Pulp & Paper Canada, Vol. 103, No. 2, pp. 40–43, 2002.

State management for process monitoring, diagnostics and optimization

Cluster analysis was applied to a Finnish paper mill

BY M. FEDERLEY, E. ALHONIEMI, M. LAITILA, M. SUOJÄRVI AND R. RITALA

An UNDERLYING assumption in most data analysis, monitoring and diagnostics methods is that the process considered is in a stationary state. However, in practice this assumption is often ignored due to lack of tools for the analysis of operational states. Ignoring the stationarity assumption results in poorer performance of the monitoring and diagnostic methods. In paper production, there exist many operational states that affect the performance of the process, yet are not clearly distinguishable from measured data [1].

The simplest example of a discrete state variable is product grade. It is obvious that the grade will affect the monitoring alarm limits, how diagnostics should be carried out and how to interpret the data analysis results. Usually the grade information is explicitly available together with other process data and thus it can be easily taken into account. However, as the process is dynamic with long delays, even the concept of grade is not completely clear. In addition, other state variables of which explicit data is not available, but which obviously affect the process states further complicate the analysis. Such factors include, for example, shift personnel, raw material suppliers and fabric suppliers.

Effects of operation of the process, setpoint manipulations, disturbances and operational states are all present in measurements that are the main source of information for troubleshooting activities, process performance analysis and optimization. Methods to easily distinguish the operational states and their influence on the process from measured data are needed in order to comply the assumption of a stationary state.

This paper describes a collection of software tools developed for easy extraction and identification of discrete states from measurement data. These tools, combined with a process analysis system, can be used to enhance process monitoring and diagnostics: the assumption of stationary state can be met when each process state is monitored and analysed separately. As a result, one can achieve more reliable monitoring and diagnostic results.

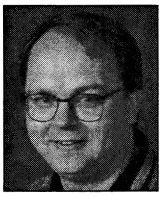
MANAGEMENT OF OPERATIONAL STATES

A lot of effort is put into better detection of short-term disturbances in the papermaking process. However, when implementing, for example, a monitoring or diagnostics system for improved fault detection, a problem frequently arises: are the data used as reference for future process operation collected from a stationary state and is this state achievable repeatedly? Also, the stability of the process — and thus the effect of a same kind of disturbance — may greatly vary depending on various circumstances, which are not connected to the origins of the disturbance. Thus, another question concerning the operational states of a process is: If the same product can be produced in various ways, which is the most profitable? So far, simple tools for seeking answers to these questions from paper mill data have not been available. As a result, operational states other than the grade have seldom been taken into consideration when data are analysed.

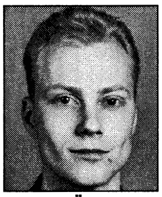
Cluster analysis has been successfully applied for wide variety of goals and subjects [2]. For an effort to distinguish discrete states from high-dimensional measured data the approach appears to be most useful: Clustering (or classification) means grouping of similar objects into clusters of their own. The grouping is based on some similarity or distance measure between the objects. The clustering algorithm produces a partitioning of the data into clusters according to a criterion.

M. FEDERLEY
KCL Science
and Consulting
Espoo, Finland

E. ALHONIEMI
Helsinki University
of Technology
Hut, Finland



M. LAITILA
UPM-Kymmene
Valkeakoski,
Finland



M. SUOJÄRVI
KCL Science
and Consulting
Espoo, Finland

R. RITALA
KCL Science and Consulting
Espoo, Finland
risto.ritala@kcl.fi

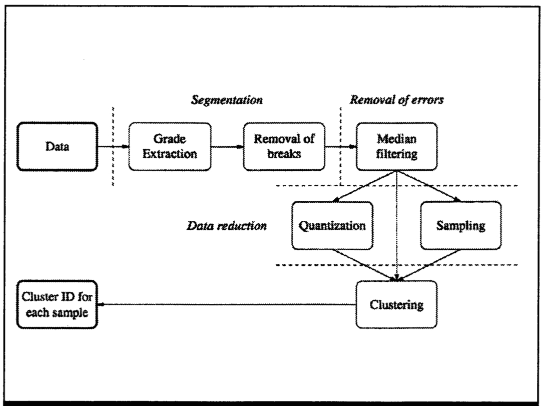


FIG. 1. The clustering procedure.

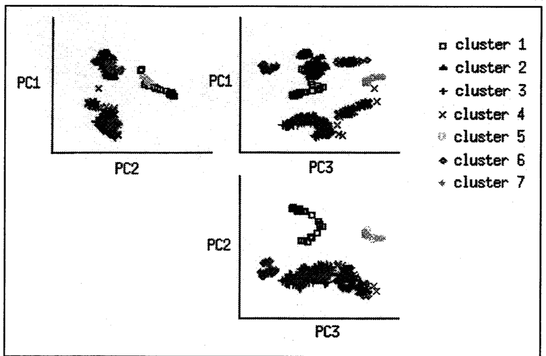


FIG. 2. Plots of the data projected on the three first principal components of the data. The symbols and colours indicate the cluster memberships.

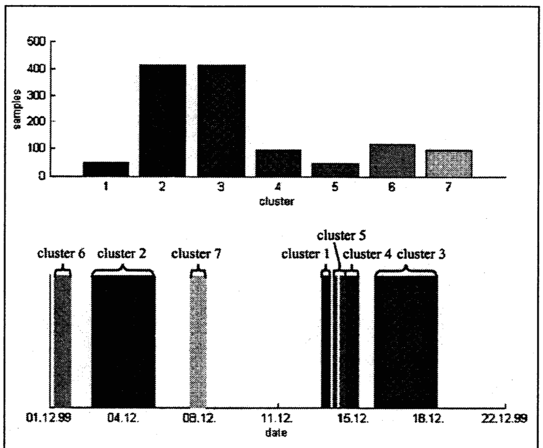


FIG. 3. Sizes and locations of the clusters in time.

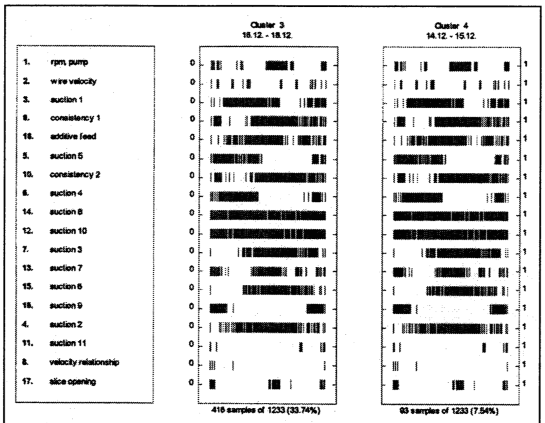


FIG. 4. Differences between the clusters 3 and 4 variable by variable.

CLUSTERING PROCEDURE

The clustering procedure is illustrated in Fig. 1. The goal of the clustering is to partition the data set into a number of separate clusters. In other words, after the clustering each data sample is assigned the identifier of the cluster it belongs to.

Data pre-processing: After process data retrieval, the measurement signals are segmented. Data describing the grade of interest are extracted and process breaks are removed. Median filtering eliminates short-term deviations in measurements. A description of the utilized filtering method is presented in [3]. Then, if the clustering algorithm to be used is computationally heavy, the number of data points must be reduced using vector quantization methods [4], or by sampling the data set. In the other case, all the data points are clustered directly.

Before clustering, the data are scaled. In the experiments, the variance of each variable was scaled to unity, which gives each variable equal weight in clustering.

Clustering: There are several different clustering algorithms [5,6], each having particular characteristics. However, a problem that is common for all of them is how to find out the number of clusters in the data. Even though several validity indices for the purpose have been proposed [7], it is advisable that also a human analyst validates the clustering result.

In partitive clustering algorithms, the number of clusters is fixed in advance. The algorithm carries out partitioning of the data points into the clusters according to an error criterion. The most commonly used such a method is the k-means. Another widely used approach is the Gaussian Mixture Model (GMM), which assumes the clusters to be Gaussians [8]. The clustering consists of two steps: estimation of the parameters of the Gaussians and assignment of each sample to the cluster (that is, Gaussian) with highest probability.

Hierarchical algorithms produce a clustering tree, a dendrogram, which can be cut at any level to obtain a desired number of clusters. The hierarchical algorithms can be further divided into agglomerative and divisive algorithms that correspond to bottom-up and top-down strategies to build a clustering tree. The former start with each data point in a cluster of its own; at each step, the two clusters that are closest to each other are combined. The latter methods take the entire data set and typically split it into two clusters, which are then further divided in a similar way.

Interpretation of the clustering results: Although the partitioning of data can be almost completely automated, careful inspection of the clustering results is always important. The interpretation requires interactive tools that, for instance, are able to answer the following questions:

- Is there any real cluster structure in the data? (In the literature this data property is often called the clustering tendency.)
- Which variables most affect the formation of each cluster, i.e., why are the clusters considered as clusters?
- What is the size of each cluster?
- What are the differences and similarities between the clusters?

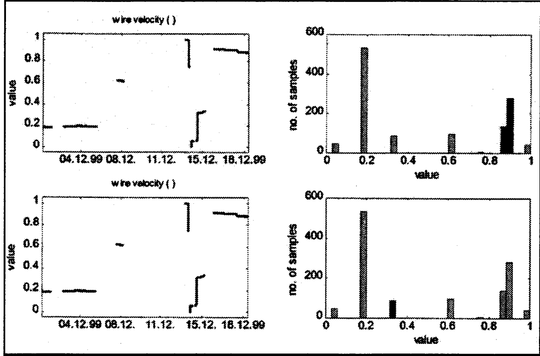


FIG. 5. Comparison of the values of the wire velocity in clusters 3 and 4 (dark areas in the histogram).

• What is the time period that each cluster represents?
• Has something unusual occurred in the process during the measurement period?

CASE STUDY: CLUSTERING AT A PAPER MILL

A case study was carried out at a Finnish paper mill to test the applicability of clustering in practice and also to further develop the tools. For the study, 18 measurements that reflect operators' actions to control the wire section were selected. The target was to study the differences in operational procedures within the most frequently produced grade. The case presented here is based on the data with 10 minutes' sampling interval and runs from one month's operation. All the measurement values presented are scaled.

After the removal of breaks and other grades from the data set, 1233 measurement points were left for clustering. The data were quantized to decrease the computational load of the clustering algorithm. In the clustering, a hierarchical divisive method was used [6].

Principal component analysis (PCA) [9] can be utilized to present the majority of the variability of the data with a couple of variables. The xy-plots of the calculated PCA-variables are often illustrative for examining the clustering tendency of the data (item 1 in the list of the previous section), and for determining the number of clusters to be examined. This is not, however, always the case since real, distinct clusters may exist even if there are no remarkable correlations between the variables. Thus, PCA-visualizations and other tools only offer guidelines; the cluster structure of the data and the selection of number of clusters relevant to the case have to be validated by an analyst with process knowledge. In this case study, the data formed clear groups in the plots of the three first principal components, Fig. 2.

Consequently, it can be concluded that there exists cluster structure in the data. Division of the data into seven clusters resulted in each of the distinct groups of samples in the PCA-plots being a cluster of its own. This was selected as the number of the clusters to be examined in this study, since the PCA-plots indicate existence of evident differences between the clusters.

For each cluster, size and location in time is shown in Fig. 3 (items 3 and 5). The data set consists of six separate runs of the grade. In the figure it can be seen that each run forms a cluster of its own. In addition to these, the run from December 14 to 15 is divided into two clusters, 4 and 5. The longest continuous runs form the largest clusters.

The original variables must be studied in order to find out the dissimilarities between the clusters (items 2 and 4). In Fig. 4, the clusters 3 and 4 are compared. The variables with the greatest differences between the clusters are shown on the top. All the values of each measurement that are comprised in the data set are presented by grey tick marks in the figure. The scale is from the minimum value of the data set (on the left; scaled to zero for each variable) to the maximum (on the right; scaled to one for each variable). The values that occur within the examined cluster are marked by black colour. For example, variable "suction 1" gets values near the minimum in cluster 3, and values near the maximum in cluster 4. Comparison of the plots side by side offers a quick way to discover the most remarkable differences and gives a general idea of distinctness of the clusters.

For more detailed examination of the data, a time series plot and a histogram can be displayed for any single variable, Fig. 5. Data points that belong to the cluster of interest are marked by black, others by a grey colour. The upper figures represent cluster 3, lower ones cluster 4.

In the case study, the machine speed and magnitudes of some suctions on the wire section had most effect on formation of the clusters. From Fig. 5, it can be seen that the machine is operated on six different speed levels. Each forms a cluster. Depending on a case, the machine speed could thus be handled as a state variable together with the grade information when inspecting stationarity.

The explanation for the division of the fifth run into two clusters can also been found by examining the original variables: there has been a maintenance shutdown on December 14. Even though an extended period has been removed from the data due to the shutdown as the data were pre-processed, the shutdown preparations and the start-up of the process form a cluster that differs clearly from the normal operation.

Typically, it seems that the operators interfere little in the control of the wire section during a run. There are, however, remarkable differences in operational parameters between the runs. Nevertheless, the product quality has remained roughly the same thorough all the clusters (or runs) despite the differences in the wire section operation. Based on that, it would seem that the same quality could be produced in various ways. However, a much more detailed study, including profound knowledge of all kinds of factors affecting the process during these periods, would be needed to be able to draw conclusions and to optimize the operational procedures. A thorough analysis of primary causes and consequences of clusters must be carried out by mill personnel, since a lot of information that is not measured is also needed (item 6).

CONCLUSION

The cluster analysis is very useful approach to reveal and study discrete states of processes. With visual, interactive and simple-to-use tools both the analysis and the interpretation of the results are easily performed. Although the theory of the methods may be complicated, the results can be easily understood and brought to analyst in a readable form.

The clustering concept, and thus the developed tools, are generic: they can be applied as such on wide variety of data and for multiple targets, for example:
• Analysis of operational procedures;
• Analysis of long-term variations;
• Study of stationarity,
• Optimization; and
• Improvement of process control by taking operational states into account.

However, process analysis that is purely based on measured data has its own difficulties. For example, selection of variables may at first glance seem very simple. As the work is carried out in practice, it is often noted that some of the variables are irrelevant with regard to the problem to be solved or some necessary measurements are so unreliable that they have to be ignored. The nature of the cluster analysis is iterative: modifications to the original data set need to be done several times before the results are satisfactory.

Data pre-processing always depends on the characteristics of the input data and the goal of the analysis. For example, in this study small fluctuations with duration less than 30 minutes could be removed by median filtering. In some other study they might as well have been the interesting features of the data.

Results of most clustering algorithms are affected by scaling of variables. The methods are usually based on distances between points in high-dimensional space. A standard approach, which was also used in the experiments, is to scale the variance of each variable to unity. This is practically the only thing one can do unless the relative importance of all the variables are known, which is rarely the case.

Different clustering algorithms were previously discussed in more detail. The selection of the "best" algorithm in a particular application depends on computational load and the expected shape of the clusters. Typically, different clustering algorithms favour different kinds of clusters. For example, the commonly used k-means is fast to compute and seeks for spherical clusters, which are roughly of equal size. In this study, the k-means algorithm produced poor results due to the fact that the clusters are typically of different size.

Validation of the clustering result is not very straightforward when the data dimension is high. However, it can be significantly aided by visual displays that were presented in the experimental part of this paper. One of the most important things in the validation is the determination of number of clusters in data. For this purpose, several methods (called cluster validity indices) have been suggested, and can be used to give guidelines in the selection of the number of clusters.

The clustering tool developed contains already several clustering algorithms as well as methods for interpretation of results. In the future, more clustering algorithms and a set of cluster validity indices will be added to the package and benchmarked. The methods will be tested with real process data, and be further used in the analysis and optimization of the wire section of the case mill.

In the case presented here, the main emphasis was on studying the applicability of cluster analysis for distinguishing operational states of a paper machine from measured data and to develop tools for easy utilization of cluster analysis. The applicability of the method and the procedure developed were confirmed by the results from the case study: it seems also intuitively reasonable that the production rate would affect the state of the process. The study of the operational states of the wire section is continued with more complicated cases to determine factors affecting the operational procedures and, on the other hand, to find out the effects of the operational procedures on the quality of the paper and the performance of the process. In these studies the effects of the state variable identified here, the production rate, will be eliminated so that less obvious factors will be revealed.

The cluster-analysis tool set has been received with interest by the case mill personnel: it is a new informative tool that can be used to describe (in this case) the differences in wire section running conditions between the runs of the same product. Better understanding of the wire section behaviour gives more possibilities to optimize the operational procedures and to manage the state the machine is operated in.

The algorithms and tools are implemented in Matlab environment. The cluster-analysis tool will be added at later point to KCL-WEDGE process-analysis system where large amount of data is readily obtainable. This will simplify the analysis significantly. The information of the process states will then also be available in other analysis, monitoring and diagnostics tasks that are carried out with KCL-WEDGE.

Résumé: Avant d'employer presque n'importe quelle méthode d'analyse des données, de surveillance ou de diagnostic, on doit s'assurer que les données ont été recueillies à partir d'un procédé à l'état stationnaire. La présente communication présente les résultats d'une étude visant distinguer l'état discret des mesures d'un procédé de fabrication du papier à l'aide d'une analyse typologique. Les outils développés peuvent être utilisés pour confirmer la stationnarité et aussi pour l'analyse de procédé et l'analyse comparative.

Abstract: Before employing almost any data analysis, monitoring or diagnostics method one should verify whether data have been collected from a process that is in a stationary state. In this paper, results from a study to distinguish discrete states from measurements of a papermaking process using cluster analysis are presented. The tools developed may be utilized to confirm the stationarity as well as for process analysis and benchmarking.

Reference: FEDERLEY, M., ALHONIEMI, E., LAITILA, M., SUOJÄRVI, M., RITALA, R. State management for process monitoring, diagnostics and optimization. *Pulp Paper Can.* 103(2): T48-51 (February 2002). Paper presented at the Control Systems 2000 Conference in Victoria, BC, on May 2 to 4, 2000. Not to be reproduced without permission of PAPTAC. Manuscript received March 3, 2000. Revised manuscript approved for publication by the Review Panel October 29, 2001.

Keywords: DATA PROCESSING, PROCESS CONTROL, PAPER MAKING, STATIONARY PROCESS, ANALYSIS, BENCHMARKING, MONITORING, OPTIMIZATION, DIAGNOSIS.

LITERATURE

1. RANTA, J., OLLUS, M., LEPPÄNEN, A. Issues and problems of flexible paper production: On the interaction of products, production and organization. Preprints, Control System '90, Helsinki, appendix (1990).
2. SEBER, G.A.F. Cluster Analysis: Multivariate observations. New York: John Wiley & Sons, 347-394 (1984).
3. PRATT, W.K. *Digital image processing.* New York: John Wiley & Sons (1978).
4. GERSHO, A., GRAY, R.M. Vector Quantization and Signal Compression. Norwell: Kluwer Academic Publishers (1992).
5. EVERITT, B.M. Cluster Analysis. London: Arnold (1993).
6. KAUFMAN, L., ROUSSEEUW, P.J. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons (1990).
7. MILLIGAN, G.W., Cooper, M.C. An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika* 50(2): 159-179 (1985).
8. McLAHLAN, G.F., BASFORD, K.E. Mixture Models: Inference and Applications to Clustering. New York: Marcel Dekker (1988).
9. JOLLIFE, I.T. Principal Component Analysis. New York: Springer-Verlag (1986).