36 Intelligent Process Data Analysis

**Olli Simula, Jussi Ahola, Esa Alhoniemi,
Johan Himberg, Pekka Hippeläinen, Jaakko Hollmén,
Juha Parhankagas, Jukka Parviainen and Juha Vesanto**

Analysis and control of complex nonlinear processes constitutes a difficult problem area in many practical applications. In complicated systems it is not possible to model the system analytically. In this case analysis of the available process data is the only possible approach. The data analysis begins with acquisition of all available data describing the system. The quality of the data is improved by removing noise and clear-cut errors. After this, based on process knowledge some variables may be combined to form new ones which are more useful from the problem point of view. The process, depicted in Figure 59, is repeated iteratively in order to find the most important variables relevant to the problem.

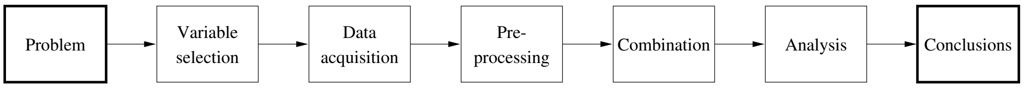


Figure 59: Data analysis process. In practice this sequential schema is iterated over and over again.

The SOM has proven to be a powerful tool to aid in the analysis due to its ability to form a visual illustration of data with respect to the selected set of variables. The SOM has the desirable feature of describing the nonlinear relationships between the large number of parameters and variables phenomenologically. Because the SOM algorithm performs a topology preserving mapping from the high-dimensional space to map units, it can also serve as a clustering tool of high-dimensional data. The SOM has also a capability to generalize, i.e. the network can interpolate between previously encountered inputs.

In this research project, several industrial processes have been analyzed in close cooperation with industrial partners. These include companies in steel and forest industry as well as design and consulting. Part of the work has been carried out in the TEKES Technology Program on Adaptive and Intelligent Systems Applications. In addition, work has been carried out in the Brite-Euram project, Application of Neural Networks Based Models for Optimization of the Rolling Processes (NEUROLL), which concentrates on improving steel manufacturing processes: using the methodology presented above, it is possible to investigate complex dependencies between incoming raw materials, process parameters at different stages, and quality parameters of the final product. Some case studies will be described in more detail below.

36.1 Pulp Mill

In a case study, behavior of a continuous pulp digester was analyzed. An illustration of the digester and separate impregnation vessel is shown in Figure 60. Wood chips and 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 starts. During the cook, the chips slowly 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 carries out washing of the chips.

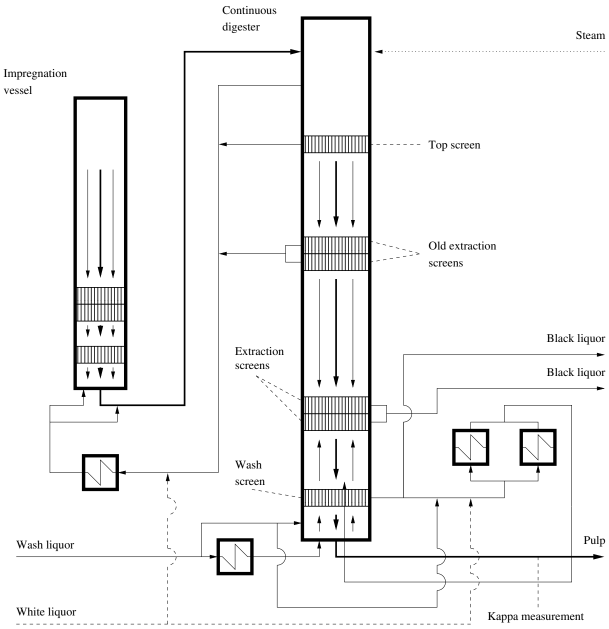


Figure 60: The continuous digester and the impregnation vessel. The cooking and wash liquor flows are marked by thin lines and the chip flow by thick line.

Digester operation problems indicated by drops of pulp consistency in the digester outlet were the starting point for the analysis. In those situations, end product quality variable (kappa number) values were smaller than the target value.

Measurement data were obtained from the automation system of the mill. The analysis was started with several dozens of variables which were gradually reduced down to six most important measurements during the data analysis process. The data used in the experiments consisted of three separate measurement periods during more than one month of normal pulping operation.

The periods were segmented by hand in such a way that they mainly consisted of faulty situations of the process. The production speed was required to be constant. During the measurement periods there were no significant errors in the measurements. Process delays between signals were compensated using beforehand known digester delays. In Figure 61 the six signals and production speed of the fiber line are shown. The three segmented parts are shown by solid line and the parts that were left out of the analysis by dotted line.

In Figure 62, the component planes of a 17 by 12 units SOM trained using signals of Figure 61 are presented. Five of them depict behavior of the digester and the last one is the output variable, the kappa number. The most problematic process

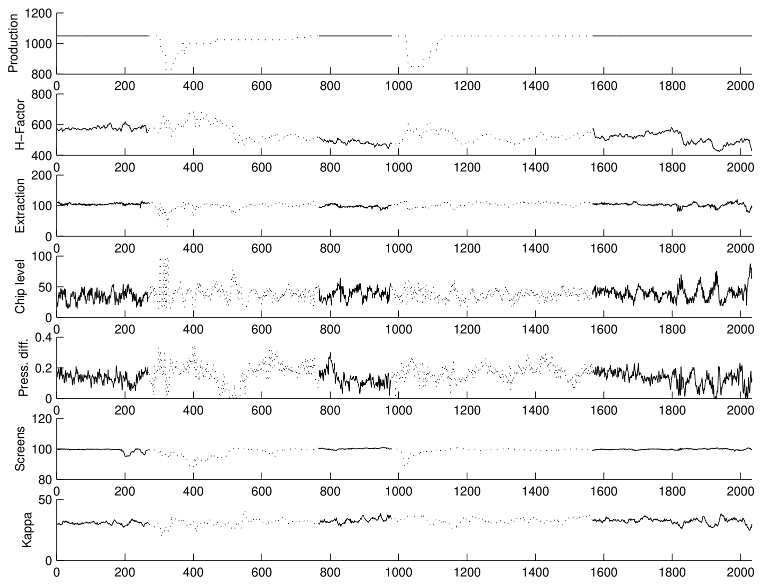


Figure 61: Measurement signals of the continuous digester. The analyzed parts are marked by solid line and the parts that were ignored by dotted line.

states are mapped to top left corner of the SOM: the model vectors in that part of the map have too low kappa number value.

Correlations between the kappa number and other variables were studied using Figure 63, where the SOM of Figure 62 has been presented using continuous color coding. The colors assigned to map units are shown in the top left corner of Figure 63. The five scatter plots are based on *model vector component values* of the SOM. They all have the values of kappa number on the x-axis and the other five variables on y-axis.

The scatter plots indicate that *in the faulty states* denoted by violet color there is weak correlation between kappa number and H-Factor, which is the variable used to control the kappa number. Otherwise there is a negative correlation as might be expected. On the other hand, the variables *Extraction* and *Chip level* seem to correlate with the kappa number. Also, the values of *Press. diff.* are low and the value of variable *Screens* (which during the analysis was noticed to indicate sensitivity of digester faults) is high.

The interpretation of the results is that in a faulty situation, the downward movement of the chip plug in the digester slows down. The plug is so tightly packed at the extraction screens that the wash liquor cannot pass it as it should. 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: in the H-factor computation, cooking time is assumed to be constant, while in reality it becomes longer due to slowing down of the chip plug movement.

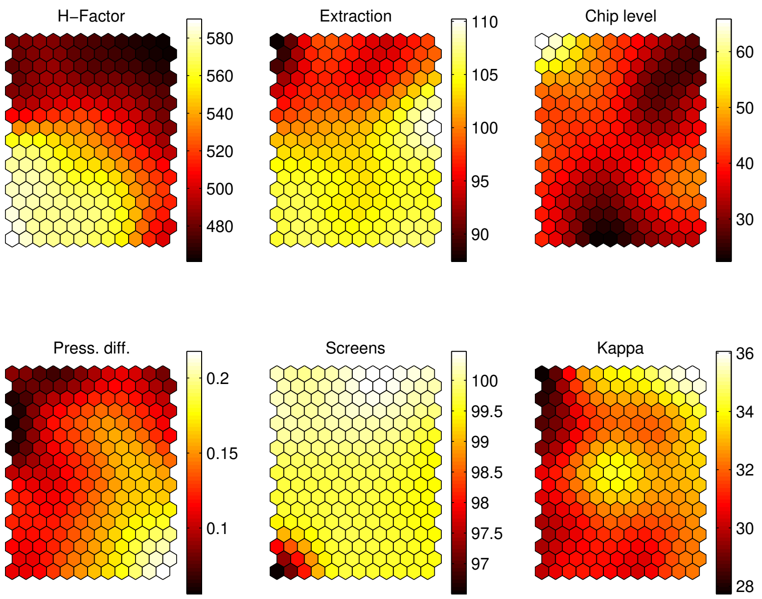


Figure 62: Component planes of the SOM trained using six measurement signals of the digester.

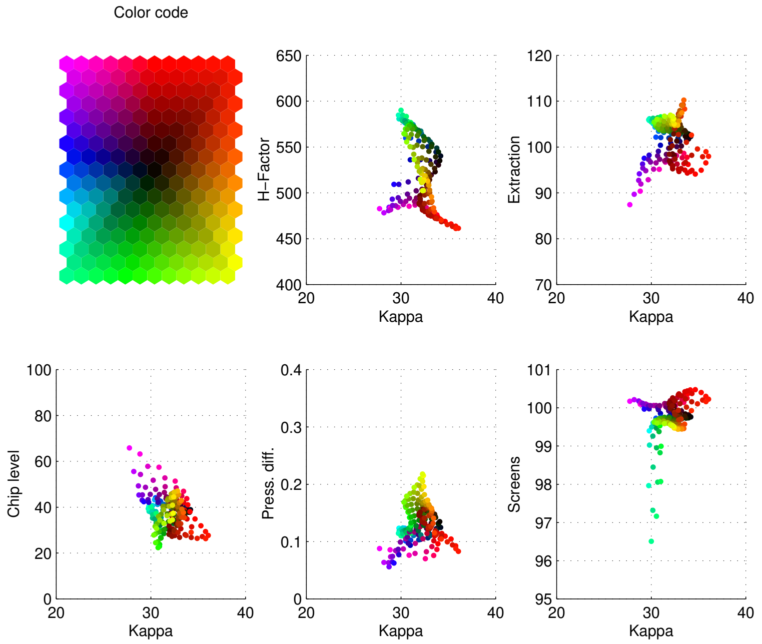
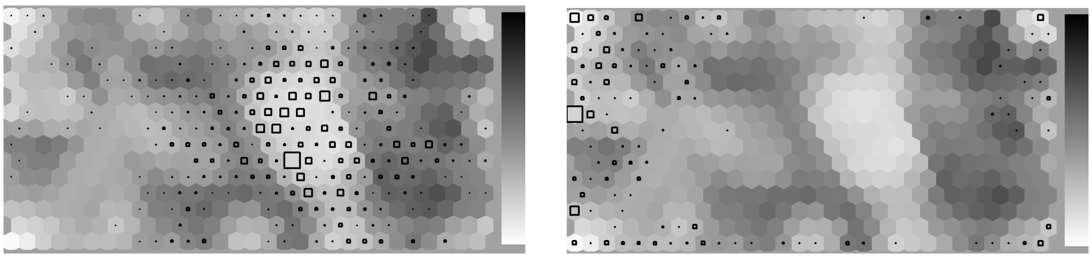


Figure 63: Color map and five scatter plots of model vectors of the SOM. The points have been dyed using the corresponding map unit colors.

36.2 Pulp and Paper Mills Technology

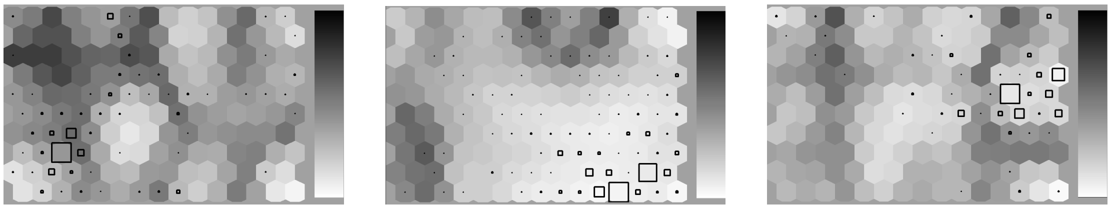
In this case study, the technology of pulp and paper mills all over the world was studied. The data was divided to three separate sets: information of the mill itself, its paper machines, and its pulp production. One SOM was trained for each of the three sets, and a fourth SOM was built using the combination of BMU coordinates of the input data on the three low level maps.
Figures 64a and 64b show the unified distance matrix (u-matrix) of the combined map, and the distribution of Chinese and Scandinavian paper mills on the map. The two mill sets are easily separable, although there was no geographic information present in the data. It can also be seen that in the area where the Chinese mills are, the values of the u-matrix are very low. That is, the variation between weight vectors in that area is low, which means that most of the Chinese mills resemble each other.
Figures 65a, 65b and 65c show the distribution of Chinese paper mills on the three low level maps. Also from these figures it is apparent that most of the Chinese mills are centered on a single area of the map. Taking a look at the weight vectors in these areas, it can be seen that the typical Chinese paper mill has small capacity, a large number (e.g. 4) of paper machines and it most probably produces printing/writing paper. The paper machines are small, slow and the paper weight is low. The pulp is produced chemically. Scandinavia, on the other hand, represents a technologically advanced region. The mills are new, they have big-capacity paper machines and the majority produces printing/writing papers or pulp.



(a) Chinese

(b) Scandinavian

Figure 64: Chinese (a) and Scandinavian (b) paper mills on the u-matrix of the combined map.



(a) Mills (b) Paper machines (c) Pulp production

Figure 65: Chinese paper mills on the three low level maps: mills (a), paper machines (b) and pulp production (c).

36.3 Other Applications

In a project completed in 1998, a process control software, SOM-Line, based on SOM was implemented. The software was a tool for monitoring functions of a galvanizing line of steel. The SOM is first trained by classified measurements. After training the software is capable of monitoring change in process status illustratedly in real time. If necessary the process can be adjusted based on information supplied by the software.

In a project with Suomen Perusmetalli continuous casting is being investigated. Continuous casting is a large scale industrial method for producing steel slabs for further refinement, e.g. for hot rolling. There may come up surface ruptures on the steel slab during the casting. The ruptures lower the product quality, and sometimes they may cause a breakthrough of liquid steel that leads to a long maintenance pause in the casting plant. The casting process is, however, monitored by thermocouples inside the mold. The obtained temperature data is analyzed using SOM-based visualization tools in order to get efficient features for rupture and breakthorugh warnings in the automation system.

The NEUROLL ("Application of Neural Network based Models for Optimisation of the Rolling Process") in an EU-financed project, the objective of which is to improve the efficiency of the production and the end quality of the hot rolled steel products. This is achieved by detecting complex relationships between measured or calculated process parameters and input/output variables, using statistical data analysis methods. Furthermore, the process is modeled at some level and methods are developed for the process state visualization and monitoring, as well as for diagnosis purposes.

In our laboratory the study has concentrated in finding correlations between quality parameters (e.g., width and thickness deviation and surface defects) and other process parameters using traditional correlation analysis and the SOM-based methods. This task is almost done and the results are utilized in improving the process in the general level and finding the variables for the process modeling. So, this task is also in progress and several, mostly SOM-based, approaches for process state visualization and monitoring have been considered.

In a project with Metsäteho Oy, the Finnish forest research organisation, data collected by forest harvesters is analysed. Data consists of measurements from individual trees and forest areas as a whole. The main goal is to apply unsupervised clustering techniques to group similar forest areas together. This clustering could prove to be useful when controlling the resource management for Finnish sawmills.

References

[1] E. Alhoniemi, J. Hollmén, O. Simula, and J. Vesanto. Process Monitoring and Modeling Using the Self-Organizing Map. *Integrated Computer Aided Engineering*, volume 6, pages 3–14, 1999.

[2] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995.

[3] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas. Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84(10):27 pages, October 1996.

[4] 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. Elsevier, Amsterdam, 1995.

[5] O. Simula, P. Vasara, J. Vesanto and R-R. Helminen. *Industrial Applications of Neural Networks*, chapter 4, The Self-Organizing Map in Industry Analysis, eds. L.C. Jain and V.R. Vemuri. CRC Press, 1999.

[6] J. Vesanto. SOM-Based Data Visualization Methods. *Intelligent Data Analysis*, April 1999 (to appear).