

Use of multivariate data analysis for lumber drying process monitoring and fault detection

Mouloud Amazouz and Radu Pantea

Abstract—Process monitoring refers to the task of detecting abnormal process operations resulting from the shift in the mean and/or the variance of one or more process variables. To successfully operate any process it is important to detect and diagnose any process upsets, equipment failures or other events that may have significant impact on energy consumption and productivity. In most manufacturing processes, it is difficult if not impossible to detect abnormal operation by simply tracking some physical variable such as temperatures and pressures. Lumber drying (batch process) performance depends on more than 200 variables making the process very difficult to model and control using classical methods. Multivariate data analysis (MVDA) as a data mining technique makes the task easier and allows early fault detection thus allowing acting well before process goes out of control. MVDA techniques (PCA, PLS) were successfully applied on historical data of a sawmill operation to develop a multivariate statistical process monitoring (MSPM) of the wood kiln drying. A multivariate statistical process monitoring (MSPM) was developed using the SIMCA-P commercial software and applied offline on batches which went out of control (also known as outliers). The method was proven very powerful to detect the abnormalities and then diagnosis the faults. A database of faults and diagnosis is under development and an expert system will be developed for on-line fault detection and diagnosis.

The purpose of this paper is not to discuss the theory behind of multivariate data analysis but rather to demonstrate the applicability and the usefulness of the technique.

I. INTRODUCTION

EXPERIENCE has shown that many abnormal or merely sub-optimal situations occurring within a process could be identified and diagnosed by adequate analysis of existing data. However, the main obstacles in adequately addressing these situations are as follows:

- Large numbers of recorded process variables
- Significant and complex interactions (linear and non-linear) between process variables
- Process noise

Multivariate statistical process monitoring (MSPM) attempts to overcome these obstacles by directly addressing the issue of inter-dependence between the various process

variables. More specifically, MSPM reduces the information contained within all of the process variables down to two or three composite metrics through the application of statistical modeling. These composite metrics can then be easily monitored in real time in order to monitor process performance and highlight potential problems, thereupon providing a framework for continuous improvements of the process operation.

MSPM technology addresses many key business drivers:

- Maximization of throughput.
- Minimization of running costs (energy and raw material usage).
- Minimization of environmental impact.
- Improvement and maintenance of product quality.

The quality of a product at the end of a batch is dependent on the initial feed conditions and the trajectories of all manipulated variables over the course of the batch. By attempting to optimize the manipulated variable trajectories and end time of the batch, one can decrease variation and possibly improve product qualities while also increasing production rate of the product

Multivariate statistical models, developed within the framework of MSPM depict underlying inter-relationships that exist between various process variables. Such models can then be used in various functionality blocks of the overall process automation scheme:

- Soft sensors
- Inferential Control
- Sensor validation
- Performance Benchmarking
- Detection and Diagnosis of Abnormal Process Behavior
- Advanced Process Control

MSPM techniques can be integrated with knowledge-based systems in order to deliver comprehensive process monitoring solutions. This field of potential applications has been extensively studied in recent years (1,2,3,4).

This paper describes the approach used in the application of PLS to monitor and detect faults in a lumber mill, and gives partial results obtained on historical data.

II. PROCESS DESCRIPTION AND DATA ACQUISITION

M. Amazouz is with Canmet Energy Technology Centre, 1615, Lionel Boulet Blvd, Varennes, (Qc) J3X 1S6, Canada (phone: 450-652-6809; fax: 450-652-0999; e-mail: mamazouz@nrcan.gc.ca).

R. Pantea is with Canmet Energy Technology Centre, 1615, Lionel Boulet Blvd, Varennes, (Qc) J3X 1S6, Canada (e-mail: rpantea@nrcan.gc.ca).

Wood processing consist of sawing, drying and planing operations. Wood is stored before and after each operation. It is submitted to external conditions that have an influence on its physical and psychometric conditions (figure 1). Air-drying and air-equilibration are controlled by Mother Nature, while drying process is controlled by operators. The quality of the product and its manufacturing cost are directly related to these conditions.

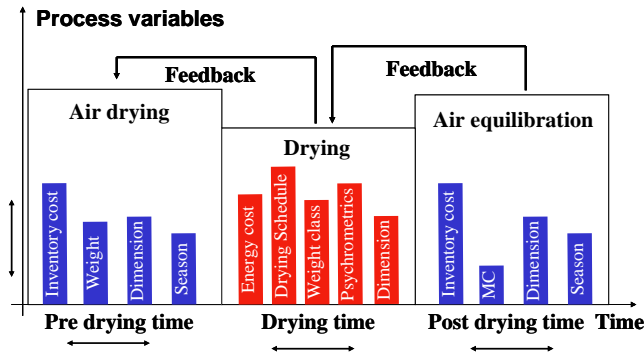


Fig. 1. Lumber drying stages.

Kiln control is based on empirical drying schedules. The kiln operator sets forced drying conditions based on these empirical schedules. This practice is not optimal and depends on operator’s personal judgment to set drying conditions and ending drying time. A market study ([pdf document](#)) performed in the lumber manufacturing industry in the province of Quebec (Canada) showed that more than 16% of the manufactured product is over dried and 6% are under dried. This translates into higher production cost and revenue losses.

Attempts to develop a physical model based control of wood kiln by several researchers failed due to the non-homogeneity and anisotropy of the material. In addition, lumber is dried in stacks and loads made of thousands of boards. Each board is different from the other (moisture content, physical properties). Statistics based approach can provide better understanding of the behaviour of wood during its processing and further improve the process. This approach needs data on the physical properties and process operation. An information system composed of a PI system was implemented in a sawmilling plant.

The plant information system gathers data on variables that affect the quality of the processed wood. Consequently, weather conditions and drying conditions are gathered and monitored. In addition of common instruments used on this industry, it was necessary to develop a raw material flow tracking system. Therefore, for each bundle the history could be retraced. A lumber bundle labeling system was implemented. Labels contain information about lumber dimension, initial moisture content, and production date. A database was developed for further analysis. Figure 2 describes schematically the material as well the information

flow among the lumber transformation process.

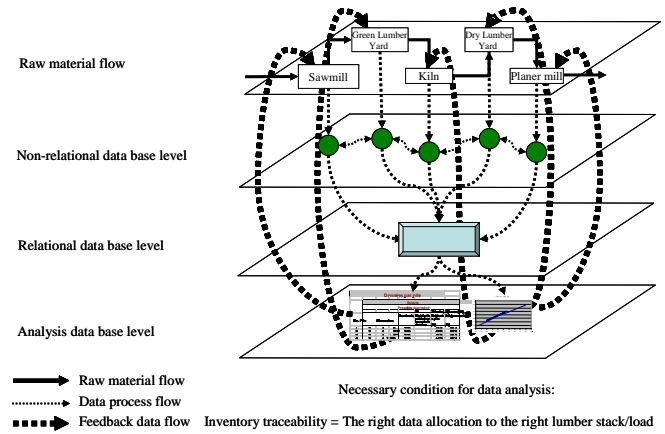


Fig. 2. Raw material and data flow information schema for a lumber mill operations.

III. DEVELOPMENT OF THE MSPM

A. Methodology

Two multivariate statistical modeling techniques were used to analyze data from kiln operation:

- Principal Component Analysis (PCA) for trends and patterns overview.
- Partial Least Squares (PLS) for prediction

Both of these methods utilize the presence of cross-correlations between process variables. As a result, problem complexity is reduced and impact of noisy measurement is minimized. Projection methods are applied to reduce the complexity of the problem. It is then possible to invoke sophisticated classification analysis techniques to describe ‘envelopes’ of normal process operation.

B. Data preparation

A one year historical data was gathered through the plant information system. Using the mill process and instrumentation diagram key data tags were identified. Raw operation data were pre-processed by visual inspection, preliminary statistics analysis, outlier detection, data reconciliation, missing data replacement and basic transformation for batch analysis. The data has been unfolded with respect to variable trajectory. The data matrix is composed of 36500 rows (observations) and 52 columns (main variables). The available variables that were used in the analysis can be grouped in five main categories:

- Drying energy
- Ventilation
- Psychrometrics
- Psychrometrics control
- System response

C. Data exploration

The PCA analysis was performed on the prepared data using the SIMCA-P software. The objective was to explore and to understand the data set variability and to identify the reference batches (good batches) that will become the data set for the PLS modeling. This analysis defined four different groups of major variables depending on the board initial moisture content. In each group, three seasons were identified (summer, winter and spring/fall). The results permitted the formulation of two modeling directions:

- The global modeling of all the groups and seasons together.
- The modeling of each group and season separately.

If the first one gives a wider acceptance behavior range, the second should represent a more precise approach capable of emphasizing easily peculiar process behavior. However, for practicality reasons, the first direction has been explored.

IV. PROCESS MONITORING AND FAULT DETECTION

The next step was to develop the PLS global model of all the groups and all the seasons together and to evaluate it as a fault detection tool. The scores of the model are aligned and reorganized so that the scores of one batch form one row vector (t1 followed by t2, followed by t3, etc.) in a matrix. When batches have phases, the alignment is done by phases using their respective maturity variable. In our case the batch undergoes ten processing steps and the maturity variable is the gas consumption. From the score matrix, the averages and standard deviations (SD) of the scores are calculated, (average trace of normal batches (modeled reference) and control intervals ± 3 SD). In this case the $+3SD$ is UCL, while $-3SD$ is the LCL. This means that 99% of the observations lie inside the interval of $\pm 3SD$.

As an example, Figure 3, shows that the monitored batch, begins to behave differently with the third phase of the drying batch process. This behavior continues in the fourth and the fifth phase of the process. The circled phase (the fifth one), shows how different the behavior of the monitored batch is compared to the reference batch or the modeled reference. In fact, the monitored batch evolves in the LCL (lower control limit) space, between the modeled reference batch and the LCL reference, while the reference batch evolves in the UCL (upper control limit) space, between the modeled reference and the UCL reference

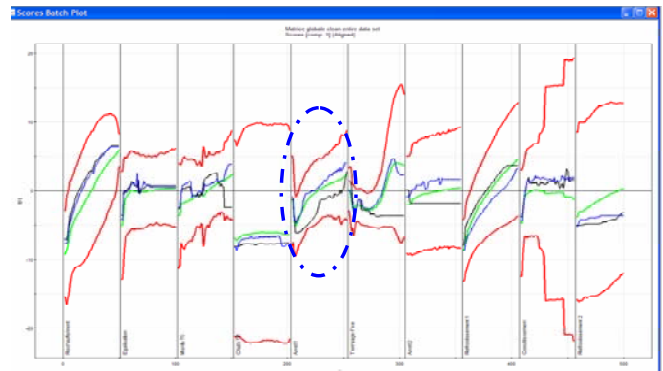


Fig. 3. Batch trajectory visualization.

Therefore, the monitored batch should be furthermore investigated. A more detailed representation of the fifth phase of the drying process is showed in Figure 4.

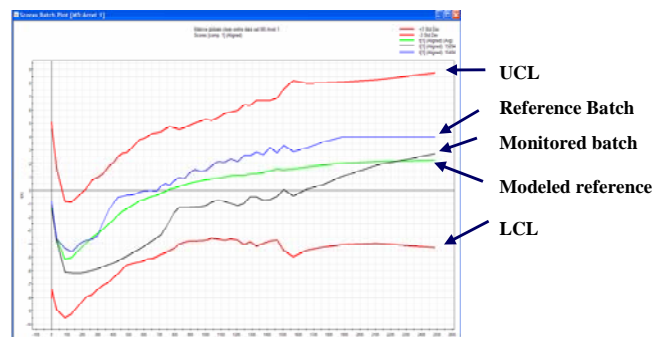


Fig. 4. Detailed view of the process behavior, during the batch's fifth phase.

The next step is to determine which process variable or variables are responsible for the deviation of the monitored batch from the reference batch. The contribution plot helps to identify those variables that have a strong influence on the system behavior (see Figure 5).

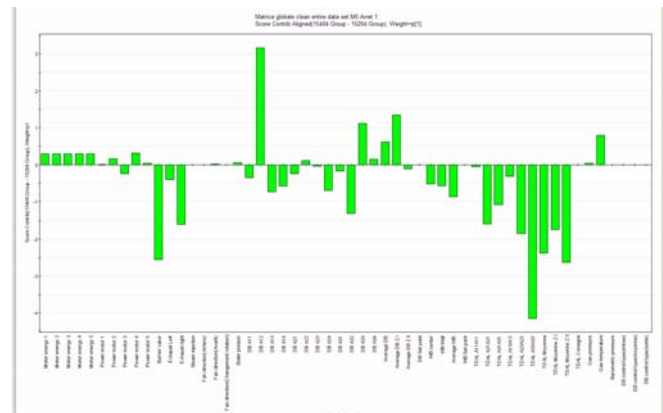


Fig. 5. Contribution plot of the variables responsible for the first score

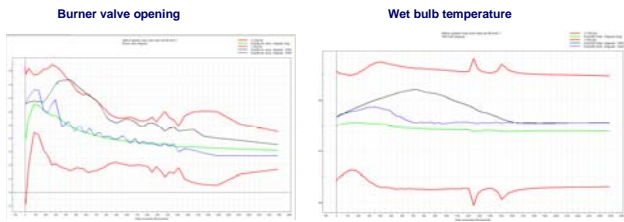


Fig. 6. Trajectory plots of variables under investigation.

A further investigation through a detailed visualization of different process variables (see Figure 6) allowed the recognition of an unexpected behavior for the fifth phase of the process: an increase in energy consumption combined with an unexpected increase in air humidity.

The analysis showed that the burner valve opening is bigger than that of the reference batch and the wet bulb temperature is higher than that of the reference batch. After deeper analysis of the process control logic, it was concluded that the first event is a consequence of the second one. The controller reacted to the increase in the wet bulb temperature by increasing the burner valve opening in order to maintain the first one at the set-point level. At this stage of analysis, it could be concluded that the fault is either due to the wet bulb temperature sensor failure or to a malfunction of the steam injection system. However, the fact that the wet bulb temperature rejoined the values of the reference batch shows that most probably the fault is due to a malfunction of the steam injection system. When the controller changed the set-point to “close injection steam valve” the actuator didn’t close it entirely. Therefore, for almost half of the fifth phase steam was injected in the kiln, and the controller reacted by bringing more drying energy into the kiln. This resulted in an increased drying time combined with an increase in energy consumption.

V. CONCLUSION

The study demonstrated the applicability of multivariate statistical methods as a data mining tool for process performance monitoring. The modeling has been encouraging despite limitations of the data that includes missing data on important variables, non-measured and/or unmeasured, but important, variables and the combining of spot and composite samples. The results can be used as a feedback to process engineers to enhance process knowledge, improve faulty data identification, and identify important variables to be additionally monitored and pinpoint the underlying cause of a specific abnormal operation.

MSPM was shown to be very useful and powerful to detect process abnormalities. Best practices could be also identified and used to improve process operation. Online implementation of such tool could increase productivity,

save energy and plan for maintenance.

The next step of the work will be oriented toward the testing of other data mining techniques such as modified statistical methods or/and neural networks. A knowledge data base and an expert system will also be developed for automatic fault diagnosis.

ACKNOWLEDGMENT

This work was supported by PERD (Canadian Program for Energy Research and Development). Special thanks go to our partners in the project; the Leduc sawmill and the Canada’s wood products Research Institute (Forintek Canada Corp.) in Quebec

REFERENCES

- [1] Jesus Flores-Cerillo et al, “Latent variable MPC for trajectory tracking in batch processes” *Journal of process control*, 15, 2005, 651-663.
- [2] Henk-Jan Ramaker et al. Fault detection properties of global, local and time evolving models for batch process monitoring. *Journal of Process control*, 15 (2005) 799-805
- [3] Ji-Hoon Chao et al. Fault identification for process monitoring using Kernel principal component analysis, *Chemical engineering science*, volume 60, 2005, 279-288
- [4] Theodora Kourti (2005). Abnormal situation detection and projection methods -- industrial applications. October 28-29, 2003. Hamilton, Ontario, Canada, *Chemometrics and Intelligent Laboratory Systems*, 76, 215-220