

Using multivariate regression techniques to analyse the performance of a steam heated drying process

M. J. Wade and J. Balderud

Abstract—This paper explores the use of a multivariate statistics approach to monitor and diagnose the performance of a steam heated drying process. The results presented in this paper are a first stage investigation into the capabilities and limitations of these methods when applied to this particular process. The results show that, despite the non-linear behaviour of the drier process, the model predictions correspond well with the process measurements. The results also indicate ability to detect process failures, however, accurate diagnosis of the failures remains a subject for future studies.

I. INTRODUCTION

A modern paper machine is a complex non-linear process comprised of a large number of unit operations. The interactive and non-linear nature of these, combined with a large number of unmeasurable process variables, make efficient operation of the paper making process a non-trivial task. Traditionally, 'optimal' operation of the paper making process has, therefore, relied more on the operator's knowledge and experience of the process behaviour and suitable operating conditions than on conventional control algorithms. However, increased competition between paper mills and progressively stricter environmental regulations have had the result that modern paper machines are constantly evolving and improving, with frequent changes to the process behaviour as a result. To retain operational efficiency under these conditions, operators must quickly adapt to the changing behaviour of the process. On-line tools for monitoring and diagnosis of the process variables are essential to achieve this objective.

This paper explores the use of a multivariate statistics approach to monitor the performance of the drier part of the paper making process. The drier is, due to lack of suitable measurements, normally operated in an open loop manner with very little knowledge about its internal states. This is unfortunate, since efficient operation of the drier generally equates to large energy savings and, thereby, lower production costs. The objective in this paper is to model the moisture content of the paper as it travels along the drier. If such a measurement can be reliably estimated, it will not only solve common operational problems, such as over-heating and under-heating of the paper sheet, but also potentially allow for energy minimization.

Multivariate statistics have been emergent methods for large-scale industrial process monitoring since the early

Dr M. J. Wade is with the Institute of Grassland and Environmental Research, North Wyke Research Station, Okehampton, Devon EX20 2SB, UK, Phone: +44 183 788 3564, E-mail: matthew.wade@bbsrc.ac.uk

Dr J. Balderud is with the Industrial Control Unit, University of Strathclyde, Graham Hills Building, 50 George Street, Glasgow G1 1QE, UK, Phone: +44 141 548 2666, Fax: +44 141 548 4203, E-mail: j.balderud@eee.strath.ac.uk

1990s, even though they have been widely utilised in other applications over a much longer period. The common terminology describing the union of industrial applications with these statistical tools is Multivariate Statistical Process Control (MSPC), which is a catch-all term covering process performance monitoring [1], fault detection and diagnosis [2], [3], soft-sensors [4] and advanced control [5], [6].

The most commonly used MSPC methods are Principal Component Analysis (PCA) and Projection to Latent Structures or Partial Least Squares (PLS), which are both dimensionality reducing, data-driven techniques employed, in process terms, to handle common cause variation where standard process control (i.e. univariate statistics) methods cannot [7]. PCA is associated with capturing the underlying trends in large, multivariate systems and is, therefore, suited to detecting and identifying process faults, whereas PLS is more suited to modelling the relationship between independent and dependent process variables (i.e. input-output), providing inferential models or soft-sensors that can be used for prediction of quality variables and process monitoring.

As MSPC has become increasingly prevalent in the process industries, so the requirements for more robust, efficient and practical formulations of the core algorithms have been sought. For example, given that all processes are non-linear to some degree, then non-linear derivations of MSPC methods have been proposed, generally incorporating theory from Neural Networks [8].

Perhaps the most important advance in MSPC in recent years has come with the development of on-line, real-time applications methods handling changes in the intrinsically dynamic data structure of the process plant. Adaptive or Dynamic PCA, for example, has been applied both to theoretical plant simulations [9] and real world processes [10].

PLS is widely used in the chemical industry, especially in association with spectroscopic techniques, for prediction of chemical reaction parameters, circumventing the need for costly, prolonged and repetitive laboratory analysis, as well as utilising the benefits of the non-destructive spectroscopy.

In the near future, MSPC research is likely to become increasingly more focused on its integration with actual automatic control and plant supervision. This is especially pertinent when it is considered that most high-end processes rely heavily on large numbers of automatic controllers and their associated control loops. From simple on/off control to advanced Model Predictive Control (MPC), the potential for MSPC to enhance the supervision (control and monitoring) of industrial processes is now being realised. For example, [6] proposed using the MPC framework and the score space

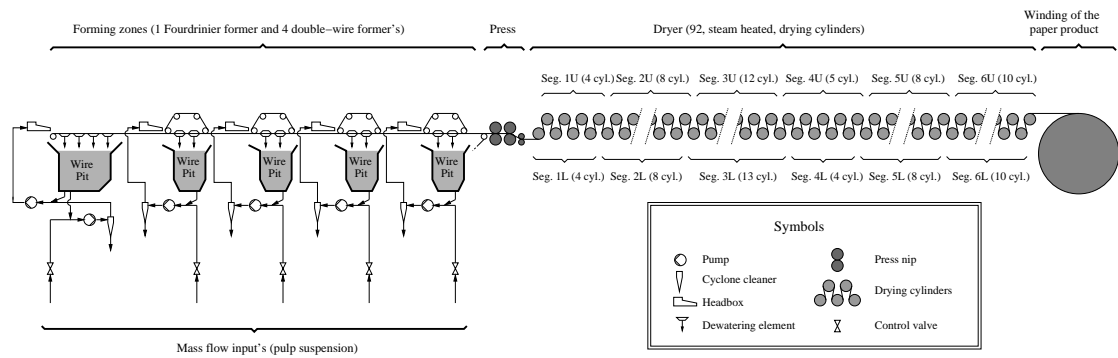


Fig. 1. Simplified process flow diagram, describing a multi-ply paper board machine.

generated by PLS to build a multivariate statistical controller to sit on top of a bank of PID controllers. However, very few papers in this area have been published subsequently, although [11] does present a detailed overview of the current state of MSPC related control.

In this paper, conventional MSPC is applied to the data generated from a paper machine model as a first stage investigation into the capabilities and limitations of the methods as applied to this process. It is envisaged that once a satisfactory statistical model has been identified, further work into novel applications of MSPC with grey-box modelling and automatic control may be initiated, with the simulation model in the first instance and with real plant data, subsequently.

II. PROCESS DESCRIPTION

The production of paper normally starts with a pulp suspension consisting of water, chemicals and cellulose fibres. The pulp suspension is fed to the paper machine, which forms a wet fibre web. Subsequently, by subjecting the wet fibre web to different dewatering processes, the paper machine progressively decreases the water content of the fibre web to ultimately obtain a dry sheet of paper.

The wet fibre web is formed by dispersing the pulp suspension onto a moving nylon mat, which acts as a forming table. The nylon mat is perforated and, by using vacuum suction boxes mounted beneath the nylon mat, water can be drained from the fibre web as it moves along the paper machine. The target paper machine, which is shown in Fig. 1, employs five forming zones, which are fed by five separate pulp feeds. This particular paper machine is used for the production of liquid packaging board products (end uses for this paper product include juice and milk containers), which is a five-layer composite paper product.

The water content of the fibre web, as it leaves the forming zones of the target paper machine, is between 15-20 %. Three press nips further increase the dryness of the fibre web to approximately 40 %, before the fibre web enters the drier. The drier removes much of the remaining water by letting the fibre web travel along the surfaces of steam heated drying cylinders. The target paper machine is equipped with 92 steam heated drying cylinders, divided into 12 groups. The

steam temperature, heating each group of cylinders, can be individually controlled.

A first principles model of the paper machine shown in Fig. 1 was previously developed and validated in [12]. This model tracks the hydraulic transport and mixing dynamics for five mass components (water, filler and three types of fibres), includes a detailed description of the drainage dynamics occurring while the fibre web travels along the forming zones, as well as incorporating a detailed description of the drier process. This model is used in this work to emulate the dynamics of a real multi-ply paper board machine.

III. MULTIVARIATE STATISTICS

As numerical formulation of the standard Multivariate statistics are regularly cited in the literature, only a brief description of the methods employed in this paper are presented. A comprehensive summary of the use of these methods in continuous process monitoring can be found in [13].

A. PLS

Given i observations (samples) described by k dependents (measurement variables), forming a predicted matrix $\mathbf{Y} \in \mathbb{R}^{i \times k}$ and j predictor variables, forming a predictor matrix $\mathbf{X} \in \mathbb{R}^{i \times j}$, the aim of PLS regression is to predict the values of \mathbf{Y} from any given \mathbf{X} . The PLS algorithm performs a simultaneous decomposition on both the \mathbf{X} and \mathbf{Y} matrices, explaining the maximum covariance between them for a given number of latent variables. These latent variables are analogous to components in Principal Component Analysis (PCA), which are orthogonal dimensions explaining the variance within the original dataset in descending order of magnitude. Hence, the first component (latent variable) captures the most information and the k :th component captures the least information, where k is the total number of observations in the original matrix. Typically, an optimum number of components are selected to maximise the amount of original information retained in a PCA or PLS model, without capturing unwanted artefacts such as noise, generally found in the lower order components. Cross-validation is a common method for selecting the number of latent variables to retain in the model, although a number of variants have been proposed [14].

There exists two derivations of the standard PLS algorithm, Nonlinear Iterative Partial Least Squares (NIPALS), [15], and the speedier SIMPLS algorithm [16]. The PLS algorithm (SIMPLS method) can be summarised as follows [16]:

Algorithm 1:

For $l = 1, \dots, a$ (where a is the number of LVs to be retained in the model)

Given the initial conditions:

$$\mathbf{A}_0 = \mathbf{X}^T \mathbf{Y}, \mathbf{D}_0 = \mathbf{X}^T \mathbf{X} \text{ and } \mathbf{C}_0 = \mathbf{I}$$

Calculate:

\mathbf{q}_l , the eigenvector of $\mathbf{A}_l^T \mathbf{A}_l$ associated with the largest eigenvalue.

$\mathbf{w}_l = \mathbf{A}_l \mathbf{q}_l$, $c_l = \mathbf{w}_l^T \mathbf{D}_l \mathbf{w}_l$, $\mathbf{w}_l = \mathbf{w}_l \sqrt{c_l}$, with \mathbf{w}_l stored in the l :th column of the matrix \mathbf{W} .

$\mathbf{p}_l = \mathbf{D}_l \mathbf{w}_l$, with \mathbf{p}_l stored in the l :th column of the matrix \mathbf{P} .

$\mathbf{q}_l = \mathbf{A}_l^T \mathbf{w}_l$, with \mathbf{q}_l stored in the l :th column of the matrix \mathbf{Q} .

$\mathbf{v}_l = \mathbf{C}_l^T \mathbf{p}_l \rightarrow \mathbf{v}_l = \mathbf{v}_l / \|\mathbf{v}_l\|$.

$\mathbf{C}_{l+1} = \mathbf{C}_l - \mathbf{v}_l \mathbf{v}_l^T$ and $\mathbf{D}_{l+1} = \mathbf{D}_l - \mathbf{p}_l \mathbf{p}_l^T$.

$\mathbf{A}_{l+1} = \mathbf{C}_l \mathbf{A}_l$.

The scores matrix can then be computed as the product of the original predictor matrix and the weights matrix ($\mathbf{T} = \mathbf{XW}$) and the regression matrix for \mathbf{Y} on \mathbf{X} is the product of the weights matrix and the loadings matrix ($\mathbf{B} = \mathbf{WQ}$).

B. PCR

Principal Components Regression (PCR) is similar to standard PLS regression, but prior to regression, PCA is performed on \mathbf{X} , before applying the multiple linear regression algorithm with the scores (\mathbf{T}) as the predictor variables and \mathbf{Y} as the predicted variables.

IV. ANALYSING THE PERFORMANCE OF THE DRIER

The target paper board machine produces a wide range of liquid packaging board products, ranging from 200 g/m² grades up to 450 g/m² grades. The training and validation datasets comprise data collected from the paper board machine process during normal production. Each dataset encapsulates a reduced production cycle consisting of selected grades that span the range of grades produced on the paper machine. A slightly different operational strategy has been employed when generating each dataset. This is to capture the effect on the process variables from the different operational strategies that the paper machine normally is operated under.

In total, three cycles were used for the training datasets and a fourth cycle for the validation dataset.

Each dataset include measurements of basis weight (at the end of the machine), moisture (at the end of the machine), machine speed, the five mass flows (for each pulp feed) and the twelve steam pressures (for each steam group). In addition, each dataset also include measurements of the paper moisture at 183 positions along the machine direction (MD)

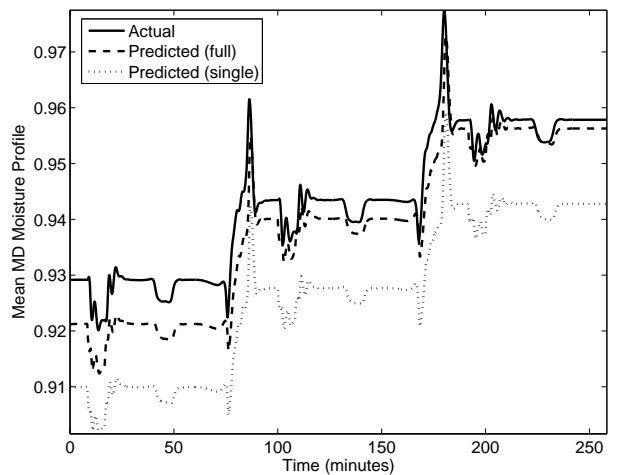


Fig. 2. Mean MD Moisture Profile as predicted using two different training sets, compared with the actual Profile.

of the drier. Note that these measurements are normally not available in the on-line measurement database. However, they are collected regularly by the maintenance team and therefore available for off-line analysis exercises, such as model identification and validation.

The training predictor matrix ($\mathbf{X}_t \in \mathbb{R}^{4653 \times 20}$), which include measurements of basis weight, moisture, machine speed, mass flows and steam pressures, and the predicted matrix ($\mathbf{Y}_t \in \mathbb{R}^{4653 \times 183}$), which include measurements of the MD moisture profile along the drier, were used to build a PLS regression model using the SIMPLS algorithm, with autoscaling (data is mean centred to zero with a standard deviation of one) and seven latent variables determined by leave-one-out cross-validation. The SIMPLS algorithm was preferred to the NIPALS algorithm for speed and increased predictive robustness [16].

This model was then validated using the fourth dataset ($\mathbf{X}_v \in \mathbb{R}^{1551 \times 20}$, $\mathbf{Y}_v \in \mathbb{R}^{1551 \times 183}$), using the same scaling options. As a comparison, a supplementary model was generated using only one cycle of the training set and this model was validated in the same manner.

The two fault datasets are similar to the training and validation datasets in that they consist of a reduced production cycle. However, unlike the training and validation datasets, the heat transfer coefficient, describing the heat transfer between the steam and cylinder shell, has been significantly reduced for some of the cylinders. Reducing the heat transfer coefficient emulates the effect of condensate build-up inside the cylinder; a fault that frequently occurs during paper making. For Fault 1, the heat transfer coefficients for cylinders 48, 50, 52, 54 and 56 have been reduced, while for Fault 2, the heat transfer coefficient for cylinder 52 has been reduced.

V. RESULTS

The accuracy of the model was assessed by averaging the model output (MD moisture profiles for all positions), giving $\hat{\mathbf{Y}}_v$ (actual), $\hat{\mathbf{Y}}_m$ (complete model), $\hat{\mathbf{Y}}_{m1}$ (single training dataset model), as shown in Fig. 2. The corresponding Root

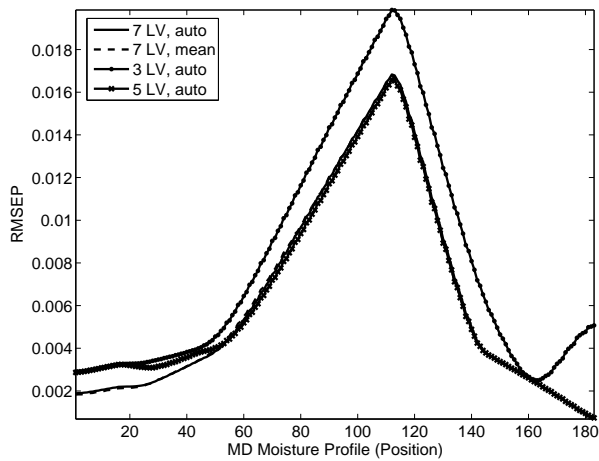


Fig. 3. RMSEP of the MD Moisture Profile for several model configurations.

TABLE I

AVERAGE RMSEP FOR EACH OPERATING CONDITION FOR THE TWO MODELS

	NOC	Fault 1	Fault 2
PLS	0.0025 ± 0.0006	0.0206 ± 0.0097	0.0049 ± 0.0020
PCR	0.0026 ± 0.0006	0.0231 ± 0.0116	0.0054 ± 0.0023

Mean Square Error of Prediction (RMSEP), an indicator of model accuracy, is shown in Fig. 3, for four different model configurations; the optimal model (7 LVs, autoscale), mean centred scaling and two sub-optimal LV selections.

As can be seen, the model generated using only 3 LVs performs poorly in comparison to the other configurations, as would be expected. This is because the 3 LV model does not capture enough information to allow for robust prediction. The RMSEP for the autoscaled and mean centred configurations are almost identical but, as for all configurations, the largest RMSEP values occur in the central MD moisture profile positions.

PLS and PCR models were both developed using the optimal parameters, but the PLS model was shown to give slightly better predictions, as shown in Table I and, subsequently, Fig. 2 and 3 relate to the PLS model only.

The optimal model was used to in a pseudo-real-time manner to predict the output variable Y under the three operating conditions. The model was run in real-time, but the data was analysed off-line at the end of the run. The mean error (measured - predicted) was calculated for each condition and the results are shown in Fig. 4, where NOC is the Normal Operating Condition. As can be seen, the model performs well when no fault is present, but its predictive ability deteriorates to a degree dependent on the severity of the fault (i.e. poor performance with fault 1).

The scores plot shown in Fig. 5 displays the predicted model scores on the first latent variable versus the second latent variable for each operating condition, with the 95%-ile confidence limit ellipse. The NOC scores are entirely

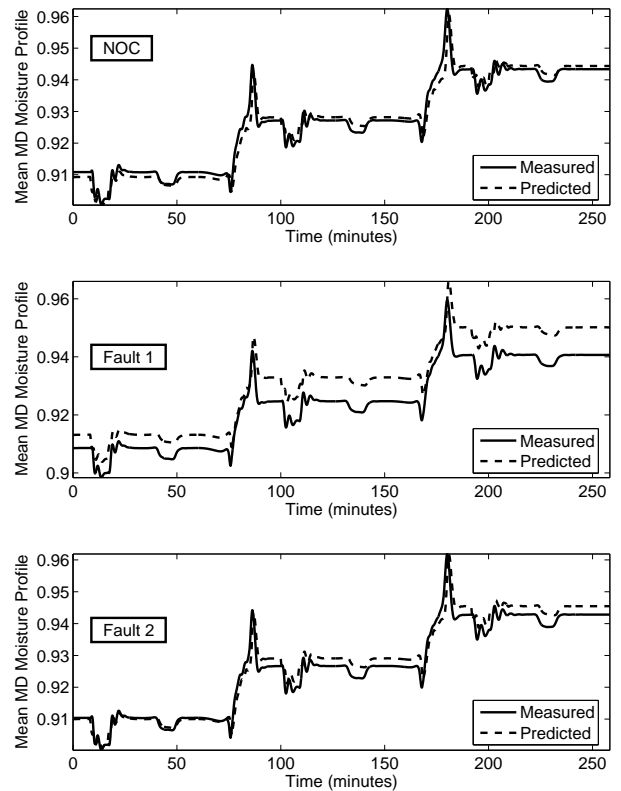


Fig. 4. Comparison of Predicted and Actual MD Moisture Profiles for the three process runs (Normal (NOC), Fault 1 and Fault 2).

within the confidence limit on the first LV, whereas the fault scores drift outside the limit. However, some scores for all conditions fall outside of the limit on the second LV. This suggests that LV 1 describes scores related to faults, whereas LV 2 is related to model prediction errors. These prediction errors are associated with variations due to step changes in the model occurring at approximately 10, 100 and 190 minutes. Improved model robustness is required to account for these step changes.

The model was investigated to determine whether it was capable of providing further information about the faults. The two most commonly used measures of model variation are the Squared Prediction Error (SPE), referred to as the Q Residual and Hotelling's T^2 Statistic. When confidence limits are applied to these values, some indication of the status of the process can be determined. The Q Residual measures variation outside of the model, providing information on the variation in the prediction errors, whereas the T^2 Statistic measures variation in the model subspace as a function of the distance from the centre of the training dataset.

The T^2 Statistic with 95%-ile confidence limit for each plant operating condition is shown in Fig. 6, for both the PLS and PCR models, whilst only the PCR model is used to display the Q Residual (and 95%-ile confidence limit), as shown in Fig. 7. It was discovered that, for Q in the PLS model, variable 7 was masking the true variation, i.e. the contribution of variable 7 to the Q Residual was close

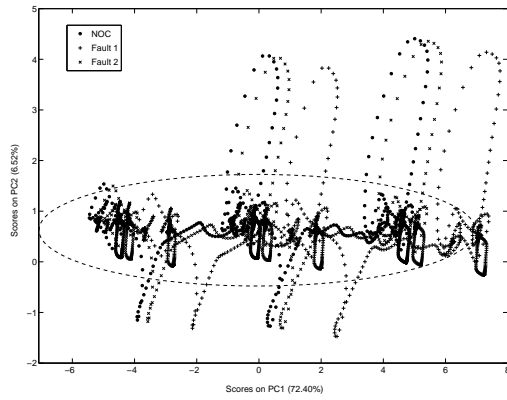


Fig. 5. Scores plot for each operating condition, with 95%-ile Confidence Limits.

to 100%. Examination of the raw data revealed that both variables 3 and 7 were constant mass flows. Removing these variables from the model, however, resulted in negligible Q values (near zero). The PCR method appeared to handle this problem, although removing the two constant variables resulted in a lowering of the Q values to a negligible level. In this regard, the Q Residual should only be used as an indicator of problems associated with generating a model with good closeness-of-fit.

There are two significant observations that can be made regarding the monitoring plots. Firstly, the step changes are detected by both monitoring statistics for all conditions and the magnitude of the statistics increases for each subsequent step change. The fact that this phenomenon is detected by both the Q and T^2 statistics, as well as during the NOC, suggests that more training data is required to generate a model that can better account for normal process changes. Although the model is fairly well conditioned, the step change is detected both inside and outside of the model subspace and, thus, is highlighted as a significant process change. As mentioned, caution should be taken when regarding the Q Residual as an indicator of model error, as the variable 3 and 7 appear to contribute to the unusually high values shown in Fig. 7.

The second observation relates to the differences in the statistics between NOC and the two faults. The T^2 statistic for Fault 1 is uniformly above its confidence limit, for Fault 2 sporadically and for the NOC, only during the step changes. It should also be noted that the PCR model gives lower T^2 values, which make it harder to distinguish whether there is a true fault, especially in the case of Fault 1. It appears that the PCR model is more robust to particular phenomenon in the process data. Further work is suggested to explore the role of these models in fault detection and to find methods of diagnosing particular faults by looking at the variable contributions to the monitoring statistics, especially with regard to improving the robustness of these statistics.

The standard prediction errors for model predictions during normal operating conditions are shown in Fig. 8. This

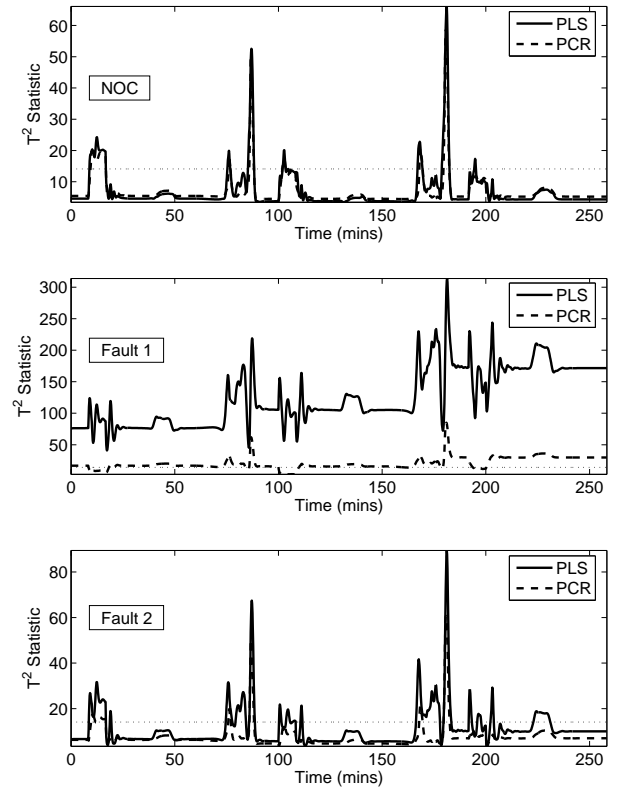


Fig. 6. T^2 Statistic for each operating condition (PLS and PCR models), with 95%-ile Confidence Limits.

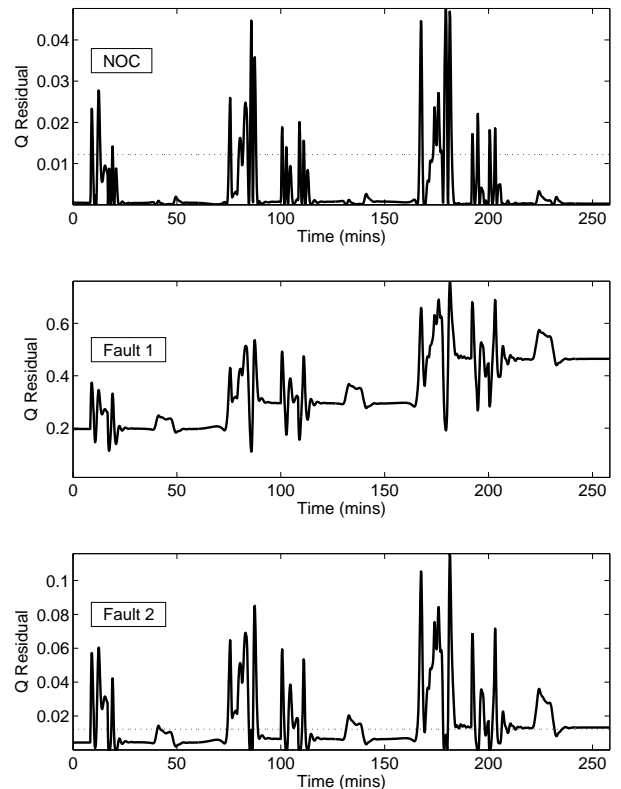


Fig. 7. Q Residual for each operating condition (PCR model), with 95%-ile Confidence Limits.

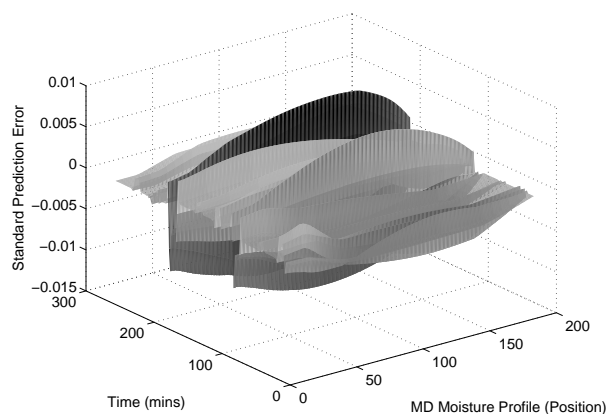


Fig. 8. Standard Prediction Error over time for each MD Moisture Profile Position (Normal case).

figure essentially describes the accuracy, by which the model can predict the MD moisture profile at particular time instants and particular positions along the drier. As shown, the prediction errors are overall small (natural variation of the process is in reality in the order of ± 0.01), and without structure, suggesting high accuracy over a relatively wide range of operating conditions. Exceptions from this occurs at $t \approx 75$ minutes and $t \approx 175$ minutes, which coincide with large step changes that excite the non-linearities of the process. Thus, under a very limited and rare set of operating conditions, where the non-linearities of the process are continuously excited, the accuracy of the predictions is expected to rapidly deteriorate.

VI. FURTHER WORK

Clearly, the potential for expanding on the work presented is high, given both the limitations of the techniques used and the evolving nature of MSPC technologies/Chemometrics. It is proposed that an iterative program of work will be undertaken in order to achieve improved paper machine plant monitoring and control.

The key areas that form the basis of the work have been identified as:

- 1) Fault handling Expanding on the current work to include several artefact classes (i.e. Measurement errors, Disturbances, Controller dynamics). Improved modelling, based on hybrid techniques (i.e. Discriminant-PLS), will be required to distinguish between and diagnose different artefact classes;
- 2) Real-time Implementation of an adaptive monitoring scheme for the plant to enable truly real-time monitoring. This will include on-line model updating of both the model and the monitoring statistics [10].
- 3) Control Implementation of MSPC with the advanced control (MPC) scheme existing in the current plant model. Expansion of the monitoring and control framework, with the intention of making the methodology viable for the paper making industry and real-world plants.

It is anticipated that each area is presented in separate papers, with a final paper collating the results into a cohesive document detailing the final MSPC-MPC framework.

REFERENCES

- [1] J. Kresta, J. MacGregor, and T. Marlin, "Multivariate statistical monitoring of process performance," *The Canadian Journal of Chemical Engineering*, vol. 69, pp. 35–47, 1991.
- [2] R. Dunia, S. Qin, T. Edgar, and T. McAvoy, "Identification of faulty sensors using principal component analysis," *AIChE Journal*, vol. 42, no. 10, pp. 2797–2812, 1996.
- [3] R. Isermann, "Supervision, fault-detection and fault-diagnosis an introduction," *Control Eng. Practice*, vol. 5, no. 5, pp. 639–652, 1997.
- [4] Z. Han, S. Shah, C. Robson, R. Patwardhan, and J. Pakpahan, "Development of softensors for a pulp bleaching process using partial least squares," in *Proc. IEEE Advanced Control Workshop*, Vancouver, Canada, April 2003.
- [5] M. Piovoso and K. Kosanovich, "Applications of multivariate statistical methods to process monitoring and controller design," *Int. J. Control*, vol. 59, no. 3, pp. 743–765, 1994.
- [6] G. Chen, T. McAvoy, and M. Piovoso, "Multivariate statistical controller for on-line quality improvement," *J. Process Control*, vol. 8, no. 2, pp. 139–149, 1998.
- [7] B. Ogunnaike and W. Ray, *Process Dynamics, Modeling and Control*. New York: Oxford University Press, 1994.
- [8] M. Kramer, "Nonlinear Principal Component Analysis using Autoassociative Neural Networks," *AIChE J.*, vol. 37, no. 2, pp. 233–243, 1991.
- [9] K. Chow, K.-J. Tan, J. Z. H. Tabe, and N. Thornhill, "Dynamic Principal Component Analysis using integral transforms," in *AIChE Annual Meeting*, Miami Beach, November 1999.
- [10] M. Wade, A. Sanchez, and M. Katebi, "On real-time control and process monitoring of wastewater treatment plants: Real-time process monitoring," *Trans. Inst. Meas. Cont.*, vol. 27, no. 3, pp. 173–193, 2005.
- [11] T. Kourti, "Application of latent variable methods to process control and multivariate statistical process control in industry," *Int. J. Adaptive Cont. Signal Proc.*, vol. 19, pp. 213–246, 2005.
- [12] J. Balderud, "Model-assisted optimization of a multi-ply paper board machine," PhD Thesis, Karlstad University, Sweden, Department of Chemical Engineering, 2004.
- [13] P. Goulding, B. Lennox, D. Sandoz, K. Smith, and O. Marjanovic, "Fault detection in continuous processes using multivariate statistical methods," *Int. J. Sys. Sci.*, vol. 31, no. 11, pp. 1459–1471, 2000.
- [14] J. Jackson, *A User's Guide to Principal Components*. Wiley, 1991.
- [15] H. W. Jackson, "Estimation of principal components and related models by iterative least squares," in *Multivariate Analysis*, P.R. Krishnaiah, Ed. Academic Press, 1966, pp. 391–420.
- [16] S. de Jong, "Simpls: an alternative approach to partial least squares regression," *Chem. Lab. Intell. Sys.*, vol. 18, pp. 251–263, 1993.