the inlet pressures of fuel gas (P2 and P3) drop. The reduced supply of gas can explain the decrease of the burner outlet temperature. Moreover, it can be seen that the behaviour of the fuel gas pressure is different along the supply line; up to P1 the pressure increases and beyond that it decreases. This indicates the flow of gas was stopped after P1.
Step 6		Diagram FG 1 (row above) is searched for the elements that can influence the flow between P1 and the next sensors, P2 and P3. Two valves are identified and the actuator tags are selected.	The data for the actuator tags is not present in this dataset. The conclusion of the analysis is that attention should be focused on why valves were closed.

4.3 Discussion

This root cause analysis demonstrated how combining information from different sources can be used effectively to analyse the process operation in a large-scale system.

Another strength of this approach was also illustrated. From step 3 onwards, there were no records of alarms because the main alarm log is overwritten periodically; still the integration of the process data and diagrams allowed the analysis to continue. Indeed, missing information is a real problem faced by analysts, and the integration of different sources can help by providing redundancy and robustness to the decision-making process.

The analysis involved mainly two types of tasks; one related to searching the data sources to visualise the selected tags, and the other using the connectivity information and process knowledge to analyse the data and alarms. The approach is still manual but the case study showed that it lends itself to automation; hence the next section details the functional requirements for this purpose.

5. FUNCTIONAL SPECIFICATIONS FOR SEMI-AUTOMATED TOOL

This section defines the functional requirements to partially automate the integrated analysis described in the previous section. One of the motivations for this tool is to liberate the expert from the tasks consisting of the interaction with the databases; these tasks are time-consuming and do not add value to the analysis. The other reason is to implement the structured approach of navigating the system towards the root-cause.

In brief, the proposed tool should perform certain operations on a selected envelope of tags and time. The outputs of these operations should be the efficient visualisation of trends, alarms and diagrams, and suggestions for the next envelope of analysis. The decision for this next envelope should remain with the user because it involves process knowledge; this way the user can use his insights to evaluate the suggestions given, and be confident with the final outcomes.
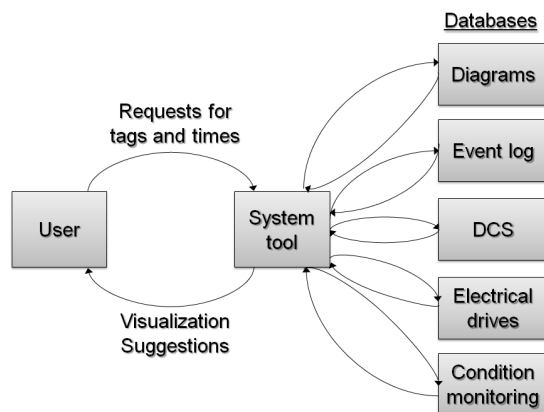

Fig. 1. Context diagram of the proposed tool representing its interactions with the main external entities.

Figure 1 shows a context diagram of the tool. A context diagram represents the relationships between the tool (in the centre) and the main external entities. As evidenced, the tool should interact with the various data sources to search for particular tags. Then, it has to interact with the user, who requests/decides the envelope of tags and times to analyse, and receives visualisation and suggestions.

The tool should be fit for two categories of users, defined by their experience with the plant and the process of analysis. The experienced user may only use level 1 functions, which mainly concern the interaction with the databases and visualisation; while the inexperienced user should also require level 2 functions, which analyse together the different sources of information in order to provide suggestions.

5.1 Level 1 functional requirements

On the whole, level 1 functions should facilitate the interaction between the user and the databases, by attaining and presenting the data and diagrams corresponding to the tags and times selected by the user. The efficient visualisation of data may require data processing functions. Specifically, the tool should:

- produce the diagram from the tag name and highlight the queried tag in the diagram;

- produce the diagram from part of its name;

- allow the selection of one or more tags in diagrams, to bring out the corresponding data, alarms or diagrams;

- produce time trends from the tags name and time instant;

- allow simultaneous visualisation of multiple time histories;

- offer different techniques of visualising time histories;

- produce a scrollable alarm list centred around a queried time instant;

- allow the selection of an alarm tag, to bring out the corresponding data or diagram.

5.2 Level 2 functional requirements

Broadly, level 2 functions combine the analysis of data and alarms with connectivity information and some rules based on first principles in order to suggest the most likely envelope to approach the root-cause. Specifically, the tool should:

- extract the plant connectivity from the diagrams, distinguishing the different types of connections, namely mass, heat, momentum and information;

- find the alarm tag(s) closest to a queried time instant and a queried tag, where for the latter evaluation the tool should use the extracted connectivity information;

- detect the time instant t when a tag changed from steady-state operation to unsteady from the current window of time history, to use the output (t, tag) to query the alarm list;

- extract the features related to cause-effect from a set of time histories using signal analysis to determine a data-based causality path within that set;

- compare the data-based causality path to the connectivity-based causality path to determine if there is a mismatch and, if so, between which two tags, named boundary tags;

- in case of mismatch, use the connectivity information to find the external influences (mass, heat, momentum and information) between the two boundary tags, and extract the name of the first tag in those other systems;

- in case of no mismatch, or if an alarm query gives no results, expand the set of analysed tags upstream and downstream of the current set (along mass streams).

6. CONCLUSIONS AND COMMENTS

This paper proposed the integration of different information sources for the analysis of process performance, in particular to address the challenge of large-scale systems.

The background section showed that until now the data used for the analysis has been limited to process measurements. The paper identified additional sources of information and compared their technical characteristics. The dissimilarities identified are challenges to the integration and represent directions for research.

In section 4, the proposed concept of integrated analysis was applied to diagnose an unplanned plant shut-down, using an alternative/original approach to large-scale processes. The process had over 7500 measurement tags and the analysis was able to focus attention on a particular area. This result supports the value of the integrated analysis approach to the area of process performance analysis.

Finally, the paper defined functional requirements to semi-automate the proposed integrated analysis approach. Some of the required functions belong to the domain of informatics, and are already trivial in commercial software, *e.g.* clicking on a name on a figure and producing some information related to that entity.

Conversely, the core requirements, directly concerned with the analysis of information, need to be addressed with new technical contributions and represent directions for research. This research is timely because industries are manifesting interest in having all plant data together, and thus it is more likely that the data from different systems will be available synchronised, in the same database. Furthermore, there are already some contributions in extracting connectivity from process diagrams (see background section), valuable for the integration of diagrams in the analysis.

ACKNOWLEGEMENTS

REFERENCES

Bauer, M., & Thornhill, N. F. (2008). A practical method for identifying the propagation path of plant-wide disturbances. *Journal of Process Control , 18* (7-8), 707-719.

Cecilio, I. M., Chen, S. L., & Thornhill, N. F. (2011). Importance of auxiliary systems for process fault detection and diagnosis. *IEEE 19th Mediterranean Conference on Control & Automation*, (pp. 952-957).

Chiang, L. H., & Braatz, R. D. (2003). Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemometrics and Intelligent Laboratory Systems , 65* (2), 159-178.

Choudhury, M. A., Kariwala, V., Thornhill, N. F., Douke, H., Shah, S. L., Takada, H., et al. (2007). Detection and diagnosis of plant-wide oscillations. *Canadian Journal of Chemical Engineering , 85* (2), 208-219.

Fiske, T. (2009). Aegis Analytical Delivers Process Intelligence from Disparate Data Sources. *ARC Advisory Group*.

Jiang, H., Patwardhan, R., & Shah, S. L. (2009). Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control , 19* (8), 1347-1354.

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2003). A Systematic Framework for the Development and Analysis of Signed Digraphs for Chemical Processes. 2. Control Loops and Flowsheet Analysis. *Industrial & Engineering Chemistry Research , 42* (20), 4811-4827.

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2005). Fault Diagnosis by Qualitative Trend Analysis of the Principal Components. *Chemical Engineering Research and Design , 83* (9), 1122-1132.

Qin, S. J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics , 17* (8-9), 480-502.

Qin, S. J., Valle, S., & Piovoso, M. J. (2001). On unifying multiblock analysis with application to decentralized process monitoring. *Journal of chemometrics , 15* (9), 715-742.

Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000). *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes* (1 ed.). Springer.

Thambirajah, J., Benabbas, L., Bauer, M., & Thornhill, N. F. (2009). Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history. *Computers & Chemical Engineering , 33* (2), 503-512.

Thornhill, N. F. (2007). Locating the source of a disturbance. In A. Ordys, D.Uduehi, M.Johnson (Eds.), *Process Control Performance Assessment: From Theory to Implementation* (pp. 199-225). London: Springer.

Thornhill, N. F., Shah, S. L., Huang, B., & Vishnubhotla, A. (2002). Spectral principal component analysis of dynamic process data. *Control Engineering Practice , 10* (8), 833-846.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering, 27* (3), 293-311.

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003b). A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering , 27* (3), 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003c). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering , 27* (3), 327-346.

Westerhuis, J., Kourti, T., & MacGregor, J. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of chemometrics , 12* (5), 301-321.

Wold, S., Kettaneh, N., & Tjessem, K. (1996). Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of chemometrics , 10* (5-6), 463-482.

Yim, S. Y., Ananthakumar, H. G., Benabbas, L., Horch, A., Drath, R., & Thornhill, N. F. (2006). Using process topology in plant-wide control loop performance assessment. *Computers & Chemical Engineering , 31* (2), 86-99.