Using Correlation in MFM Model Design and Validation

Tord Bergquist, Jonas Ahnlund, Jan Eric Larsson, and Lambert Spaanenburg

Lund Institute of Technology, Department of Information Technology
P.O. Box 118, 221 00 Lund, Sweden

Abstract: *Industrial process signals do affect one other. However, causal dependencies, or correlation, and thereby consequences of control actions cannot always be understood by the operator, which may lead to accidents. Models explain dependencies and are mostly built on process knowledge and blueprints, while causalities are visible in the measurements and ought to be used to increase process understanding. In this paper we will discuss why the cross-correlation technique is insufficient for causality detection in industrial processes and we will present a new method that, for instance, can be used to validate the design of multilevel flow models.*

Keywords*: Correlation, multilevel flow model, fault detection, alarm analysis.*

1 INTRODUCTION

Large industrial plants are difficult to fully understand and operate. Modern control systems provide large quantities of process measurements, but are less equipped to interpret the data in a way useful for the operator. The existence of unknown or hidden cross-dependencies between the process signals is one of the major causes to upsets and accidents in the process industry today. The operators cannot always understand the consequences of their actions and the alarm systems are inadequate to handle intelligent reasoning and analysis. Therefore additional techniques are required.

Fault detection, fault diagnosis, root cause analysis, and several other diagnostic tasks can be efficiently implemented using models. One such modeling concept is multilevel flow modeling, (MFM). Here, the objectives, the functions, and the causality between different functions of a system are explicitly described.

The main difficulty with model-based systems is that the models themselves require an effort to design and, especially, to validate. If the model does not agree with the real system, the diagnostic results may be wrong, and indeed, it may be better to avoid using the diagnostic system. Thus, models must be correct and validated at the design stage, and during the whole lifetime of the system.

This paper describes a correlation algorithm, which can, in a general sense, detect whether there is a dependency or causality between two measured signals in, for example, an industrial plant.

By using this algorithm, it is possible to validate a causal model, such as an MFM model, both during design and operation. By comparing the MFM model with the result of the correlation algorithm, it is possible to detect inconsistencies, meaning that either the model is incorrect, or something is wrong in the real system.

The main contribution of the paper is a description of an algorithm to detect correlated features within groups of time-dependent signals. There are, of course, well-known correlation algorithms in literature, but simply taking industrial plant signals and feeding them into a standard correlation algorithm does not, as we will see, provide useful results.

Section 2 serves as an introduction to multilevel flow models and describes the test plant and its model. In Section 3 we explain why we need to create a new correlation algorithm, which is described in this section, followed by some experimental results in Section 4.

2 INTRODUCTION TO MFM

Multilevel flow models (MFM) are graphical models of goals and functions of technical systems. The goals describe the purpose of a system or subsystem, and the functions describe the capabilities of the system in terms of flows of mass, energy, and information. MFM also contains the relation between the goals and the functions that achieve those goals, and between functions and the sub-goals, which provide conditions for these functions. MFM was created by Morten Lind at the Technical University of Denmark, see [12]. Jan Eric Larsson at Lund Institute of Technology has contributed several new algorithms and implementations, see [4][5][6][7].

MFM provides a good basis for diagnostic algorithms. The work of Larsson [6] presents several algorithms based on MFM. *Measurement validation* checks consistency between redundant sensor values, and can discover flow leaks, sensor failures, and other measurement errors. The *alarm analysis* algorithm analyses any (multiple) fault situation and can tell which faults are primary and which faults that may be consequences of the primary ones. *Fault diagnosis* uses sensor values and queries to the operator to discover the faults of the target system. Other algorithms have been developed later. The *failure mode analysis* uses MFM with added timing information to predict the consequences of failures. It can be used both during the design phase of a plant and in real-time during actual operation. The *fuzzy alarm analysis* works in a way similar to the discrete alarm analysis, but is based on fuzzy logic, which makes it more robust when faced with noisy signals close to decision boundaries, see [3][8].

2.1 An Example of an MFM Model

Here, the basics of MFM modeling will be illustrated in a small example: a part of the main systems of a bio-fueled district heating plant. A simplified process graph is shown in Figure 1 and the corresponding MFM model is given in Figure 2. The topmost flow describes the flow of thermal energy from the boiler via the dome and the heat exchanger to the district heating system. Below it is a description of the mass flow of water through the plant's circulation system, the so-called primary flow. There is also a small flow describing the generation of carbon monoxide under less than optimal burning conditions, and finally, a description of the mass flows involved in the burning itself, an example of an exothermal chemical reaction.

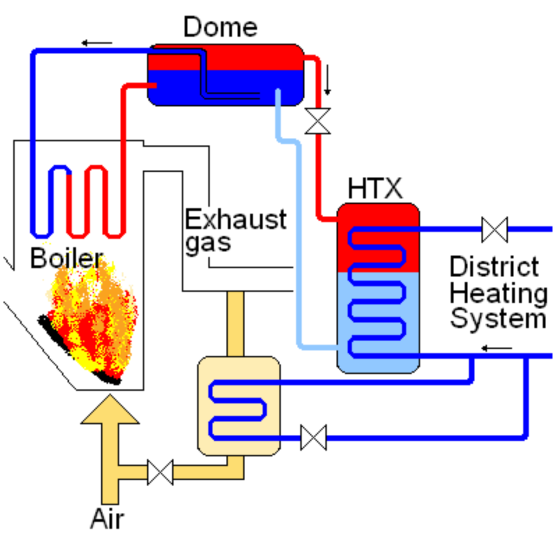

Figure 1. *A process graph of the main systems of the test plant.*

It should be noted that MFM describes how different flows enable each other, and which flows, volumes, etc., depend on each other. For example, the MFM model in Figure 2 shows the following:

- The temperatures of the dome and the heat exchanger are closely dependent.
- The temperature of the heat exchanger and the incoming district heating water are not dependent.
- The boiler temperature depends on the oxygen concentration in the boiler.

It is also important to observe that the given example is a very small illustration. MFM is designed specifically to handle large models with thousands of objects.

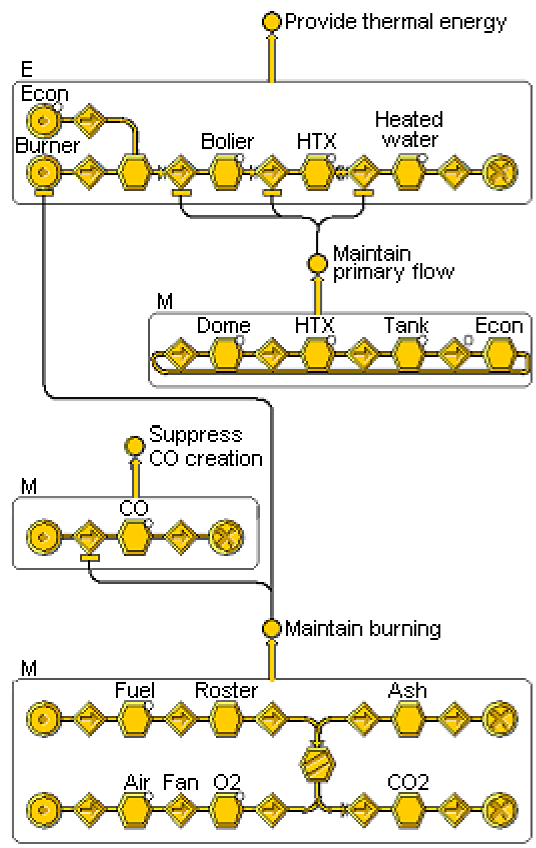

Figure 2. *A MFM model of the power plant.*

2.2 Advantages of MFM

All algorithms described in [6] are based on discrete logic where the "sensor" values are low, normal, or high, and the resulting values are consistent or inconsistent, working or failed, primary or consequential, etc. In other words, MFM uses a linguistic interpretation of logic variables, just as rule-based expert systems and systems based on fuzzy logic. In addition, the MFM algorithms all operate by searching in fixed graphs and together with the discrete logic, explicit means-end concepts, and graphical nature of MFM, this gives several advantages:

- The explicit description of goals and functions gives a *small semantic gap* between the diagnostic task formulation and the knowledge representation.
- The graphical representation provides strong support for knowledge base overview and consistency.
- The high level of abstraction makes knowledge acquisition, knowledge engineering, and knowledge base validation and support orders of magnitude easier than with standard rule-based reasoning.
- The graphical nature of the models allows the algorithms to have good real-time properties, such as an easily computed worst-case time, low memory demands, and high efficiency.
- The high level of abstraction allows the algorithms to be very fast. A worst-case fault diagnosis on a large nuclear power plant model takes less than 4 milliseconds, (worst case).

MFM has been used for medical diagnosis, [6][10][11][9] and nuclear power, [13][14]. The research of Larsson has spawned the spin-off company GoalArt, which has performed several successful projects, in medical equipment, conventional and nuclear power, and electrical and control systems.

3 A NEW CORRELATION METHOD

To examine if two signals (time series) $x(t)$ and $y(t)$ are correlated, one can use the mathematical cross-correlation function defined by

$$r_{xy}(k,l) = E\{x(k)y(l)\}$$

and the cross-covariance function defined by

$$c_{xy}(k,l) = E\{(x(k) - m_x(k))(y(l) - m_y(l))\}$$

where m_x and m_y are the mean values. However, there are drawbacks with these methods. The result of the cross-correlation is dependent on several variables, for example, the length of the series, the magnitude, and the zero-level, and therefore it is sensitive to scaling, but foremost a major disturbance in the signal has a major impact on the result. In process signals, the most interesting disturbances to detect are small variations that are precursors to larger problems.

A new correlation method has been developed for use on process signals off line. The method was designed with the following in mind:

- It should provide a *measure* of how strongly dependent two signals are, or in other words, how strong the causal connection between them is.
- The causal connection should provide a basis for the *understanding* of the process.
- If the measured causality disagrees with the model description, it can point out the need to update the model to match the reality, that is, the correlation algorithm can be used to *maintain the model* in a validated state.

In order to provide these properties, the correlation algorithm must be designed in a special way. We are not interested in a general signal correlation, but in the correlation between specific signal properties, such as, small variations that could indicate larger problems in the future, large variations that indicate process upsets, how often the signals move outside safe levels, so called alarm limits, etc.

The method is called Multiple Local Property Correlation (MLPC) and is based on the following, easy derived, general signal properties:

- Local maximum value (magnitude).
- Local minimum value.
- Local change in the variance of the signal.
- Local maximum change in the difference of two consecutive samples.

To handle negative correlation the local maximum and the local minimum values are merged into one property called *maximum distance*.

One assumption has to be made: a change in one of the properties above is only of interest within a short period. The validity of this assumption depends on process characteristics but is normally true. A change in a pressure measurement should not be correlated with the opening of a valve one week before. Instead, we search for the matching properties within a small time interval (typically a few hours or less), see Figure 3.

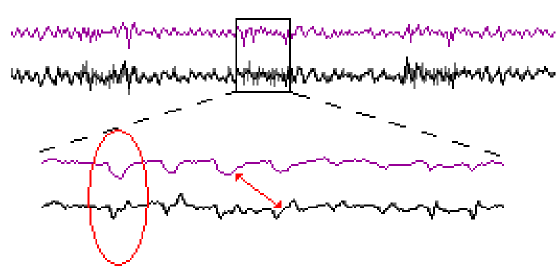

Figure 3. *The small variations appear at the same time in both signals. The match inside the circle is of interest but matches that are far from each other are of less importance.*

It is important that the method is capable of finding small changes even though the signal's measurements are clearly different in mean values, amplitude of variance, and so on. Therefore, we cut the two signals into equally sized segments. Different properties are then extracted from each segment and compared with corresponding properties in the other signal. These comparisons are then used to create one bar chart for every property. Strongly correlated signals will have almost the same distribution in all of their bar charts. More weakly correlated signals will have different distributions, but the correlated properties will still be detectable.

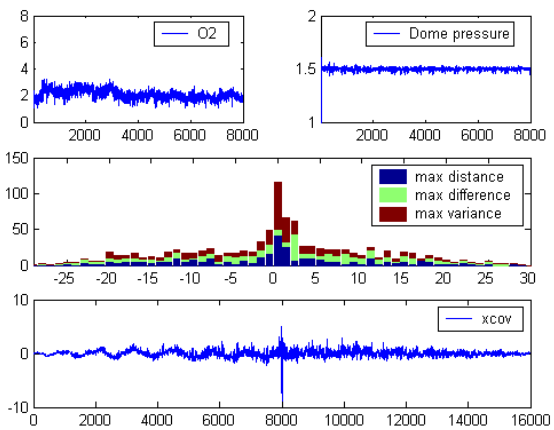

Figure 4. *Two correlated steady state process signals, their MLPC correlated bar charts and Matlab's xcov function.*

A comparison between MLPC and the mathematical cross-covariance is used to illustrate the new method's capabilities. Here, the two functions are applied to two process signals from the test plant, the oxygen level inside the boiler

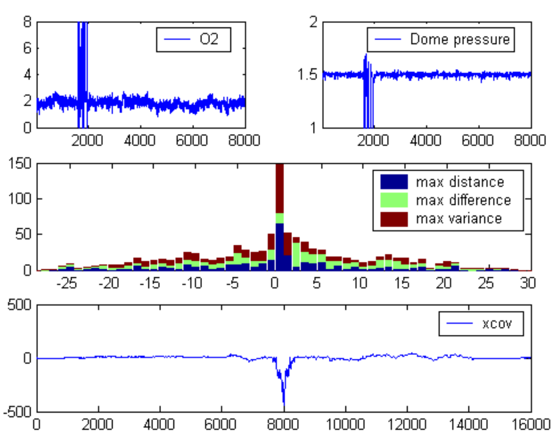

Figure 5. *Two correlated process signals suffering of a major disturbance, their MLPC correlated bar charts and Matlab's xcov function. Note the scale change (from Figure 4) in the cross-covariance plot. There is only a small change in the MLPC.*

and the pressure in the dome. They have been sampled during two weeks in December 2002. First we apply the functions on a period of stable production, see Figure 4. One week later, there was a major disturbance in both signals due to a plant trip, see Figure 5. The disturbance has a significant impact on the cross-covariance function, but has only a modest effect on the MLPC method.

In the search for dependencies in two signals, it is crucial to be able to handle the uncorrelated case as well as the correlated. In Figure 6 we use the oxygen level again but correlate it with normal distributed noise. The absence of a clear peak shows that the two signals are uncorrelated

MLPC is robust against noise. In Figure 7, the two process signals have an additive noise with the same variance as the signal themselves. For industrial applications this is a high signal to noise ratio.

Again, it can be seen that MLPC is quite robust in its detection of dependencies.

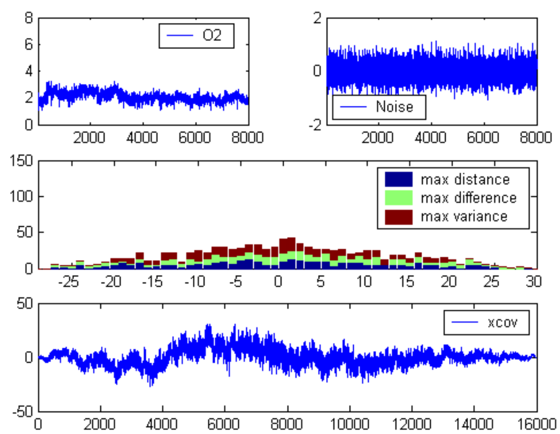

Figure 6. *A process signal and normal distributed noise. The absence of peaks shows that the signals are uncorrelated.*

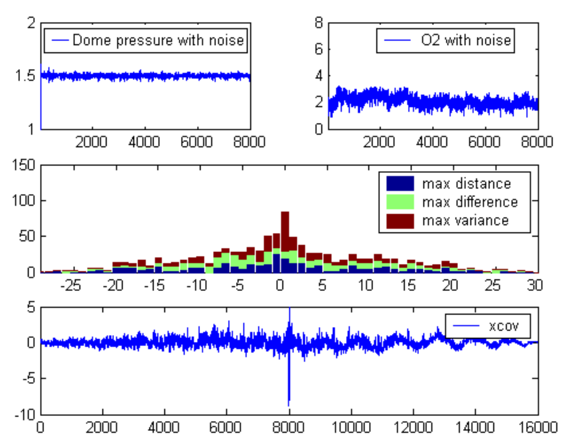

Figure 7. *The two process signals with added noise. The noise only has a small effect on the MLPC.*

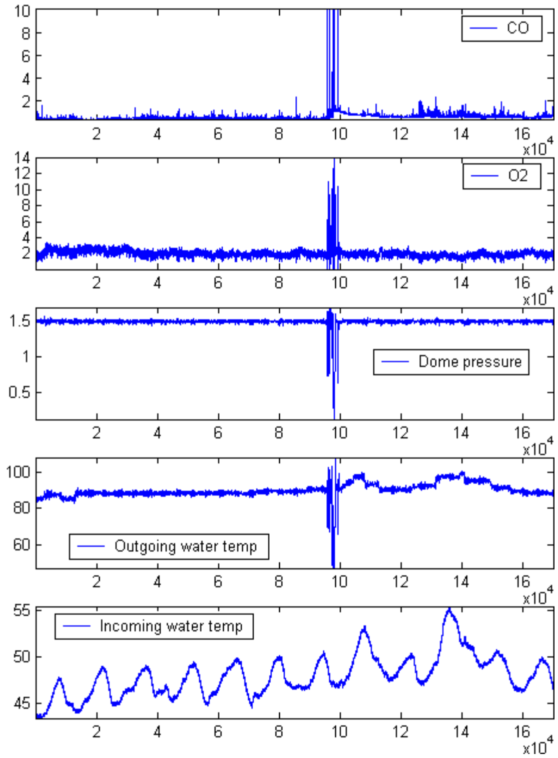

Figure 9. *Five signals from the model.*

One of the main reasons for developing a new algorithm was the difficulty of making an automatic and reliable decision on whether two signals are correlated or not using the unbounded cross-correlation. With MLPC, the number of segments, N, and the number of samples in every segment, L, are known a priori. The probability that a feature will occur at a specific index, i, is

$$p(i) = \frac{1}{L}.$$

In case that the signals are independent and uncorrelated, the probability of the time shift, t, between the indices in two corresponding segments is

$$p(i - j = t) = (L - |t|)\frac{1}{L^2}$$

This probability can be used to set boundaries for when two signals are to be considered correlated. In Figure 8 and throughout section 4 we will use

$$l(t) = 2N(L - |t|)\frac{1}{L^2}$$

as a limit.

The number of segments bounds the output value, and the bar charts can be merged using a fuzzy system that enable us to put different weights on the properties. We can present the result for instance as normalized values or with colors. The advantage of using a color scale is that the result can be presented in a color matrix, which would give a better overview when the method is applied to all signals within a plant in a batch manner.

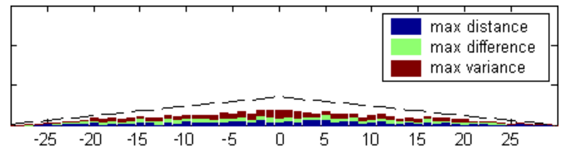

Figure 8. *The correlation of two random (noise) signals with the threshold l(t) marked.*

The introduced method is already functional and ready for industrial applications. It is easy to append more functions detecting additional properties, for instance, changes in applied adaptive filter parameters or neural network weights.

4 EXPERIMENTAL RESULTS

The new correlation method has been applied to process signals from a bio-fueled District Heating Plant (FFC) in Malmö, Sweden. A part of the same process is also described by the MFM-model in Section 2.1. We have focused on five signals from the model: the oxygen level, the carbon monoxide level, the dome pressure, and the incoming and outgoing water temperatures, see Figure 9. The signals show a large disturbance in the middle of the period, that affects all signals except the return temperature. The signals do not seem to have any other properties in common, but according to the MFM-model, there should be correlation between some of the signals or the model description is wrong.

From the model we know that the following signal pairs should be correlated:

- The oxygen level and the carbon monoxide level.
- The oxygen level and the dome pressure.
- The dome pressure and the outgoing water temperature.

Furthermore, the dome pressure should only be slightly correlated with the incoming water temperature. The incoming water temperature should affect the outgoing temperature, but not the other way around. The reason for this is that there are many other contributors to the district heating system. Even if the plant is shut down, the district heating system will still work.

The result of the new correlation method is shown in Figure 10. It is clear that the dome pressure is strongly correlated to the outgoing water temperature. All the three properties (distance, difference, and variance) have high values in the center, (time delay = 0).

The oxygen level and the carbon monoxide level also seem to be correlated. However, the max difference property is not concentrated to the center. Instead it has its highest value with a one-minute delay, with no contribution at all in the center.

The correlation between the oxygen level and the dome pressure is barely noticeable. Again, the max difference property is delayed, this time with two minutes.

As we suspected, the correlation between the dome pressure and the incoming water temperature is marginal. The center value comes mostly from the max variance property.

There seems to be a strong center max variance property value in all the signals, even from signals that should not be correlated. This probably comes from small but global disturbances in the process, that is, process noise.

The new correlation method can find correlation in process signals, which are not detectable with the ordinary cross-covariance function. With the use of our method, we can validate the model, that is, point out inconsistencies with real measurement values.

The presented method has already been used in model validation. When the MFM model of the district heating plant, Figure 2, was first developed, the authors assumed that the temperature in the local heating network would depend on the temperature in the heat exchanger. This is a natural assumption, because the plant is used to provide thermal energy to the network.

However, the local network is supported by several heating plants and well controlled, and variations in temperature in a single plant do not propagate into the network. In fact, limited variations cannot even be observed in the network temperature. Furthermore, the flow of this heat energy described by the model is a combination of the water temperature and water flow, which is regulated with a shunt-valve. It is therefore not enough to compare the two temperatures without taking the water-flows in consideration. We first realized these facts when we applied the MLPC and failed to discover any correlation. After further analysis, we realized that there is in fact no dependency, and the model was updated. Thus, the model in Figure 2 has been validated by our method and we have already discovered and corrected one model fault.

5 DISCUSSION

An algorithm that can be used to detect causal connections in industrial process signals has been developed. The MLPC is more sensitive to small changes than, for instance, the cross-correlation function, and it is also quite robust against noise. It

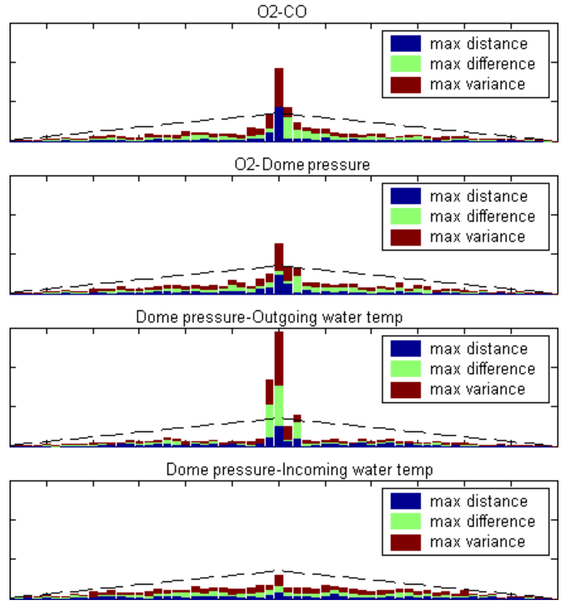

Figure 10. *The correlation result of signals from the MFM-model.*

is therefore suitable to use as a tool when trying to detect dependency in real process signals. The simple structure and the merging of detected properties by a fuzzy system make it easy to enhance the MLPC with other signal properties.

There is a need for a simple way to validate a model's design at an early stage. Using MLPC to create a dependency model from the process signals and compare its consistency with an MFM model is straightforward and will be further evaluated.

The correlation method can also be used in other applications. In previous papers [1][2] the authors have described methods for reducing the number of faulty alarms from an industrial plant by analyzing signal history and moving the alarm limits and adding filters, a so called alarm cleanup. The MLPC could be used to guide the selection of new candidate signals for cleanup. If one signal is giving off many alarms, while a correlated one gives no alarms, this is an indication that at least one of the signals has wrongly set alarm limits.

The aim of the presented method is to give a measure of how well a model and an industrial process match. In this way, the model of the process can be validated and kept correct. With a validated model, the method will also be able to detect fault in the process, that is, to provide fault detection. Finally, the method provides process understanding for designers, operators, and service technicians.

6 ACKNOWLEDGEMENTS

We want to express gratitude to Professor Gustaf Olsson for his excellent ideas and encouragement. Special thanks go to Ann-Britt Östberg and Tomas Erlandsson, Carl Bro Energikonsult AB, Johnny Pettersson, Sydkraft Värme Syd AB, Johan Söderbom, and Lennart Gårdman, Vattenfall Utveckling AB, for their support.

The project is sponsored by VINNOVA, the Swedish Agency for Innovation Systems under the project number AIS-33, and by Värmeforsk, the Swedish Thermal Engineering Research Institute under the project number P4-312.

7 REFERENCES

[1] Ahnlund J. T. Bergquist and M. Råberg, "Larmsanering med Generella Metoder," Värmeforskrapport P4-211, Värmeforsk, Stockholm, Sweden, September, 2003.

[2] Bergquist T. J. Ahnlund, and J. E. Larsson, "Alarm Reduction in Industrial Process Control," Proceedings of the ETFA2003, Lisbon, Portugal, September, 2003

[3] Dahlstrand, F., "Alarm Analysis with Fuzzy Logic and Multilevel Flow Models", Proceedings of the 18th Annual International Conference of the British Computer Society Special Group on Expert Systems, ES98, Cambridge, England, pp.173–188, 1998.

[4] Larsson, J. E., *Knowledge-Based Methods for Control Systems*, Doctor's thesis, TFRT–1040, Department of Automatic Control, Lund Institute of Technology, Lund, 1992.

[5] Larsson, J. E., "Diagnostic Reasoning Strategies for Means-End Models," *Automatica*, vol. 30, no. 5, pp. 775–787, 1994.

[6] Larsson, J. E., "Diagnosis Based on Explicit Means-End Models," *Artificial Intelligence*, vol. 80, no. 1, pp. 29–93, 1996.

[7] Larsson, J. E., "Diagnostic Reasoning Based on Means-End Models: Experiences and Future Prospects," *Knowledge-Based Systems,* vol. 15, no. 1-2, pp. 103-110, 2002.

[8] Larsson, J. E. and F. Dahlstrand, "New Algorithms for MFM Alarm Analysis," invited paper, Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Diego, California, 1998.

[9] Larsson, J. E. and B. Hayes-Roth, "Guardian: An Intelligent Autonomous Agent for Medical Monitoring and Diagnosis*," IEEE Intelligent Systems*, vol. 13, no. 1, pp. 58–64, 1998.

[10] Larsson, J. E., B. Hayes-Roth, and D. M. Gaba, "Goals and Functions of the Human Body: An MFM Model for Fault Diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, no. 6, pp. 758–765, 1997.

[11] Larsson, J. E., B. Hayes-Roth, D. M. Gaba, and B. E. Smith, "Evaluation of a Medical Diagnosis System Using Simulator Test Scenarios," *Artificial Intelligence in Medicine*, vol. 11, pp. 119–140, 1997.

[12] Lind, M., "Representing Goals and Functions of Complex Systems", Tech. Rep. 90-D-381, Department of Automation, Lyngby, Denmark, 1990.

[13] Öhman, B., "Code Generation for Alarm Analysis with Multilevel Flow Models," Proceedings of the Second International Symposium on Engineering of Intelligent Systems, EIS '2000, Paisley, Scotland, 2000.

[14] Öhman, B., "Alarm Analysis on Large Systems Using Multilevel Flow Models," Proceedings of the IFAC Symposium on Artificial Intelligence in Real-Time Control, Budapest, AIRTC–2000, Hungary, 2000.