SOM Based Analysis of Pulping Process Data

Olli Simula and Esa Alhoniemi

Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 5400, Finland

Abstract. Data driven analysis of complex systems or processes is necessary in many practical applications where analytical modeling is not possible. The Self-Organizing Map (SOM) is a neural network algorithm that has been widely applied in analysis and visualization of high-dimensional data. It carries out a nonlinear mapping of input data onto a two-dimensional grid. The mapping preserves the most important topological and metric relationships of the data. The SOM has turned out to be an efficient tool in data exploration tasks in various engineering applications: process analysis in forest industry, steel production and analysis of telecommunication networks and systems. In this paper, SOM based analysis of complex process data is discussed. As a case study, analysis of a continuous pulp digester is presented. The SOM is used to form visual presentations of the data. By interpreting the visualizations, complex parameter dependencies can be revealed. By concentrating on the significant measurements, reasons for digester faults can be determined.

1 Introduction

Modeling and control of industrial processes usually requires that an analytic system model can be built. However, in large industrial systems, global models cannot always be defined. In such cases, discovering of complex relationships between system variables is often problematic and modeling should be based on experimental knowledge. Modern automation systems produce masses of measurement data which, however, may be very difficult or even impossible to interpret. In many practical situations, even minor knowledge about the characteristic behavior of the system might be useful. For this purpose, the measurements need to be converted into some simple and comprehensive display which would reduce the dimensionality of measurements and simultaneously preserve the most important metric relationships between the data.

Artificial neural networks have successfully been used to build system models directly based on process data. They provide means to analyze the process without explicit physical model. The Self-Organizing Map (SOM) [7] is one of the most popular neural network models. It is especially suitable for system analysis due to unsupervised learning and topology preserving properties. The SOM algorithm implements a nonlinear mapping from high-dimensional input data space onto a two-dimensional grid or net of neurons. The mapping preserves the most important topological and metric relationships of the input data. The net roughly approximates the probability density function of the data and, thus, inherently clusters the data. Various visualization alternatives of the SOM are helpful, e.g., in hunting correlations between measurements and in investigating the cluster structure of the data.

The SOM based data exploration has been applied in various engineering applications such as pattern recognition, text and image analysis, financial data analysis, process monitoring and modeling as well as control and fault diagnosis [8, 12]. In addition, the SOM has been used in analysis of telecommunications environment, e.g., in discrete-signal detection and adaptive resource allocation problems [10, 13].

The ordered signal mapping property of the SOM algorithm has proven to be especially powerful in analysis of complex industrial processes [1]. In this paper, analysis of a pulping process is considered. The SOM based approach has been utilized to determine the reasons for situations where the pulp quality variable, kappa number, becomes too low. Similar approach has been used also in the analysis of steel rolling process.

2 Process Data Analysis

In modern process automation systems, it is possible to collect and store huge amounts of measurement data. On the other hand, the development of the computer capacity has made it possible to efficiently carry out analysis of the data. In Figure 1, different approaches to treatment of process data are presented.

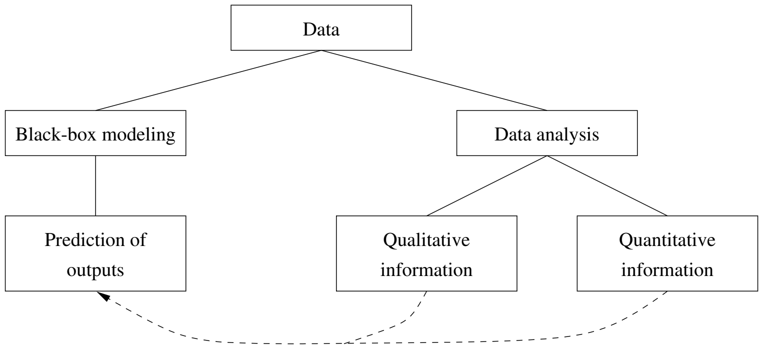

Fig. 1. Different possibilities of process data treatment.

Black-box models build a regression model between inputs and outputs of a system based on measurement data. After that, they can be used to predict system outputs if the inputs are known. For example, in [11, 2, 9], artificial feed-forward neural networks are suggested for prediction of pulping processes. The disadvantage of black-box models, however, is that they give little information about the dependencies of the variables and it is difficult to get general view of the system behavior.

In data analysis, useful information of process data is extracted without modeling the system. The approach needs to be distinguished from process experiments, where inputs are intentionally varied to find out the effect of the changes in the output. Data analysis could merely be used to find out reasonable setups for the experiments.

The application areas of data analysis are such processes or phenomena which — due to their complexity or nature — are impossible to model analytically. The advantage of the analysis is that neither process modifications nor experiments are required and general methods independent of application can be used.

Two different types of information may be acquired. *Qualitative* information can be obtained using such a data representation that is easy to understand and interpret. This is usually done using data visualization techniques. An example of *quantitative*, i.e., numerical information is the correlation coefficient depicting the strength of linear dependency between two variables.

Because the whole analysis is based on data, the measurements have to be reliable. Signal noise is usually not a problem, because it may be reduced using filtering techniques. Real problems are due to sensor faults. A good example is signal non-stationarity caused by slow drifting of dirt on the surface of the measurement sensor.

Analysis of data originating from a complex system is an iterative process. In Figure 2, different stages of data analysis are presented. After data acquisition, the variables of interest are selected. Common preprocessing operations include removal of erroneous measurements, noise reduction, and compensation of delay between different variables. The analysis may directly lead to conclusions, but typically several variable sets need to be considered; also the preprocessing may need to be altered. Usually the data analysis is started using as many variables as possible which are then reduced to the most significant ones.

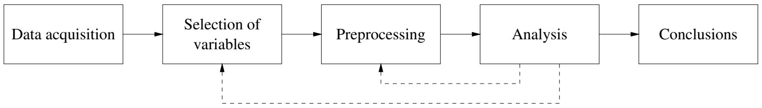

Fig. 2. A schematic illustration of data analysis.

It should be emphasized, that the whole analysis process requires the presence of a process expert, whose assistance is valuable in selection of variables, in choosing methods and their parameters for preprocessing as well as in interpretation and checking of the results. Expertise is — of course — needed also in making the correct conclusions based on the results.

3 The Self-Organizing Map in Data Analysis

As stated earlier, the SOM is a neural network structure consisting of neurons arranged to a two-dimensional regular grid. Each neuron is presented by a model (prototype) vector. The SOM algorithm carries out a nonlinear, topology preserving mapping from a high-dimensional input space onto model vectors. In other words, the algorithm simultaneously does two things: vector quantization and projection of the prototype vectors into two dimensions.

The topology preserving property of the mapping makes it possible to apply several different visualization methods to study the dependencies between variables in different parts of the input space. In the following, the visualizations used in the case study are discussed in detail.

3.1 SOM visualization

The greatest advantage of the SOM in analysis of measurement data is efficient visualization of the data [4, 6, 16]. The visual presentation can be obtained by visualizing the model vectors of the map or the training data using the model vectors. In this research project, following visualization methods have been applied [1].

- *Component plane representation* shows the distribution of one model vector component, i.e. one variable, on the map. By displaying all component planes

[1] The color images of this article can be viewed at URL `http://www.cis.hut.fi/projects/ide/publications/fulldetails.html#iwann99`

(distributions) at the same time, one may roughly observe relationships between variables and distinguish different operating states of the process (for example, see Figure 4). In our visualizations, the magnitude of each variable is presented using a color scale.

- *Ordering of the component planes* makes it easier to simultaneously investigate several component planes [14, 15]. The basic idea is to arrange the component planes in such a way that similar planes (correlating variables) lie close to each other. This reorganization is also carried out using the SOM algorithm. For example, in Figure 5, the component planes of Figure 4 have been rearranged.
- In *continuous coloring*, each map unit is assigned a color, which is similar in neighboring units. In our experiments, the HSV color system is used so that the value of component H varies with direction from the map center, S is constant and V is proportional to the map unit distance from the map center.
  Following techniques based on the coloring are used in this article:
  - Two variables are selected. Corresponding components of the model vectors are drawn in a scatter plot, and each point is dyed using the corresponding map unit color. The color coding makes it possible to study dependence of two variables in different process states. For an illustration, see Figure 7.
  - The coloring can also be used in time series visualization: the best-matching unit (BMU) of each sample is determined, and the corresponding points in the series are dyed using the color of the BMU (see Figure 8).

4 Case Study: Analysis of a Continuous Digester

4.1 The Continuous Digester

In the case study, the behavior of a continuous pulp digester of a pulp mill was studied. An illustration of the digester and the separate impregnation vessel is shown in Figure 3.

Presteamed wood chips together with cooking liquor are fed into the impregnation vessel. After the impregnation, the chips are fed into the digester. At the top of the digester, they are heated to cooking temperature using steam, and the pulping reaction — removal of lignin — begins. During the cooking, the chips move downwards the digester. The cooking ends at extraction screens by displacement of hot cooking liquor by cooler wash liquor, which is injected to the digester through bottom nozzles and bottom scraper. The liquor moves counter-current to the chip flow and performs washing of the chips.

4.2 Analysis of Digester Measurements

Problems in the digester operation, where pulp consistency in the digester outlet dropped, were the starting point for the analysis. In those situations, end product quality variable — the kappa number — values were smaller than the target value.

The test material consisted of measurements made in the digester during one week with constant production speed. At the end of the period, the digester ended up into such a faulty situation that the production speed of the line had to be dropped in order to restore control of the digester.

The signals depicting operation of the digester were collected from the automation system of the mill. The test period was selected so that there were no significant errors in the measurements. In preprocessing, the signals were delayed with respect to each other using known digester delays. Because the signal values were already averages of ten measurements made once a minute, no further noise reduction by filtering was required.

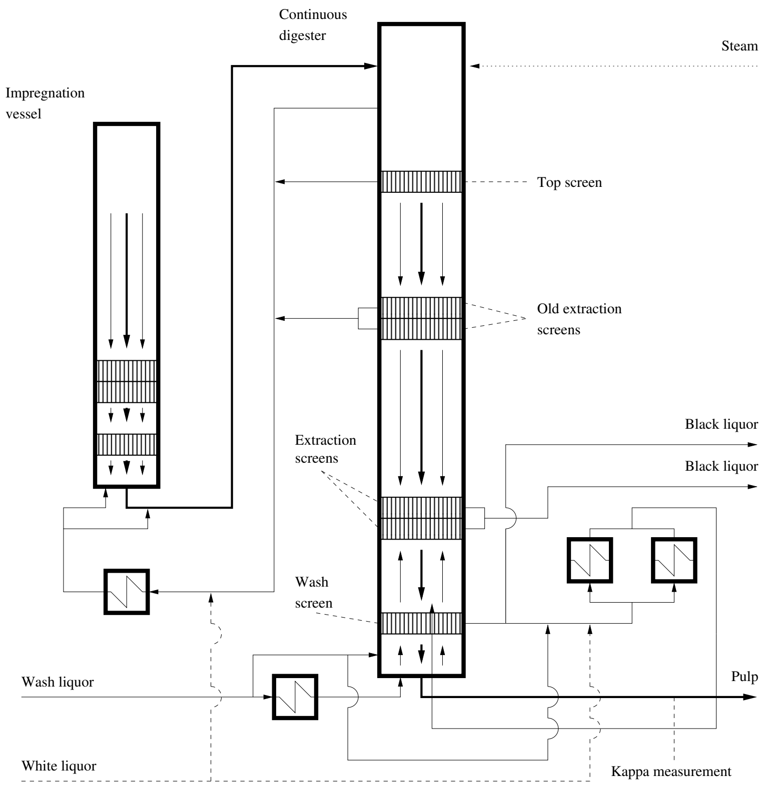

Fig. 3. The continuous digester and the impregnation vessel. The cooking and wash liquor flows are marked by thin lines and the chip flow by thick line. The four square-shaped symbols are heat exchangers.

In Figure 4, the component planes of a SOM trained using 24 measured signals are presented. In the lower part of the map, especially in the right corner, the value of kappa number is low. This means that the problematic states are mapped into that part of the map.

The component planes of Figure 4 are rearranged in Figure 5 so that component planes resembling each other, i.e., the correlating ones, lie near each other: this aids in interpretation of the numerous component planes.

Using the reorganization of the component planes in Figure 5, the 11 variables best correlating with kappa number were selected for further investigation while the 12 others were rejected. In Figure 6, component planes of an another SOM trained using the selected variables only are shown. Also in this case, the problematic process states with low kappa number were mapped into the bottom right corner of the map. To illustrate correlations between the kappa number and several other variables, each map node is assigned a hue (see Figure 7 top left corner). Then, xy-plots with the kappa number on the y-axis and the other ten variables on x-axis were produced using the map weight vector values.

The scatter plots indicate that *in the faulty states* there seems to be very little correlation between kappa and H-Factor, which is the variable used to control the kappa number. On the other hand, the variables “Dig_chip_l”, “Dig_liq_l”, “Black_liq”, “Screens”, and “Press_d” seem to correlate with the kappa number.

The explanation for this is that in a faulty situation, the downward movement of the chip plug in the digester slows down. This is due to the fact that the plug is so tightly packed at the extraction screens that the wash liquor cannot pass it.

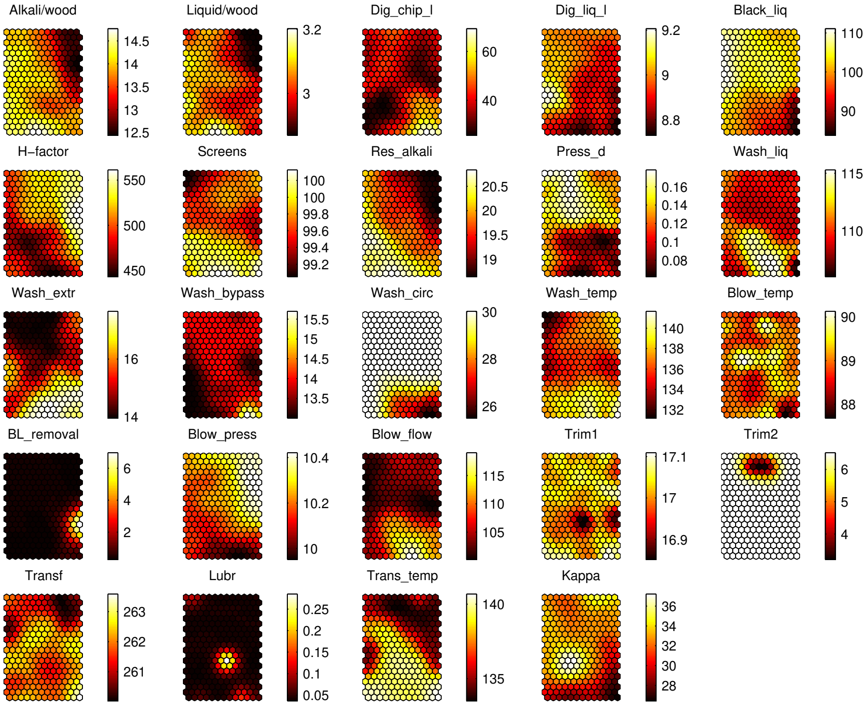

Fig. 4. Component planes of the SOM trained using 24 measurement signals of the digester.

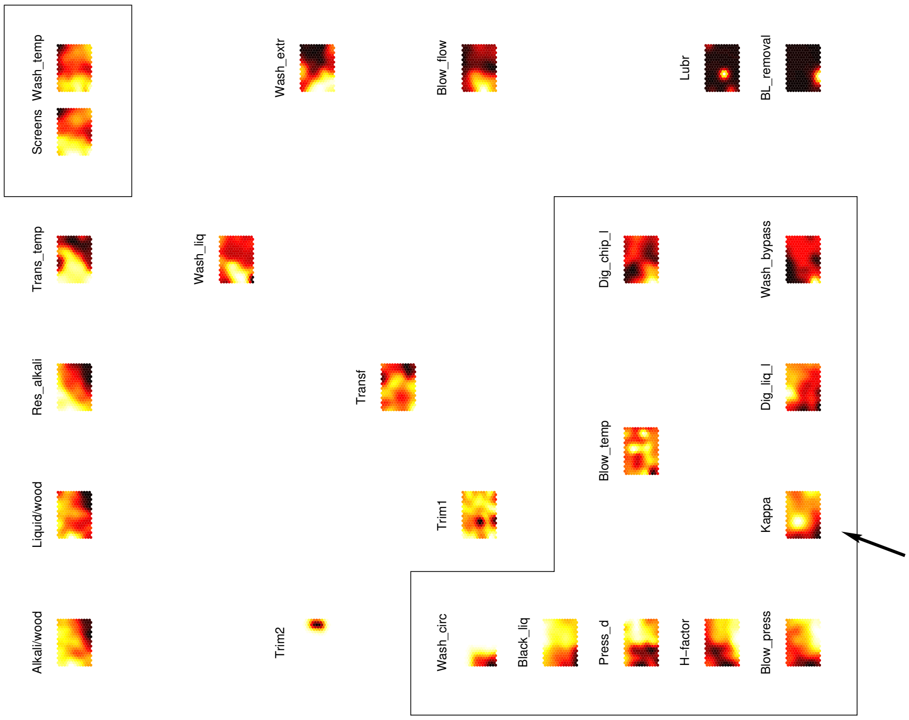

Fig. 5. Rearranged component planes. The variables selected for further investigation are surrounded by black line. The output variable of interest, the kappa number, is indicated by black arrow.

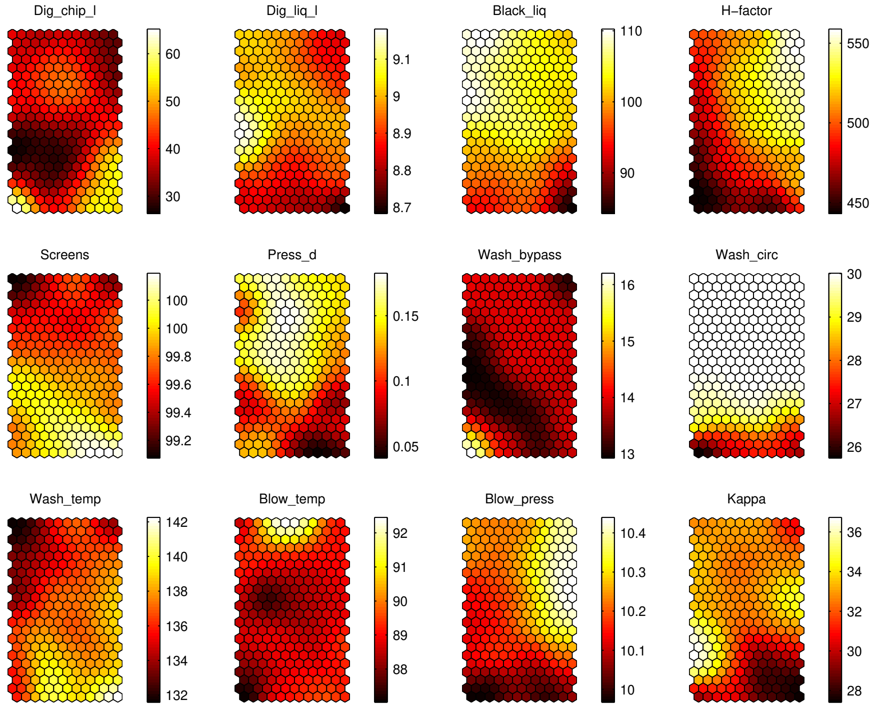

Fig. 6. Component planes of the selected variables.

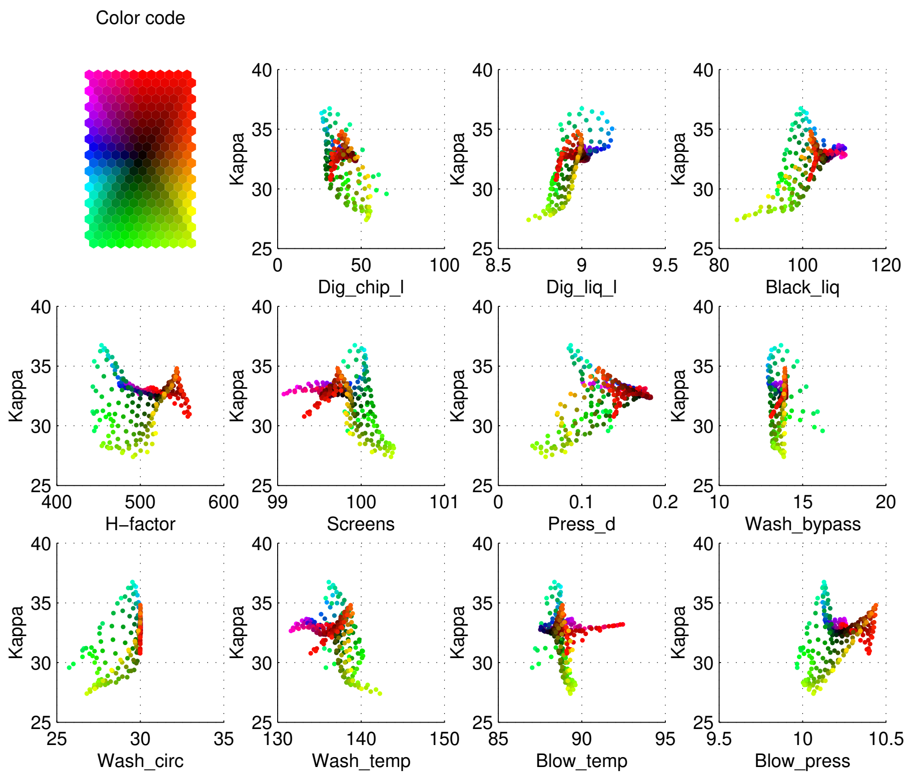

Fig. 7. Color coding on the SOM and 11 scatter plots.

There are two consequences: the wash liquor slows down the downward movement of the plug and the pulping reaction does not stop.

Because the cooking continues, the kappa number becomes too small. In addition, the H-factor based digester control fails, because in the H-factor computation, assumed constant cooking times for chips become longer due to slowing down of the chip plug movement.

Finally, in the Figure 8, the signals that were used to train the SOM of Figure 6 are colored using the coding presented in Figure 7. Now it can be clearly seen that the process operates normally until sample 700. From that on, problematic situations dyed by yellow and green hues appear every now and then. The last variations in the kappa number are so alerting that the operators have to slow down the production rate of the digester (not shown).

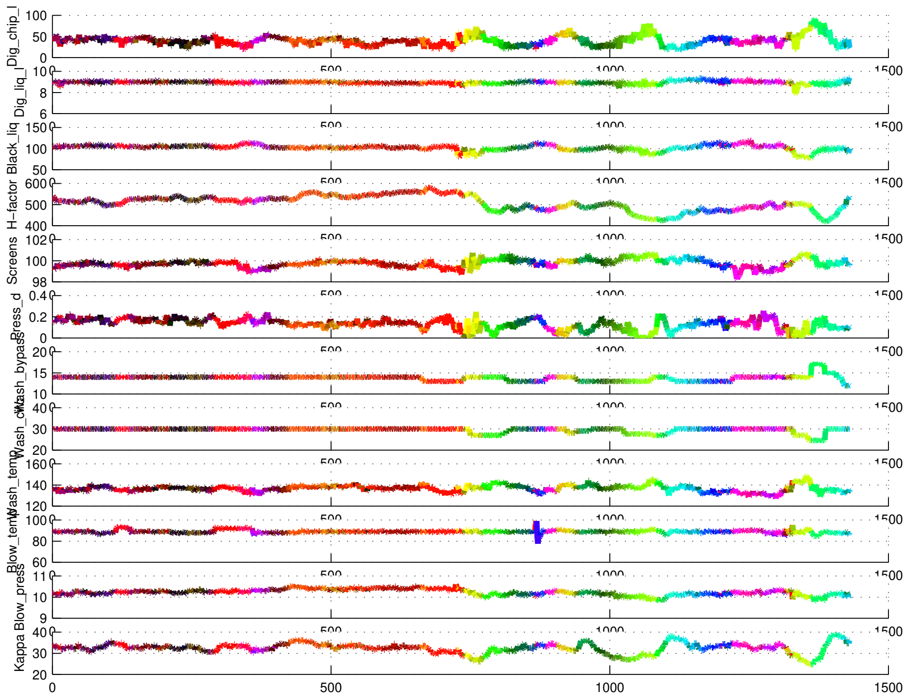

Fig. 8. Color coding of signals used to train the SOM with 12 variables.

5 Conclusions

Data analysis approach is useful in problems where the system or phenomenon of interest is difficult to analytically deal with due to its complexity or nature. In normal operation, the behavior of the continuous digester can be modeled based on first principles [5, 3]. In the faulty states, however, the analysis of the pulping data is the only way to investigate the behavior of the system.

The Self-Organizing Map can be effectively used to find and visualize correlations between process variables in different operational states of the process. In faulty states of the continuous digester, variables that normally do not have much effect on the pulp quality (chip level etc.) seemed to affect the pulping. This was due to the fact that they correlated with chip plug movement in the digester. In the problematic situations, the H-factor based kappa number control failed due to increased residence time of the chips in the digester, which is assumed to be approximately constant at constant production speed. In other words, the H-factor was bigger than the control system expected it to be.

5.1 Acknowledgments

The authors wish to thank UPM-Kymmene Wisaforest pulp mill for the pulping data and UPM-Kymmene Pulp Center for aid in the interpretation of the results. Financial support by Technology Development Centre of Finland and UPM-Kymmene is gratefully acknowledged.

References

1. E. Alhoniemi, J. Hollmén, O. Simula, and J. Vesanto. Process monitoring and Modeling Using the Self-Organizing Map. *Integrated Computer-Aided Engineering*, 6(1):3–14, 1999.
2. B. S. Dayal, J. F. MacGregor, P. A. Taylor, and S. Marcikic. Application of feedforward neural networks and partial least squares for modelling kappa number in a continuous kamyr digester. *Pulp & Paper Canada*, 95(1):26–32, 1994.
3. R. R. Gustafson, C. A. Sleicher, W. T. McKean, and B. A. Finlayson. Thoretical model of the kraft pulping process. *Industrial & Engineering Chemistry Process*, 22(1):87–96, Jan. 1983.
4. J. Himberg. Enhancing SOM-based data visualization by linking different data projections. In L. Xu, L. W. Chan, and I. King, editors, *Intelligent Data Engineering and Learning*, pages 427–434. Springer, 1998.
5. E. Härkönen. A mathematical model for two-phase flow in a continuous digester. *Tappi Journal*, 70:122–126, Dec. 1987.
6. S. Kaski, J. Venna, and T. Kohonen. Tips for Processing and Color-Coding of Self-Organizing Maps. In G. Deboeck and T. Kohonen, editors, *Visual Explorations in Finance*, Springer Finance, chapter 14, pages 195–202. Springer-Verlag, 1998.
7. T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995.
8. T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas. Engineering Applications of the Self-Organizing Map. *Proceedings of the IEEE*, 84(10):1358 – 1384, 1996.
9. M. T. Musavi, D. R. Coughglin, and M. Qiao. Prediction of wood pulp k# with radial basis function neural network. In *Proceedings of the 1995 IEEE International Symposium on Circuits and Systems*, volume 3, pages 1716–1719, Piscataway, 1995. IEEE.
10. K. Raivio, J. Henriksson, and O. Simula. Neural detection of QAM signal with strongly nonlinear receiver. *Neurocomputing*, 21:159–171, 1998.
11. J. B. Rudd. Prediction and control of pulping processes using neural network models. In *80th Annual Meeting, Technical Section*, volume B, pages 169–173, Montreal, Quebec, Canada, Feb. 1994. Canadian Pulp & Paper Association.
12. O. Simula and J. Kangas. *Neural Networks for Chemical Engineers*, volume 6 of *Computer-Aided Chemical Engineering*, chapter 14: Process monitoring and visualization using self-organizing maps, pages 371–384. Elsevier, Amsterdam, 1995.
13. H. Tang and O. Simula. The optimal utilization of multi-service scp. In *Intelligent Networks and New Technologies*, pages 175–188. Chapman & Hall, 1996.
14. J. Vesanto. SOM-Based Data Visualization Methods. *Intelligent Data Analysis*, 1998. Accepted for publication.
15. J. Vesanto and J. Ahola. Hunting for Correlations in Data Using the Self-Organizing Map. Accepted for publication in *International ICSC Symposium on Advances in Intelligent Data Analysis*.
16. J. Vesanto, J. Himberg, M. Siponen, and O. Simula. Enhancing SOM Based Data Visualization. In T. Yamakawa and G. Matsumoto, editors, *Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems*, pages 64–67. World Scientific, 1998.