Data-driven Development and Maintenance of Soft-Sensors

Conference Paper · January 2014

DOI: 10.1109/SAMI.2014.6822414

3 authors:

János Abonyi
University of Pannonia, Veszprém

Barbara Farsang
University of Pannonia, Veszprém

Tibor Kulcsár
University of Pannonia, Veszprém

Data-driven Development and Maintenance of Soft-Sensors

Janos Abonyi, Barbara Farsang and Tibor Kulcsar
Department of Process Engineering
University of Pannonia, Veszprem, Hungary, POB 158
janos@abonyilab.com, farsangb@fmt.uni-pannon.hu, kulcsart@fmt.uni-pannon.hu

***Abstract*—Product quality related process variables have significant role in advanced process control (APC). Online analyzers and software sensors can provide accurate and timely information for APC systems. In this paper we give an overview of data based soft-sensor development. We show that soft-sensor models of APC require maintenance and demonstrate that statistical quality control (SQC) techniques can be effectively used to automatize the related fault detection tasks.**

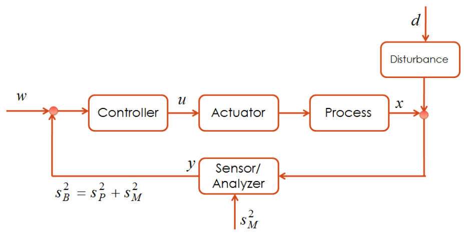

Fig. 1. Soft-sensors and online analysers in feedback control.

I. Introduction

Advanced process control systems should have the following functionalities to ensure and continuously improve product quality [1]:

- prediction of product quality from operating conditions,
- optimisation of operating conditions to improve product quality, and
- detection of faults or malfunctions for preventing undesirable operation.

These functionalities require timely and accurate information about process variables characterizing and influencing product quality. Offline laboratory tests mostly take more than two hours which time delay can cause control problems resulting economic loss. Online analysers have faster response time (1-4 minutes) and can eliminate the dependence on laboratory data by providing timely information for corrective action or real time control (see Fig. 1) [2]. However, due to the high instrumentation and maintenance costs and low reliability of online analyzers there is a need for an easily implementable, maintainable and robust alternative.

The infrasturcure of the research has been supported by the European Union and the Hungarian Republic through the project TAMOP-4.2.2.A-11/1/KONV-2012-0071. The contribution of the authors was supported in the frames of TAMOP 4.2.4. A/2-11/2012-0001 National Excellence Program- Elaborating and operating an inland student and researcher personal support system.

In order to enable continuous process monitoring and efficient process control in such cases, soft-sensors are usually used to estimate these difficult-to-measure process variables.

A soft-sensor or virtual sensor is a common name for software where several measurements are processed together. The interaction of the signals can be used for calculating new quantities that need not be measured. soft-sensors are especially useful in data fusion, where measurements of different characteristics and dynamics are combined. It can be used for fault diagnosis as well as control applications. Soft-sensors can estimate unmeasured, but important variables from other easily measured variables using computational models. The advantage of these inferential measurements compared to the lab measurement is that estimation happens in real time so information is continuously available to improve process control and optimization and increase process safety.

A very comprehensive review of the applications of soft-sensors in process industry can be found in [3], [4], and [5]. Soft-sensors are widely used in hydrocarbon industry. E.g. in a fluid catalytic cracking process soft-sensors are used to estimate catalyst circulation rate and heat of reaction to support model predictive control [6]. Soft-sensors are particularly widespread in pharmaceutical industry in

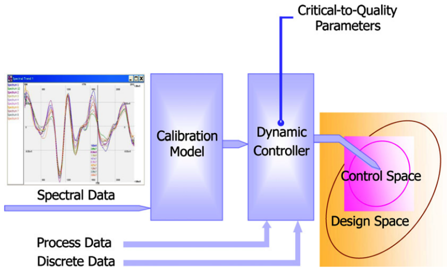

Fig. 2. Scheme of NIR based process control [9].

connection with Process Analytical Technology (PAT). A typical application is is shown in [7]. It is important to note that in PAT soft-sensors are used to generate on-line quality estimate based on on-line analytical measurements. An example for this concept can be seen in [8]. This PAT approach is important in NIR based process control. Fig. 2 shows an interesting application where the mapped spectral space is used to define the control space representing good product quality.

Application of soft-sensors and online analysers for (advanced) process control is not trivial. The book [10] deals with some key points of the soft sensors design procedure, starting from the necessary critical analysis of rough process data, to their performance analysis, and to topics related to on-line implementation. We give an overview of the intelligent techniques that can support the development and maintenance of data-driven models used in soft-sensors. We show that soft-sensor models of APC require maintenance and demonstrate that statistical quality control techniques can be effectively used to automatize the related fault detection tasks.

II. Modeling for soft-sensor development

According to the information used for soft-sensor development first-principle or data driven soft-sensors can be distinguished [11]. First principle models (also referred as white-box, mechanistic or *a priori* models) are based on balance equation (mass, component, energy) and contain detailed physical-chemical information about the system. Unfortunately, in many practical applications, processes are often too complex and uncertain, so detailed mechanisms of the system are not sufficiently well understood for detailed model development, so

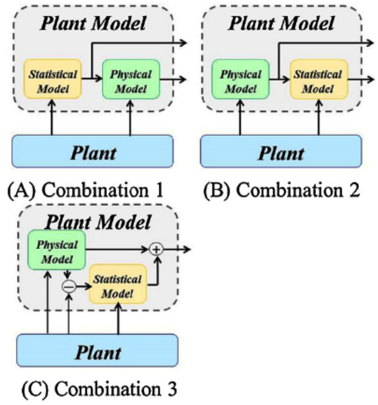

Fig. 3. Soft-sensor structures according to how property variables are estimated by the combination of white and black box models [13].

the applicability of this approach is very limited [5]. Furthermore, the development of first-principle models of complex industrial process is very difficult and time-consuming procedure. In particular, it is difficult to build precise first-principle models that can explain why defects appear in products. This is a critical issue since product life cycles are getting shorter and the time available for improving of product quality and yield requires fast and adaptive solutions [12].

Data driven (black-box or *a posteriori*) model based soft-sensors are built when no detailed knowledge is available about the process. In this case process data is used to build statistical models to determine the relationship between inputs and outputs. Statistical regression methods have become increasingly popular techniques for process modeling and they are used for fault detection and quality estimation.

Data and *a priori* model driven approaches can have several synergistic combinations (see Fig. 3).

- In the first case the statistical model is the input of physical model in the form of differential or algebraic equations or a complex flowsheeting simulator. In this case statistical model is used to estimate difficult to model parameters and phenomena.
- Combination 2 shows the case when outputs of physical model are transformed by a statistical model.
- In the third option the difference between measured and calculated variables are the

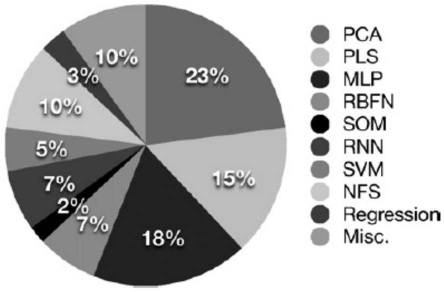

Fig. 4. Distribution of model types used in soft-sensors. PCA: Principal Component Analysis, PLS: Partial Least Squares regression, MLP: Multi Layer Perceptron - Neural Network, RBFN: Radial Basis Function Networks, SOM: Self-Organising Neural Models, SVM: Support Vector Machines [5].

	Methodology							
Process	Phys	MRA	PLS	O.L.	ANN	JIT	Gray	Total
Distillation	20	256	41	6	0	5	3	331
Reaction	5	32	43	0	0	5	1	86
Polymerization	0	4	8	0	3	0	5	20
Others	0	1	1	0	0	0	0	2
Total	25	293	93	6	3	10	9	439

Fig. 5. Numbers of known applications of soft-sensors. Phys: White box models, MRA: PCA and multi linear regression models, ANN: Neural Netwoks. [14]

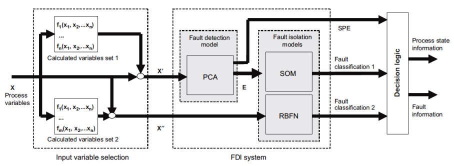

Fig. 6. Hybrid soft-sensor model as combination of PCA, SOM and RBFN models. PCA is used for fault detection while classifier models are used for fault isolation [15].

inputs of statistical model used for correction. Model based data reconciliation techniques are similar to this approach.

The most popular data-driven soft sensor models are the multivariate statistical techniques, i.e. the PCA and PLS, which together cover 38 % of the applications (see Fig. 4 and Fig. 5). Soft-sensors are also often used for decision support. In these applications appropriate classifier models have also be tuned. Fig. 6 shows an example for this model structure. It is interesting to note that this scheme is also an example for hybrid soft-sensor model, where different model types (PCA, SOM and RBFN) are combined.

Kalman filters are often used for model update because of its good performance characteristics and simple implementation. In this way all the process information can be included in the model leading to more precise predictions [5].

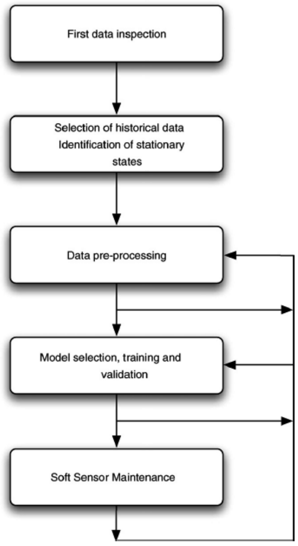

Fig. 7. Main steps of data-driven soft-sensor development [5].

Generally, building a high performance soft-sensor is very laborious task, since input variables and samples for model construction have to be selected carefully and parameters have to be tuned appropriately. In the following session we overview the related model development methodologies.

III. Data-driven Model Development

The general procedure of data-driven soft-sensor development is shown in Fig.7.

- *Data inspection* In the first step it is necessary to specify the most important unmeasured variables and how we can calculate them from measured data. We should get an overview of the data structure and define problems which may be occurring. We collect which types of soft-sensor can be used during the next steps: simple regression model or complex model (e.g. PCA) or more complex (e.g. neural network).
- *Selection of historical data and identification of stationary states* In the second step, data to be used for the training and evaluation of the model are selected. Next, the stationary segments of the data have to be identified and selected. The identification of the stationary process states is usually performed by manual annotation of the data, [1], but time-series segmentation

algorithms can also be used to automatize this procedure [16].

- *Data preprocessing* Since measured variables have got different order of magnitude data pre-processing includes normalization of data (e.g. in case of PCA it is inevitable). Usual steps are the handling of missing data, outlier detection, selection of relevant variables (i.e. feature selection and transformation) and data reconciliation. As shown in Figure 7 outlier removal and feature selection and transformation steps are repeatedly applied until the model developer considers the data as being ready to be used for the training and evaluation of the actual model [1].
- *Model selection, training and validation* Model selection, training and validation are critical steps in development of soft-sensors. Selection depends on the application problem, nature of data and personal preferences, since developers prefer models which are in their field of expertise. But it is accepted if it is possible use simple model and gradually increase model complexity as long as significant improvement in the model's performance can be observed.
- *Validation and testing* After finding the optimal model structure and training the model, the evaluation of the model performance is needed.
- *Maintanance* Even when a soft-sensor is developed successfully, its performance deteriorates when process characteristics changes. For example in chemical processes equipment characteristics can be changed by catalyst deactivation or adhesion. Soft-sensors should be updated to follow these changes. Manual and repeated construction of models should be avoided due to its heavy workload [14] .

IV. Sustaining APC benefits - Continuous Quality Verification

Not only soft-sensors but all the models of APC (advanced process control) solutions require maintenance. If nothing in the plant ever changes, then almost no maintenance and no model updates are required. However, when significant changes are made to the process, or if feedstock characteristics change significantly, then the APC models must be made ”aware” of these changes. When model updates are not performed, or when regular controller maintenance is not carried out due to significant process and instrumentation changes, the performance of the APC system starts to degrade [9]. More details about the reasons for performance degradation can be found in [17].

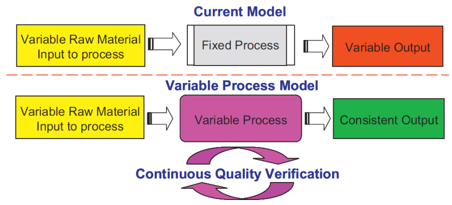

Fig. 8. Continuous development of APC solutions [9].

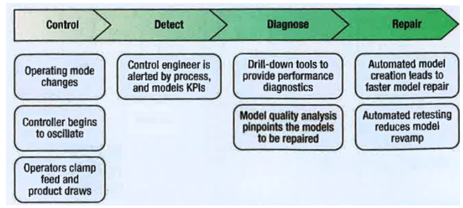

Fig. 9. Automated APC maintenance workflow [17].

The typical manual APC maintenance workflow is labor-intensive [17]. The control engineer must often manually extract process data to isolate the root cause. After the nature of the problem has been determined, manual model-building has to be performed. Usually this takes significant time to correct the problem and return the controller to full service. This approach is inefficient, , it is largely reactive and not proactive, and it causes a loss of benefits that can be as high as 50%-60% during the four- to five-year application life cycle. With supporting automation, this workflow can be significantly streamlined. and the time and effort needed to keep the controllers at peak efficiency can be reduced.

A proper APC maintenance methodology should have the following characteristic [17]:

- Minimizes effort by automating and simplifying maintenance tasks.
- Uses proper baselines, as well as key performance indicators (KPIs) covering both controllers and models.
- Uses automated reports to rapidly detect changes in performance.
- Employs diagnostic rules to isolate root causes of performance degradation.

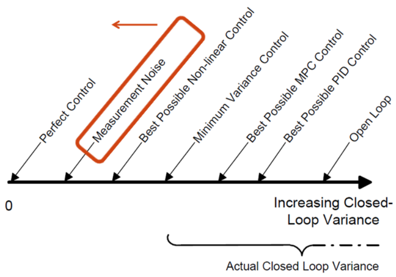

Fig. 10. Theoretical analysis of the variance of a control loop. The performance of the sensor and analyzer determines the available best control performance.

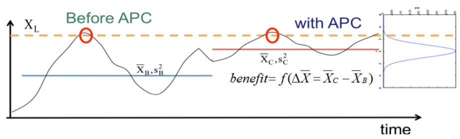

Fig. 11. Statistical Process Control focuses to the reduction of variance of the control loop.

- Uses automatic step testing to quickly generate high-quality data for improved model.
- Prepares data for modeling using preprocessing rules, establishes automated data-cleaning tasks, and minimizes the need to manually slice data.
- Automatically generates new models without extensive engineering effort.
- Proactive instead of reactive.

Tools supporting detection, diagnostics and repair of online analyzers are described in [17]. In the following session we will show that statistical quality control (SQC) techniques can also support automated soft-sensor maintenance.

V. SQC BASED PERFORMANCE MONITORING

In steady state operation the economic benefit is often increasing by shifting the steady state operation point closer to the constraints of the process Fig. 10. To reach this goal, the reduction in variance of the key process variables are necessary with e.g. re-tuning the controller or even re-design the existing control strategy. As Fig. 11 illustrates the variance of the sensors and analyzers has significant effect to the variance of the closed loop controller, so the performance of online analyzers has significant economic impact.

Statistical Quality Control (SQC) techniqes can be effectively used for the diagnosis of

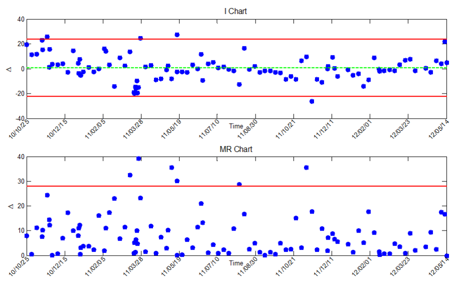

Fig. 12. I and MR charts show out of control samples of the gas chromatograph of an industrial distillation column.

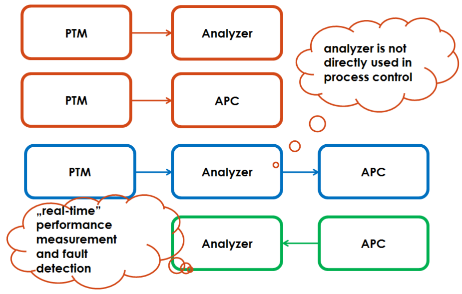

Fig. 13. The proposed concept of simultaneous and indirect validation of online analyzers and soft-sensors.

soft-sensors. The D 6299 ASTM standard (Applying Statistical Quality Assurance Techniques to Evaluate Analytical Measurement System Performance) provides guidelines for the design and operation SQC tools to monitor and control the stability, precision and bias performance of analytical measurement systems. The key idea of this procedure is to use quality control samples to establish and monitor the precision of the analytical measurement system. This sample is selected as a stable and homogeneous material having physical or chemical properties similar to those of typical samples tested by the analytical measurement system. Control charts and other statistical techniques are presented to screen, plot, and interpret test results to ascertain the in-statistical- control status of the measurement system. The ASTM standard D6122-10 (Standard Practice for Validation of the Performance of Multivariate Online, At- Line, and Laboratory Infrared Spectrophotometer Based Analyzer Systems) extends these techniques to online analyzers. As Fig. 12 shows, this technique also compares analyzer and primary test method results (PTM) by plotting their difference on control charts. When the difference is not within control limits, then the result for the analyzer is invalid.

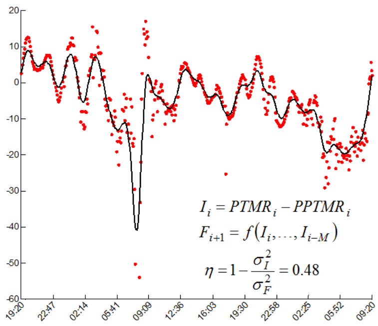

Fig. 14. A modified Harris Index used to measure the performance of a soft sensor of the studied distillation column.

We propose a methodology for indirect periodic continual validation tests. As Fig. 13 shows PTM based validation can be directly used to validate online analyzers and soft-sensor models of APC systems, which is a labor intensive and costly procedure. We suggest an indirect validation procedure, when the soft-sensor model is continuously validated based on the measurements of the online analyzer. The statistical analysis of the residuals can give information about the reliability of the soft-sensor and the online analyzer. For such purpose we propose a modified version of Harris index [18] suggested to assess controller performances by means of the so-called Minimum Variance Control (MVC) benchmark (see Fig. 14). As this measure compares the standard variation of the residuals of the online analzer and the soft-sensor to a theoretical value, it can directly used to monitor of the whole APC system.

VI. Conclusions

Soft-sensors are models used to calculate difficult to measure process variables to provide useful information for monitoring, control and optimization of industrial systems. We gave an overview of data-driven modeling approaches of soft-sensor design. To ensure that soft-sensor retains its precision, continuous maintenance (adaptation) mechanism has to be implemented. We proposed a statistical quality control based approach for this purpose.

References

[1] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Computers and Chemical Engineering*, vol. 32, no. 12, pp. 12 – 24, 2008.

[2] K. K. Gidwani and R. F. Beckman, "Evaluation of refinery control systems," *ISA Transactions*, vol. 33, no. 3, pp. 217 – 225, 1994.

[3] G. Gonzalez, J. P. Redard, R. Barrera, and M. Fernandez, "Issues in soft-sensor applications in industrial plants," in *Industrial Electronics, 1994. Symposium Proceedings, ISIE '94., 1994 IEEE International Symposium on*, 1994, pp. 380–385.

[4] G. Gonzalez, "Soft sensors for processing plants," in *Intelligent Processing and Manufacturing of Materials, 1999. IPMM '99. Proceedings of the Second International Conference on*, vol. 1, 1999, pp. 59–69 vol.1.

[5] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers and Chemical Engineering*, vol. 33, no. 4, pp. 795 – 814, 2009.

[6] S. Yang, X. Wang, C. M. (Fellow), and Q. Chen, "Soft sensor based predictive control of industrial fluid catalytic cracking processes," *Chemical Engineering Research and Design*, vol. 76, no. 4, pp. 499 – 508, 1998.

[7] C. Undey, S. Ertuni, T. Mistretta, and B. Looze, "Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control," *Journal of Process Control*, vol. 20, no. 9, pp. 1009 – 1018, 2010.

[8] C. Cimander, T. Bachinger, and C.-F. Mandenius, "Integration of distributed multi-analyzer monitoring and control in bioprocessing based on a real-time expert system," *Journal of Biotechnology*, vol. 103, no. 3, pp. 237 – 248, 2003.

[9] Z. Chen, D. Lovett, and J. Morris, "Process analytical technologies and real time process control a review of some spectroscopic issues and challenges," *Journal of Process Control*, vol. 21, no. 10, pp. 1467 – 1482, 2011.

[10] L. Fortuna, *Soft sensors for monitoring and control of industrial processes*. Springer, 2007.

[11] J. Abonyi, "Fuzzy model identification," in *Fuzzy Model Identification for Control*. Birkhäuser Boston, 2003.

[12] R. Grbic, D. Sliskovic, and P. Kadlec, "Adaptive soft sensor for online prediction based on moving window gaussian process regression," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2, 2012, pp. 428–433.

[13] M. Nakaya and X. Li, "On-line tracking simulator with a hybrid of physical and just-in-time models," *Journal of Process Control*, vol. 23, no. 2, pp. 171 – 178, 2013.

[14] M. Kano and M. Ogawa, "The state of the art in chemical process control in japan: Good practice and questionnaire survey," *Journal of Process Control*, vol. 20, no. 9, pp. 969 – 982, 2010.

[15] P. Kampjarvi, M. Sourander, T. Komulainen, N. Vatanski, M. Nikus, and S.-L. Jamsa-Jounelac, "Fault detection and isolation of an on-line analyzer for an ethylene cracking process," *Control Engineering Practice*, vol. 16, pp. pp. 1–13, 2008.

[16] J. Abonyi, B. Feil, S. Nemeth, and P. Arva, "Modified gath–geva clustering for fuzzy segmentation of multivariate time-series," *Fuzzy Sets and Systems*, vol. 149, no. 1, pp. 39–56, 2005.

[17] S. Lodolo, M. Harmse, A. Esposito, and A. Autuori, "Use adaptive modeling to revamp and maintain controllers," *Hydrocarbon Processing*, June 2012.

[18] T. Harris, C. Seppalaa, and L. Desborough, "A review of performance monitoring and assessment techniques for univariate and multivariate control systems," *Journal of Process Control*, vol. 9, no. 1, pp. 1 – 17, 1999.