PROCESS HISTORY BASED MULTIVARIATE STATISTICAL MODEL OF HEAT TREATMENT FURNACE FOR CONTROL OPTIMIZATION

Božidar Bratina[1], Nenad Muškinja[1], Boris Tovornik[1]

[1]University of Maribor, Faculty of Electrical Engineering and Computer Science, 2000 Maribor, Smetanova 17, Slovenia

bozidar.bratina@uni-mb.si (Božidar Bratina)

Abstract

Systems for supervisory control and data acquisition are being constantly upgraded to be able to satisfy production and customer needs. Lately, improvements are towards trend monitoring, fault detection, diagnosis and prognostics, where monitoring software includes advanced signal processing methods such as simple statistic analysis tools, sensor fusion, soft computing methods, etc. Since SCADA systems emerged many industrial processes have been collecting process data while monitoring the process operation. These process history databases are great source of information for data based modeling techniques such as multivariate statistical analysis. Furthermore, the model can be used for control optimization or fault detection and diagnosis tasks. In the paper some interesting modeling aspects are overviewed and tested by using process history data sets, multivariate statistical methods, neural networks, etc. By comparison of commercial SCADA system and Matlab we wish to test, if and how far modern SCADA systems are suitable for development and realization of data based process models and optimization tasks. Case study models used for testing are heat treatment furnace model, combustion chamber of biomass boiler, and tree tank system.

Keywords: modeling, neural networks, fault detection and isolation, prognostics, multivariate statistical analysis, SCADA.

Presenting Author's biography

Božidar Bratina, is currently working as researcher at the Faculty of Electrical Engineering and Computer Science, University of Maribor. In 2009 he finished PhD thesis from the field of automation and treatment of fault detection and diagnosis in industrial systems. Beside fault detection, isolation, diagnosis and prognosis, he is also interested in system modeling, identification and simulation, SCADA systems, intelligent buildings.

1 Introduction

Rapid development and functionality add-ons of SCADA systems causes many process operators headaches due to continuous upgrades, however gain gives them new ways for solving many however it also gains powerful tools for signal processing and data analysis. During development phase many advanced processing and system analysis tasks were addressed by special mathematical tools such as Matlab/ Simulink, Mathematica, etc, where results were transferred to operator's level (industrial software). During past decades the desire to overcome this problem resulted in development of tools directly for industrial systems. More and more developers integrate simple data analysis techniques into their software to keep advantage before competition on the market. Advanced methods are mostly developed at research institutes and faculties with appropriate knowledge and support. The development expanded many special areas of process monitoring, such as control optimization, fault detection and diagnosis, tolerant fault control and recovery, etc. However design engineers possess specific know-how and experience during years of practice which can serve as a base for development of specific control or operation tasks or improvements. The software enables them to use collected history process data to execute analysis tasks, control optimization tasks, etc.

As mentioned, a special case of process monitoring can be process fault detection, isolation and diagnosis. For successful operation of FDID systems, process must adequately modeled including fault scenarios and limitations of the system. Control loops contribute to rather difficult interpretation of faults, due to controller functionality, where it changes output in case of large or un-allowed deviations. A typical plant with control and diagnostic system presents Fig. 1.

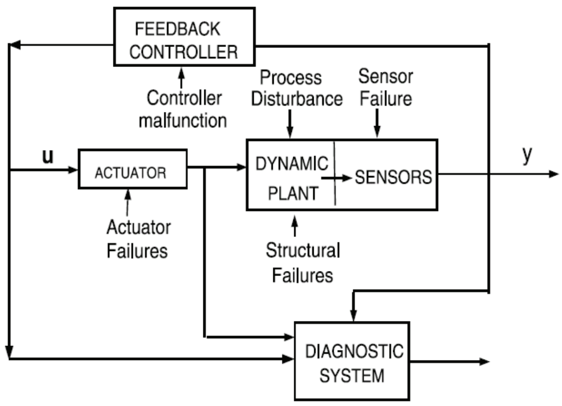

Fig. 1 Control scheme with diagnostic system

During past years the next logical step became systems for failure prognostics for the role of improved system condition based maintenance. These systems are also known as health management systems (HMS), they can be used for achieving longer and reliable operation, prediction of maintenance intervals, remaining useful life of components, needed reconfiguration, controller optimization, etc. The term prognosis is yet becoming popular as a discipline, and differentiates from fault detection and isolation objective since it can predict remaining time of failure to occur. From technical or production point of view any information that describes potential undesired process deviations is very important for operator to prevent un-necessary process down-time (reduced money loss, customer penalty, safety violation, reduced production plan).

Techniques for system modeling and later prediction can be achieved by many modeling techniques with its own advantages and disadvantages (transparency, complexity, implementation). In the paper data driven methods are discussed, since they can be used for many SCADA systems incorporating hidden process knowledge in history database for modeling. Most of them come from the field of artificial intelligence, statistics. In survey paper [1] many discussed algorithms are from model-based and data-based group of algorithms, similar to FDID concepts. Very popular are multivariate statistical methods, derivations of Monte Carlo method, support vector machine learning algorithms, Kalman filters, neural networks, fuzzy logic, etc.

In the paper various concepts of multivariate statistical methods are used for control optimization, precise fault detection and prediction of sensor degradation. For this task principal component analysis (PCA), its nonlinear extension and data reconciliation methods are used and realized in Matlab or commercial SCADA software. Case study scenarios are heat treatment furnace model, combustion chamber of biomass boiler and three-tank hydraulic model. In case of unpredicted process behavior, the monitoring system informs the operator about potential threats hence visual inspection of root causes is possible.

2 Multivariate statistics

By definition it encompasses simultaneous observation and analysis of more than one statistical variable. The application of multivariate statistics is multivariate analysis. Many methods exist: Multivariate analysis of variance, Principal components analysis (PCA), Factor analysis, Canonical correlation analysis, Linear discriminant analysis, Artificial neural networks extend regression methods to non-linear multivariate models, Clustering systems, etc.

2.1 Classic PCA

Principal component analysis is a mathematical transformation of a number of possibly correlated variables into a smaller number of uncorrelated variables. These are called principal components. The first principal component presents the highest variability in the data samples, and each succeeding component the remaining variability in the data sample. In practice it is mostly used as a tool in exploratory data analysis and for making predictive models. Its operation can reveal the internal structure of the data in a way which best explains the variance in the data, and by visualization in a high-dimensional data space (1 axis per variable), PCA generates a lower-dimensional picture with most informative viewpoint [2].

Multivariate statistical methods proved to be easy to implement and satisfactory for basic industrial tasks, therefore by modern SCADA systems with implemented data acquisition services, statistical model of the process can be obtained upon large process history datasets. The linear PCA method does not require much of a processing power therefore has been widely used in many areas: image compression, fault detection, dimensionality reduction of data (gene expression, medicine), etc. It can handle high dimensional and correlated process variables, provides a natural solution to the errors-in-variables problem and includes disturbance decoupling. However, main drawback lies in its linearity therefore a lot of research was invested to nonlinear extension.

2.2 Euler angles and PCA

To be able to detect un-predicted process deviations statistical measures e.g. Hottteling T2 or Q-norm are used to define residual bounds for detection of nominal point deviation. However complementary to such distance-based measure a quantitative analysis can be used to visually monitor and predict degraded process performance. For this task the Euclidean concept of distance can be very useful [3] when considering angles in 2 or 3 dimensional space. Building on distance between points, the Euclidian angle between points a and b with vertex at the origin, can similarly be defined for higher dimensions using vector products:

$$\cos\left(\theta_E\right) = \frac{(a' \cdot b)}{\left(\|a\| \cdot \|b\|\right)} \tag{1}$$

The angle definition is adjusted as weighted distance and the Mahalanobis angle between a and b through the origin can be defined:

$$\cos\left(\theta_M\right) = \frac{(a' \cdot D^{-1} \cdot b)}{\left(d\left(a,0\right), d\left(b,0\right)\right)} \tag{2}$$

By using the Mahalanobis distance for points a and b:

$$d\left(a,b\right) = \sqrt{\left(a-b\right)' \cdot D^{-1} \cdot \left(a-b\right)} \tag{3}$$

where D is a dispersion. A constant Mahalanobis angle around the line joining point a with the origin is a hyperconical surface, with distortion D. In such way a simple interpretation of the angle is possible through PCA. Rescaling the scores of PCA so each has equal variance is done by the Mahalanobis distance measure, distorting the ellipsoid described by the scatter of data observations into a sphere. Fig. 2 shows first three principal components presented in three dimensional space, where position (centre) and direction of PCs are changed according to different operation regimes of the process. In case of batch process monitoring such visual inspection can help predicts batch quality by using only small portion of measurements (Fig. 3). Decision regarding batch rejection is possible even before the batch is finished.

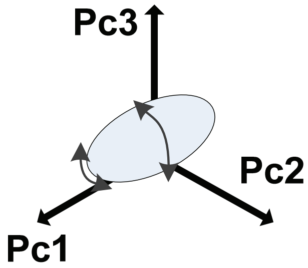

Fig. 2 Different PCA models for different process operating regimes.

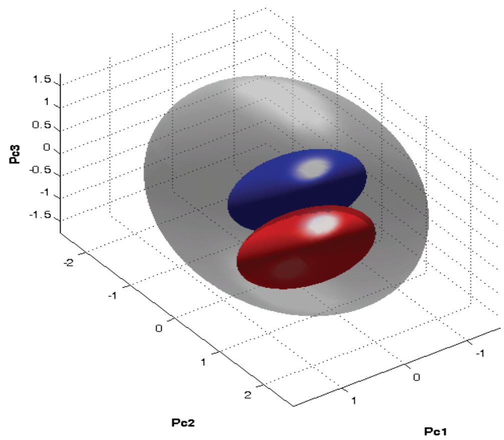

Fig. 3 Prediction of batch quality using small part of measurement data.

2.3 Nonlinear PCA

PCA enables quick and rough results. To improve statistical model of the process a nonlinear extension of PCA model is used. NLPCA can be achieved by using different techniques: neural networks, fuzzy logic, genetic algorithms, etc. Most applications is realized by neural networks, which enable extraction of nonlinear principal components that can be used for monitoring of process deviation.

In 1991, Kramer [4] presented nonlinear PCA method by using auto-associative neural network (AANN). Kramer presented a feed-forward neural network to perform identity mappings, where network inputs are reproduced at the output layer. Kramer's NLPCA is a classic PCA generalization, however the difference between NLPCA and PCA is that NLPCA allows nonlinear mappings. To perform NLPCA, the neural network contains three hidden neuron layers and input/output layer. Properly designed AANN is capable of capturing nonlinear characteristics therefore a good data-driven model can be obtained.

Nonlinear model can be used to observe small and precise process deviation or control issues. By extracting and observing changes of the curve, possible system component degradation or failure can be predicted.

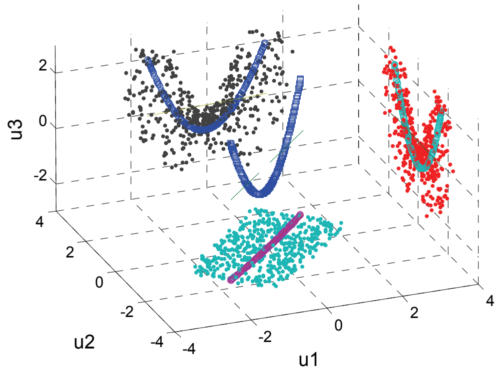

Fig. 4 Extracted nonlinear principal component

2.4 Virtual sensor

Another interesting application of AANN were introduced by Hines et al. [5] where instead of just observing changes in the process by nonlinear components, a virtual sensor can be designed to reconstruct sensor data measurement from statistical or neural network model. By using his data reconciliation scheme each system component is modelled upon history data of the process and compared to the real measured value (Fig. 5). Deviation or the trend of deviation can be analyzed to predict time to fault/failure of the system component or system operation.

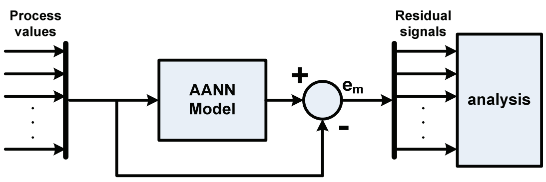

Fig. 5 Data reconciliation scheme (model and system output comparison).

In case of highly dynamic processes (measurements) a dynamic or even recurrent neural network structure is suggested.

3 Case scenarios

Process industry demands reliable operation so many times implementation of various algorithms is a difficult task. However the trend of modern SCADA platforms is moving towards implementation of advanced tools for development of various algorithms. Today these platforms usually include basic statistical methods with simple pre-processing algorithms to obtain system health information or small insight into the process behaviour. Presented multivariate statistical methods are tested on different simulation and laboratory models (heat treatment furnace temperature, combustion chamber temperature, three-tank hydraulic model level control) and in different development software (Matlab/Simulink and GE FANUC Intelligent Platforms solutions – Proficy Troubleshooter).

3.1 Heat treatment furnace

Furnaces are widely used in industry for heat treatment of material, products, etc. The dynamics of the temperature control of furnaces are dominated by heat transfer process. Modeling can be based on physical principles or identification procedure. For the task of PCA realization, a gas-fired industrial furnace [6] was used where zone temperature deviations were inspected.

The discrete state space model is defined:

$$\begin{aligned} x(k+1) &= A^0 x(k) + Bu(k) + A^1 u_1(k) x(k) \\ y(k) &= Cx(k) \end{aligned} \tag{4}$$

$$u = \begin{bmatrix} v & 1 \end{bmatrix}^T$$

$$A^0 = \begin{bmatrix} 0 & -0.6955 & 0 & 0.2402 \\ 1 & 1.6381 & 0 & -0.1315 \\ 0 & 0.3250 & 1 & -0.6081 \\ 0 & -0.2285 & 1 & 1.5588 \end{bmatrix}$$

$$A^1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.0013 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.0012 \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} 0 & 0 \\ 1.4563 & -56.864 \\ 0 & 0 \\ 1.4004 & -53.580 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The model of furnace has two temperature zones, where only small deviations upon model changes were observed. The discrete state space model was realized in Matlab/Simulink and PCA model with appropriate visualization was constructed in Proficy Troubleshooter [7] upon exported data values (measurements) from Matlab (Fig. 6).

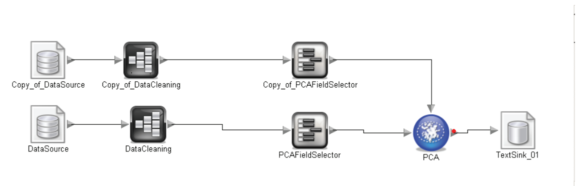

Fig. 6 PCA model and data pre-processing in Proficy Troubleshooter.

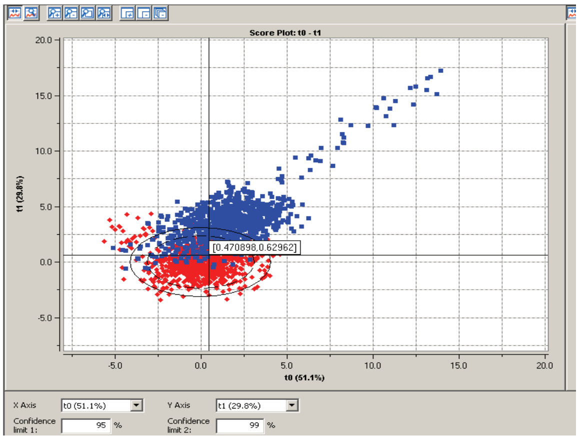

Fig. 7 PCA model visualization in Proficy Troubleshooter; normal (red) and with temperature drift (blue) operation.

3.2 Combustion chamber temperature

The model of combustion chamber temperature process [8] has four input and two output variables as depicted in Fig. 8. Process values are stored in SCADA history datasets which were used to model the process.

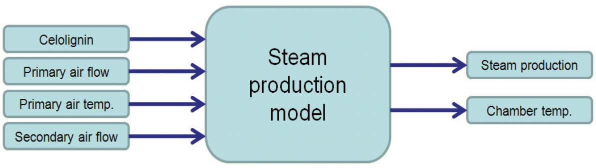

Fig.8 Steam production model with corresponding inputs and outputs

In order to obtain a dynamical model of the process NLPCA model was developed. A special type of neural network can be used called an auto-associative neural network (AANN) which is a feed-forward neural network that performs identity mappings (Fig. 9); network inputs are reproduced at the output layer.

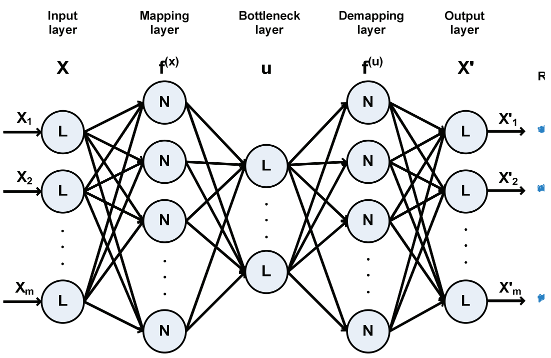

Fig.9 The structure of Auto-associative ANN

Several training procedures and neural network parameters settings were tested to obtain desired results. The sampling time used was 5s, due to reduced computational complexity of training and a large number of data samples from process measurements (fairly slow process changes). After successful training, the AANN was defined by 15 neurons in mapping and de-mapping layer, bottleneck layer had 5 neurons, and input/output layers were measured values of the process. 70% of measured data samples were used for modeling and 30% were used for validation.

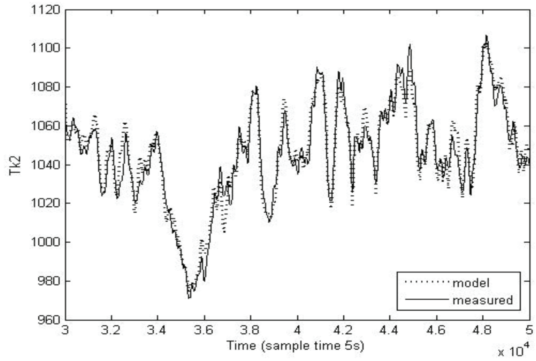

Fig.10 Combustion temperature – model and process output comparison

Steady state model obtained by AANN (NLPCA) structure is very accurate for a wide range of data, however it doesn't describe dynamics of the process adequately as variables has changeable delays, sampling times, dynamics, etc. The number of neurons in the input and output layers is determined by the structure of the process model. Comparison between neural model output and measurement data are shown in Fig. 10.

3.3 Batch level control

Process output quality is monitored through the whole process stages. Visualization and prediction of quality are very important in time delay systems. Imagine a batch in a pharmacy plant where the growing process of test cells takes a few months. Through the whole process cells have to be maintained in a certain environment conditions (temperature, pressure) to comply to world regulations and standards. By using accurate models and implementing prognostic methods with online batch monitoring, undesired process deviations can be monitored and analyzed in-line before the batch is finished.

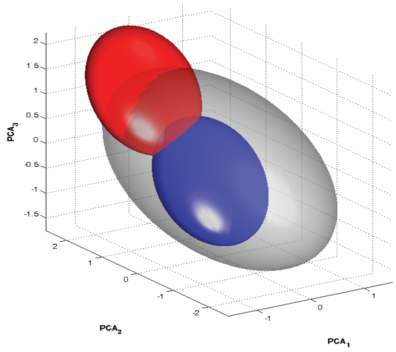

Fig. 11 PCA batch process prediction (grey-normal; blue-acceptable; red-rejected)

In case of unpredicted behaviour a prediction potential scenarios can be analyzed, or time to solve the faults can determined before the batch will have to be rejected, etc. Similar scenario can be introduced also to other industrial processes. The presented concept is tested on laboratory three-tank hydraulic model, where main task is level control, and undesired small changes around working point are observed. In case of any faults the batch quality changes over time.

In the laboratory three-tank model an actuator fault (pump malfunction) was introduced to the system for certain process operation regime. In Fig. 11 acceptable (blue) and rejected (red) sets of data measurements (process operation regime) are shown, where in case of pump malfunction the red data measurements reflect undesired or unsatisfactory output quality which demands immediate operator reaction.

3.4 Virtual sensor

Any sensor degradation (due to aging) can bring the system into unstable operation, and degradation is usually hard to detect due to control loop effect. For the purpose of sensor degradation detection on three-tank laboratory model, a nonlinear principal components model was realized in Matlab/Simulink as shown in Fig. 12. By using data reconciliation scheme an artificial degradation of level sensor is shown (Fig. 13). Scheme was tested for very small degradation (2-4% of measured signal).

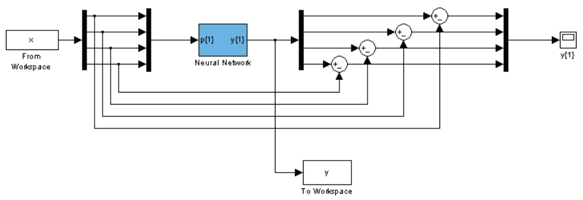

Fig. 12 Matlab scheme of process sensor data reconciliation

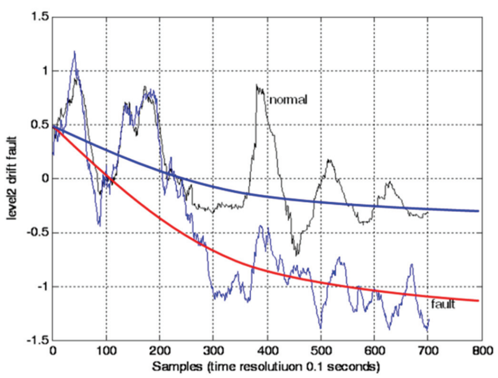

Fig. 13 Level sensor degradation detection.

4 Conclusion

In the paper some of the multivariate statistical methods and their practical use are presented to achieve visual information and analysis of potential process deviations. Presented concept were tested on simulated model of heat treatment data and furthermore realized in commercial GE Proficy Troubleshooter environment. The test showed that industrial software is developed so far, it can help operators to design their own simple signal processing operations while plants are running online. Nevertheless, advanced methods are still in the domain of special software, such as scheme with nonlinear principal components and virtual sensors, which were realized in Matlab.

The development of monitoring systems will soon enable quality advanced tools such as prognostics, optimization and energy management tasks, where artificial intelligence and soft computing methods can offer great results, especially if combined into hybrid platforms. Our research will continue in direction of advanced algorithms for prognostics that can be easily implemented into commercial process industry equipment or software that enables advanced mathematical computations to achieve desired results.

5 References

[1] M. Schwabacher, K. Goebel. A Survey of Artificial Intelligence for Prognostics. *The Intelligence Report: AAAI 2007 Fall Symposium*, NASA, 2007

[2] I.T. Jolliffe. Principal Component Analysis. Second Edition. Springer, New York. 2004.

[3] A. Raich, A. Cinar. Diagnosis of process disturbances by statistical distance and angle measures, *Computer Chemical Engineering*, 21:661-673, 1995.

[4] M.A. Kramer. Nonlinear principal component analysis using auto-associative neural networks. *AIChE Journal*, 37:233-243, 1991.

[5] J.W. Hines, R.E. Uhrig, D.J. Wrest. Use of auto-associative neural networks for signal validation. *Journal of Intelligent and Robotic Systems*, 21:143-154, 1998.

[6] L. Ding-Li. Diagnosing simulated faults for an industrial furnace based on bilinear model. *IEEE transactions on control systems technology*, 8:435-442.

[7] GE FANUC Proficy Troubleshooter (http://www.ge-ip.com/products/3391)

[8] B. Bratina, N. Muškinja, B. Tovornik. Recurrent auto-associative artificial neural network model of biomass steam boiler system. *Preprints of IFAC Workshop on Programmable Devices and Embedded Systems.* IFAC, 2009:214-219.