

An Identification Tool for Uncertainty Estimation of Process Measurements

Riku-Pekka Nikula, Esko Juuso, Kauko Leiviskä

University of Oulu, Control Engineering Laboratory, forename.surname@oulu.fi

KEYWORDS identification, modelling, sensor validation, soft sensor

ABSTRACT

This paper presents a tool, which is used in the identification of models for the estimation of the uncertainty of measurements in an on-line process monitoring application. The tool has three separate parts. The first part is used to identify simple linear regression models and limits for the measurements. The second part is used in the generation of inequality equations between the variables. The third part is used to identify soft sensors which have the basis in physical relationships between the variables. To estimate the overall uncertainty of a measurement, the models from the tool need to be used in conjunction with each other and together with other methods. The identification tool is built in the Matlab platform and the necessary historical data, which is needed for identification, is fetched from a database using SQL. The performance of the tool is demonstrated with an industrial data set from a combined heat and power plant. The main benefit of the tool is the fast implementation of the identified models into the database and for the monitoring purposes. The use of the tool presumably simplifies and speeds up the implementation of the models into the on-line application compared with manual implementations without the tool.

1 INTRODUCTION

Industrial plants have numerous measurements, which provide data that is used for operational, reporting and analysis purposes. Critical parts such as control loops and performance monitoring continuously use real time information from the process. To maintain a high level of economic or environmental efficiency and safety, it is essential that this information is correct, reliable and up-to-date. The measurement uncertainty is caused by internal and external reasons. Internal uncertainty originates from the measurement method, parts and materials, cabling and mounting of the sensing device, and I/O-card. Therefore, a single measurement has several possible sources of errors. The increased wear of parts, the corrosion of equipment, the fouling of sensors, calibration difficulties and other challenges that originate from the harsh industrial environments are examples of external reasons for uncertainty. The uncertainty can be decreased by the introduction of high quality sensing systems with regular maintenance and through advanced monitoring of the measurements themselves using computational methods. This study concentrates on the latter one.

Sensor validation, consisting of the tasks of sensor failure detection, isolation, and accommodation, has its basis in the redundancy of several sensors or in a sensor's own health information /3/. The redundancy approaches can be coarsely divided into physical and analytical approaches. Physical redundancy is often used in safety critical areas such as aviation and it is based on the utilization of redundant sensors measuring the same parameter of the system /3/. This kind of redundancy is best supplied as a part of the original system installation, because the retrofit work could require system downtime. Analytical redundancy approaches are model-based validation methods. The models are equations derived from the first principles /2/, /20/ or empirically derived mathematical relations estimated using data-driven system identification approaches /12/, /7/, /17/. The analytical redundancy approaches are commonly regarded as soft computing approaches or soft sensors /9/. Redundancy can also be approached with knowledge-based methods /13/, where problems are solved with qualitative rule-based models. Temporal redundancy of a sensor value is obtained through repetitive measurements of the sensor values at regular intervals /11/. The information originating from the redundant sensors or the derived models is compared with the original sensor values and the congruence is then analyzed to clarify the differences. The actual sensor faults can be classified into bias, drifting, precision degradation and complete failure /16/. The sensor faults have to be distinguished from process changes which can also be abnormal and influence the measurements. Several approaches to process monitoring and process fault detection have been proposed in the application area of soft sensors /9/.

In recent years, many research projects have developed software applications for the analysis and monitoring of process data and fault detection also in power plant environments. The general goal of such approaches is to improve the overall reliability of the process, and therefore, they themselves must be very reliable. The computational effort must be reasonable and the used methods should be understandable and tunable. An unsupervised neural network algorithm, called Self-Organizing Map (SOM), has been a popular choice in

software applications on purposes such as process state monitoring and optimization /5/, process fault detection /8/ and a knowledge-based decision support system for failure management /19/. Neural networks and other data mining algorithms such as principal component analysis, regression, classifiers and decision trees have been also used in applications for identifying sensor faults and reconstructing the erroneous measurements /10/, /1/. Challenges on such purely data-driven black box models are, for example, the lack of transparency in the models and the model robustness outside the range of the data used for training /18/. Hence, the inclusion of first principles models into approaches aiming at the reduction of uncertainty in measurements is reasonable. At the moment, there is an ongoing project for the development of an on-line data reconciliation application, which has been introduced in /6/. This paper presents three methods that can be used in the uncertainty estimation of process measurements in the on-line reconciliation application. An identification tool using the methods in order to identify models to be used in the uncertainty estimation and to implement the models into a database is developed. Data from the Salmisaari CHP plant of Helsinki Energy is used to demonstrate the software tool.

2 METHODS AND MATERIALS

The uncertainty estimation consists of several parts in the on-line reconciliation application. The parts include gross error detection, check for dead measurements, soft sensors, inequality equations, and the correlation of measurements with equation inconsistency etc. The overall uncertainty is estimated by the combination of these parts. This study presents a tool which takes the historical data of process measurements as inputs and defines models to be used as a part of the uncertainty estimation of the measurements. The software application is under constant development and it consists of three main parts:

- definition and evaluation of relative control limits based on the simple linear regression models,
- definition of inequality equations and evaluation of their correctness on the historical data,
- definition and evaluation of soft sensors based on the linear, quadratic, and cubic relationships of the process variables.

2.1 Relative control limits

Relative control limits are defined primarily to detect exceptionally high and low values in measurements in the varying states of operation. The exceptional values originate from abnormal process conditions, from unusual measurement values or from unusual behaviour of the sensing systems. Such a value indicates an increased uncertainty of the measurement. The method uses historical data from the period of normal operation in the process. The early version of the method was proposed in /15/. The approach can be used on the variables with strong linear relationships and similar dynamical behaviour. For each measurement, other measurements are sorted in the order of the magnitude of linear correlations. A certain limit, for instance correlation coefficient $R_{xy} > 0.8$, can be used to define the significant relationship. Simple linear regression models for a process variable are automatically defined in the least squares sense using linearly significant variables (key variables) one by one as the input variable. The original measurement values are divided by the estimated values from the model chosen by the user to get ratios. The distribution of the resulting ratios is estimated and used to define the relative limits. Because the used data is assumed to represent the normal operation, the limits are taken from the ends of the distribution (0.5th and 99.5th percentiles). The distances of these percentiles from the assumed median ratio (=1) are computed and multiplied with specific coefficients to define the scaled values in the range [-5 5]. The range is linear between the consecutive round numbers. The mentioned percentiles correspond to scaled values ± 2 . Currently, the scaled values ± 4 correspond to an alarm. During monitoring, the ratio of a measurement is scaled linearly into the defined range and variations in the range [-2 2] are considered normal.

2.2 Inequality equations

The inequality equations are based on the physical dependences between measurements. For instance, water temperature measurement value should be larger after a heat exchanger compared with the value measured before the heat exchanger if heating is the goal. An inequality equation of size $n=4$ measurements, $x_1 > x_2 > x_3 > x_4$, for example, is presented as a matrix,

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

where 1 corresponds to the larger side of inequality and -1 to the smaller side. The rows correspond to inequality symbols; the columns correspond to the measurements. Each row of the matrix is summed in such a way that all the possible inequality combinations between the measurements have been generated. In the presented matrix, this results in four new rows (equations). The amount of correct equations from all the equations of a particular measurement is checked to monitor the uncertainty of the measurement. The software tool includes the implementation of the equations in the database and the possibility for monitoring the historical data. Tolerances of the measurements can be attached in the equations.

2.3 Soft sensors with physical dependences

Linear, quadratic, and cubic dependences are expected to exist between the process variables y and x . An estimate for measurement y is computed using

$$\hat{y} = A \cdot x^C + B. \quad (1)$$

A simplex search method originally proposed by Nelder and Mead in 1965 /14/ is used to define the parameters A and B in the models. The method minimizes the sum of the model estimate deviation from the measured values on the chosen period. The parameter C is 1, 2, 3, $\frac{1}{2}$, or $\frac{1}{3}$. The tool includes also models with two inputs. The approach is based on approximate relationships termed affinity laws /4/ which are generally associated with dynamic pumps and fans. However, these models are not further discussed herein due to the lack of available data for publication.

3 RESULTS

The use of the tool is demonstrated with measurement data from the CHP plant of Helsinki Energy in Salmisaari. Data from January to February 2012 is used in the relative control limit tests. Data from December 2012 to January 2013 is used in other tests. One hour average data from the normal operation of the plant is used. The data for inequality equations is manually modified to test the performance. Otherwise, data is not modified. Only the computational results of the methods are shown. The operational principles of the user interface are mainly put aside.

Fig. 1 demonstrates the identification of the relative limits for the axial flue gas fan input power 1NR01E901:av. The blade angle of the fan 1NR01S104:av is chosen to be the input variable for the linear regression model. The limits were defined with data from a period of 26 days. Ratios falling into 1 ± 0.15 were taken into account, which is shown in Fig. 1. The performance was tested on data from a period of 10 days. The results on the right upper corner indicate that the model performs very well on the period. The percentiles (0.5, 50, and 99.5) that are used in the determination of the limits are shown in the right lower corner of Fig. 1. Values of the blade angle (%) are presented on X-axis and the input power on Y-axis (kW).

Fig. 2 demonstrates the performance of the inequality equation monitoring approach. District heating line including three heat exchangers is monitored by consecutive temperature measurements. Measurement 1UN00T003:av is located before the first heat exchanger. Measurement 1UN00T005:av is located before the second heat exchanger. Measurements 1UN00T008:av and 1UN00T009:av are two physically redundant measurements located before the third heat exchanger. The supply water is typically heated on each heat exchanger and the measurement values increase in the flow direction. Therefore, a logical equation is $1UN00T008:av > 1UN00T009:av > 1UN00T005:av > 1UN00T003:av$. Two redundant measurements should measure values of the same magnitude. Therefore, tolerances of $\pm 2^\circ\text{C}$ were included in the computations for each measurement. The value of 1UN00T008:av was artificially decreased in six consecutive 50-hour periods in the order 0, 2, 10, 30, 60 and 60 $^\circ\text{C}$. The smallest temperature measurement 1UN00T003:av was increased 35 $^\circ\text{C}$ at the end of the test period. The changes are shown on the left side of Fig. 2 and the uncertainty estimates for each measurement are shown on the right side. The uncertainty of 1UN00T009:av, for example, seems to increase on the period although the measurement values are correct. The reason is the increasing amount of incorrect equations the measurement belongs to. From the point 151 onwards the uncertainty of 1UN00T008:av becomes the largest. When 1UN00T003:av is artificially altered (point 251), its uncertainty is the same as with the measurement 1UN00T005:av although the latter has the correct measurement value.

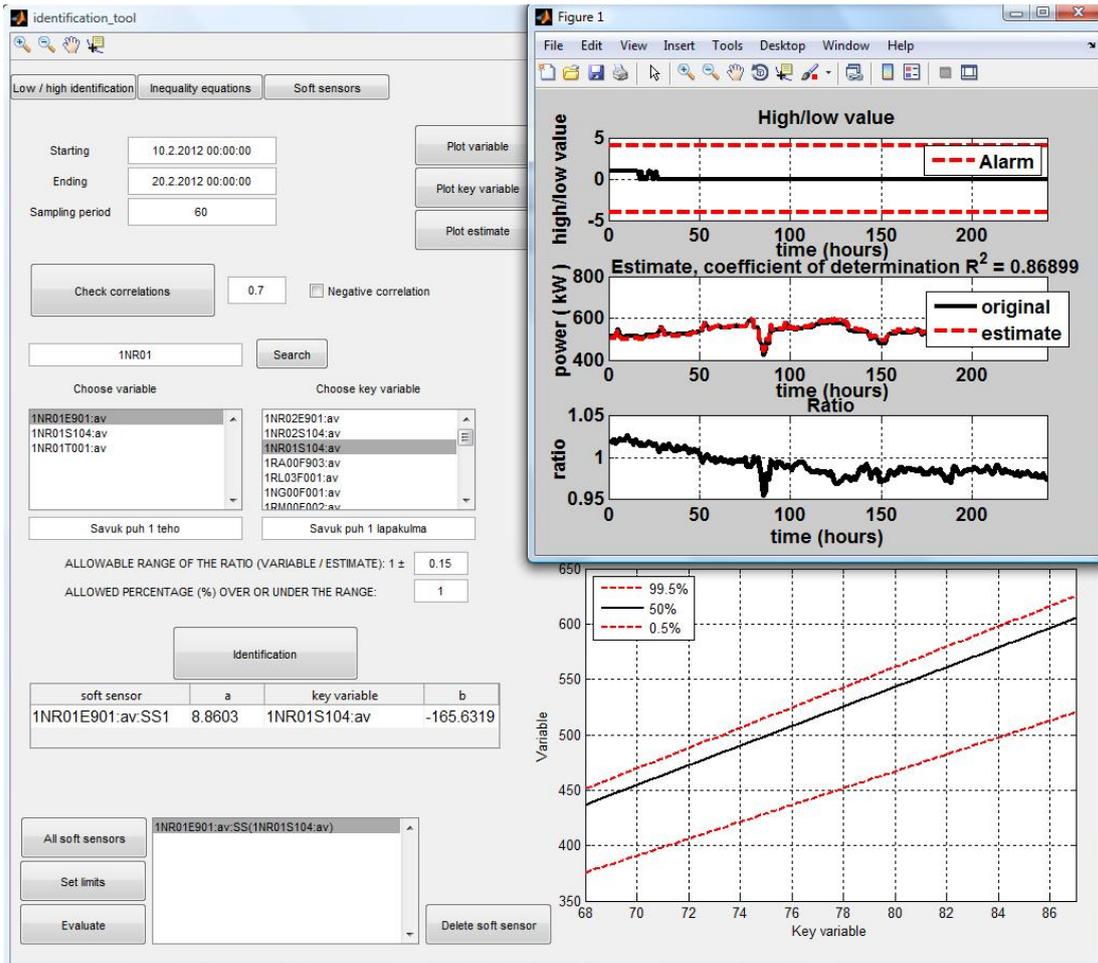


Fig. 1 Definition of the relative limits and monitoring of the ratio with the interface.

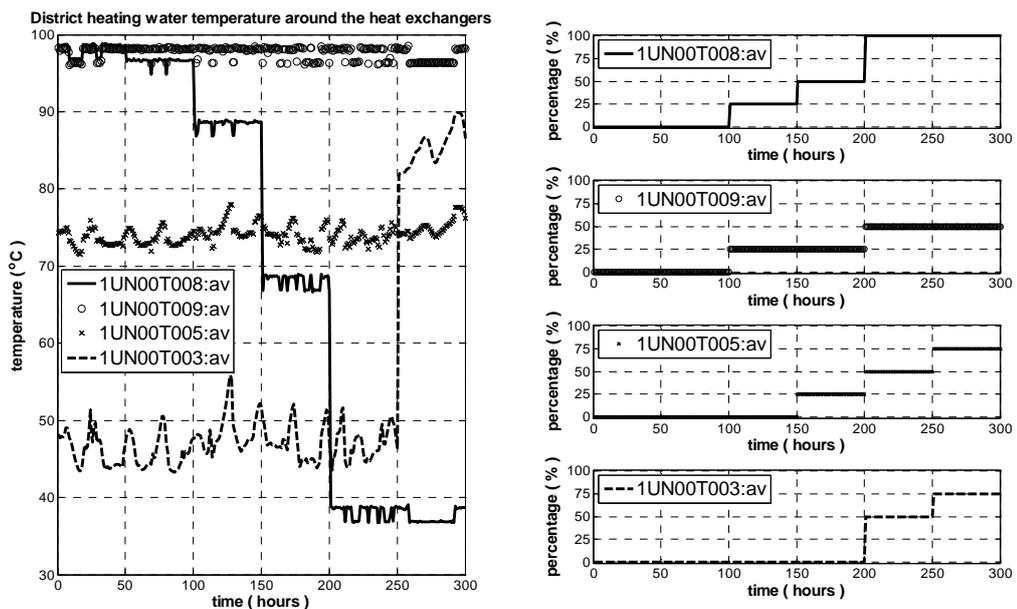


Fig. 2 District heating water temperature measurements with artificial errors (left), and the corresponding percentages of uncertainty based on the inequality equations (right).

Fig. 3 shows the Salmisaari desulphurization plant fan blade angle model performance using (1) with the power of the axial fan as the input. After testing all the possible values for C , the lowest root mean squared error (RMSE) and the highest coefficient of determination (R^2) were reached using $C = \frac{1}{3}$ (training: RMSE = 0.6237, $R^2=0.9958$; testing: RMSE = 1.5088, $R^2 = 0.9321$). The estimated parameters were $A = 11.5938$, and $B = -66.9673$. Right side of Fig. 3 shows that the estimated angle is generally slightly larger than the measured angle during the test period. The model with $C = 3$ performed the worst (testing: RMSE = 4.1818, $R^2=0.5322$).

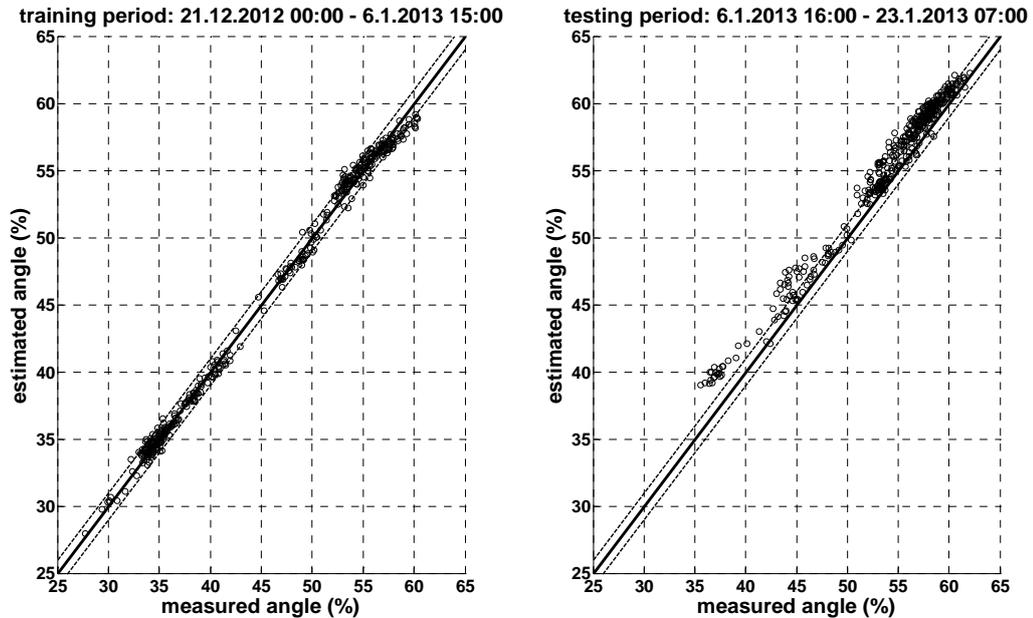


Fig. 3 The measured values of the desulphurization plant flue gas fan blade angle have been plotted on the X-axis and the estimated angles on the Y-axis. In a perfectly performing model, the circles in the figure should lie on the diagonal line. Two outer lines are 1 % away from the theoretical optimum.

4 CONCLUSIONS

The identification tool is used to identify models for the estimation of measurement uncertainty and to implement the models into a database from which an on-line application gets the requisite information for real time monitoring. The tool can be used to analyze the performance of the identified models on historical process data. The different approaches to uncertainty estimation are used in conjunction with each other. The approaches are not suitable for being used alone, which can be seen on the results of inequality equations, for instance. Systems with non-linear dynamics and without adequate amount of redundant variables are the possible challenges for the presented methods and the tool. However, the software can be updated with the mathematical representations of a system if such representations are available. Such models can be used as the basis for the soft sensors. The identification of soft sensors with realistic cases and the incorporation of the process state and performance analysis tools into the software are relevant topics for further development.

ACKNOWLEDGEMENTS

The authors would like to thank the Finnish Funding Agency for Technology and Innovation (TEKES) for funding the Measurement, Monitoring and Environmental Assessment (MMEA) programme managed by Cleen Ltd. Indmeas and Helsinki Energy are acknowledged for their contribution.

REFERENCES

1. Athanapoulou C., Chatziathanasiou V.: Intelligent system for identification and replacement of faulty sensor measurements in thermal power plants (IPPAMAS: Part 1). *Expert Systems with Applications*, 36(2009)5, 8750–8757.

2. Deckert J. C., Desai M. N., Deyst J. J., Willsky A. S.: F-8 DFBW sensor failure identification using analytical redundancy. *IEEE Transaction on Automatic Control*, AC-22(1977)5, 795–803.
3. Goebel K., Yan W.: Correcting sensor drift and intermittency faults with data fusion and automated learning. *IEEE Systems Journal*, 2(2008)2, 189–197.
4. Hall S.: *Rules of Thumb for Chemical Engineers*, 5th Edition, Butterworth Heinemann, 2012, 456p.
5. Heikkinen M., Hiltunen T., Liukkonen M., Kettunen A., Kuivalainen R., Hiltunen Y.: A modelling and optimization system for fluidized bed power plants. *Expert Systems with Applications*, 36(2009)7, 10274–10279.
6. Huovinen M., Laukkanen V., Korpela, T.: On-line data reconciliation as a tool to improve the environmental efficiency assessment in power plant environment. *Preprints of 8th IFAC Symposium on Power Plants and Power Systems Control*, Toulouse, France, September 2–5, 2012.
7. Ibarguengoytia P. H., Sucar L. E., Vadera S.: Real time intelligent sensor validation. *IEEE Transactions on Power Systems*, 16(2001)4, 770–775.
8. Jämsä-Jounela S., Vermasvuori M., Endén P., Haavisto S.: A process monitoring system based on the Kohonen self-organizing maps. *Control Engineering Practice*, 11(2003)1, 83–92.
9. Kadlec P., Gabrys B., Strandt S.: Data-driven soft sensors in process industry. *Computers and Chemical Engineering*, 33(2009)4, 795–814.
10. Kämpjärvi P., Sourander M., Komulainen T., Vatanski N., Nikus M, Jämsä-Jounela S.: Fault detection and isolation of an on-line analyzer for an ethylene cracking process. *Control Engineering Practice*, 16(2008)1, 1–13.
11. Lee S. C.: Sensor value validation based on systematic exploration of the sensor redundancy for fault diagnosis KBS. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(1994)4, 594–605.
12. Li F., Upadhyaya B. R., Coffey L. A.: Model-based monitoring and fault diagnosis of fossil power plant process units using group method of data handling. *ISA Transactions*, 48(2009)2, 213–219.
13. Milne R.: Strategies for diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17(1987)3, 333–339.
14. Nelder J. A., Mead R.: A simplex method for function minimization. *The Computer Journal*, 7(1965)4, 308–313.
15. Nikula R., Laukkanen V., Juuso E., Leiviskä K.: Key variable based detection of sensor faults in a power plant case. *MATHMOD 2012 – 7th Vienna International Conference on Mathematical Modelling*, Vienna, Austria, February 15–17, 2012.
16. Qin S. J., Li W.: Detection, identification, and reconstruction of faulty sensors with maximized sensitivity. *AIChE Journal*, 45(1999)9, 1963–1976.
17. Razak R. A., Bhushan M., Belur M. N., Tiwari A. P., Kelkar M. G., Pramanik M.: Data reconciliation and gross error analysis of self powered neutron detectors: comparison of PCA and IPCA based models. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 4(2012)1–2, 91–115.
18. Shahin M. A., Jaksa M. B., Maier H. M.: Recent advances and future challenges for artificial neural systems in geotechnical engineering applications. *Advances in Artificial Neural Systems*, 2009, 1–9.
19. Sirola M., Lampi G., Parviainen J.: SOM Based Decision Support in Failure Management, *IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Sofia, Bulgaria, September 5–7, 2005, pp. 468–473.
20. Walker N. D. Jr., Wyatt-Mair G. F: Sensor signal validation using analytical redundancy for an aluminum cold rolling mill. *Control Engineering Practice*, 3(1995)6, 753–760.