university of groningen / faculty of mathematics and natural sciences / chemistry and chemical engineering

Development of a predictive model for the moisture content in a paper mill

This report contains proprietary information and is confidential

Master Thesis Chemical Engineering
Author: B.K. Dijkstra
Student number: 1580450
Supervisor: prof. dr. ir. B. Roffel
Second corrector: prof. dr. F. Picchioni
External supervisors: J. Dercksen & R. van Wieringen (Sappi Nijmegen)
Groningen, April 2011

Abstract

In the production of paper, one important paper quality parameter is the moisture content. However controlling paper quality is a difficult task as the measurement of this quality is affected by a large dead time in the production process.
In order to overcome this problem, one possibility is to develop a predictive model based on the key variables affecting the paper quality, that are measured more frequently. In this way it should be possible that the paper quality parameter can be determined at an early stage of the production process. This thesis describes the development of such a predictive model for the Sappi paper mill in Nijmegen. Analysis of the raw process data with the use of principal component analysis yielded that the process data was arranged in clusters of consecutive data. The cause of this arrangement was found in the variation of the average values of the process input variables. Although many modelling techniques are available, in this thesis a linear regression technique (PLS) as well as a non-linear technique (Robust LSSVM) was used for the development of a predictive model. It was found that both techniques showed comparable results for the prediction of the paper quality, therefore for the predictive models, PLS was preferred because of its simplicity. The initial developed PLS models in combination with a bias update showed promising results regarding the goodness of fit. However the models used a too small range of variables close to the measurement of the paper quality, therefore these models are not real predictive models. However these models can be used as a substitute measurement of the paper quality in case the real measurement fails. The final PLS model is based on manually controllable variables and important process values, in combination with an online update of the bias term in the PLS model. Due to the nature of these variables, this model is more likely to be used for controlling the paper quality. The final model showed a good fit both with and without the online update incorporated. One remark has to be made however. The developed model is only tested for the production of paper with a gram weight of 90 g/m^2. As the paper mill produces a wider range of gram weights, it is required to perform the analysis presented in this thesis again for the other paper gram weights.

Preface

The last year of the Chemical Engineering master program at the University of Groningen consists of a self chosen research project in one of the fields of chemical engineering present in the department. Although the department offers a variety of topics in product engineering, I was more interested in topics in process engineering.
During the bachelor program Chemical Engineering at the University of Groningen I attended the courses Process Dynamics and From Data to Model, given by prof. dr. ir. B. Roffel. During these courses multiple modelling techniques were explained on the basis of their use in (chemical) processes. During these courses it intrigued me what the effects of a well-trained model are for the overall operation of a process.
Furthermore with a research project in the discipline of Process Dynamics and Control, it was possible to combine my common interest in both chemical engineering and IT.

Fortunate enough prof. dr. ir. B. Roffel had an opportunity to perform a research project in the discipline of Process Control. This research project focused on the development of a predictive model in a paper mill. This research project appealed to me immediately, as the results of such a research project can be applied in real life and do not only exist in theory. Ultimately this thesis is the result of this research project.

At this point I would like to thank prof. dr. ir. B. Roffel for his guidance during this research project and the opportunity to perform a research project under his supervision. Whenever advice was needed or alternative approaches were required I could revert to his experience.
Furthermore I would like thank Joost Dercksen and Rene van Wieringen from Sappi Nijmegen for the possibility to perform a research project at the Sappi paper mill in Nijmegen, the supply of process data and their feedback on the results of this research project.

Overall I can say that this research project was a great experience.

20th of April, 2011

Bart Dijkstra

Index

1. Introduction

1.1 The history of paper making

In today's world, paper has become a common material used in a variety of applications. It is, for example, used as:

- Copy paper for printers, copying machines and writing
- Newsprint
- Wrapping and packaging
- Hygienic tissues
- Paper currency

These numbers of applications are in sharp contrast with the beginning of the paper production when paper was only used for calligraphy. The first known production of paper is ascribed to the Chinese Cai Lun and dates back to 105 AD. The first paper was made from mortared bark of trees, hemp, rags and fishnets. In subsequent years the process was improved for use with other raw materials, however, paper making was still only performed in China. It took more than five hundred years before the paper making started to spread around the world. It took another five hundred years before the paper making became known in Europe. From then on the process became increasingly mechanized in order to obtain a higher production and reducing the production costs. In 1808 this resulted in a design (figure 1.1) of a paper machine which still counts as a reference for the basic design of current paper machines.

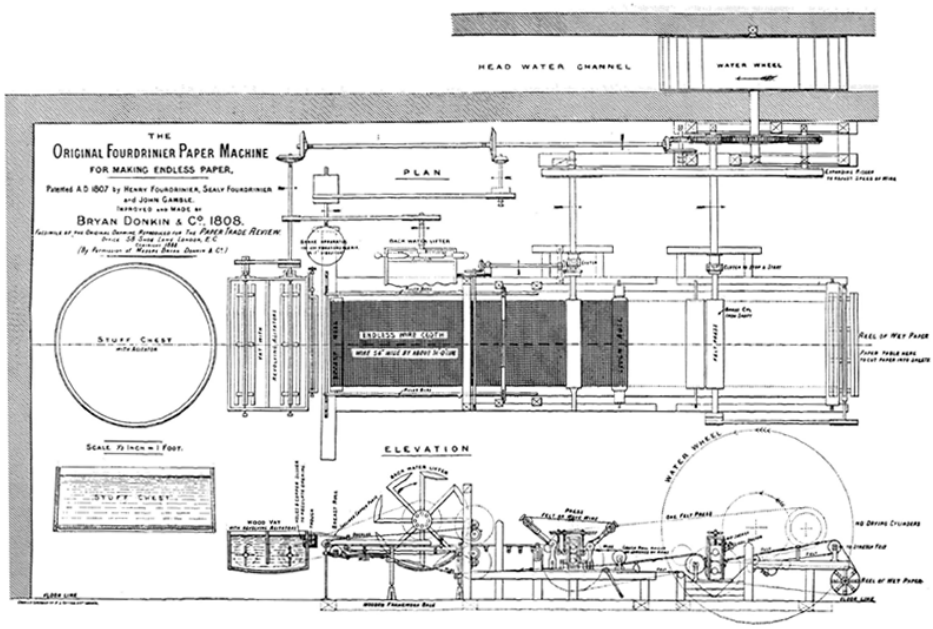

Figure 1.1 Basic paper machine design dating from 1808 [19]

The basic paper machine was equipped with additional features (i.e. refiners, dryers, coaters and calendars) in subsequent years, mainly to obtain an increased production at lower costs but also to produce a more consistent product and a larger variety of products. [8]
Modern paper machines have evolved to advanced pieces of machinery with large production capacities, however, paper mills are still facing difficulties.

On the one hand, customers of paper mills become more and more demanding when it comes to price/quality ratio of the products. On the other hand the paper mill has to meet stricter governmental and environmental regulations while also maintaining an economic viable process operation. These difficulties can partially be overcome by automating the production process, thus reducing expensive manual labour and maintaining a more constant production process. The latter is, however, already accomplished on modern paper machines as these machines have been equipped with evolving advanced computer control since the 1960s.
The current trend to overcome the difficulties is to detect deviations from normal process operation at an early stage of the process with physical and/or statistical models.

1.2 Literature review

Research on physical and statistical models in the paper making industry has primarily focused on three topics in recent years, i.e. multivariate process monitoring, model based prediction of sheet breakage and model based control of the dryer section. Although research is performed on each topic individually, one will see an overlap in the results of the studies described in the upcoming sections.

1.2.1 Multivariate process monitoring

Modern paper machines are equipped with different computer controls that generally operate individually or in a limited number of control loops. The consequence of such a distributed control system is that large amounts of information are presented to operators at the same time. This can lead to an information overload that might result in a wrong, or even no response of an operator in case of deviant process operation. It would therefore be advantageous if the vital process information could be presented to operators in workable amounts. This implies that the dimension of the original process data has to be reduced. A commonly used technique for data reduction in the process industry is principal component analysis (PCA). The basic idea behind PCA is to translate the original process data into a new linear set of variables (principle components) that describe maximum variability of the process data. As PCA will be used in this thesis a more thorough discussion will be given in a subsequent chapter.
In recent years, Bissessur *et al.* [3] and Skoglund *et al.* [17][18] have performed research on the development of multivariate process monitoring models.
The PCA model in the research of Bissessur *et al.* was developed for a dataset consisting of 60 process variables and 565 measurements representing a good production run. With PCA, the number of variables could be reduced to eight new variables that described 80% of the original variability. For monitoring of the process, it is suggested to use the scores (i.e. information how measurements relate to each other) of the lower order sixth and eight principal component. The underlying idea for this suggestion is that higher order principal components describe the major sources of variability, whereas the lower order principal components can already describe small process deviations before the deviations become a major source of variability. In addition to a score monitoring plot, Bissessur *et al.* suggest the use of the squared prediction error (SPE) of the residuals (or Q-residual) as a second monitoring parameter. As the name indicates, this parameter is calculated as the square of the difference between the new observation and the reference predicted value. For a trained PCA model, a specific Q-residual value is calculated which represents the boundary of normal process operation. In case of an unusual process deviation the Q-residual value will move out of this boundary. Bissessur *et al.* demonstrated, after a validation of the model with data from a process malfunction, that the developed model was capable of detecting the process deviation. In combination with contribution plots of the original variables it was even possible to identify the process variable that caused the deviation. Although the demonstration yielded the expected result, Bissessur *et al.* emphasized that for a more efficient process monitoring a new model should be developed for data that spans the full space of process operation.

The initial research of Skoglund *et al.* also focused on the development of a PCA monitoring model. For the development of the model a dataset consisting of 177 variables and 268 measurements was used. With PCA, the number of variables could be reduced to six principal components that described 55% of the original variability. Monitoring with this model is based on the score plot of the first two principal components and the deviation of each process variable to the PCA model. A demonstration of the model in a production facility showed that the model yielded stable predictions for all products produced, although the model was developed with data from a few of these products. Skoglund *et al.* imply that the latter is the result of the 'grey-box' nature of their model which indicates that both empirical data as well as process knowledge have been used in the development of the model. For a consistent robust model, Skoglund *et al.* emphasize that the model should be updated regularly in order to compensate for possible changes in the process operation. In the subsequent research of Skoglund *et al.* the original PCA model is extended with a knowledge based system (KBS). The most basic form of a knowledge based system consists of a number of predefined rules that take action when a certain criterion occurs. According to Skoglund *et al.* the performance of monitoring increases with the implementation of a KBS as variables can be selected or ignored more accurately for the PCA model. In addition the KBS can select a different model based on the measurements, thereby selecting the optimal model for each situation. The latter will help to improve the detection of deviant process behaviour, thus a more consistent process operation can be achieved.

1.2.2 Model based prediction of sheet breakage

Increasing the production of paper mills was partly accomplished by increasing the machine speed. However due to this high speed and the large number of cylinders in the paper machine, the paper sheet is subjected to high shear- and tear forces. These forces may lead to breakage of the paper sheet in case of weaknesses in the paper sheet. Weaknesses are, for example, excessively dry or wet paper sheet, too small fibres or uneven distribution of the fibres in the sheet.

The occurrence of a sheet breakage involves down time of the production process as well as loss of material, thus resulting in large costs. It is estimated that sheet breakage account for 2 – 7 percent of the total production loss for modern paper machines. [2] Therefore reducing the number of sheet breaks is an important factor for the performance (i.e. production) of a paper mill. Modern paper mills are equipped with a visual surveillance system that tracks the paper sheet in a number of critical points. In this way sheet breaks can be detected and it might be possible to find the cause of the break, but it is not possible to indicate whether a paper sheet will break or not. Predicting the occurrence of sheet breakage at an early stage of the production process would therefore be advantageous. As far as known, Li [9] was the only researcher who developed a successful multivariate predictive model for sheet breakage. In the research of Li, Partial Least Square (PLS) regression is used for the development of the predictive model. The basic idea behind PLS is similar to PCA, however with PLS a distinction is made between the process input variables and the process output variable(s). Ultimately PLS results in a linear equation of new variables (latent variables) which describe both maximum variability of the process input variables as well as maximum correlation between the process input variables and the process output. As PLS will be used in this thesis a more thorough discussion will be given in a subsequent chapter. In addition, Li used PCA for the development of a monitoring tool which is combined with a KBS in order to make correct decisions in case of abnormal process operation. Li developed a PLS model based on 43 variables and 3180 measurements. With PLS, the number of variables could be reduced to three latent variables that describe 92.67% of the variance in the Y (process output)-block and 74.20% of the variance in the X (process inputs)-block. Although this PLS model predicted the sheet breaks in the validation dataset correctly, Li found that some of the used variables are not true predictors of sheet breakage. These variables vary

simultaneously with the predicted variable thus yielding a strong correlation which is the basis of a PLS model.
In order to overcome this problem, Li used a pre-treatment in which all data in and during a sheet break was given the value of the corresponding set point. Subsequently a new PLS model was developed consisting of three latent variables that described 60.13% of the variance in the Y-block. With this PLS model Li was able to find the four most important predictor variables for sheet breakage. The variables consisted of the dryer steam pressure, dryer differential pressure and sheet formation rolls loadings. Li implied that this combination of variables ultimately is a representative of the moisture content of the paper sheet and thus a plausible reason for the occurrence of sheet breakage. In the validation of the PLS model with new data, as can be seen in figure 1.2, Li showed that the model was capable of predicting 92% of the sheet breaks correctly.

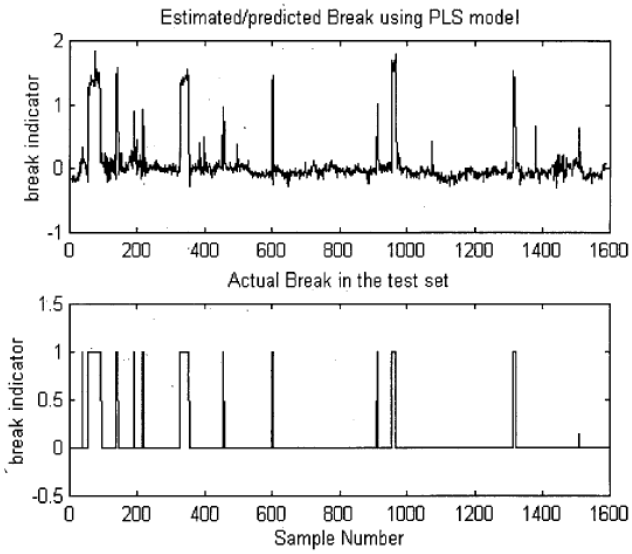

Figure 1.2 Results of the validation with new data in the study of Li. [9]

1.2.3 Model based control of the dryer section

Paper making requires large quantities of water. For one kilogram of paper, approximately 100 litres of water is required. [14] The larger part of this water is removed from the paper sheet by natural dewatering and mechanical dewatering. In terms of moisture content of the paper sheet it is reduced from roughly 97% to 45% - 55% in these steps. Product specifications require, however, a moisture content of 2% - 5% of the paper sheet. Although it appears that a large amount of water has to be removed in the dryer section based on the percentages, in reality this corresponds to an amount of 1 kg – 2 kg of water.[10] As this water is retained between the fibres through hydrogen bonds, it requires a large amount of energy to break these bonds and subsequently evaporate the water. Energy is applied to the paper sheet with the use of steam heated cylinders. The steam supply is currently controlled in a cascaded control loop. With a cascaded control loop the temperature of the most important drying cylinder is controlled based on the moisture content measurement and subsequently the temperature of the other drying cylinders is adjusted generally according to specified ratios. In this way it is possible to obtain the desired temperature gradient throughout the dryer section.
In case of an increase or decrease in the moisture content of the paper sheet, the set point of the drying cylinder has to be changed. As it takes some time to heat or cool the cylinder this change will not have an immediate effect on the temperature profile. Ultimately this will result in off-spec product or an increased probability of sheet breakage.

As direct quality measurements are not available in the dryer section, research has primarily focused on the development of a physical model.
As physical models describe a process based on physical and chemical laws of conservation (e.g. energy or mass balance); such a model can be used for better understanding of important factors influencing the dryer section.
Slätteke *et al.* [19] have developed a physical model for a steam heated dryer cylinder. With the knowledge obtained, Slätteke *et al.* first optimized the currently used control loop for the steam pressure. In addition two extensions of the control loop were suggested. The first suggestion involves a feedback control system based on the airflow through the drying section to assist the steam pressure control.
By controlling the air flow, the uptake of evaporated water can be controlled by means of dew point variations. The dew point measurement is a good representative of the moisture content of the paper sheet and has a fast response time to changes compared to the response time of the steam pressure control. Ultimately this results in a faster adjustment to a new situation thus, reducing the time where off-spec product is produced. The second suggestion involves a feed forward control system based on a surface temperature measurement. On the basis of a number of these measurements, it is possible to determine deviations in moisture content of the paper sheet. As these measurements are suggested to be performed in the first stage(s) of the dryer section it is possible to make an adjustment for the final stages, thus preventing the production of off-spec product. As can be seen in figure 1.3, the effect of a change in moisture content can be suppressed with the use of one or both of the suggested control systems.

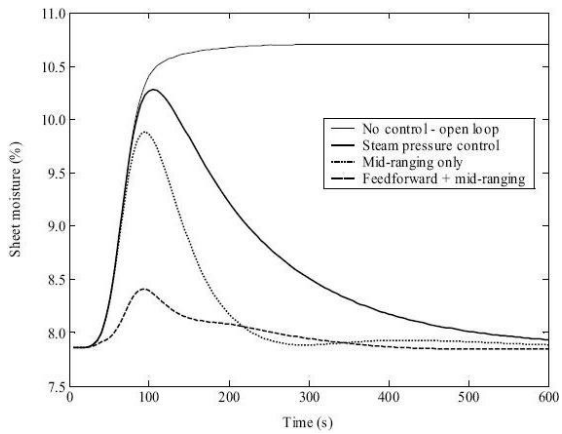

Figure 1.3 Comparison of the different suggested control systems in the study of Slätteke *et al.* [19]

Although these control system have not yet been tested in a real life situation, it showed a high potential for subsequent research.
In the research of Li *et al.*[10] another alternative feed forward control system is described on the basis of the vacuum dewatering boxes in combination with the actual moisture content measurement. Li *et al.* suggest the use of air flow measurements of the vacuum dewatering boxes as a substitute measurement for the moisture content of the paper sheet. The underlying idea of this control strategy is to fully utilize the vacuum dewatering boxes in order to minimize the energy requirement of the dryer section. An indicative validation of the model obtained, showed that the model could compensate for an error in the incoming moisture content up to 10% and a subsequent error of 8% in the middle of the process. However these results were not reliable and accurate enough to conclude that the model was an improvement for the control strategy. Li *et al.* emphasized that the model requires more optimization before it can be concluded that this model is an improvement for a real control scheme.

Although the validation yielded positive results, Li *et al.* emphasized that the model requires more optimization before it is robust and accurate enough for use in a real control scheme.
Wade *et al.* [23] performed an exploratory study on the use of multivariate regression techniques for the prediction of the moisture profile in the dryer section. Wade *et al.* used twenty online measured variables ranging from the basis weight of the paper sheet to steam pressures in the dryer. In addition 183 offline measurements of the paper moisture were included in the training data sets. The optimal PLS model consisted of 7 latent variables. Wade *et al.* found that under normal operating conditions, the model could predict the moisture content well (figure 1.4).

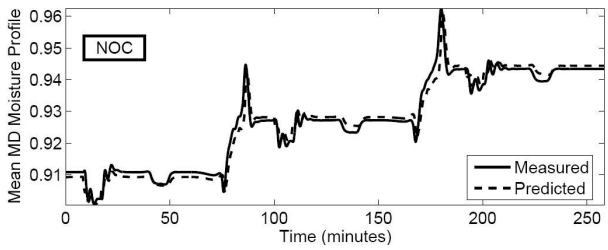

Figure 1.4 Predicted and measured moisture profile at normal operation conditions. [23]

Wade *et al.* used two additional dataset in which, deliberately, respectively five and one fault(s) were incorporated. The results thus obtained (figure 1.5) showed that the predictive ability of the model deteriorated dependent on the severity of the fault.

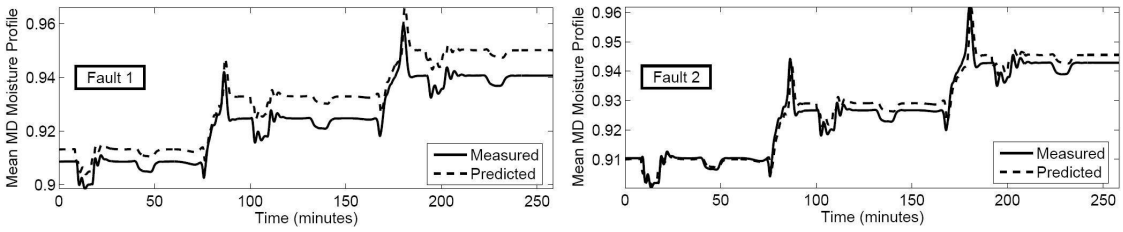

Figure 1.5 Predicted and measured moisture profile in case of faults present in the dataset.
Fault 1 contained five faults; Fault 2 contained one fault. [23]

Based on the results obtained for normal operation conditions Wade *et al,* however, concluded that the model showed a high potential for subsequent research.

1.3 Thesis scope

The literature review showed that multivariate statistical modelling is already used for a number of applications in the paper industry.
From the literature review it can be concluded that one of the most important paper quality parameters in the production process of paper is the moisture content. Furthermore it can be concluded that the moisture is also one of the most difficult parameters to control in the production process. The latter is mainly caused by the large dead time which arises from the location of the measurement of the moisture content almost at the end of a paper machine.
It would be advantageous to measure the moisture at additional stages of the process. On the one hand this requires a large investment for the equipment while on the other hand the placement of the equipment turns out to be difficult. This is caused by the fact that the paper sheet has to move freely for this type of measurement and this is not possible at all stages of the process.
The other option is to determine a statistical model that can predict the moisture content, or paper quality, based on a number of key variables at an early stage of the production process which have influence on the paper quality.

This thesis will show the steps involved in the development of such a predictive model. The steps performed in this thesis can be summarized as follows:

1. Perform a multivariate statistical analysis to determine whether the process operation is in control. In case the process is out of control an additional analysis is required to determine a possible cause.
2. Determine the variables that influence the paper quality.
3. Develop a predictive model on the basis of the variables obtained in step 2.

Ultimately the development of a predictive model should result in a better understanding of which variables influence the paper quality the most. Due to this it should be possible to control the paper quality more consistently thus resulting in a more consistent process operation which in turn should translate into lower production loss and a reduction in energy consumption.

2. Process Description

In this chapter an overview of the production facility will be given from which data is used for the development of a model for the paper quality. Furthermore an overview will be given of the production process applied in this production facility. Finally an overview will be given of the process data used for the development of the predictive model.

2.1 Sappi paper mill Nijmegen

The research presented in this thesis has been performed in association with the Sappi paper mill in Nijmegen. The Sappi group has its origin in 1936 and currently consist of eighteen paper mills that collectively produce 6.8 million tons of coated fine paper per annum. This makes Sappi the number one of the leading producers of coated fine paper globally. Apart from this, the Sappi group also produces paper pulp, chemical cellulose, uncoated paper and packaging paper. [15] The Nijmegen paper mill has its origin in 1908 and since 1997 it is part of the Sappi group. [7] The paper mill operates a single paper machine (PM 7) with a maximum production capacity of 300.000 tons per annum. The Nijmegen paper mill exclusively produces reels of two-sided double coated fine paper with gram weights ranging from 90 g/m^2 to 170 g/m^2 in three different types of finish. Paper with these finishes is known under the brand name Royal Roto with the suffix Matt, Silk or Brilliant Plus for the finish used. [13] Throughout this thesis these finishes will be denoted as grade A, B and C. An overview of the production facility in Nijmegen is shown in figure 2.1.

Figure 2.1 Production facility of the Sappi paper mill in Nijmegen. [13]

2.2 Process description

The production process of paper used in the Nijmegen paper mill can be divided into seven sections as schematically shown in figure 2.2.

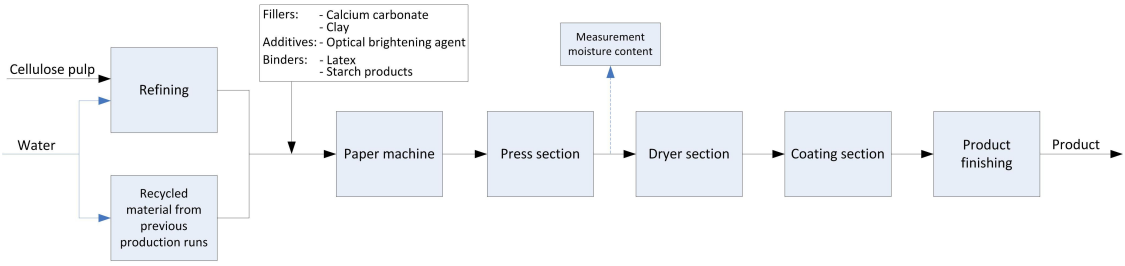

Figure 2.2 Schematic representation of the paper making process in the Nijmegen paper mill.

The starting material for the paper making process consists, on one hand, of recycled material from previous production runs which is re-pulped in water. On the other hand the starting material consists of bleached pulp of cellulose from a variety of coniferous trees and deciduous trees.
On entering, the cellulose pulp is grounded into a suspension of fibres in water in order to obtain a larger surface area which is translated into better properties of the final paper sheet.
Before entering the paper machine auxiliary materials (i.e. fillers, additives and binders) are added to the suspension of fibres and recycled material.
In the paper machine the suspension is evenly distributed over the entire width of the wire. Along the wire section the paper sheet is formed since water can fall through the wire, whereas the fibres and auxiliary materials accumulate on the wire. Initially dewatering only occurs through gravitational forces. Once the natural dewatering has taken place, mechanical dewatering (i.e. suction boxes) is applied on both sides of the paper sheet. Subsequently the paper sheet is led through a series of presses for another dewatering step and thickness control. The remaining water in the paper sheet is removed in the dryer section by means of vaporisation. The required heat is applied through direct contact of the paper sheet with a number of heated cylinders.
The second last step comprises the application of a coating on both sides of the paper sheet. The coating mainly consists of pigments and binders and is used to enhance both the visual and tactile characteristics of the paper sheet as well as the printability. In the product finishing section the paper sheet is led through a calendar where the actual visual and tactile characteristics are determined by applying pressure and/or heat to the coated paper sheet. Once the required visual and tactile characteristics have been obtained the paper sheet is winded onto a reel and subsequently cut into the desired size before it is shipped to the customer. [14]

2.3 Process data

Like other modern paper mills, the Nijmegen paper mill is equipped with a large number of sensors along the production line to measure a variety of parameters. Measured parameters include for example temperature, pressure, flow, base weight of the paper sheet and most important for this thesis the moisture content of the paper sheet.
The raw data from the paper mill consists of 140 selected variables, measured throughout the production process up to the dryer section, with a sampling rate of one minute. In total, continuous process data from seven production runs is available for the development of predictive models. An overview of the available datasets is given in table 2.1.

Table 2.1 Raw process data obtained from the Nijmegen paper mill

Dataset	Grade(s)	Gram weight	Run period	Measurements
1	C	90 g/m^2	23-01-2010 17:01 – 24-01-2010 14:00	1260
2	B	90 g/m^2	14-02-2010 12:00 – 15-02-2010 05:00	1021
3	B	90 g/m^2	20-02-2010 22:00 – 21-02-2010 22:00	1441
4	A, B, C	90 g/m^2	17-04-2010 14:00 – 19-04-2010 07:00	2461
5	A, B, C	90 g/m^2	26-04-2010 02:00 – 26-04-2010 20:18	1099
6	A, B, C	90 g/m^2	30-04-2010 14:00 – 01-05-2010 08:18	1099
7	A, B, C	90 g/m^2	17-05-2010 14:00 – 18-05-2010 08:18	1099

3. Data Processing and Modelling Techniques

Before process data can be used for the development of a model, it has to be verified that the data represents correct process operation. With the relatively large amount of available data, it would be expected that it is rather easy to ascertain correct process operation. However in reality the high dimensionality of the data (large number of measured variables) is the key factor that makes it difficult to extract useful process information from the data. In this chapter the techniques used for processing of the process data and the techniques for the development of a predictive model will be described.

3.1 Visual inspection

Before applying data reduction techniques on the process data, a visual inspection of this process data is useful in order to detect measurements from failing or offline sensors. As will be shown in the upcoming sections, high measurement variance is the key factor in the development of a predictive model. If a sensor is offline or fails during the complete range of the dataset, the measurements are more or less constant thus resulting in a low variance. This will not give difficulties in the data reduction techniques, as these techniques ignore low variance variables. It will, however, give rise to deviant results in the development of a predictive model as these faulty measurements distort the predictive capacity of the model.

3.2 Principal component analysis

As it is not possible to visualize more than 3 variables at once, numerous combinations of variables have to be checked in order to visually ascertain correct operation of the process. An additional factor that complicates visual ascertaining of the process operation is the high correlation that usually exists in process data. Such correlations account for interactions between variables which are difficult to detect with a visual inspection.
To overcome these difficulties it would be advantageous if the dimension of the data could be reduced while preserving essential features of the data.
Principal Component Analysis (PCA) is a tool for reducing the dimension of data by identifying patterns in the data (i.e. correlated variables) and expressing this data in a smaller set of new uncorrelated variables. In order to perform a PCA, the raw process data is presented as an n by m data matrix, where each column (n) represents a variable and each row (m) represents a measurement.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \tag{3.1}$$

For the decomposition of the data matrix X, PCA starts with the covariance matrix of X which is defined as:

$$\mathrm{cov}(X) = \frac{X^T X}{m-1} \tag{3.2}$$

Precondition for this notation of the covariance matrix is that the columns of X have been mean-centered or auto-scaled.
Mean-centering comprises of subtracting the column average from the columns. Mean-centering is however best used for analogous data as it does not incorporate numerical variance.

As process data consist of various measurements (i.e. temperature, pressure, flow) with different magnitudes and thus numerical variance, mean-centering will give distorted results. The better option for process data is auto-scaling as this procedure involves a normalizing step where the mean-centered column is dived by its standard deviation. Due to this normalizing step, equation 3.2 gives the correlation matrix instead of the covariance matrix.

The next step of the decomposition relies upon the eigenvector translation of the covariance/correlation matrix. With this translation the covariance/correlation matrix can be written into a number of eigenvector (p_i) and eigenvalue (λ_i) combinations.

$$\mathrm{cov}(X)\, p_i = \lambda_i p_i \tag{3.3}$$

In this translation p_i gives information how the variables relate to each other, hence the name loading vector. The eigenvalue is a measure for the variance captured. As i increases, the eigenvalues decrease
The number of combinations can at maximum be equal to the number of original variables. However, in practice it is found that the data can be adequately described with far fewer combinations than original variables. The appropriate number of combinations is generally determined by the eigenvalue. Combinations with an eigenvalue > 1 are selected (k), whereas the other combinations (generally represent noise in the dataset) are consolidated into a residual matrix (E).
Finally, information on how the measurements relate to each other is required. This information is described in the score vector (t_i) which is determined using the following equation.

$$t_i = Xp_i \tag{3.4}$$

The data matrix X can now be described as a linear combination of the outer product of the score vectors and loading vectors for k eigenvector/eigenvalue combinations and the residual matrix (E).

$$X = t_1 p_1^T + t_2 p_2^T + \ldots + t_k p_k^T + E \tag{3.5}$$

As the score vectors form an orthogonal set (i.e. $t_i^T t_j = 0$ for $i \neq j$) and the eigenvectors form an orthonormal set (i.e. $p_i^T p_j = 0$ for $i \neq j$, $p_i^T p_j = 1$ for $i = j$), the new variables, or principal components, are uncorrelated. Because of this each principal component describes the maximum variance in the direction given by the loading vector.

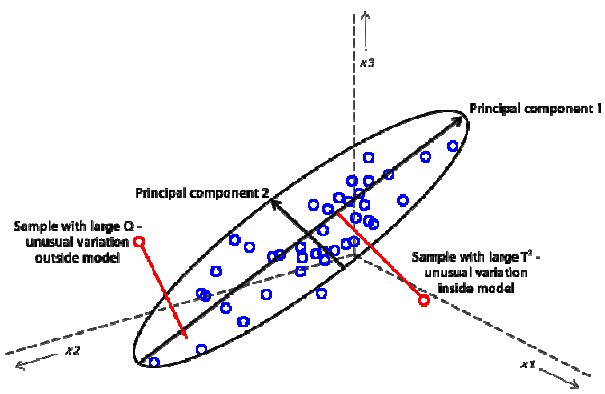

Figure 3.1 Graphical representation of a principal component analysis. [24]

In practice, the first two principal components capture most of the variance in the dataset. In case of correct process operation the majority of the data points fall within the 95% confidence limit of the model which is denoted by the ellipse figure 3.1 and is based on the first two principal components. In case of a disruption in the process operation, data points will deviate from the data points for correct process operation. For a given data point, this variation can either lead to an excessive error contribution to the principal component model or an excessive variation inside the variable measurement. The error contribution (Q_i) for each data point can be calculated from the square of the corresponding row in the residual matrix (E) denoted as e_i.

$$Q_i = e_i e_i^T \tag{3.6}$$

The total lack of fit statistic (Q-residual) is simply the sum of all Q_i contributions. A large Q_i contribution thus indicates that a data point lacks a good fit to the principal component model. The variation contribution (T_i^2) for each data point can be calculated from the multiplication of the inversed diagonal matrix containing the eigenvalues (λ^{-1}) and the square of the corresponding row in the score vector (T_k) denoted as t_i.

$$T_i^2 = t_i \lambda^{-1} t_i^T \tag{3.7}$$

The sum of these contributions is known as the Hotelling T^2 statistic and is a measure for the variation within the principal component model. A large T_i^2 contribution thus indicates that a data point has a fit to the principal component model but with unusual variation compared to the other data points.

Data points with unusual variation that fall out of the 95% confidence limit are called outliers. Excessive outliers have to be removed as these points will distort the predictive capacity of the predictive models that will be developed. [12][24] Once it has been verified that the excessive outliers have been removed the data can be used for subsequent processing.

3.3 Correlation coefficient analysis and data synchronisation

Although the 140 variables have been selected on the basis that these variables are likely to have an influence on the moisture content, one can, however, imagine that not all variables influence the moisture content equally.
Another aspect that can not be observed from the dataset nor the principal component analysis is the time delay before a change in the process input has maximum impact on the chosen process output. Based on the process description it is imaginable that a change in the beginning of the process will not have an immediate effect on the moisture content.
For the development of a predictive model with robust predictive capacities, it is required to have a quantitative representative to decide which variables influence the moisture content. In addition it is required to obtain a value for the time delay of each variable in the process, in order to synchronize the dataset.
By calculating the cross-correlation coefficient (r_{xy}) for each process input variable for k preceding measuring intervals ($k = 0,1,2,\ldots$), both a quantitative representative for the importance of variables as well as an estimate of the time delay of each variable is obtained.

$$r_{xy}(k) = \frac{c_{xy}(k)}{\sqrt{c_{xx}(0)\, c_{yy}(0)}} \tag{3.8}$$

The term c_{uy} represents the cross-covariance coefficient between the process input variable and the process output and is a measure for the similarity between these variables.

$$c_{xy}(k) = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(y_{t+k} - \bar{y}) \tag{3.9}$$

Where $\bar{x}$ and $\bar{y}$ are respectively the mean of the process input variable and the process output, x_t and y_{t+k} are respectively the measured value of the process input variable and the process output at time t, N is the total number of data points and k is the value of the preceding measurement interval.

The terms c_{xx} and c_{yy} represent respectively the auto-covariance of the process input variable and the process output. These terms are used to normalize the cross-covariance coefficient in order to obtain the cross-correlation coefficient.

$$c_{xx}(0) = \frac{1}{N}\sum_{t=1}^{N}(x_t - \bar{x})^2 \tag{3.10}$$

$$c_{yy}(0) = \frac{1}{N}\sum_{t=1}^{N}(y_t - \bar{y})^2 \tag{3.11}$$

Due to the normalizing step with the auto-covariance terms, the cross-correlation coefficient will be a value in the range of -1 to 1. A cross-correlation coefficient close to -1 or 1 implies a strong similarity between the process input variable and the process output, where a coefficient of -1 implies an opposite similarity. A cross-correlation coefficient close to 0 implies, on the other hand, that there is virtually no similarity between the process input variable and process output. For the determination of the process input variables that influence the moisture content, the above mentioned boundaries are not explicit enough. To elucidate this determination it is necessary to define a confidence limit below which no similarity between the process input variable and process output is assumed.
The confidence limit is generally a value in the range of ±0.15 to ±0.30, depending on the size of the dataset used. [12] As the dataset used in this thesis are relatively large the confidence limit was set at ±0.25.
By plotting the obtained cross-correlation versus the time delay values, a plot similar to the one in figure 3.2 is obtained.

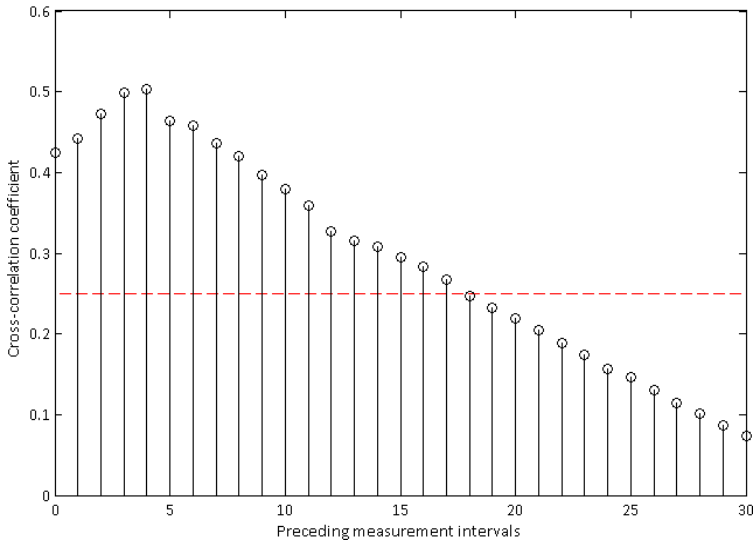

Figure 3.2 Example of cross-correlation coefficient plot.

From figure 3.2 it can be seen that numerically the maximum impact on the process output for this process input variable occurs after four measuring intervals. This implies that the measurements of this process input variable have to be shifted four measuring intervals forward in order to synchronize this process input variable with the process output. The preceding exercise can be repeated for each process input variable in order to obtain a synchronized dataset.

3.4 Partial least squares regression

While the previous sections described data processing and data reduction techniques, this section will describe the technique used to develop a predictive model.
Partial Least Squares (PLS) is a modelling technique that attempts to find factors (latent variables) that capture variance in the data while at the same time achieves maximum correlation between the output variable and input variables.
Since the pioneering work of Wold in the late 1960s, numerous algorithms have been developed for the calculation of the PLS model. However, the original Non Iterative Partial Least Squares (NIPALS) algorithm is still one of commonly used algorithms. With PLS regression the prediction of the process output (*Y*) is described by a linear combination of the process input variables given in a data matrix *X*.
First both auto-scaled *X* and *Y* are decomposed in a similar way as with PCA into a scores matrix (*T* and *U* respectively for *X* and *Y*) and loading matrix (*P* and *Q* respectively for *X* and *Y*) and a residual matrix (*E* and *F* respectively for *X* and *Y*), resulting in two so-called outer relations.

$$\begin{aligned} X &= t_1 p_1^T + t_2 p_2^T + \ldots t_n p_n^T + E \\ X &= TP^T + E \end{aligned} \tag{3.12}$$

$$\begin{aligned} Y &= u_1 q_1^T + u_2 q_2^T + \ldots u_n q_n^T + F \\ Y &= UQ^T + F \end{aligned} \tag{3.13}$$

As the intention with PLS is to describe the process output in the best possible way (i.e. minimizing the norm of *F*) and thereby linking the process input variables to the process output in the best possible way it is necessary to define an additional, so-called, inner relation that performs this transition.

The simplest model that can perform this transition is a model in which the scores of *X* are linked to the scores of *Y* for every component with the use of a regression coefficient (*b*) as can be seen from figure 3.3.

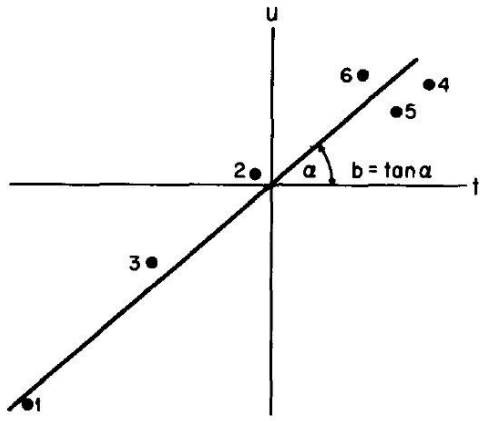

Figure 3.3 Example of simplest inner-relation in which the scores of *X* and *Y* are linked with a linear model. [6]

This model can be described by the following equation.

$$u_h = t_h b_h \tag{3.14}$$

The regression coefficient *b* is given by equation 3.15 where *h* represents a component number between 1 and *n*.

$$b_h = \frac{u_h^T t_h}{t_h^T t_h} \tag{3.15}$$

Combining equation 3.13 and 3.14 thus results in an equation for the prediction of *Y*.

$$\begin{aligned} \hat{Y} &= t_1 b_1 q_1^T + t_2 b_2 q_2^T + \ldots t_n b_n q_n^T + F \\ \hat{Y} &= TBQ^T + F \end{aligned} \tag{3.16}$$

The next objective is to find a pair of vectors (*t* and *u*) which on the one hand describe maximum covariance between the process input variables and the process output while on the other hand form mutual orthogonal pairs. Referring to PCA the major pair of vectors can be found on the basis of the first set of loading vectors (p_1 and q_1 respectively for *X* and *Y*). By projecting the *X* and *Y* data on the corresponding loading vector, the first set of score vectors (t_1 and u_1 respectively for *X* and *Y*) is obtained.
Although these score vectors will yield an equation for the prediction of *Y* with the use of the inner-relation, this prediction will not be the best possible. The latter is caused by the weak relation between the scores of *X* and *Y* which in turn is caused by the fact that *X* and *Y* are decomposed separately. A better relation is obtained when information of the scores is exchanged between each other. The latter can be achieved with the following algorithm.

Starting values comprise a score vector t_old which is any column of the matrix X and a score vector u which is any column of the matrix Y.

1. u = y
2. t_old = x_1

Calculations on the X matrix.

3. $w = (u^T * X / (u^T * u))^T$
4. w = w / ||w||
5. $t = X * w / (w^T * w)$

Calculations on the Y matrix.

in case of one Y variable

6. q = 1

Check for improvement. If the improvement is too little an additional step is required.

7. if ||t – t_old|| > threshold
   t_old = t
   return to step 3

If the calculated score vector is equal (within a specified error) to the preceding calculated score vector, the calculations can be stopped. An additional loop is required to transform the obtained score vector into an orthogonal score vector.

8. $p = (t^T * X / (t^T * t))^T$
9. p = p / ||p||
10. t = t / ||p||
11. w = w / ||p||

Determine regression coefficient b_h

12. $b_h = u_h^T t_h / (t_h^T t_h)$

Calculate residuals of X and Y with the use of the inner relation

13. $X = X - t_h p_h^T$
14. $Y = Y - t_h b_h q_h^T$

Each time the algorithm converges a $\hat{u}_h q_h^T$ combination is obtained. This combination, or latent variable, accounts for a certain part in the prediction of *Y*. The corresponding $t_h p_h^T$ combination accounts for a certain part of the *X* data used for this prediction. As these parts are already used by a latent variable, it is subtracted from the *X* and *Y* data before the algorithm is repeated. For example after the calculation of the first latent variable, the *X* and *Y* residuals become as shown in equations 3.17 and 3.18.

$$X_1 = X - t_1 p_1^T \tag{3.17}$$

$$\begin{aligned} \hat{Y}_1 &= \hat{Y} - \hat{u}_1 q_1^T \\ \hat{Y}_1 &= \hat{Y} - t_1 b_1 q_1^T \end{aligned} \tag{3.18}$$

In theory, the algorithm can be repeated until the number of latent variables equals to original number of variables (*n*). However, the calculations are generally terminated earlier because the lower order latent variables mostly describe noise present in the data thus making them irrelevant for the prediction of *Y*.

A general rule of thumb is to terminate the calculations once the eigenvector of *X* becomes one or lower. An additional rule of thumb is to terminate the calculations once the cross validation error does not improve anymore by at least 2%.

Once the PLS model has been developed, its robustness should be tested with the use of new data. [1][6][12]

3.5 Robust least squares support vector machine

Considering that papermaking involves a large number of processes of which little knowledge is available, it might be possible that these processes cannot be described by a linear modelling technique as PLS. To check whether non-linear relations form a serious threat, a second non-linear modelling technique, on the basis of Least Squares Support Vector Machines (LSSVM), will be applied to the process data. This second technique is known as Robust LSSVM.
With LSSVM the goal is to develop a function (equation 3.19) that relates the process input variables (x) to the process output (y).

$$y = f(\vec{x}) \tag{3.19}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_m \end{bmatrix} \qquad x = \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \tag{3.20}$$

For the estimation of the non-linear function, LSSVM first transforms this problem into a linear function in a high dimensional feature space (F) with the use of the regression function given in equation 3.21.

$$f(\vec{x}) = \vec{w}^T \vec{\varphi}(\vec{x}) + b \tag{3.21}$$

In this function, w is weight factor in the feature space F, b a bias term and $\vec{\varphi}(\vec{x})$ the non-linear mapping term that transform the input data points to corresponding data points in the feature space (F).

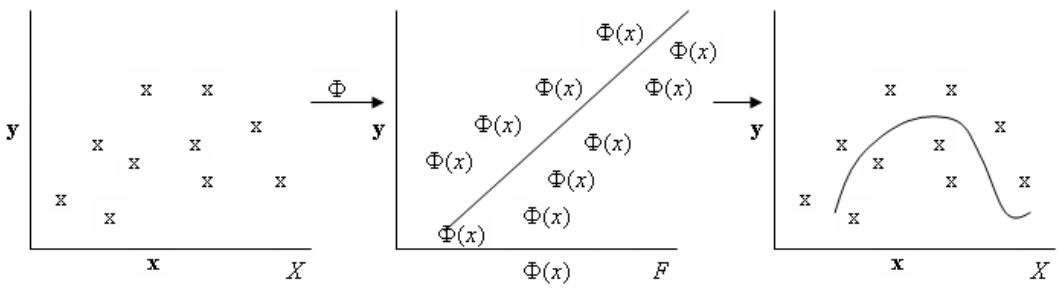

Figure 3.4 Example of the transformation of a non-linear problem to a linear problem in the feature space (F) with the use of the mapping term φ. Subsequently the problem is solved with linear regression in the feature space (F) which is equal to non-linear regression in the original space.

Subsequently LSSVM defines an optimization problem (equation 3.22) subject to the equality constraints given in equation 3.23.

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{m} e_i^2 \tag{3.22}$$

$$y_i = w^T \varphi(x_i) + b + e_i \tag{3.23}$$

Where i is a value between 1 and m, e is the error variable and γ is the regularization parameter that determines the trade-off between the fitting error minimization and smoothness of the estimated function.

To solve this optimization problem, Suykens *et al.* [20] stated that the Lagrange function with its optimality condition should give the desired result.

$$L(w,b,e_i,\alpha_i) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{i=1}^{m} e_i^2 - \sum_{i=1}^{m}\alpha_i(w^T\varphi(x_i) + b + e_i - y_i) \tag{3.24}$$

where α_i are Lagrange multipliers. The optimality conditions for this Lagrange function are given by equation 3.25.

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{m}\alpha_i\varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{m}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma.e_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i = w^T\varphi(x_i) + b + e_i \end{cases} \tag{3.25}$$

De Brabanter [5], Suykens *et al.* [20]and Vapnik [22] explained that after elimination of *w* and *e* the Lagrange function and its optimality conditions result in the linear system given in equation 3.26.

$$\begin{bmatrix} 0 & 1_m^T \\ 1_m & \Omega + \gamma^{-1}I_m \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{3.26}$$

where $Y = [y_1,\ldots,y_m]$, $1_m = [1,\ldots 1]$, $\alpha = [\alpha_1,\ldots,\alpha_m]$, *I* is an *m x m* identity matrix and Ω is the inner product of the linear mapping $\left(\Omega = \vec{\varphi}(\vec{x}_i)^T\vec{\varphi}(\vec{x}_j)\right)$.

Calculation of Ω can, however, be difficult if the dimension of the feature space (*F*) becomes very high. It would be advantageous if the inner product could be calculated directly from the input points. Such a direct calculation method is found in the kernel function method or kernel trick (*K*).

$$K(\vec{x}_i,\vec{x}_j) = \vec{\varphi}(\vec{x}_i)^T\vec{\varphi}(\vec{x}_j) \tag{3.27}$$

The advantage of using a kernel function is that it is not required to know the mapping. However, it is required, that for a particular function, to be used as a kernel function a feature space exists. It was found that when a function satisfies Mercer conditions, it can be regarded as a kernel function. [4] The commonly used kernel function for LSSVM is the Radial Basis kernel Function (RBF) as LSSVM is generally used as a non-linear modelling technique.

$$K(\vec{x}_i,\vec{x}_j) = \exp\left[-\frac{\left\|\vec{x}_i - \vec{x}_j\right\|^2}{\sigma^2}\right] \tag{3.28}$$

When the kernel trick is applied to the LSSVM optimization problem, the model for the function estimation, given in equation 3.29, is obtained.

$$f(x) = \sum_{i=1}^{m} \hat{\alpha}_i K(x, x_i) + \hat{b} \tag{3.29}$$

where α and b are obtained from solving the linear system given in equation 3.26.[5]

$$\hat{\alpha} = \left(\Omega + \gamma^{-1} I_m\right)^{-1} \left(Y - 1_m \hat{b}\right) \tag{3.30}$$

$$\hat{b} = \frac{1_m^T \left(\Omega + \gamma^{-1} I_m\right)^{-1} Y}{1_m^T \left(\Omega + \gamma^{-1} I_m\right)^{-1} 1_m} \tag{3.31}$$

Although LSSVM can handle normal error distribution in process data, it is however not capable of dealing with outliers. As process data can contain outliers due to a failing or offline sensor, it might be possible that the LSSVM model give distorted results.
This problem can be solved by removing the outliers from the process dataset; nevertheless this can be time consuming for larger datasets. Therefore it would be advantageous to develop a model that can deal with the outliers thereby retaining a good predictive quality of the model.
Robust LSSVM is the extension for the LSSVM which is capable of dealing with outliers while retaining a good predictive quality of the model. The extension on the original LSSVM model is the introduction of an additional weighing factor (v_i) which is calculated from the error variables (e_i) of the LSSVM model.
The consequence of this additional factor is that a new optimization problem arises (equation 3.32) subject to the equality constraints given in equation 3.33. (note: the symbol * denotes the unknown variable for the Robust LSSVM)

$$\min_{w^*, b^*, e^*} J(w^*, e^*) = \frac{1}{2} w^{*T} w^* + \gamma \frac{1}{2} \sum_{i=1}^{m} v_i e_i^{*2} \tag{3.32}$$

$$y_i = w^{*T} \varphi(x_i) + b^* + e_i^* \tag{3.33}$$

As a new optimization problem has been defined, this also results in a new Lagrange function with corresponding optimality conditions for solving this problem.

$$L(w^*, b^*, e_i^*, \alpha_i^*) = \frac{1}{2} w^{*T} w^* + \gamma \frac{1}{2} \sum_{i=1}^{m} v_i e_i^{*2} - \sum_{i=1}^{m} \alpha_i^* (w^{*T} \varphi(x_i) + b^* + e_i^* - y_i) \tag{3.34}$$

$$\begin{cases} \dfrac{\partial L}{\partial w^*} = 0 \rightarrow w^* = \sum_{i=1}^{m} \alpha_i^* \varphi(x_i) \\ \dfrac{\partial L}{\partial b^*} = 0 \rightarrow \sum_{i=1}^{m} \alpha_i^* = 0 \\ \dfrac{\partial L}{\partial e_i^*} = 0 \rightarrow \alpha_i^* = \gamma . v_i e_i^* \\ \dfrac{\partial L}{\partial \alpha_i^*} = 0 \rightarrow y_i = w^{*T} \varphi(x_i) + b^* + e_i^* \end{cases} \tag{3.35}$$

Ultimately the solution of the Lagrange function is given by equation 3.36, because of the additional weighing factor an additional diagonal matrix containing this weighing factor is obtained (equation 3.37).

$$\begin{bmatrix} 0 & 1_m^T \\ 1_m & \Omega + V_\gamma \end{bmatrix} \begin{bmatrix} b^* \\ \alpha^* \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{3.36}$$

$$V_\gamma = diag\left(\left[\frac{1}{\gamma \cdot v_1}, ..., \frac{1}{\gamma \cdot v_m} \right] \right) \tag{3.37}$$

The weighing factors (v_i) are obtained from the error variables (e_i) of the normal LSSVM model and the robust scale estimator ($\hat{s}$).

$$e_i = \frac{\alpha_i}{\gamma} \tag{3.38}$$

$$\hat{s} = \frac{IQR}{2 \times 0.6745} \tag{3.39}$$

The Inter Quartile Range (IQR) is defined as the difference between the 75th percentile (3rd quartile) and 25th percentile (1st quartile) of the error (e_i) distribution.
The weighing factor (v_i) can now be determined based on the following criteria.

$$v_i = \begin{cases} 1 & \text{if} \quad \left| \dfrac{e_i}{\hat{s}} \right| \le c_1 \\ \dfrac{c_2 - \left| \dfrac{e_i}{\hat{s}} \right|}{c_2 - c_1} & \text{if} \quad c_1 \le \left| \dfrac{e_i}{\hat{s}} \right| \le c_2 \\ 10^{-4} & \text{otherwise} \end{cases} \tag{3.40}$$

Where the constants c_1 and c_2 are typically chosen as $c_1 = 2.5$ and $c_2 = 3$ according to Suykens *et al.* [20]

At this point it is again possible to perform the kernel trick, thus resulting in the function estimation for the Robust LSSVM model, as given in equation 3.41.

$$f(x) = \sum_{i=1}^{m} \hat{\alpha}_i^* K(x, x_i) + \hat{b}^* \tag{3.41}$$

where α^* and b^* are obtained from solving the linear system given in equation 3.37.

$$\hat{\alpha}^* = \left(\Omega + V_\gamma\right)^{-1}\left(Y - 1_m \hat{b}^*\right) \tag{3.42}$$

$$\hat{b} = \frac{1_m^T \left(\Omega + V_\gamma\right)^{-1} Y}{1_m^T \left(\Omega + V_\gamma\right)^{-1} 1_m} \tag{3.43}$$

Figure 3.5 shows the results obtained in a comparative study on LSSVM and Robust LSSVM performed by Valyon. [21]

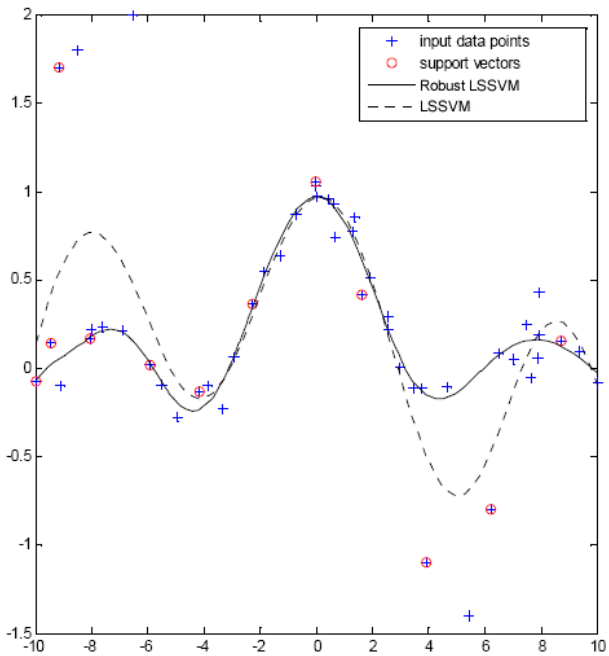

Figure 3.5 Results of comparative study on LSSVM and Robust LSSVM performed by Valyon. [21]

As can be observed from figure 3.5, the Robust LSSVM method was able to reduce the effect of outliers significantly in comparison with the LSSVM method.

4. Predictive modelling of paper quality

In this chapter an overview will be given of the steps involved in the development of the predictive model for the paper quality. At first the findings on the condition of the process operation will be discussed on the basis of a principal component analysis for each dataset used. Subsequently the results of the different predictive models, developed in this thesis, will be discussed.

4.1 Data pre-processing

As described in chapter 3, raw data from any industrial process has to be pre-processed before it can be used for the development of a model.
Principal component analysis was used in this thesis as the data pre-processing tool, after the data was first visually inspected for large deviations. Subsequently the condition of the process operation was determined on the basis of the score scatter plot of the first two components of the principal component analysis.

4.1.1 Analysis dataset 1

A visual inspection of dataset 1 showed that the measured variables x95 and x108 were offline during the production run and that the variables x59, x62, x71, x96, x120, x121, x129 and x133 had a constant value.
The variables x95 and x108 were removed from the dataset in order to prevent difficulties in the development of a predictive model. The variables with a constant value were left as is, since these variables will automatically be ignored in the development of predictive models due to the lack of variance.
The principal component analysis resulted in a model with 17 principal components that described 71.9% of the variation present in the dataset. The score scatter plot of the first two principal components is given in figure 4.1.

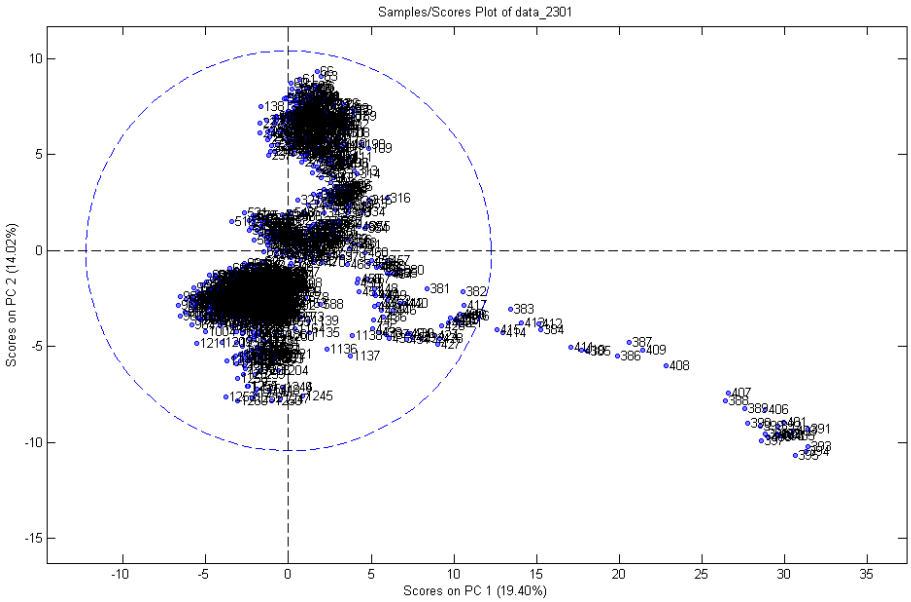

Figure 4.1 Score scatter plot of the first two principal components of the initial PCA model of dataset 1.

What strikes the most in figure 4.1 is the sequence of outliers at right side of the figure. An examination of the contribution of the variables in this sequence is shown in figure 4.2.

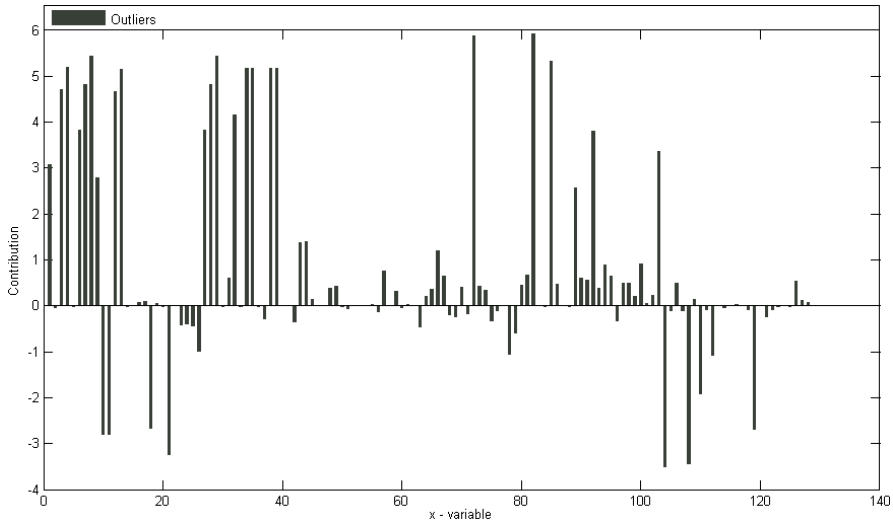

Figure 4.2 Variable contribution plot of the outlier sequence.

From figure 4.2 it can be observed that the outlier sequence is primarily influenced by variables in the refining section, variables of retention aids and variables in the press section.
In consultation with Sappi, it was found that this deviation in process operation was caused by sheet breakage. With this additional information it becomes clear why the mentioned variables deviate. As a new sheet has to be made, after a sheet breakage, more new fibres a required thus explaining the deviation in the refiner variables. In order to prevent another sheet breakage, a more firm sheet it desired. One way to achieve this is the addition of larger quantities of retention aids. Finally the press section has to be adjusted to prevent another sheet breakage, thus these variables shown a different behaviour compared to the normal operating conditions.
As this sequence will distort the predictive capacity of future models, the affected data was removed from the dataset. The consequence of this removal is that the principal component analysis has to be repeated. The new analysis resulted in a model with 18 principal components that described 0,704 of the variation present in the dataset. The corresponding score scatter plot of the first two principal components is shown in figure 4.3.

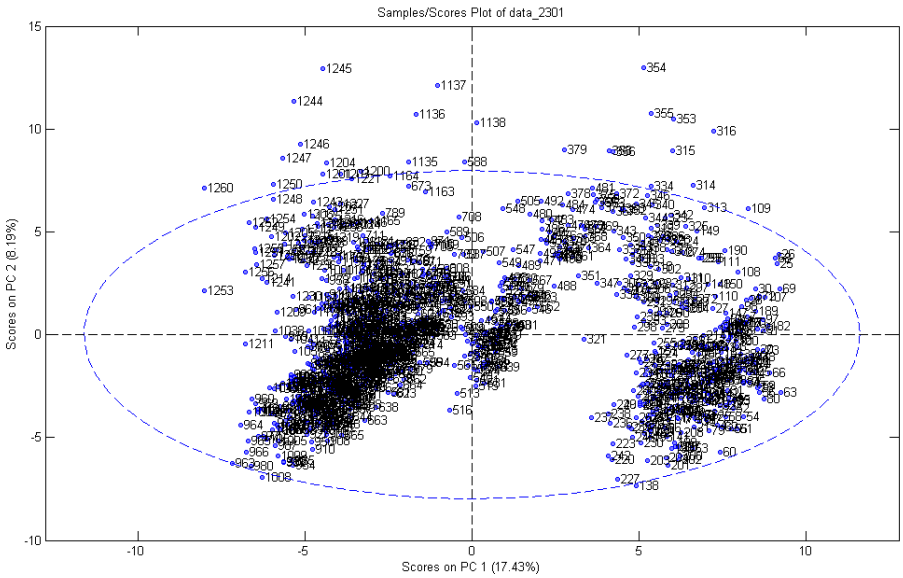

Figure 4.3 Score scatter plot of the first two principal components of the second PCA model of dataset 1.

From figure 4.3 it can be observed that most of the data points now fall within the 95% confidence limit. A detailed inspection showed that the number of outliers present is in accordance with the five percent of acceptable outliers. However, one remark has to be made, i.e. the data points are not randomly distributed within the 95% confidence ellipse. The data points are actually distributed in six clusters of consecutive data. The clusters present in the dataset consist of the data points as given in table 4.1.

Table 4.1 Distribution of the data points in the cluster in dataset 1.

Cluster	Data points	Period
1	1 – 299	23-01-2010 17:01 – 23-01-2010 21:59
2	300 – 350	23-01-2010 22:00 – 23-01-2010 22:50
3	351 – 379	23-01-2010 22:51 – 23-01-2010 23:19
4	471 – 579	24-01-2010 00:51 – 24-01-2010 02:39
5	580 – 1199	24-01-2010 02:40 – 24-01-2010 12:59
6	1200 - 1260	24-01-2010 13:00 – 24-01-2010 14:00

In order to determine the cause of the clustering, the contribution of the variables in each cluster was determined. The result is shown in figure 4.4.

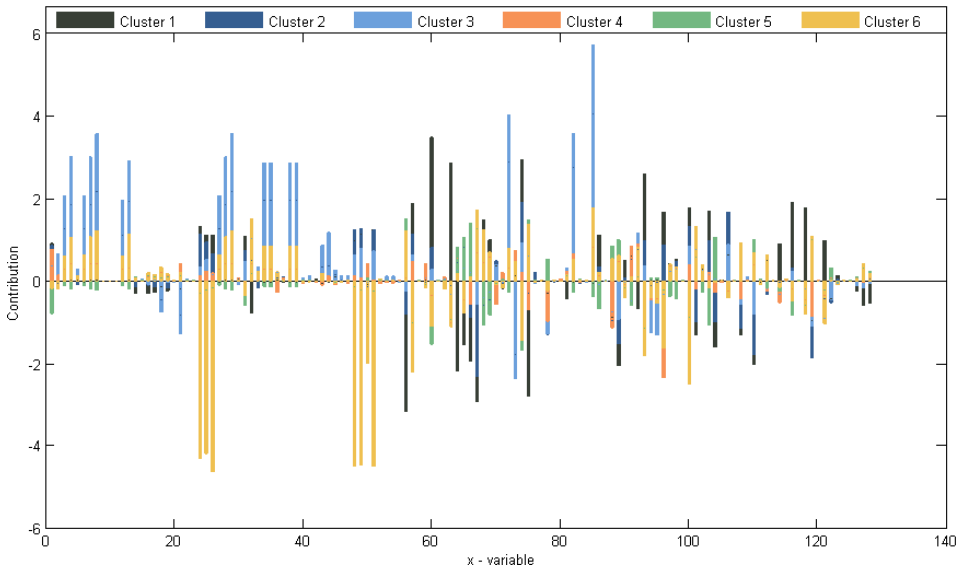

Figure 4.4 Variable contribution in each cluster in dataset 1.

From figure 4.4 it can be observed that most deviations in variable contribution exist in cluster 1, 3 and 6. Furthermore it can be observed that most deviations occur in the section of the variables x5 to x40 and in the section of variables x60 to x90. From Appendix I it can be found that these variables represent the refiner section and the additive section. When the contributions were analysed in more detail, it was found that most of the deviations in variable contribution were caused by variables of flows, temperatures and levels of storage tanks. For these variables it is not unusual to deviate over time, therefore it may be assumed that the data in this dataset represent normal process operation.
The placement of the clusters in the score scatter as shown in figure 4.4 can most likely be explained by the sign of the sum of the deviation of each variable average in a cluster in relation to the corresponding variable average used in the auto-scaling.

It is assumed that based on the auto-scaling of the data the scores should be placed around the origin (0,0) in the score scatter plot. However if all variables in a cluster have a higher average value compared to the average value used in the auto-scaling, all variables will have a positive contribution. The latter will be translated into a shift to the right, of the cluster, in the score scatter plot. If all variables in a cluster have roughly the same average value as in the remaining clusters, the cluster will be placed around the origin. If all variables in a cluster have a lower average value than the value used on the auto-scaling, all variables contribute negatively which is translated into a shift to the left of the score scatter plot. For this dataset the sum of the cluster contributions are shown in table 4.2.

Table 4.2 Cluster contributions for dataset 1.

Cluster	Sum of contribution
1	3312
2	3447
3	1271
4	-1616
5	-1373
6	-2878

As can be observed from table 4.2 and figure 4.3, the sum of contribution is in accordance with the arrangement of the clusters in the score scatter plot (figure 4.30). The only deviating value is found for cluster 4. In figure 4.3 cluster 4 is positioned on the right side of the origin. Based on the assumption made previously, this should be translated into a positive sum of contribution. However from table 4.2 it can be observed that the sum of contribution for the fourth cluster has a negative value. The latter might be explained by the fact that this cluster still contains abnormal values of the variables due to the sheet breakage. Furthermore it can be observed from table 4.2 that all sum of contribution have a relatively large value. This can be accounted to one variable that has a wide range of measurements. Because of that the overall average of this variable is small compared to the average value in the clusters. This again translated into a large sum of contribution.

As the other datasets are analysed according to the same methodology used for dataset 1, the results of the analysis of these datasets have been combined in Appendix II.

It was found that the data in all datasets is arranged in clusters of consecutive data. An analysis of the variable contributions in the clusters of the datasets, showed that the clusters in the datasets were formed due to changes in the average value of variables that represent temperatures, flows or valve positions. As these values of these variables are likely to change over time, it can be concluded that the datasets used in this thesis represent data from normal process operation.

4.2 Development of a predictive model

After the dataset has been processed for the use in the development of a predictive model, it is necessary to divide it into three sections. The first category is training data. Training data is data for which the model is developed and should ideally span the full space of process operation. The second section is validation data. Validation data consist of similar data compared to the training data and is used to test the robustness of a developed model.

The third section is validation test data. Validation test data consist of additional new data and is used to determine how the model responds to untrained data.

Once the datasets have been categorized the final step is to determine the variables that affect the paper quality and the corresponding time delay. In this thesis a cross-correlation analysis, as described previously, was used for this determination.

For a larger reliability, both the training data and validation data were used in the cross-correlation. The obtained results were subsequently incorporated into all datasets in order to obtain synchronized datasets.

For a fair comparison of the quality of different models, it is necessary to use a quantitative measure. A commonly used quantitative measure for the quality of models is the coefficient of determination (R^2).

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \tag{4.1}$$

$$SS_{err} = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{4.2}$$

$$SS_{tot} = \sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2 \tag{4.3}$$

where n is the number of data points, Y_i is the actual measured paper quality, $\hat{Y}_i$ is the predicted paper quality and $\overline{Y}$ is the mean of the measured paper quality.
The closer the coefficient of determination is to 1.0, the better the model can predict the measured values.
In addition the mean square error (e_{MSE}) will be used as an indicative measure to compare the robustness of the model with various validation and validation test datasets.

$$e_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{4.4}$$

where n is the number of data points, Y_i is the actual measured moisture content and $\hat{Y}_i$ is the predicted moisture content. The closer the value of the mean square error is to zero, the better the fitting between the actual measurements and the prediction.

4.2.1 Grade specific PLS model

The first predictive model in this thesis was developed using the PLS regression technique. This technique was preferred, because of its relative simplicity compared to robust LSSVM. The initial PLS model was developed with dataset 2 as training data, datasets 1 and 3 as validation data and datasets 5, 6 and 7 as validation test data. The cross-correlation analysis on the datasets 1, 2 and 3 yielded 56 process input variables which were likely to affect the paper quality. The complete list of these process input variables together with the corresponding time delay can be found in appendix III.
Once the obtained time delays were incorporated into each dataset, the PLS model was developed. Although the cross-correlation analysis yielded 56 process input variables which are likely to affect the paper quality, it might be possible that certain variables contain similar information. Another possibility is that variables, in reality, have little influence on the paper quality, but where selected as they, coincidentally and only in this case, had a strong correlation with the paper quality.
By removing one variable at a time, while tracking the predictive capacity of the model, it was possible to determine the set of variables (table 4.3) that yielded the best predictive model based on the used datasets.

Table 4.3 Variables used for the development of the PLS model with cross-correlation time delay.

	Variable name
x93	Flow additive
x98	Flow suspension to wire
x103	Headbox consistency
x106	Conductivity measurement headbox
x111	Vacuum suction roll
x115	Valve position vacuumbox
x127	Valve position vacuumbox
x128	Process value vacuumbox

The resulting PLS model consisted of three latent variables which explained 74.8% of the variation in *X* and 89.5% of the variation in *Y* in the training data. An indicative parameter for the quality of the model can be found in the relation between the first *X*-score and the first *Y*-score. Ideally this relation should yield a linear plot which means that the *X* and *Y* scores are exchanged perfectly.

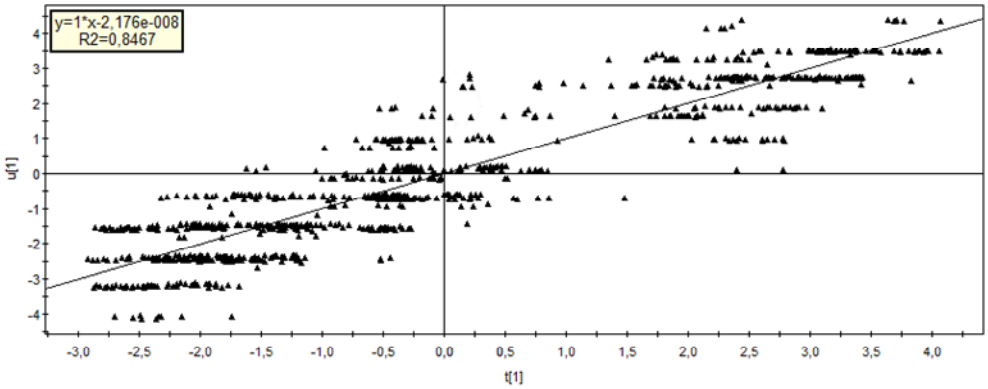

Figure 4.5 Relation between first *X*-score and *Y*-score as an indicative parameter of the quality for the PLS model with cross-correlation time delay.

As can be observed from figure 4.5 the relationship for this model does not yield a perfectly linear plot. However, the relationship can be denoted as being more linear than non-linear. The latter indicates that a model based on this data should have the ability to predict the paper quality to some extent.
The prediction of the paper quality and the actual measurement of the paper quality for the training data is shown in figure 4.6.

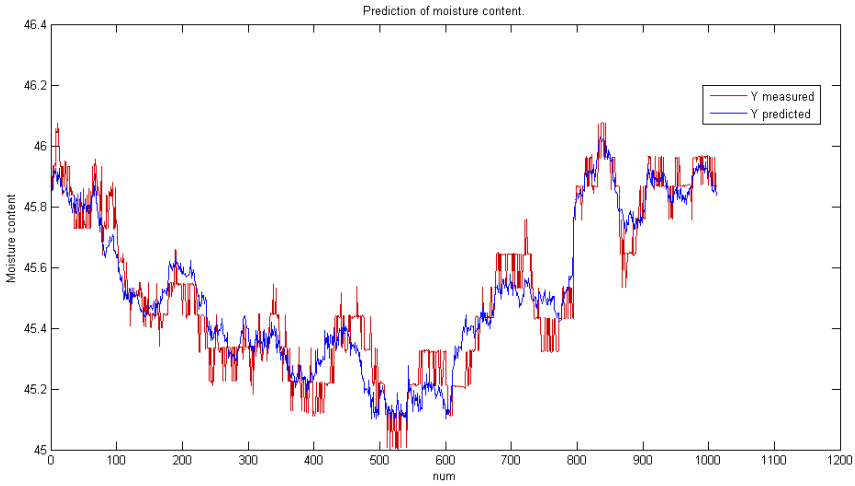

Figure 4.6 Prediction and actual measurement of the paper quality for the training dataset performed with the PLS model with cross-correlation time delay.

According to equation 4.1 the coefficient of determination is 0.8952. From figure 4.6, it can be observed that the model can predict the average trend in paper quality very well. The first test of the predictive capacity of the model was performed with the validation data. The results of this test are shown in the figures 4.7 and 4.8.

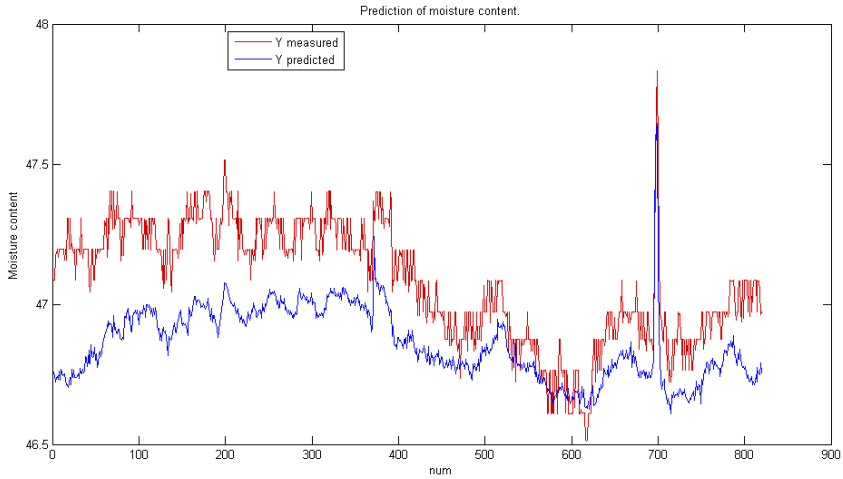

Figure 4.7 Prediction and actual measurement of the paper quality of the validation dataset 1 performed with the PLS model with cross-correlation time delay.

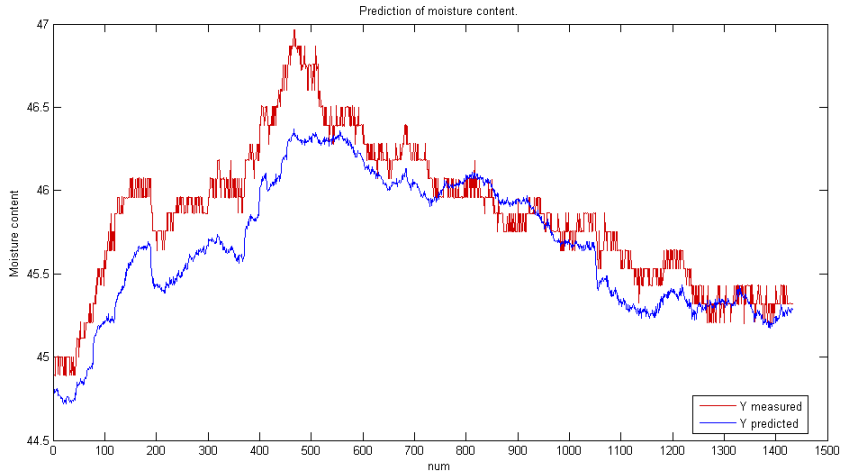

Figure 4.8 Prediction and actual measurement of the paper quality of the validation dataset 3 performed with the PLS model with cross-correlation time delay.

According to equation 4.4 the mean square error for validation data 1 is 0.0647, whereas the mean square error for validation data 3 is 0.0679. From the figures 4.7 and 4.8 it can be observed that the general trend in the paper quality is predicted well in both cases. However in both cases a slight offset between the actual measurement and the prediction of the paper quality is noticeable. The final test for this model was the test with validation test data, to check how the model responded to new untrained data. The results are shown in the figures 4.9 to 4.11.

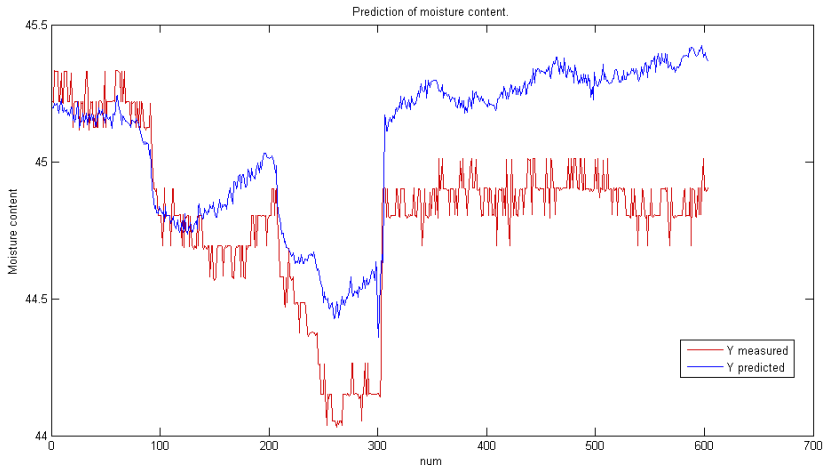

Figure 4.9 Prediction and actual measurement of the paper quality of the validation test dataset 5 performed with the PLS model with cross-correlation time delay.

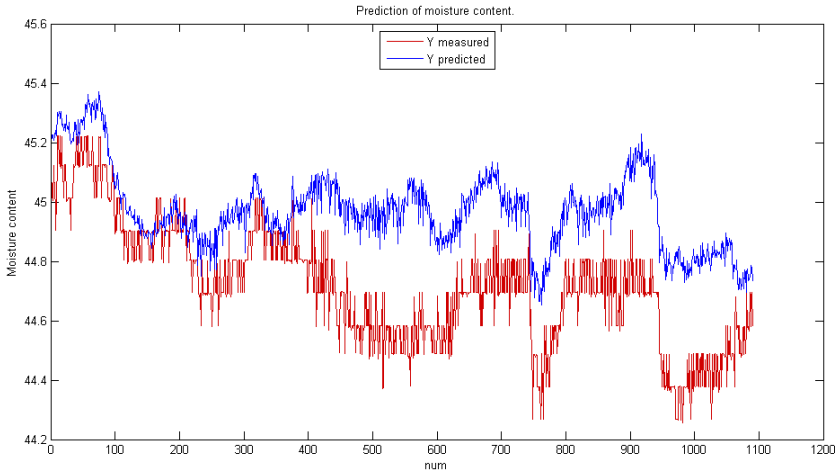

Figure 4.10 Prediction and actual measurement of the paper quality of the validation test dataset 6 performed with the PLS model with cross-correlation time delay.

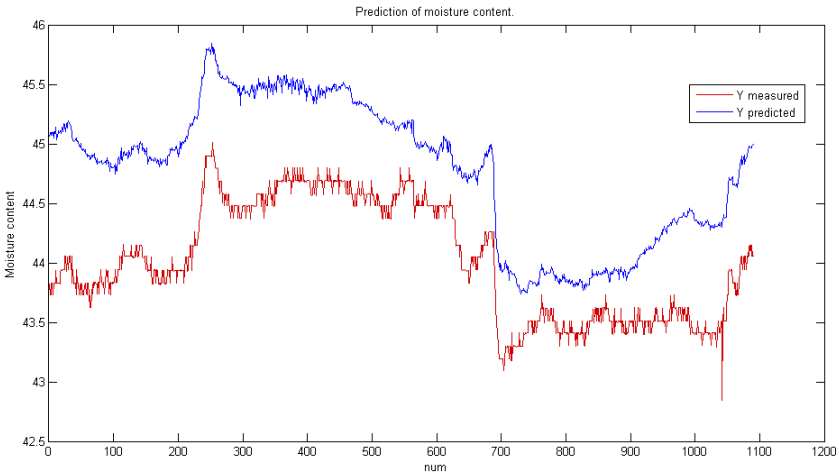

Figure 4.11 Prediction and actual measurement of the paper quality of the validation test dataset 7 performed with the PLS model with cross-correlation time delay.

According to equation 4.4 the mean square error for the validation test dataset 5, 6 and 7 is respectively 0.1132, 0.0859 and 0.6390. From the figures 4.9 to 4.11 it can be observed that the model is able to follow the general trend in the validation test up to a large extent. However also in this case the actual measurement and the prediction show a large offset, of which it is unknown what the cause is.

4.2.2 Grade specific PLS model with time delays from operating experience

In a first attempt to find a possible cause for the offset, between the actual measurements and predictions observed in the results of the initial PLS model, a different set of time delays was used. The time delays from operating experience used for the development of the new model can be found in Appendix III. As can be observed from Appendix III a large offset exists between the time delays obtained from the cross-correlation analysis and the time delays from operating experience. This offset can probably be explained by the fact that the time delays obtained from the cross-correlation analysis, represent the mathematical time delay a certain variable to have maximum influence on the paper quality. However in reality this time delay can be influenced by limitations in the process, thereby deviating from the theoretical value.

In order to make a fair comparison with the model obtained with the cross-correlation time delays, an exact copy of that model was used for the development of this new model with time delays from operating experience.
The PLS model thus obtained consisted of three latent variables which explained 74.8% of the variation in *X* and 88.3% of the variation in *Y* in the training data. The relation between the first *X-score* and first *Y-score* is shown in figure 4.12.

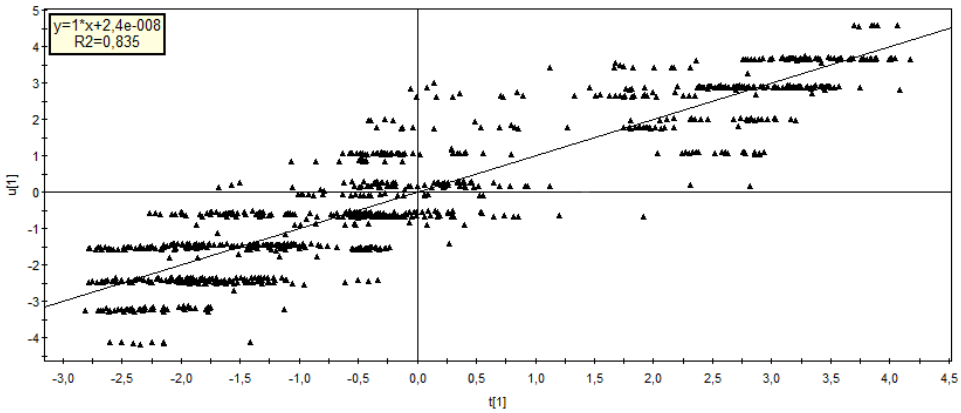

Figure 4.12 Relation between first *X*-score and *Y*-score as an indicative parameter of the quality for the PLS model with time delays from operating experience.

Again this relationship is not completely linear. But also in this case the relationship is more linear than non-linear, thus indicating that a model based on this data should have the ability to predict the paper quality. In comparison with the PLS model with cross-correlation time delay, the relationship is slightly deteriorated. However, this does not imply that the results of this model will be inferior to the results of the PLS model with cross-correlation time delay.
The result of the new PLS model for the training data is shown in figure 4.13.

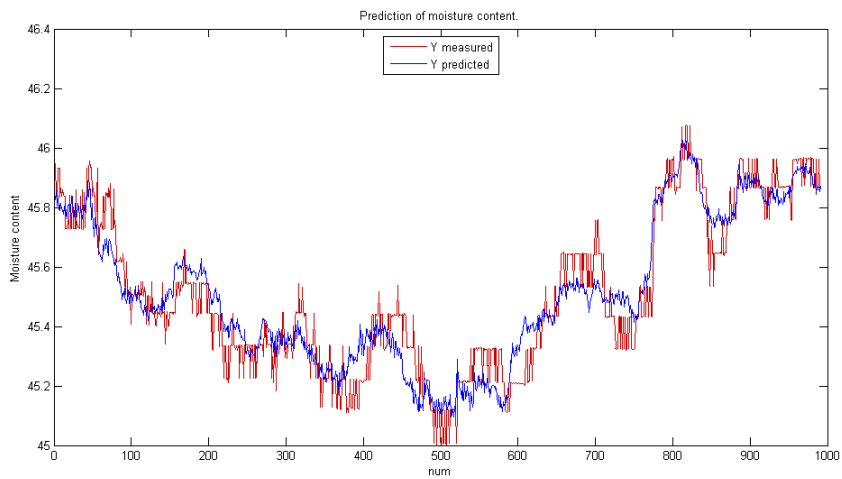

Figure 4.13 Prediction and actual measurement of the paper quality for the training dataset performed with the PLS model with time delays from operating experience.

According to equation 4.1 to coefficient of determination is 0.8832. From figure 4.13 it can be observed that this model can predict the general trend in the paper quality well. In comparison to figure 4.20, almost no differences are noticeable.

The latter implies that different time delays do not influence the predictive capacities of the model. Subsequently the obtained model was tested with the validation data. The results of this test are shown in the figures 4.14 and 4.15.

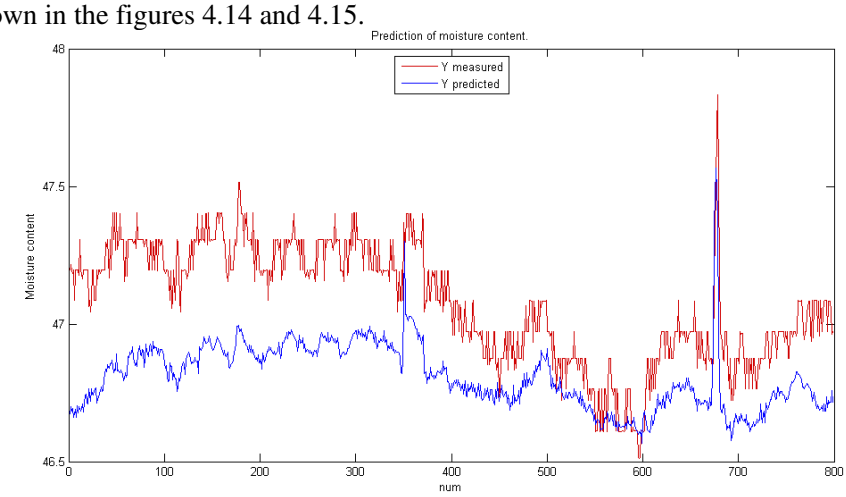

Figure 4.14 Prediction and actual measurement of the paper quality of the validation dataset 1 performed with the PLS model with time delays from operating experience.

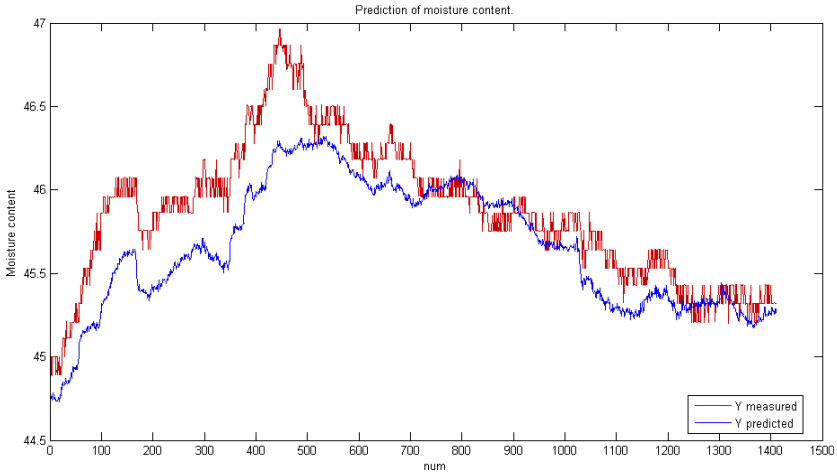

Figure 4.15 Prediction and actual measurement of the paper quality of the validation dataset 3 performed with the PLS model with time delays from operating experience.

According to equation 4.4 the mean square error for validation dataset 1 and 3 are respectively 0.0900 and 0.0817. From figures 4.14 and 4.15 there is a slight offset noticeable between the measured and predicted values of the paper quality. A comparison between the figures 4.7, 4.8 and the figures 4.14, 4.15 there is virtually no difference visible.
Finally the model was tested using the validation test data. The results of this test are shown in the figures 4.16 to 4.18.

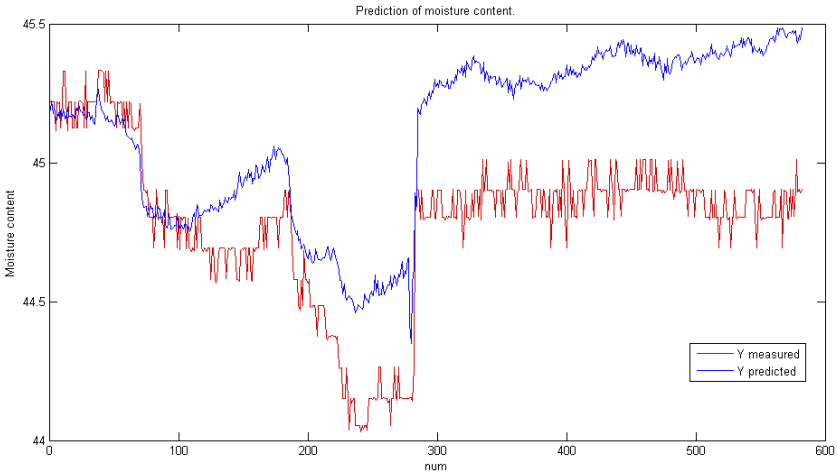

Figure 4.16 Prediction and actual measurement of the paper quality of the validation test dataset 5 performed with the PLS model with time delays from operating experience

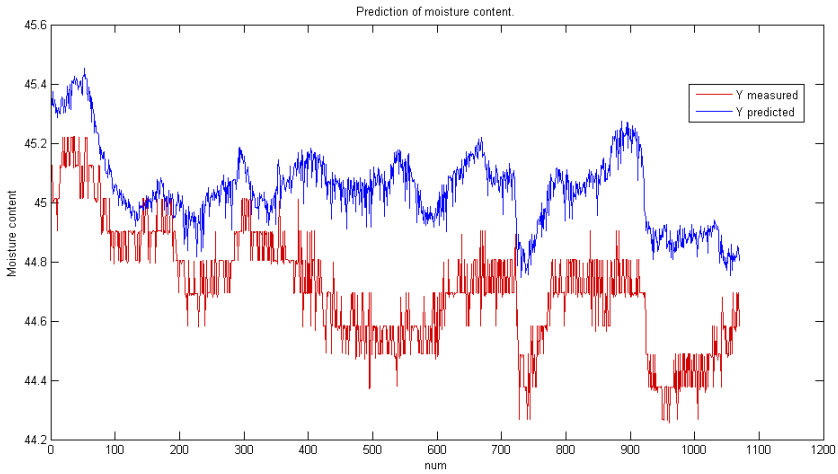

Figure 4.17 Prediction and actual measurement of the paper quality of the validation test dataset 6 performed with the PLS model with time delays from operating experience.

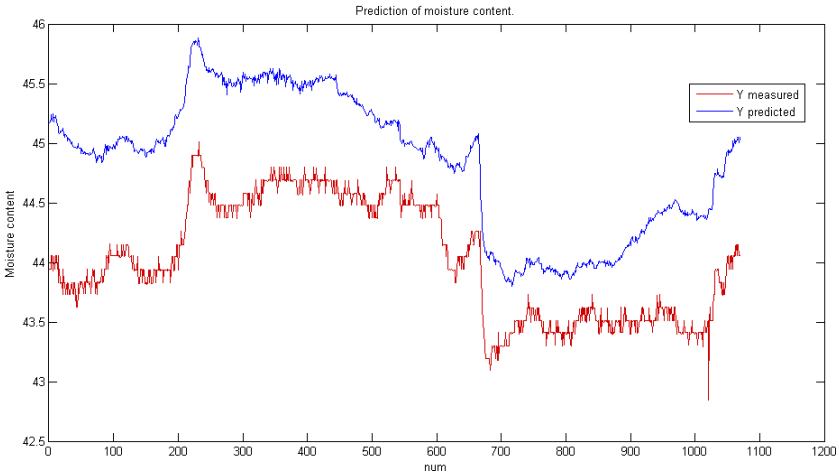

Figure 4.18 Prediction and actual measurement of the paper quality of the validation test dataset 7 performed with the PLS model with time delays from operating experience.

According to equation 4.4, the mean square error of validation test dataset 5, 6 and 7 are respectively 0.1500, 0.1352 and 0.7215. As expected the figures 4.16 to 4.18 show a large offset between the measured and predicted value of the paper quality. In addition there are virtually no differences visible between the figures 4.9 to 4.11 and the figures 4.16 to 4.18. From the preceding it can thus be concluded that the offset between the measured and predicted values of the paper quality is not affected by the time delay. In general it can be assumed that the overall predictive capacity of the PLS model is most likely not affected by the time delay at all.

4.2.3 Multi-grade PLS model

In the previous sections it was found that a grade specific PLS model was capable of predicting the general trend in the paper quality. However the model showed an offset between the measured and predicted values of the paper quality with the validation data and validation test dat. The latter issue could not be solved with different time delays.
A possible solution for the offset issue was found in the training data used for the initial models. It was mentioned that the training data should ideally span the full space of process operation. This criterion was not met for the initial model as the training data consisted of just one grade, whereas, the paper mill can produce three different grades.
Therefore in this new approach, a training dataset was used which contained data from all three grades. For the training data the datasets 5, 6 and 7 were used, as these datasets contain data of all grades and should therefore span a large space of the process operation. The validation data consisted in this new approach of the datasets 1, 2 and 3. Finally dataset 4 was used as validation test data. A new cross-correlation analysis yielded 49 process input variables which are likely to have influence on the paper quality. The complete list of these process input variables together with the corresponding time delay can be found in Appendix IV. As it was concluded from the previous analysis that the time delays from operating experience had no influence on the predictive capacity of the model, it was decided to use the time delays obtained from the cross-correlation analysis.
For the development of this PLS model the same approach was used as for the initial grade-specific PLS model. Ultimately the set of variables given in table 4.4 was obtained that yielded the best predictive model for the given datasets.

Table 4.4 Variables used for the development of the multi-grade PLS model.

	Variable name
x69	Consistency tank
x92	Flow additive
x98	Flow suspension to wire
x99	Flow excessive water from wire
x109	Topformer consistency
x111	Vacuum suction roll
x128	Process value vacuumbox

The set of variables given in table 4.4 yielded a PLS model of three latent variables that were able to explain 70.4% of the variation in *X* and 78.9% of the variation in *Y* in the training data. With the combination it should be possible to predict 78.9% of the variation in *Y*. The relationship between the first *X-score* and first *Y-score* of this model is shown in figure 4.19.

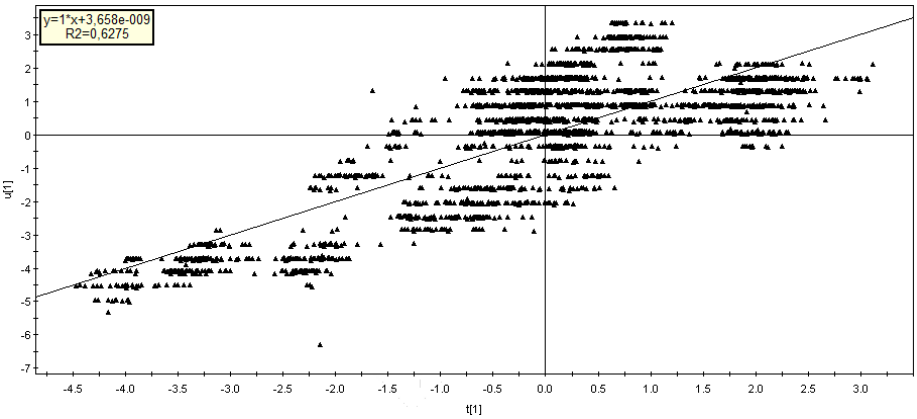

Figure 4.19 Relationship between first *X*-score and *Y*-score as an indicative parameter of the quality for the multi-grade PLS model.

As can be observed from figure 4.19, this relationship is also not completely linear. But again the relationship is more linear than non-linear which indicates that a model based on this data should have the ability to predict the paper quality to some extent.
The result of the new PLS model for the training data is shown in figure 4.20.

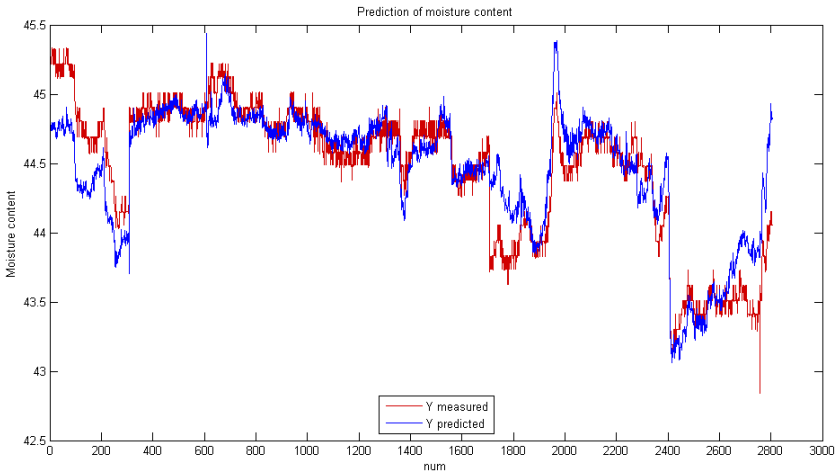

Figure 4.20 Prediction and actual measurement of the paper quality for the training dataset performed with the multi-grade PLS model.

According to equation 4.1 the coefficient of determination for this model is 0.7895. From figure 4.20 it can be seen that this model is again capable of predicting the general trend in paper quality quite well. The results of the test with validation data are shown in the figures 4.21 to 4.23.

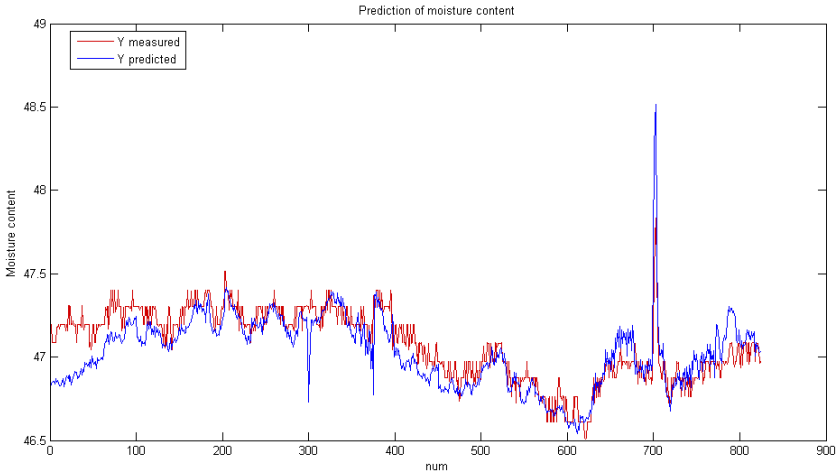

Figure 4.21 Prediction and actual measurement of the paper quality of the validation dataset 1 performed for the multi-grade PLS model.

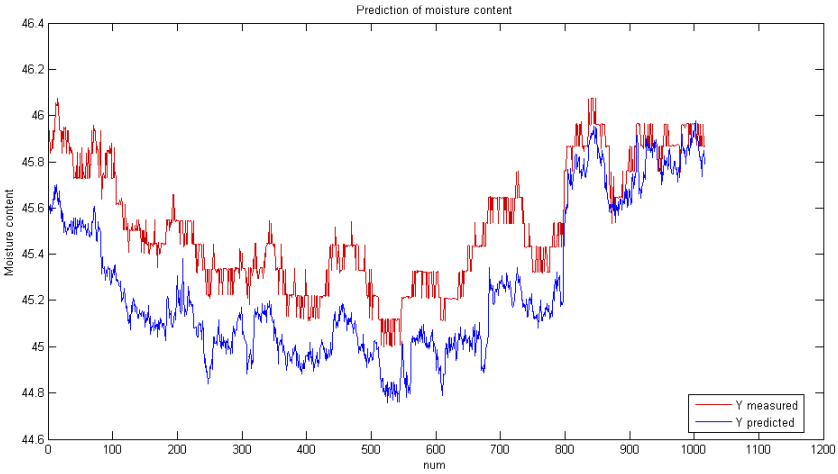

Figure 4.22 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the multi-grade PLS model.

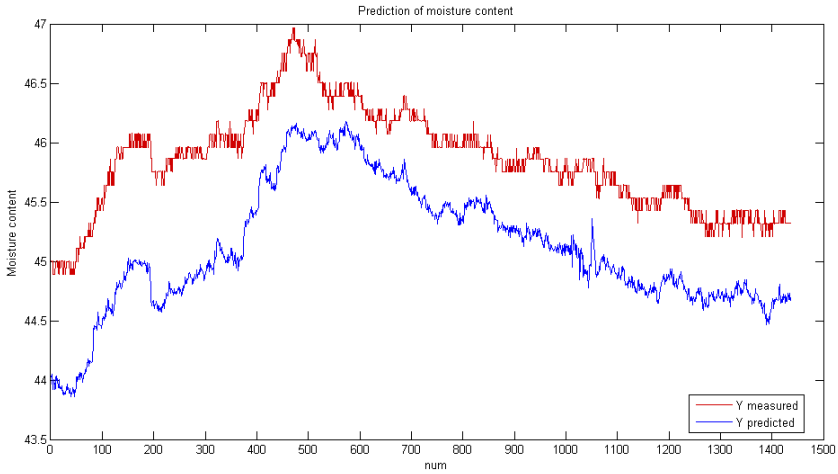

Figure 4.23 Prediction and actual measurement of the paper quality of the validation dataset 3 performed for the multi-grade PLS model.

The figures 4.21 to 4.23 showed that this model in combination with the validation data is also capable of predicting the general trend in the paper quality quite well. This is also noticeable from the mean square error values calculated using equation 4.4. For the validation datasets 1, 2 and 3 the mean square error values are respectively 0.0195, 0.0843 and 0.5935. Although the model can predict the paper quality for the first validation dataset with virtually no offset, an offset is still noticeable between the measured and predicted values of the validation datasets 2 and 3.
The result of the last test with the validation test data is shown in figure 4.24.

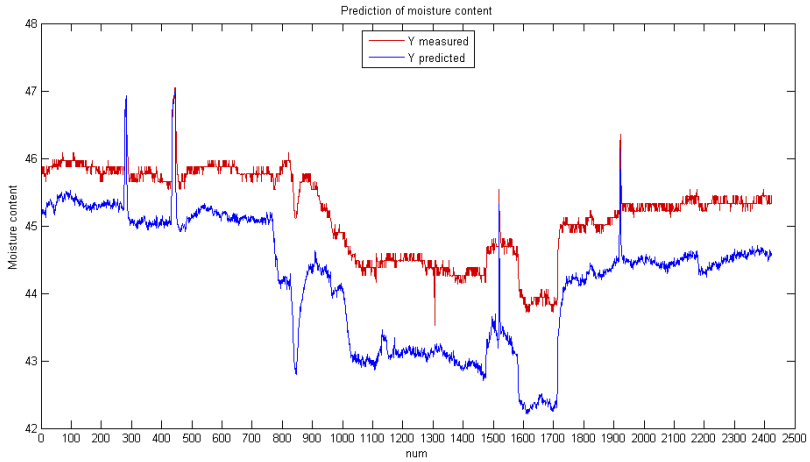

Figure 4.24 Prediction and actual measurement of the paper quality of the validation test dataset 4 performed for the multi-grade PLS model.

Just like with the validation datasets 2 and 3 this validation test data shows an offset between the measured and the predicted values of the paper quality. According to equation 4.4 the mean square error for the validation test data is 1.0910. As can be observed form figure 4.24 a large offset exists between the measured and predicted values of the paper quality for the validation test data. Nevertheless the model is still able to predict the general trend of the paper quality quite well. A comparison of the grade specific PLS model and the multi-grade PLS model showed that the prediction with the multi-grade PLS yielded a more accurate prediction.
From the preceding it can be concluded that a multi-grade PLS model is also capable of predicting the general trend in the paper quality quite good. However in most validation and validation test cases an offset between the measured and predicted values of the paper quality was still visible.

4.2.4 Robust LSSVM model

In a new attempt to find the cause of the offset found in the previous PLS models, the approach was to develop a Robust LSSVM model. With this approach it should be possible to determine whether the offset was caused by non-linearity in the process.
In order to make a fair comparison with the multi-grade PLS model, the same variables, training dataset, validation datasets, validation test data and cross-correlation time delays as for the multi-grade PLS model have been used for the development of the Robust LSSVM model.
In this thesis, the LSSVMlab v. 1.5 Matlab toolbox developed by Suykens *et al.*[20] was used for the development of the Robust LSSVM model.
Besides the training dataset and the variables, the Robust LSSVM model requires two additional parameters, i.e. *gam* and *sig2*. The parameter, *gam,* represents a regularization that determines the ratio between function smoothness and error minimization. The parameter, *sig2*, represents the kernel parameter.

The optimal parameters for the Robust LSSVM model in this thesis have been determined by trial and error. It was found that the optimal value of the *gam* parameter was 4 and that the optimal value of the *sig2* parameter was 4000.
The obtained result for the training data of the developed Robust LSSVM model is shown in figure 4.25.

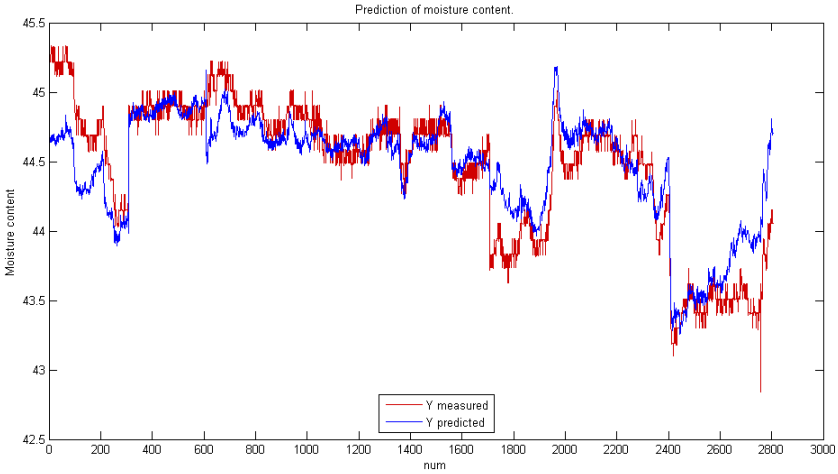

Figure 4.25 Prediction and actual measurement of the paper quality of the training data for the Robust LSSVM model.

According to equation 4.1 the coefficient of determination of this model is 0.7678. When the figures 4.20 and 4.25 are compared, there are virtually no differences visible. This can imply that the offset found in with the PLS models is not caused by non-linear relationships in the process. However the latter assumption can only be tested with the use of the validation data and validation test data. The results of the validation data are shown in the figures 4.26 to 4.28.

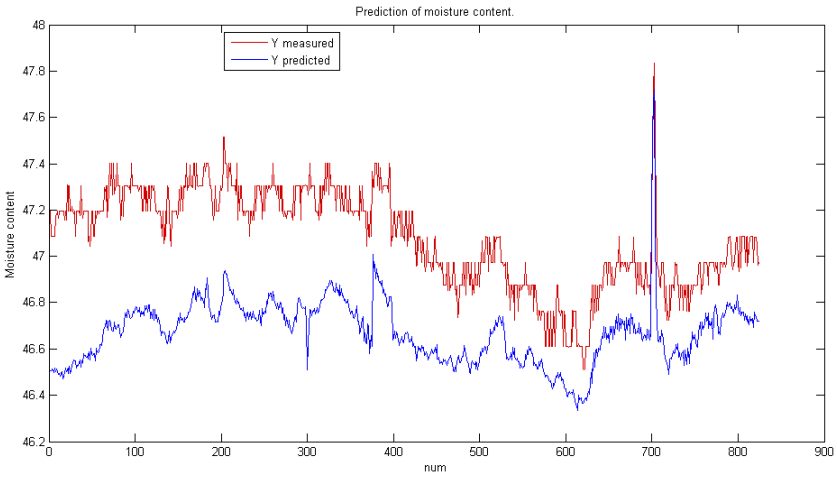

Figure 4.26 Prediction and actual measurement of the paper quality of the validation dataset 1 performed for the Robust LSSVM model.

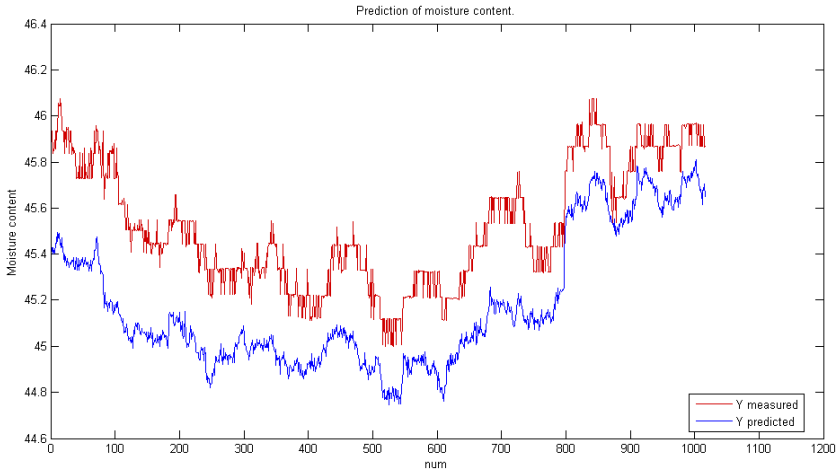

Figure 4.27 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the Robust LSSVM model.

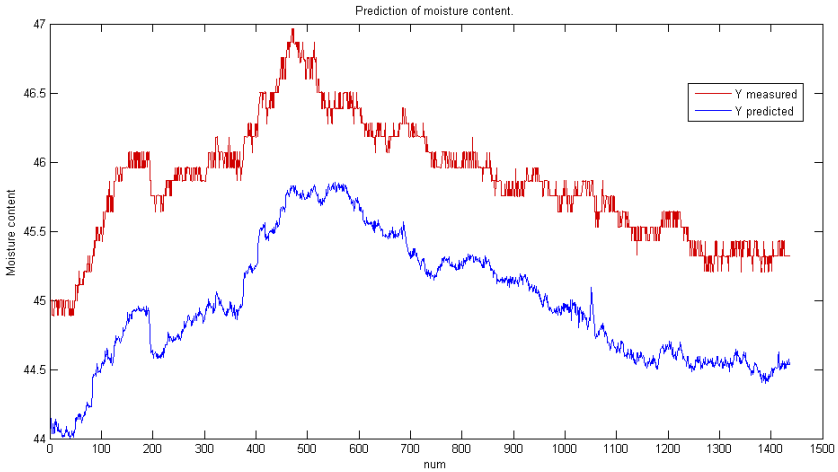

Figure 4.28 Prediction and actual measurement of the paper quality of the validation dataset 3 performed for the Robust LSSVM model.

According to equation 4.4 the mean square error for the validation datasets 1, 2 and 3 is respectively 0.1872, 0.1331 and 0.8105. When the figures 4.21 to 4.23 are compared to the figures 4.26 to 4.28 it can be observed that the results of the Robust LSSVM model show a larger offset between the measured and predicted values of the paper quality compared to the results of the PLS model. The latter again implies that the offset in the PLS models was not caused by non-linearity in the process. The final test of the Robust LSSVM was performed with the validation test data. The result of this test is shown in figure 4.29.

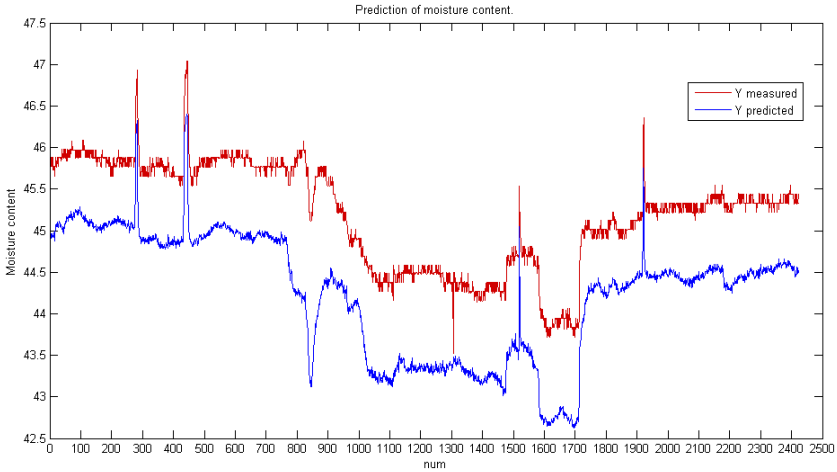

Figure 4.29 Prediction and actual measurement of the paper quality of the validation test data performed for the Robust LSSVM model.

According to equation 4.4 the mean square error for the validation test data is 1.0394. As expected the result of the validation test data shows virtually now difference compared to the result obtained with the multi-grade PLS model.
From the preceding results it can be concluded that the offset between the measured and predicted values of the paper quality could not be reduced or removed by using a Robust LSSVM model. Therefore it is most likely that the offset found with the PLS model was not caused by non-linear relationships in the process.

4.2.5 Online updating of PLS model

As the Robust LSSVM model did not yield the desired result of an offset free prediction of the paper quality, it was necessary to find another solution for the offset issue. A plausible cause for the offset issue was found in the literature. Several researchers, amongst others Sharmin [16], Mu [11] and Zhang [25], encountered the same problem when developing PLS models for various applications, amongst others in the prediction of polymer quality. All researchers came to the same conclusion for the reason of the offset between the measured and predicted values for a PLS model.
According to these researchers, the offset between the measured and predicted variables can be explained by the fact that when a PLS model is developed, the data for this model is scaled to unit variance using the average and standard deviation of each variable in that specific training dataset. New data is scaled to unit variance with the same averages and standard deviations of the training dataset, according to equation 4.5.

$$\hat{Y} = \left(b_1 \left(\frac{X_1 - \overline{X}_{1,t}}{\sigma_{X_{1,t}}} \right) + \ldots + b_n \left(\frac{X_n - \overline{X}_{n,t}}{\sigma_{X_{n,t}}} \right) + \left(\frac{\overline{Y}_t}{\sigma_{Y_t}} \right) \right) \cdot \sigma_{Y_t} \tag{4.5}$$

where y is the predicted y value, y_t is the average value of y in the training dataset, σ_{y_t} is the standard deviation of y in the training dataset, $b_1 \ldots b_n$ are the scaled model coefficients, $X_1 \ldots X_n$ are the measured value of the variables in the validation dataset, $\overline{X}_{1,t} \ldots \overline{X}_{n,t}$ are the average value of the variables in the training dataset and $\sigma_{X_{1,t}} \ldots \sigma_{X_{n,t}}$ are the standard deviation of the variables in the training dataset.

Upon rewriting equation 4.5, it can be observed that the equation actually exist of a constant part based on the actual measurements and a scaling term (bias) based on the model coefficients and the average value of the variables in the training dataset.

$$\hat{Y} = \left(b_1 \left(\frac{X_1}{\sigma_{X_{1,t}}} \right) + \ldots + b_n \left(\frac{X_n}{\sigma_{X_{n,t}}} \right) \right) \cdot \sigma_{Y_t} + \left(\left(-b_1 \left(\frac{\bar{X}_{1,t}}{\sigma_{X_{1,t}}} \right) - \ldots - b_n \left(\frac{\bar{X}_{n,t}}{\sigma_{X_{n,t}}} \right) \right) + \left(\frac{\bar{Y}_t}{\sigma_{Y_t}} \right) \right) \cdot \sigma_{Y_t} \tag{4.6}$$

The values of the averages and/or standard deviations may however change over time due to a set-point change of a variable or the presence of a disturbance. As a result of this change the average values and/or the standard deviation of the training data and validation data may differ, which in turn results in the offset between the measured and predicted values The researchers state that if the prediction of *Y* was performed using equation 4.7 where the original average values have been replaced by the average value of the variables in the validation dataset, an offset between the measured and predicted values will not occur.

$$\hat{Y} = \left(b_1 \left(\frac{X_1}{\sigma_{X_{1,t}}} \right) + \ldots + b_n \left(\frac{X_n}{\sigma_{X_{n,t}}} \right) \right) \cdot \sigma_{Y_t} + \left(\left(-b_1 \left(\frac{\bar{X}_{1,v}}{\sigma_{X_{1,t}}} \right) - \ldots - b_n \left(\frac{\bar{X}_{n,v}}{\sigma_{X_{n,t}}} \right) \right) + \left(\frac{\bar{Y}_v}{\sigma_{Y_t}} \right) \right) \cdot \sigma_{Y_t} \tag{4.7}$$

where Y_v is the average value of y in validation dataset and $\bar{X}_{1,v} \ldots X_{n,v}$ are the average value of the variables in the validation dataset.

Although equation 4.7 can also be used with adjusted values for the standard deviation of the variables, it is assumed in this thesis that only the average value of the variables will change over time. The standard deviation is assumed to be constant based on the idea that for a given variable set-point the value may differ within a set of limits. As the magnitude of these limits is assumed to be independent of the variable set-point, the standard deviation will be more or less constant.

The proposed solution of the researchers is to add an additional term to the original PLS model equation (equation 4.6) that compensates for an existing offset. For the calculation of this additional term ($bias_{calc}$), three different methods are proposed in this thesis which will be discussed in the following sections based on the multi-grade PLS model in combination with validation dataset 2 used for the multi-grade PLS model.

4.2.5.1 Simple bias update

The first calculation method is based on a relatively simple update mechanism. This method will calculate, at a certain time *t*, the difference between the previous measurement and the previous predicted value.

$$bias_{calc} = Y(t-1) - \hat{Y}(t-1) \tag{4.8}$$

The thus obtained $bias_{calc}$ is subsequently added to the new prediction of *Y* (equation 4.9), thereby reducing the offset between the measured and predicted values of the paper quality with each calculation step.

$$\hat{Y}(t) = \left(b_1 \left(\frac{X_1(t)}{\sigma_{X_{1,t}}} \right) + \ldots + b_n \left(\frac{X_n(t)}{\sigma_{X_{n,t}}} \right) \right) \cdot \sigma_{Y_t} + \left(\left(-b_1 \left(\frac{\bar{X}_{1,t}}{\sigma_{X_{1,t}}} \right) \ldots - b_n \left(\frac{\bar{X}_{n,t}}{\sigma_{X_{n,t}}} \right) \right) + \left(\frac{\bar{Y}_t}{\sigma_{Y_t}} \right) \right) \cdot \sigma_{Y_t} + bias_{calc} \tag{4.9}$$

The result for the multi-grade PLS model with validation dataset 2 and this calculation method is shown in figure 4.30.

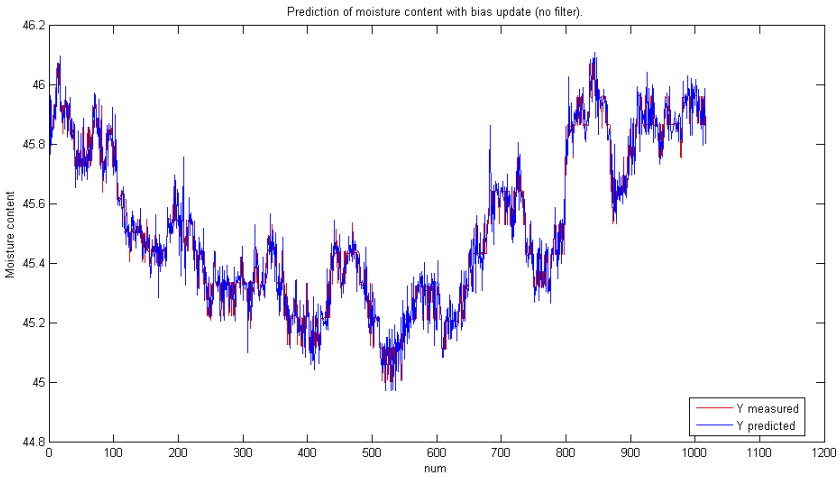

Figure 4.30 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the multi-grade PLS model in combination with a simple bias update.

According to equation 4.4 the mean square error of validation dataset 2 in combination with the simple bias update is 0.058. In comparison with the original mean square error of 0.0843, it can be stated that this bias update has improved the prediction of the paper quality slightly. On the other hand when figure 4.30 is compared to figure 4.22 it can be observed that due to the bias update significantly more noise is present in the prediction of the paper quality. The latter can be explained by the fact that the prediction is updated too often. As every new measurement is directly processed into a new $bias_{calc}$ value there is no room for error smoothing. Therefore a large error in the measurement is directly translated into excessive noise in the prediction of the paper quality.

4.2.5.2 Bias update with correction factor

The second calculation method is an extension of the simple bias update. This second calculation method is still based on the previous measured and predicted value, only now a smoothing factor is introduced to reduce excessive noise in the prediction. At time $t = 0$ no value can be calculated for the bias as there is not yet a prediction, therefore it is assumed that the bias is 0 at that specific time.
At a time $t > 0$ the error between the previous measured and predicted value can be determined.

$$error = Y(t-1) - \hat{Y}(t-1) \tag{4.10}$$

Subsequently the $bias_{calc}$ can be determined using equation 4.11.

$$bias_{calc} = \alpha \cdot bias_{old} + (1.0 - \alpha) \cdot error \tag{4.11}$$

here α is the correction factor which can have a value between 0.0 and 1.0. A correction factor of 0.0 means that the prediction is updated using the simple bias update method. The latter will thus result in significantly more noise in the prediction as can be seen from the previous section. On the other hand a correction factor of 1.0 means that the prediction is not adjusted at all.

The latter can be used if the offset is constant over a long period of time or when the correction factor is updated frequently by hand. However when looking at the figures 4.35 to 4.37 there is no constant offset visible. Therefore it is preferable to automatically update the $bias_{calc}$ term. In order to perform the latter, the correction factor has to have a value larger than 0.0 but smaller than 1.0. In this thesis a value of 0.9 was used for the correction factor. This value was chosen mainly because the data is updated with a relatively high frequency in this process. With this correction factor the bias is in fact averaged over ten measurements, which should be adequate for this process to obtain a prediction with less noise. The last unknown variable in equation 4.11 is the $bias_{old}$ term. This term represents the previous calculated bias term as calculated with equation 4.11 for time $t > 0$. At time $t = 0$, the $bias_{old}$ term can be an estimate of the offset however in this thesis the initial $bias_{old}$ term is equal to 0.
When this bias update is incorporated into the multi-grade PLS model and tested with validation dataset 2, the result as shown in figure 4.31 is obtained.

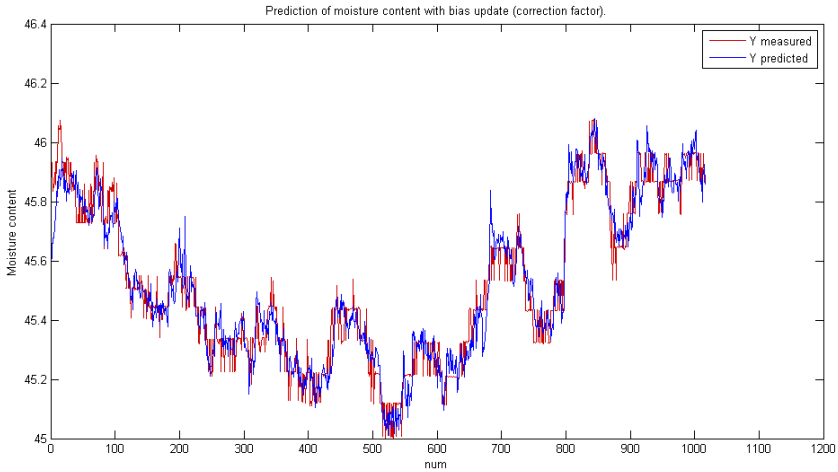

Figure 4.31 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the multi-grade PLS model in combination with a bias update with correction factor.

According to equation 4.4 the mean square error for the validation in figure 4.31 is 0.0054. In comparison with the original mean square error of 0.0843, the mean square error has become roughly ten times smaller. In addition from figure 4.31 it can be observed that with the correction factor applied the noise in the prediction virtually disappeared. Although the predicted values line up well with the measured values, one remark has to be made. Because of the averaging of the $bias_{calc}$ term, the predicted values are slightly out of sync with the measured values. The latter can be controlled with the value of the correction factor. Therefore it might be possible that a different correction factor will show an even better result. However the correction factor will always be a concession between fit and noise reduction.

4.2.5.3 Bias update with moving window

The third calculation method is the more complex solution of the three methods suggested in this thesis. This calculation method is based on the claim that equation 4.7 is an improvement of the original PLS model equation. When a prediction is made using both equations a difference will exist between these predicted values. Mathematically this difference can be derived to a $bias_{calc}$ term and a $\Delta\overline{Y}$ term based on a subtraction of both equations.

$$bias_{calc} = \left(-b_1 \left(\frac{\bar{X}_{1,v}}{\sigma_{X_{1,t}}} \right) - \ldots - b_n \left(\frac{\bar{X}_{n,v}}{\sigma_{X_{n,t}}} \right) \right) - \left(-b_1 \left(\frac{\bar{X}_{1,t}}{\sigma_{X_{1,t}}} \right) - \ldots - b_n \left(\frac{\bar{X}_{n,t}}{\sigma_{X_{n,t}}} \right) \right) \tag{4.12}$$

$$\Delta \bar{Y} = \left(\frac{\bar{Y}_v}{\sigma_{Y_t}} \right) - \left(\frac{\bar{Y}_t}{\sigma_{Y_t}} \right) \tag{4.13}$$

In order to make this solution work, it is required to define a method for the calculation of the average values of the validation data, or in a real application the new process data. The latter can be performed using a moving window average. With a moving window average, the average of a certain variable is determined on the basis of *n* previous measurements. At a certain time *t* the average value of a variable is determined according to equation 4.14.

$$\bar{Y}_v(t) = \frac{\left(Y_v(t-1) + Y_v(t-2) + \cdots + Y_v(t-n) \right)}{n} \tag{4.14}$$

At time *t + 1* the last measured value is added to the range of values used, while the last value in the range of the previous calculation is removed. (equation 4.15) This way the average value is adjusted along with the frequency of new measurements.

$$\bar{Y}_v(t+1) = \frac{\left(Y_v(t) + Y_v(t-1) + \cdots + Y_v(t-n) \right)}{n} \tag{4.15}$$

With the moving window it is possible to obtain an accurate average value for the variables of the validation data with relatively little noise because of the averaging. The amount of noise present in the average value is mainly determined by the size of the moving window. The size required is generally determined on the frequency of new measurements in the process. As the frequency of the data used in this thesis is relatively high, a moving window of ten measurements was used. Ultimately the updated prediction of *Y* can be determined using equation 4.16.

$$\hat{Y} = \left(b_1 \left(\frac{X_1 - \bar{X}_{1,t}}{\sigma_{X_{1,t}}} \right) + \ldots + b_n \left(\frac{X_n - \bar{X}_{n,t}}{\sigma_{X_{n,t}}} \right) + \left(\frac{\bar{Y}_t}{\sigma_{Y_t}} \right) + bias_{calc} + \Delta \bar{Y} \right) \cdot \sigma_{Y_t} \tag{4.16}$$

With this bias update incorporated into the multi-grade PLS model, the result as shown in figure 4.32 is obtained for the validation dataset 2.

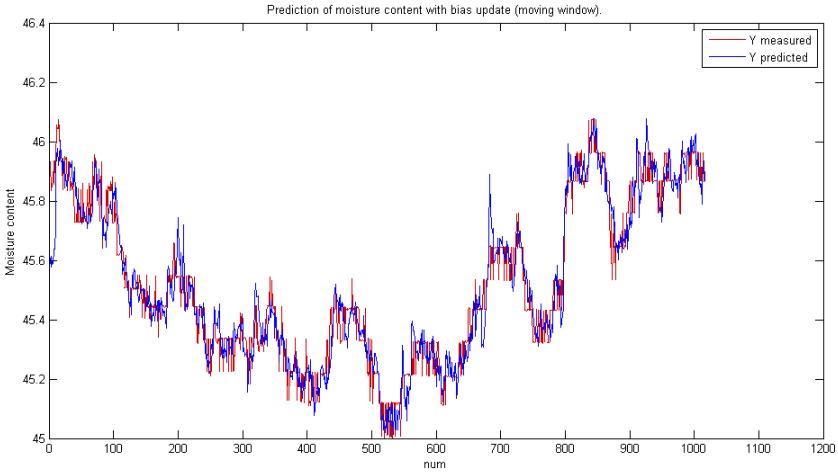

Figure 4.32 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the multi-grade PLS model in combination with a bias update with moving average.

According to equation 4.4 the mean square error for this prediction is 0.0056. In comparison with the original mean square error of 0.0843, this update method also reduces the mean square error roughly by a factor ten. From figure 4.32 it can be observed that little noise is present in the prediction because of the averaging in the moving window. For this calculation method the size of the moving window is a concession between fit and noise reduction

Comparing the result of this calculation method to the method with bias update with correction shows that these methods are comparable. As in this thesis the assumption is made that only the average value of the variables change, the bias update with correction factor is preferred because of its simplicity compared to the bias update with moving window. However in the case that a change in both the average values and the standard deviations of the variables is assumed, the bias update with moving window will probably be the better choice as that method can handle both deviations.

From the preceding it can be concluded that the offset between the measured and predicted values with the multi-grade PLS model was caused by changes in the average values of the variables in the training dataset and validation datasets. Further it can be concluded that by the addition of a bias term to the original PLS model equation that compensates for this difference, the offset can be removed from the prediction without additional noise in the prediction.

4.2.6 Cause – effect PLS model

Although the previous developed models showed good results for the prediction of the moisture content, all the models have two drawbacks. When looking at the grade-specific PLS model and the multi-grade model, the variables used fall within a relatively small range compared to the total number of variables available. When looking at the schematic process description, it can be observed that the variables used, are only found in the additive section and the paper machine section. Furthermore it can be observed that the variables used, are placed relatively close to the paper quality measurement. In a real life situation this would result in a too small time margin to take corrective actions in case of a deviation in the process operation. Nevertheless, because of the goodness of the fit of the models obtained, it is possible to use the models as a substitute measurement of the paper quality in case the real measurement of the paper quality fails. The second drawback is found in the nature of the variables used. Most of the variables used are controlled in an automated way. Due to the latter it is not surprising that these variables show a high correlation with the paper quality as this is the important factor of this control strategy. However such a control strategy makes it difficult to take a corrective action in case of a deviation in the process operation.

In order to overcome this issue, it was decided to remove this type of variables from the dataset and to end up with a dataset that only contains manual controllable variables or variables that represent a process value. As a consequence the list of variables in Appendix I is not valid anymore for this dataset. Therefore the new list of variables is shown in Appendix V.
For the development of the new PLS model the datasets 5, 6 and 7 were again used as training data as the combination of these datasets span a large space of the process operation. For the validation data the datasets 1, 2, 3 and 4 were used. For the validation test data a new dataset will be used with process raw data that has only be synchronized. This validation test data contains data from a production in the period 21-02-2011 23:24 to 22-02-2011 14:30. During this production paper was produced with a gram weight of 90 g/m^2 and the grades A, B and C.
The cross-correlation analysis on the training data and validation data yielded 49 process input variables which are likely to have influence on the paper quality. The complete list of variables in combination with the time delays can be found in Appendix VI. The obtained process input variables were subsequently further analyzed for containing equal information or information that is not relevant for the predictive model. Ultimately this resulted in the set of process input variables as shown in table 4.5.

Table 4.5 Variables used for the development of the cause-effect PLS model.

	Variable name
x5	Flow refiner
x8	Refiner 2 loading
x9	Refiner 3 loading
x20	Flow refiners
x25	Refiner 5 loading
x49	Flow tank 723
x52	Redox measurement tank 724
x55	Flow recovered material
x62	Flow binder additive
x63	Flow dewatering additive
x65	Flow binder additive
x66	Flow flocculation additive
x67	Flow retention additive
x70	Incoming consistency suspension on wire
x71	Flow suspension to wire
x72	Flow excessive water from wire
x75	Temperature tank 721
x76	Headbox consistency
x77	Headbox consistency (additives)
x81	Topformer consistency
x82	Topformer consistency (additives)
x83	Dewatering vacuumsection
x87	Process value vacuumbox 9

With the variables from table 4.5 a PLS model was developed. This resulted in a model with 5 latent variables which explained 77.9% of the variation in *X* and 91.4% of the variation in *Y* present in the training dataset. The relation between the first *X-score* and the first *Y-score* for this model is shown in figure 4.33.

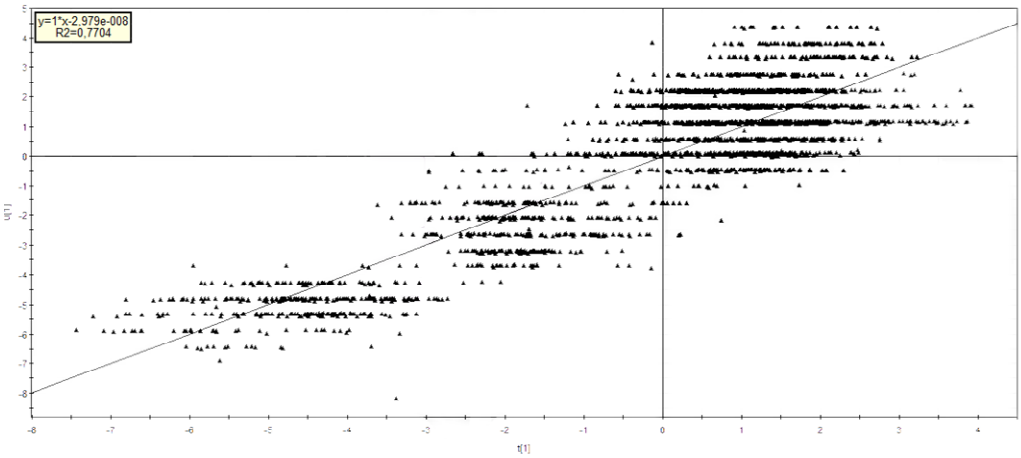

Figure 4.33 Relationship between first *X*-score and *Y*-score as an indicative parameter of the quality for the cause-effect PLS model.

As can be observed from figure 4.33 the relationship is more linear than non-linear. Based on the latter it can be assumed that this model has to ability to predict the paper quality.
The result of the PLS model for the training dataset is shown in figure 4.34.

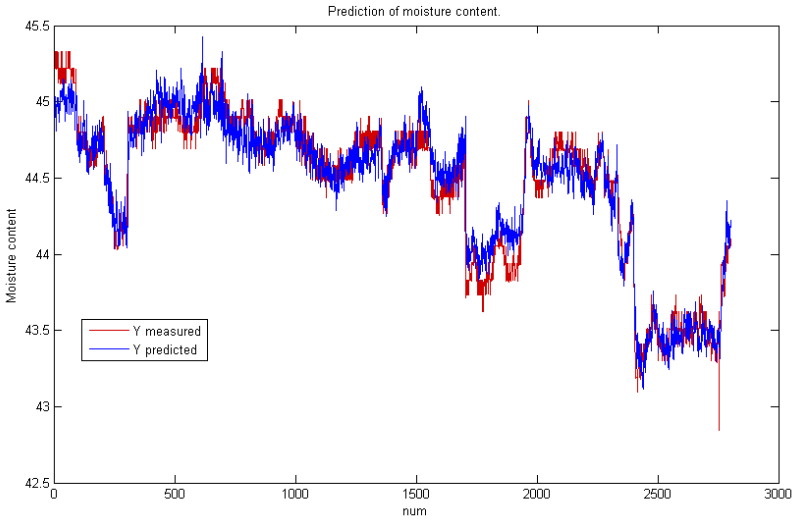

Figure 4.34 Prediction and actual measurement of the paper quality of the training data for the cause-effect PLS model.

According to equation 4.1 the coefficient of determination for this model is 0.9138. From figure 4.34 it can be observed that the general trend in the paper quality is almost perfectly predicted for the training data. However the first test of the PLS model is the test with validation data. As it is expected from previous analysis that there is an offset noticeable between the predicted values and the measured values of the paper quality, the results will be displayed both for the prediction without bias update and with bias update. The results of this test are shown in the figures 4.35 to 4.38.

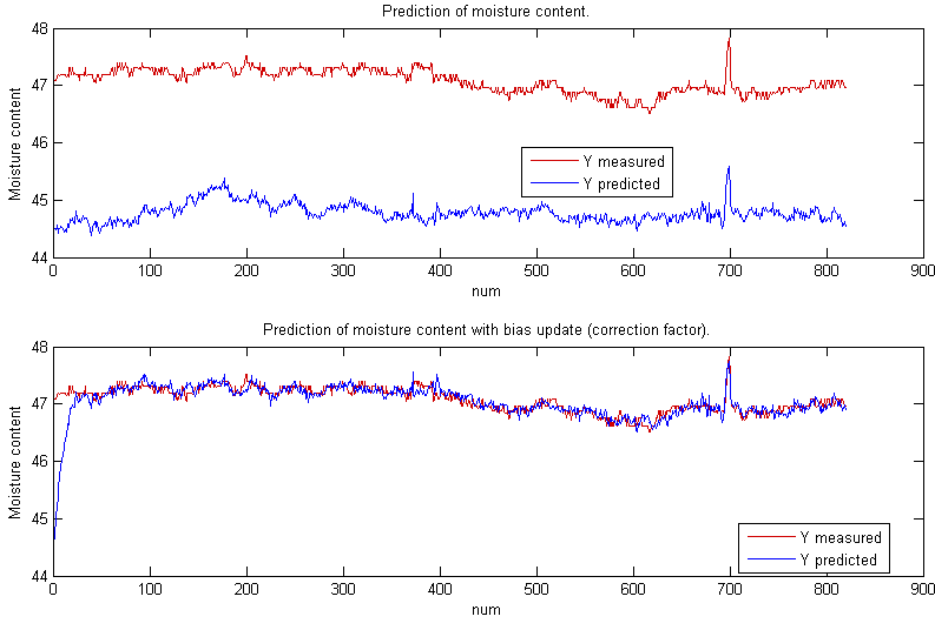

Figure 4.35 Prediction and actual measurement of the paper quality of the validation dataset 1 performed for the cause-effect PLS model without and with bias update.

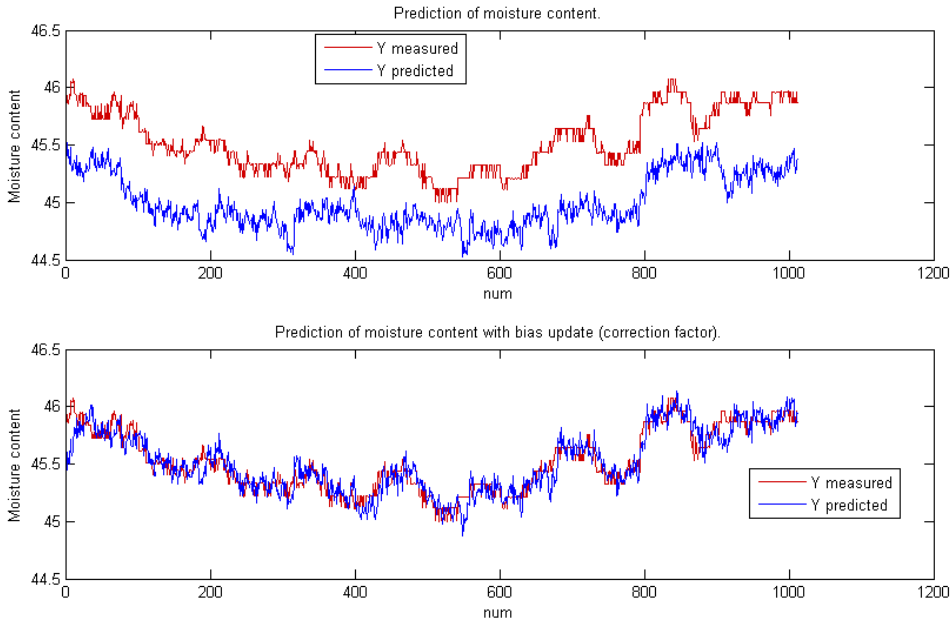

Figure 4.36 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the cause-effect PLS model with and without bias update.

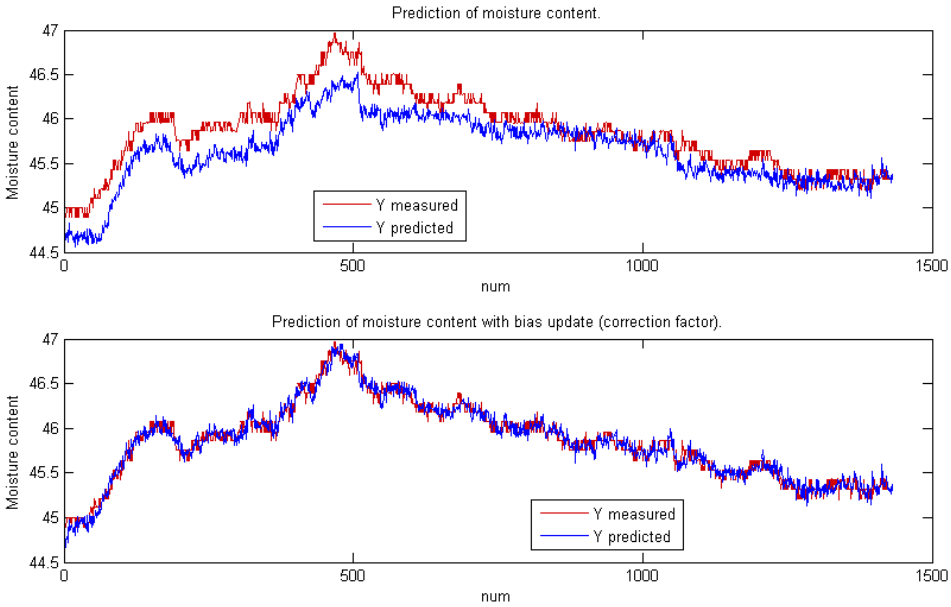

Figure 4.37 Prediction and actual measurement of the paper quality of the validation dataset 3 performed for the cause-effect PLS model with and without bias update.

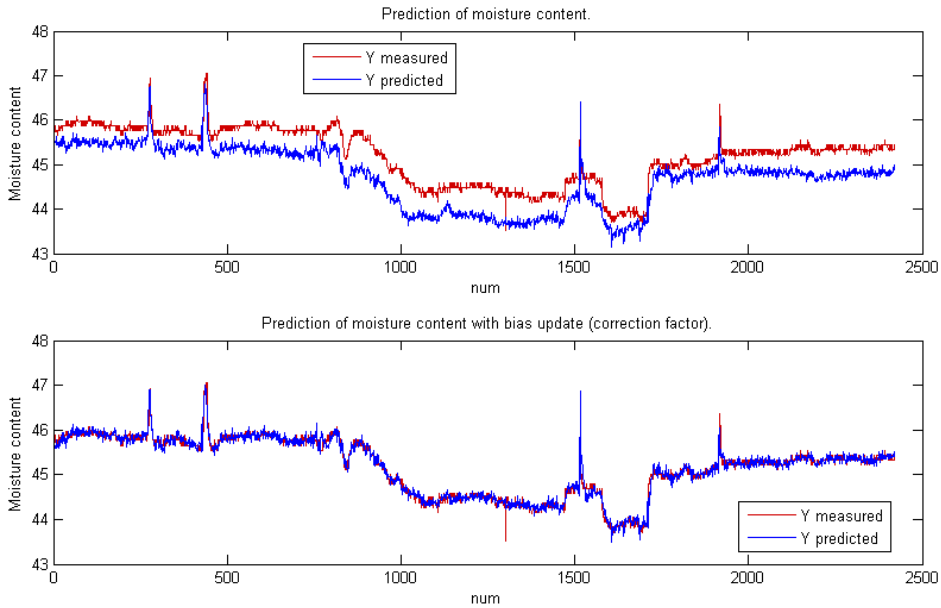

Figure 4.38 Prediction and actual measurement of the paper quality of the validation dataset 4 performed for the cause-effect PLS model with and without bias update.

According to equation 4.4 the mean square error for the datasets 1, 2, 3 and 4 without bias update are respectively 4.2800, 0.3075, 0.0690 and 0.2644. For the prediction with bias update the mean square error is respectively 0.0845, 0.0134, 0.0082 and 0.0138.
The results from the validation data show that the model is indeed capable of predicting the paper quality good up to a large extend. As expected the measured and predicted values show an offset, however, as can be seen this offset can be removed by the addition of the bias update mechanism to the prediction.

From the mean square error values it can be observed that due to the bias update, the values could be decreased by roughly a factor ten. The final test of the model is performed with the validation test data. The result of this test is shown in figure 4.39.

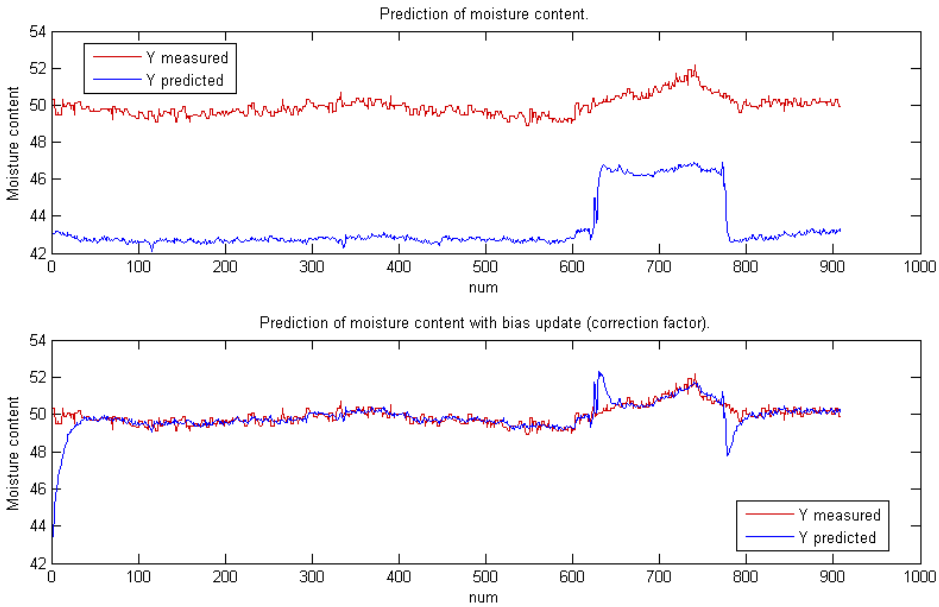

Figure 4.39 Prediction and actual measurement of the paper quality of the validation test dataset performed for the cause-effect PLS model with and without bias update.

According to equation 4.4 the mean square error for the validation test data without bias update mechanism is 43.9, whereas the mean square error for this data with bias update mechanism is 0.4205.

From figure 4.39 it can be observed that for the validation test data, the model is capable of predicting the general trend in the paper quality. There is, however, a deviant peak noticeable in the predicted values of the paper quality. The cause of this deviant peak is unknown. As the peak is partly removed with the bias update mechanism, it can be assumed that this peak is caused by deviant measurements of one of the variables used in the model.

Although the obtained model shows good results and an accurate prediction of the paper quality apart form the possible addition of a bias update mechanism, the model is based on a relatively large number of process input variables. Therefore the importance of the used process input variables was further examined. This resulted in the figure as shown in figure 4.40.

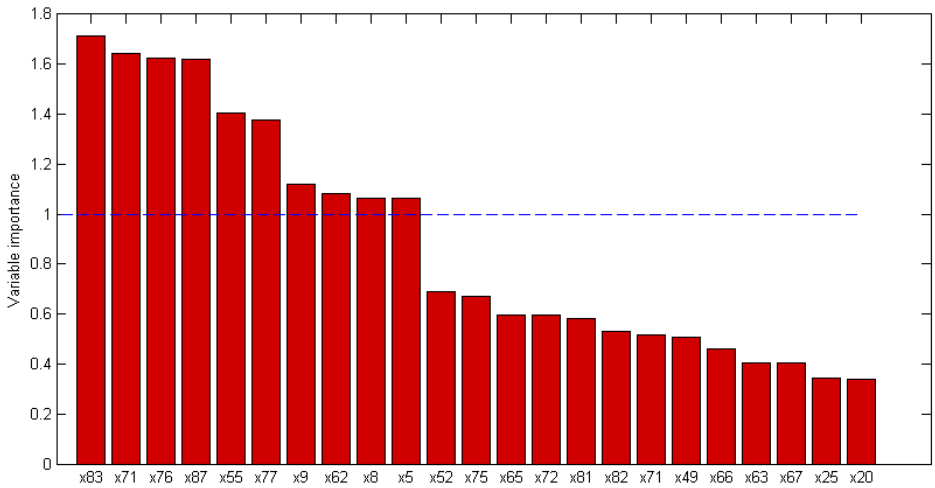

Figure 4.40 Variable importance for the variables used in the cause-effect PLS model.

In figure 4.40 the variables with a variable importance larger than 1, are considered important for the prediction of the paper quality. As can be observed from figure 4.40, this criterion is met for ten variables. With these important variables another PLS model was developed. This PLS model consisted of 4 latent variables that explained 82.8% of the variation in *X* and 86.8% of the variation in *Y* present in the training dataset. The thus obtained figure for the first *X-score* and first *Y-score* is shown in figure 4.41.

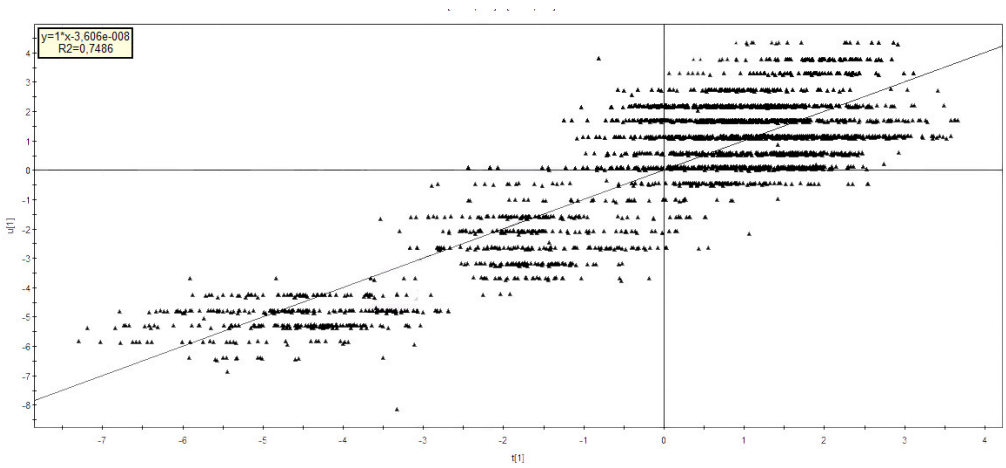

Figure 4.41 Relationship between first *X*-score and *Y*-score as an indicative parameter of the quality for the second cause-effect PLS model.

From figure 4.41 it can be observed that the relation between the first *X-score* and first *Y-score* is slightly deteriorated compared to the relation of the original cause effect PLS model. Nevertheless the obtained relation is still good and can still be denoted as being more linear than non-linear. Thus is can be assumed that this model can predict the paper quality up to a large extent. The result of the PLS model for the training data is shown in figure 4.42

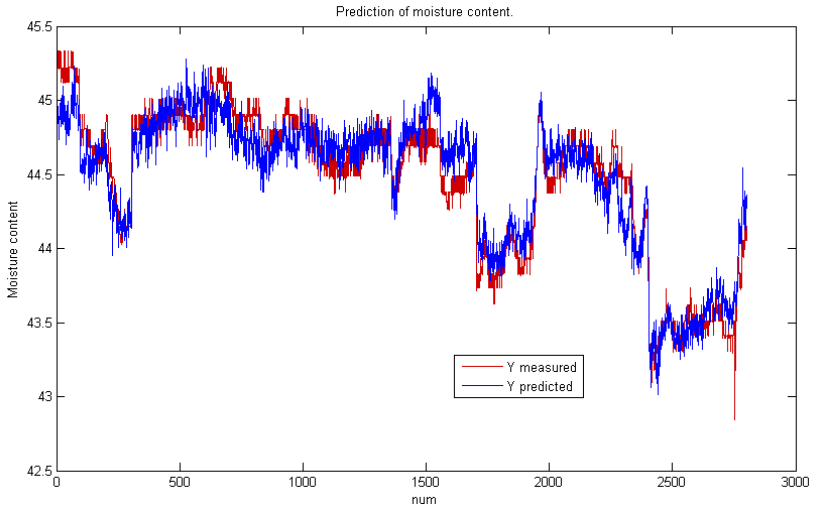

Figure 4.42 Prediction and actual measurement of the paper quality of the training data for the second cause-effect PLS model.

According to equation 4.1 the coefficient of determination for this model is 0.8684. In comparison with the original cause – effect PLS model the coefficient of determination is slightly deteriorated. This deterioration can also be observed when comparing figure 4.34 and figure 4.42. The results for the validation data of this model are shown in the figures 4.43 to 4.46.

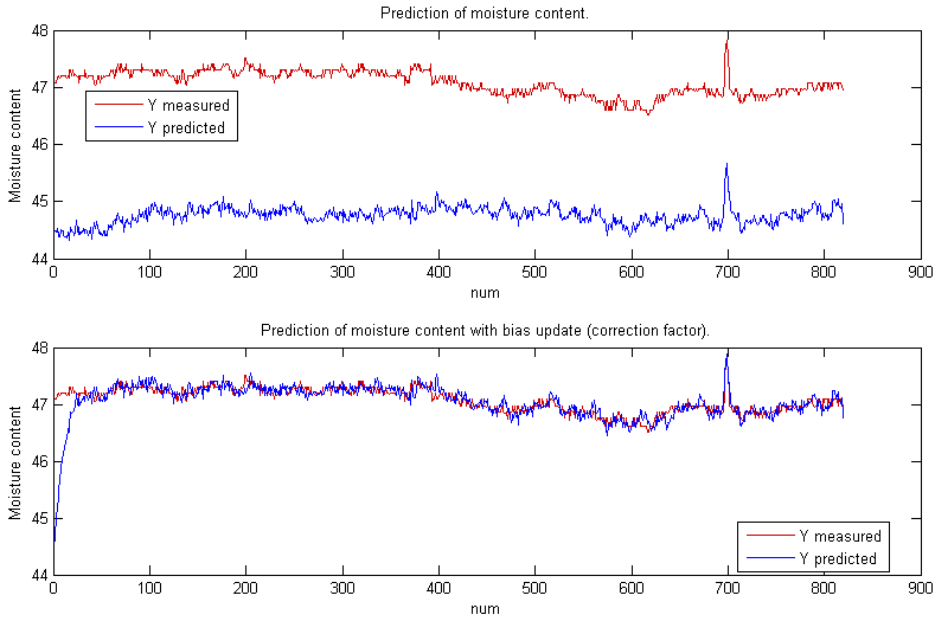

Figure 4.43 Prediction and actual measurement of the paper quality of the validation dataset 1 performed for the second cause-effect PLS model without and with bias update.

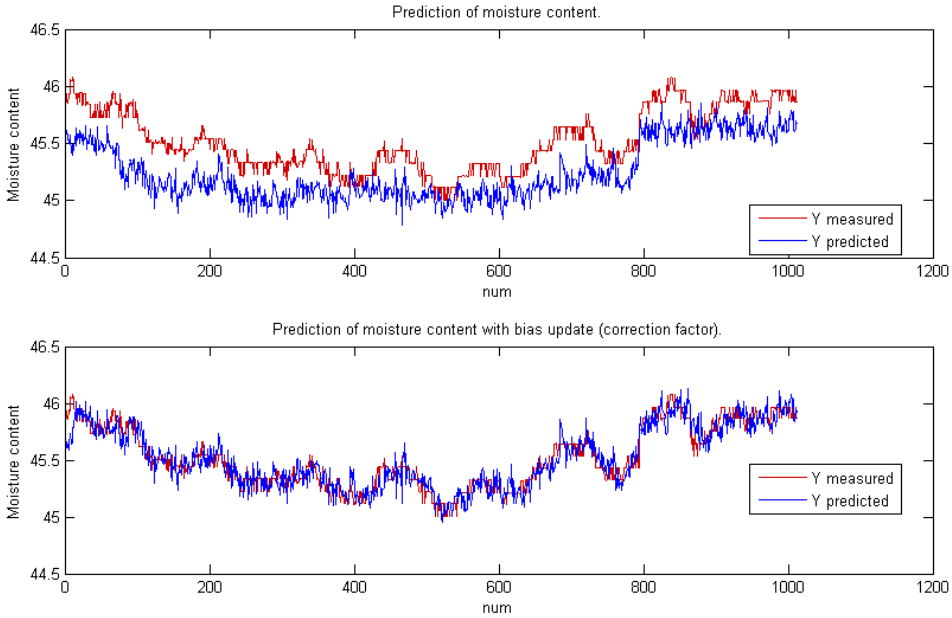

Figure 4.44 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the second cause-effect PLS model without and with bias update.

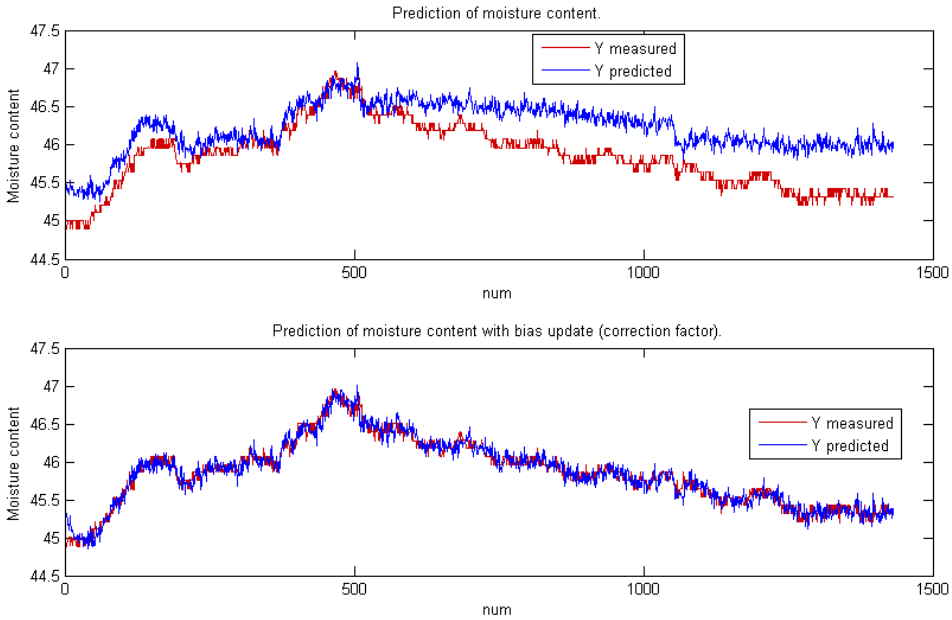

Figure 4.45 Prediction and actual measurement of the paper quality of the validation dataset 3 performed for the second cause-effect PLS model without and with bias update.

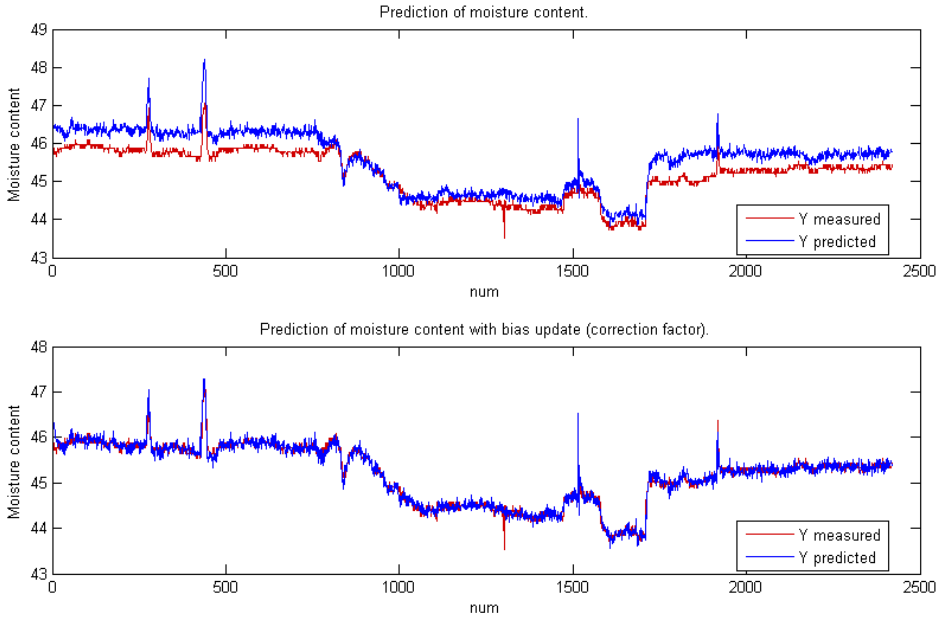

Figure 4.46 Prediction and actual measurement of the paper quality of the validation dataset 4 performed for the second cause-effect PLS model without and with bias update.

According to equation 4.4 the mean square error for the validation without bias update mechanism for the datasets 1, 2, 3 and 4 is respectively 5.4385, 0.0932, 0.1826 and 0.1831. For the validation with bias update mechanism the mean square error is respectively 0.0593, 0.0113, 0.0109 and 0.0153. Comparing these obtained mean square error values with the mean square error values of the original cause-effect PLS model shows these values are comparable. The latter can also be observed when comparing the figures 4.35 to 4.38 and 4.43 to 4.46.
One remark has to be made for validation dataset 3, which shows a deviant behaviour for the second cause-effect PLS model. This deviant behaviour is probably caused by a deviant measurement of one of the variables during the production of this dataset.
This second cause-effect PLS model was also tested with the validation test data. The result of this test is shown in figure 4.47.

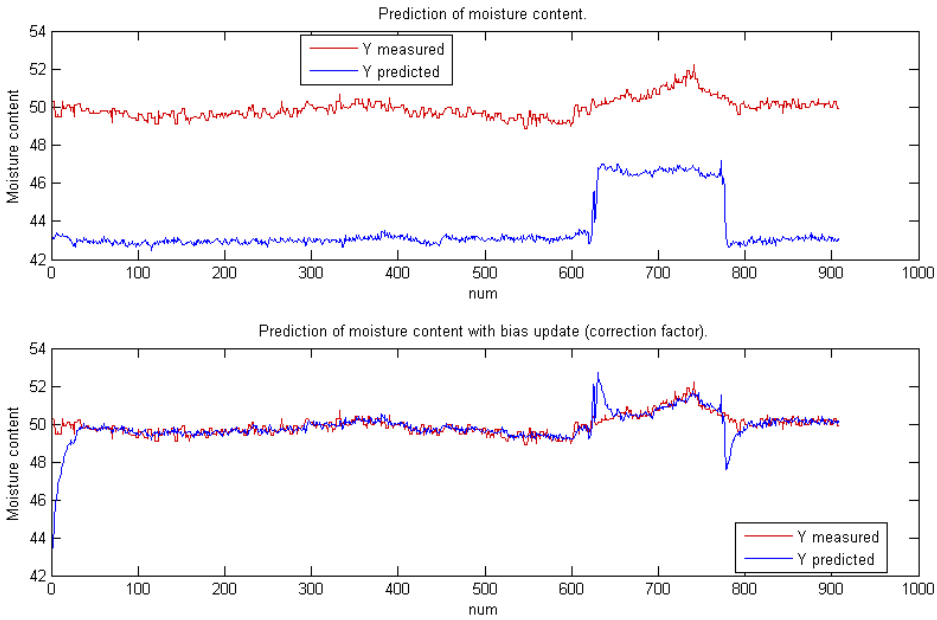

Figure 4.47 Prediction and actual measurement of the paper quality of the validation test dataset performed for the second cause-effect PLS model without and with bias update.

According to equation 4.4 the mean square error for the validation test data for this model without bias update mechanism is 41.0983, whereas the mean square error for this data with bias update mechanism 0.4249 is. Again this result is comparable with the result obtained with the first cause-effect PLS model.

Based on the obtained results it can be concluded that it is possible to develop a cause-effect type PLS model only based on manually controllable variables and indicating process values. However, the only deviant variable in the list of variables is the process value of the vacuum box. A detailed examination of this variable, however, showed that the control mechanism of this variable did not work correctly during the production runs used in this thesis, as shown for the training data in figure 4.48.

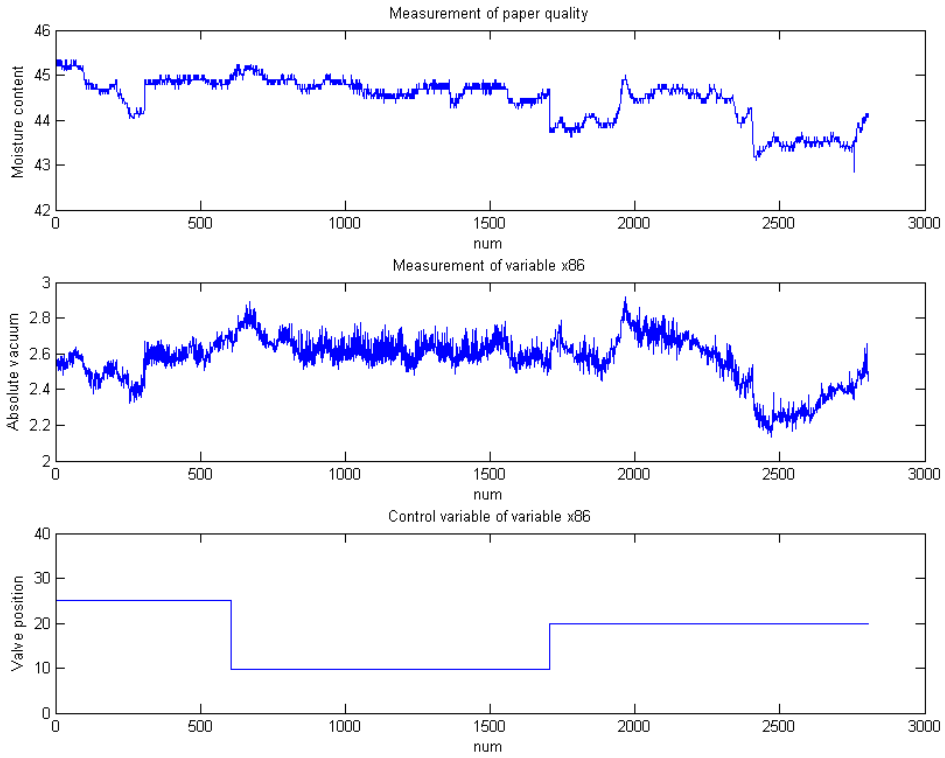

Figure 4.48 Comparison of vacuum variable (x86) behaviour in combination with control variable of x86 compared to the measurement of the paper quality.

As can be observed from figure 4.48 the control variable has a constant value during the production runs, whereas variable x86 fluctuates according to the measurement of the paper quality. From this it can be concluded that the control variable does not operate correctly. In the actual control mechanism the control variable should fluctuate with the measurement of the paper quality and should the measurement of variable x86 be constant.
If the assumption is made that this variable normally would not be incorporated in the model, as it is should have a constant measurement, and the other variables of the second cause-effect PLS model would remain the same a new model is obtained. This model consists of three latent variables which explained 80.7% of variation in *X* and 86.2% of the variation in *Y*.
The relationship between the first *X-score* and *Y-score* for this model is shown in figure 4.49.

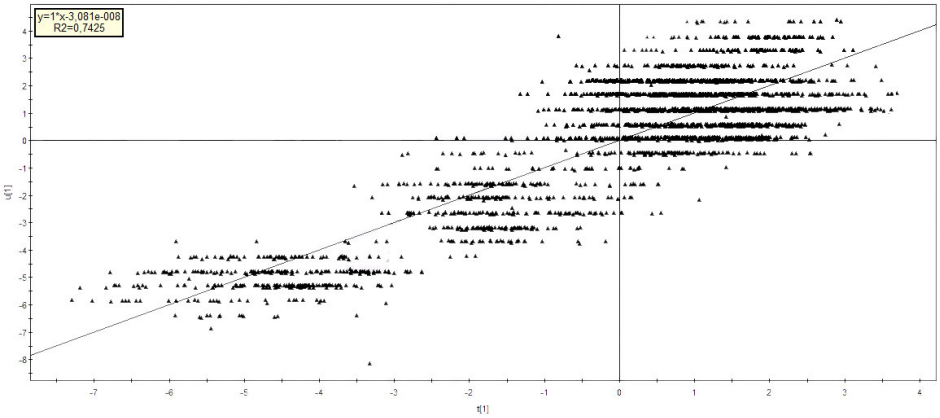

Figure 4.49 Relationship between first *X*-score and *Y*-score as an indicative parameter of the quality for the cause-effect PLS model with vacuum variable removed.

As can be observed from figure 4.48 the relationship for this is model is still good compared to the relationship shown in figure 4.41. Furthermore the relationship is still more linear than non-linear. The result of this PLS model for the training data is shown in figure 4.50.

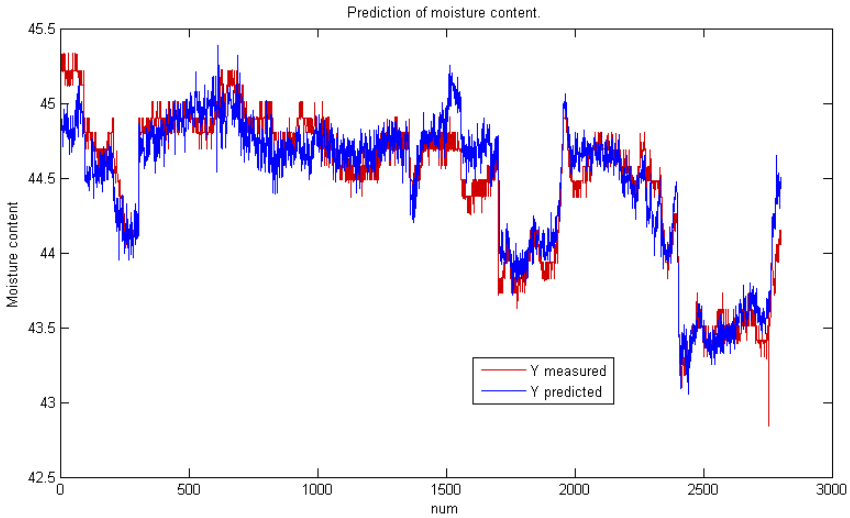

Figure 4.50 Prediction and actual measurement of the paper quality of the training for the cause-effect PLS model with vacuum variable removed.

According to equation 4.1 the coefficient of determination for this model is 0.8623. Compared to the coefficient of determination obtained for the second cause-effect PLS model, this coefficient slightly deteriorated. However from a comparison of figure 4.42 and 4.50 this effect is not noticeable. The results obtained for the validation datasets is shown in the figures 4.51 to 4.54.

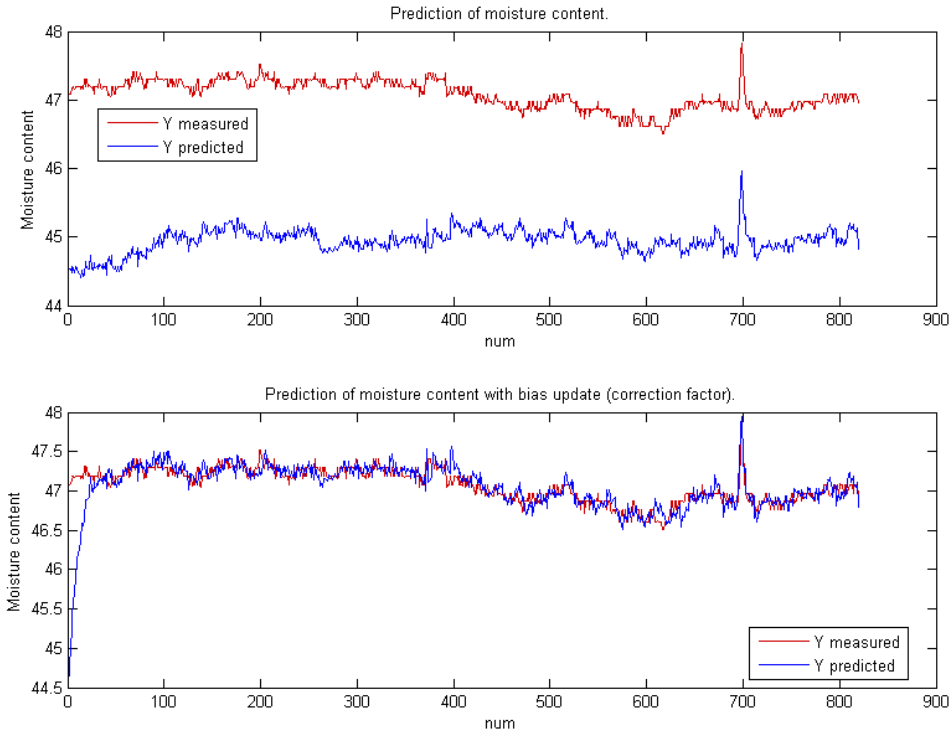

Figure 4.51 Prediction and actual measurement of the paper quality of the validation dataset 1 performed for the cause-effect PLS model with vacuum variable removed without and with bias update.

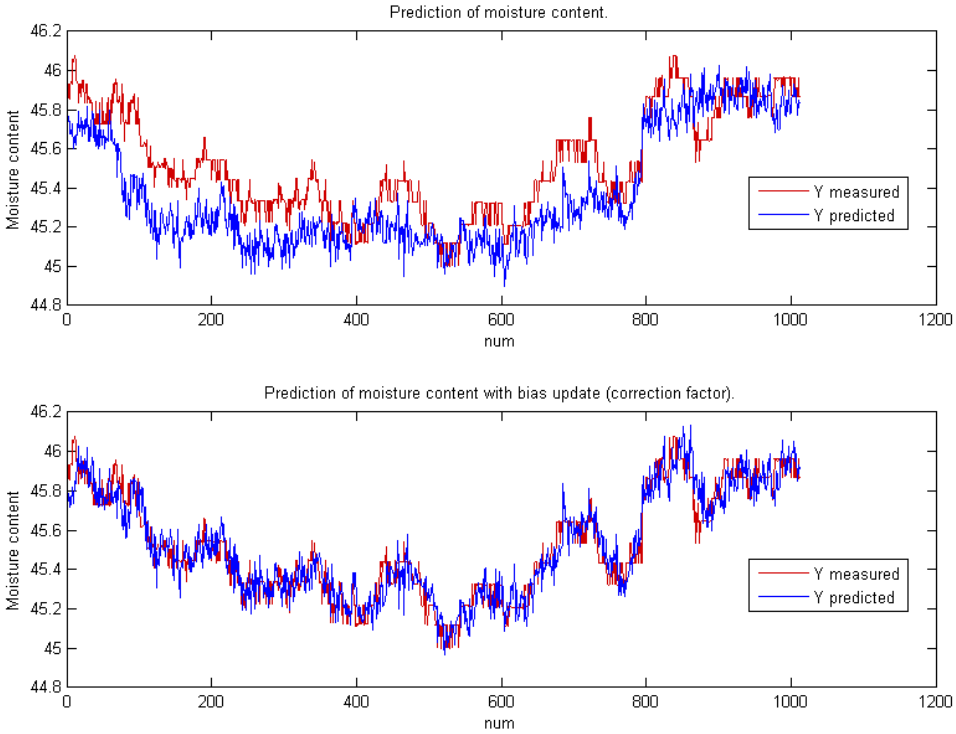

Figure 4.52 Prediction and actual measurement of the paper quality of the validation dataset 2 performed for the cause-effect PLS model with vacuum variable removed without and with bias update.

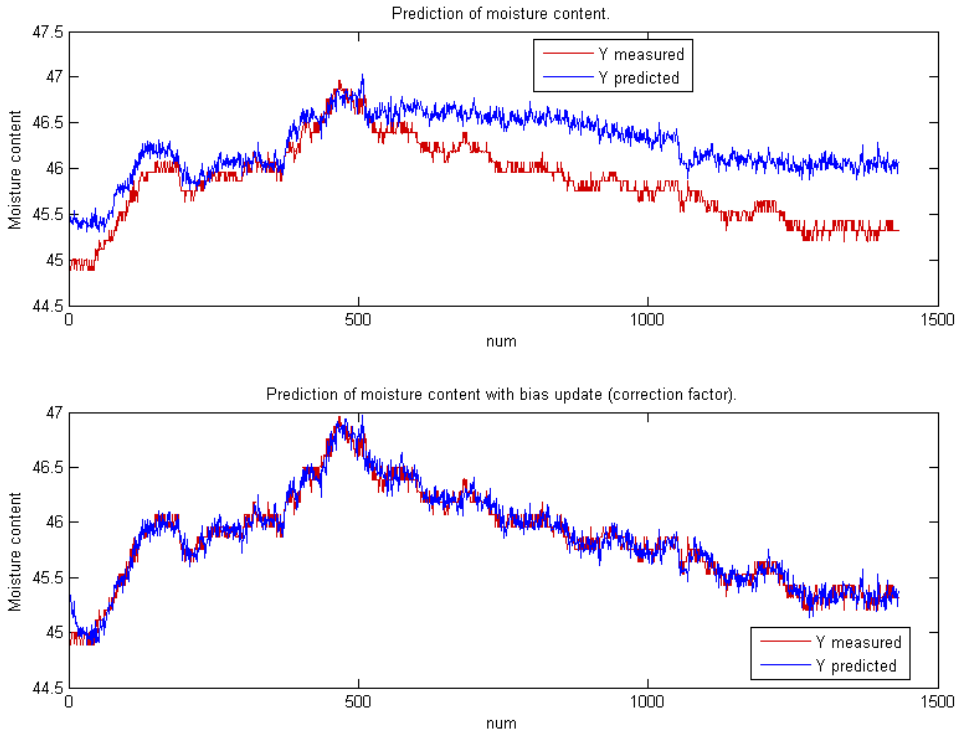

Figure 4.53 Prediction and actual measurement of the paper quality of the validation dataset 3 performed for the cause-effect PLS model with vacuum variable removed without and with bias update.

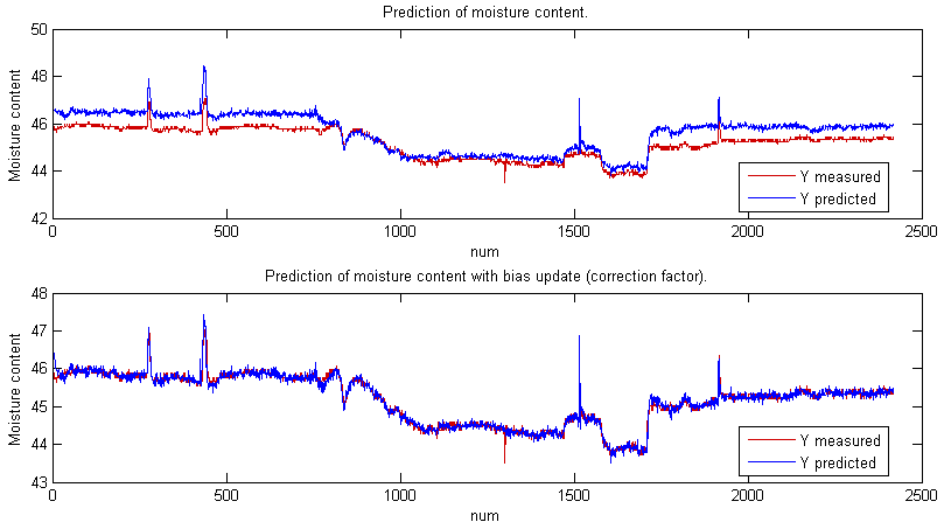

Figure 4.54 Prediction and actual measurement of the paper quality of the validation dataset 4 performed for the cause-effect PLS model with vacuum variable removed without and with bias update.

According to equation 4.4 the mean square error for the validation datasets 1, 2, 3 and 4 without bias update mechanism is respectively 4.6083, 0.0435, 0.2186 and 0.2668. The mean square error for these datasets with bias update mechanism is respectively 0.0565, 0.0091, 0.0092 and 0.0157. Compared to the results obtained for the second cause-effect PLS model, these results show a slight improvement. The latter can also be observed when comparing the figures 4.43 to 4.46 to

the figures 4.51 to 4.54. The result of the final test of the model with the validation test data is shown in figure 4.55.

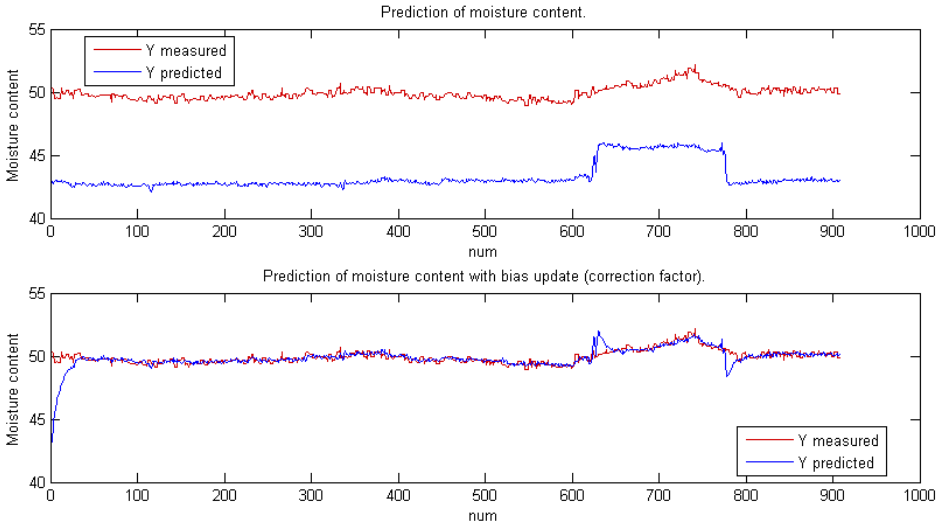

Figure 4.55 Prediction and actual measurement of the paper quality of the validation test dataset performed for the cause-effect PLS model with vacuum variable removed without and with bias update.

According to equation 4.4 the mean square error for the validation test data without bias update mechanism is 44.7151, whereas the mean square error for this data with bias update mechanism 0.3949 is. The results obtained for the validation test data are slightly deteriorated compared to results obtained for the validation test data of the second cause-effect PLS model.
However based on the obtained results for the PLS model with the vacuum variable removed it can be stated that it is likely that the vacuum variable does not belong in the predictive model. Nevertheless further research is required in order to prove this assumption.

5. Conclusion

The research project described in this thesis was aimed on developing a predictive model for the paper quality in the Sappi paper mill in Nijmegen. It was found in the literature that controlling the paper quality is a common problem in paper mills, as the measurement is affected by a large dead time. This results in a situation in which it is unknown what exact type of product is produced. It would therefore be advantageous if this paper quality could be determined at an early stage of the production process.
As it was found that additional real measurements was not an option due to a specific measuring requirement, the next best option was found in the development of a predictive model.

As raw process data may contain deviant measurements and large deviations between the magnitudes of the measured variables, the process data had to be pre-processed in order to be used in the development of a predictive model. A useful pre-processing technique was found in a principal component analysis, as this analysis both check the data for inconsistencies as well as reducing the dimension of the data while retaining vital information.
The principal component analysis for the data used in this thesis, showed that for each dataset, the data points were arranged in clusters of consecutive data. It was found that this arrangement was caused by variation of variables representing temperatures, flows or valve positions in the process. Therefore it could be concluded that all data used for the development of a predictive model represented normal operating data.

Subsequently a partial least square regression was performed to develop an initial predictive model. The developed model showed, however, an offset between the measured values and the predicted values of the paper quality. In order to find the cause of this offset, different time delays for the variables were used, as well as a different training data set that covered a larger space of the production process, in addition, the non-linear Robust LSSVM regression technique was investigated. All these alternative approaches did not yield the desired result of the removal of the offset. The solution for the removal of offset was found in the literature. It was found that the offset was probably caused by a shift in the average value of the variables over time, which in this thesis was also found in the principal component analysis of the data used.
The proposed solution consisted of an update mechanism that compensates for the offset present in the prediction. In this thesis it was found that a combination of a simple bias update based on the difference between the previous measurement and prediction corrected by a chosen correction factor was able to remove the offset from the prediction.
A more detailed analysis of the PLS model obtained showed however that the model used a small range over variables close to the actual paper quality measurement. In a real life situation this would have the consequence that there is too little time to take a corrective action in case of a deviation in the process operation. Therefore a new approach was developed in which only manually controllable variables and important measured process values were incorporated in the datasets.
This approach yielded a robust predictive PLS model with 4 latent variables that explained 82.8% of the variation in X and 86.8% of the variation in Y and showed even a good prediction for untrained data. However one remark has to be made regarding this model, i.e. one of the variables represents a process value of a vacuum box which was intended to be automatically controlled. However due to a failure in the control mechanism this control did not function during the production runs used in this thesis.

If the assumption was made that this variable normally would not be incorporated into the model, a new model was obtained. The results of this model showed that it is most likely that the vacuum variable should indeed not be incorporated into the model. However further research is required to prove this assumption.

It should also be mentioned that the models developed in this thesis were only developed using data from production runs with the gram weight of 90 g/m^2. Therefore the model obtained is in principle only valid for this gram weight. More research is required if the paper quality has to be predicted for other paper gram weights.

References

[1] Abdi H. *Partial least squares regression and projection on latent structure regression (PLS regression).* Wiley Interdisciplinary Reviews: Computational Statistics 2010; 2(1):97-106

[2] Ahola T. *Intelligent estimation of web break sensitivity in paper machines.* Oulu University; 2005 Available from: http://herkules.oulu.fi/isbn9514279573/

[3] Bissessur Y., Martin E.B., Morris A.J. *Monitoring the performance of the paper making process.* Control Eng Pract 1999 11; 7(11):1357-68

[4] Chen N. *Support vector machine in chemistry.* Singapore; Hackensack, NJ: World Scientific Pub.; 2004

[5] De Brabanter J. *LS-SVM regression modelling and its applications.* Faculteit Toegepaste Wetenschappen Katholieke Universiteit Leuven; 2004

[6] Geladi P., Kowalski B.R. *Partial least-squares regression: A tutorial.* Anal Chim Acta 1986;185:1-17

[7] Hendriks, H. [Internet] *Geschiedenis Papierfabriek "Gelderland"* Accessed: December 2010. Available from: http://www.noviomagus.nl/Gastredactie/Hendriks/Papier/Papier.htm

[8] Holik H. *Handbook of paper and board.* Weinheim: Wiley-Vch; 2006

[9] Li ISC. *Prediction and prevention of sheet break using partial least squares and an expert system.* University of British Columbia; 1997 Available from: http://hdl.handle.net/2429/6536

[10] Li P.Y., Ramaswamy S., Bjegovic P. *Pre-emptive control of moisture content in paper manufacturing using surrogate measurements.* Transactions of the Institute of Measurement and Control 2003 March 01; 25(1):36-56

[11] Mu S., Zeng Y., Liu R., Wu P., Su H., Chu J. *Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process.* J Process Control 2006; 16(6):557-66

[12] Roffel B., Betlem B. *Process dynamics and control: Modeling for control and prediction.* Chichester: John Wiley & Sons; 2006

[13] Sappi. *Environmental statement 2009 Nijmegen mill.* Available from: http://www.sappi.com/SappiWeb/About+Sappi/Sappi+Fine+Paper+Europe/Nijmegen+Mill.htm

[14] Sappi. *The paper making process. from wood to coated paper.* Sappi Idea Exchange Available from: http://www.sappi.com/SappiWeb/Tools+and+resources/Technical+brochures/Technical+brochures.htm

[15] Sappi. *Sustainability report 2010.* Available from: http://www.sappi.com/SappiWeb/Corporate+responsibility/Publications/Brochures+and+reports.htm

[16] Sharmin R., Sundararaj U., Shah S., Vande Griend L., Sun Y.-. *Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant.* Chemical Engineering Science 2006; 61(19):6372-84

[17] Skoglund A., Brundin A., Mandenius C. *Applying process monitoring with multivariate analysis through a knowledge-based systems approach to a paperboard machine.* Comput Ind 2005 6; 56(5):472-8

[18] Skoglund A., Brundin A., Mandenius C. *Monitoring a paperboard machine using multivariate statistical process control.* Chemometrics Intellig Lab Syst 2004 9/28; 73(1):3-6

[19] Slätteke O,. *Modeling and control of the paper machine drying section.* Lund: Department of Automatic Control, Lund University; 2006 Available from: http://www.lu.se/o.o.i.s?id=12588&postid=24357

[20] Suykens J.A.K. , Van Gestel T. , De Brabanter J. , De Moor B. , Vandewalle J. *Least squares support vector machines.* River Edge, NJ: World Scientific; 2002

[21] Valyon J., Horvath G., Meeting. *A robust LS-SVM regression.* 2005

[22] Vapnik V.N. *Statistical learning theory.* New York: Wiley; 1998

[23] Wade, M.J. and Balderud, J. Using multivariate regression techniques to analyse the performance of a steam heated drying process. Industrial electronics and control applications, 2005. ICIECA 2005.; 2005.

[24] Wise B.M., Gallagher N.B., Bro R., Shaver J.M., Windig W., Koch R.S. *Chemometrics tutorial for PLS_Toolbox and solo v. 5.51.* Wenatchee: Eigenvector Research, Inc.; 2006 Available from: http://www.eigenvector.com/software/pls_toolbox.htm

[25] Zhang J. *Offset-free inferential feedback control of distillation compositions based on PCR and PLS models.* Chemical Engineering and Technology 2006; 29(5):560-6

Appendix I: Variables used for the development of predictive models

	Variable name
y	Paper quality
x1	Gram weight
x2	Grade
x3	Consistency flow refiners
x4	Flow refiner
x5	Flow refiner
x6	Level tank 703
x7	Refiner 1 loading
x8	Refiner 2 loading
x9	Refiner 3 loading
x10	Refiner 1 energy
x11	Refiner 2 energy
x12	Refiner 3 energy
x13	Flow refiner
x14	Flow refiner
x15	Incoming pressure refiners
x16	Outgoing pressure refiners
x17	Incoming pressure refiner 1
x18	Incoming pressure refiner 1
x19	Outgoing pressure refiner 1
x20	Incoming pressure refiner 2
x21	Incoming pressure refiner 2
x22	Outgoing pressure refiner 2
x23	Outgoing pressure refiner 3
x24	Incoming pressure refiner 3
x25	Refiner 1 temperature
x26	Refiner 2 temperature
x27	Refiner 3 temperature
x28	Refiner 1 loading
x29	Refiner 2 loading
x30	Refiner 3 loading
x31	Level tank 702
x32	Consistency flow refiners
x33	Flow refiners
x34	Level tank 704
x35	Refiner 4 loading
x36	Refiner 5 loading
x37	Refiner 6 power
x38	Refiner 7 power
x39	Refiner 4 loading
x40	Refiner 5 loading
x41	Incoming pressure refiner 4
x42	Incoming pressure refiner 4
x43	Outgoing pressure refiner 4

x44	Incoming pressure refiner 5
x45	Incoming pressure refiner 5
x46	Outgoing pressure refiner 5
x47	Incoming pressure refiners
x48	Outgoing pressure refiners
x49	Refiner 4 temperature
x50	Refiner 5 temperature
x51	Refiner 6 temperature
x52	Refiner 7 temperature
x53	Pressure after refiner 5
x54	Pressure after refiner 4
x55	Pressure after refiners
x56	Flow tank 711
x57	Consistency tank 711
x58	Conductivity measurement tank 711
x59	Redox measurement tank 711
x60	Flow tank 712
x61	Consistency tank 712
x62	Conductivity measurement tank 712
x63	Redox measurement tank 712
x64	Flow tank 713
x65	Consistency tank 713
x66	Conductivity measurement tank 713
x67	Redox measurement tank 713
x68	Flow tank 795
x69	Consistency tank 795
x70	Conductivity measurement tank 795
x71	Redox measurement tank 795
x72	Flow tank 796
x73	Consistency tank 796
x74	Conductivity measurement tank 796
x75	Redox measurement tank 796
x76	Flow tank 723
x77	Consistency tank 723
x78	Conductivity measurement tank 724
x79	Redox measurement tank 724
x80	Charge measurement tank 724
x81	Flow material recovery
x82	Flow recovered material
x83	Consistency flow recovered material
x84	Flow recovered material (storage)
x85	Valve position tank 726
x86	Flow fixation additive
x87	Flow retention additive
x88	Flow retention additive
x89	Flow binder additive
x90	Flow dewatering additive
x91	Flow dilution water

x92	Flow binder additive
x93	Flow flocculation additive
x94	Flow retention additive
x95	Consistency natural dewatering
x96	Consistency natural dewatering (additives)
x97	Incoming consistency suspension on wire
x98	Flow suspension to wire
x99	Flow excessive water from wire
x100	Consistency natural dewatering (additives)
x101	Consistency natural dewatering
x102	Temperature tank 721
x103	Headbox consistency
x104	Headbox consistency (additives)
x105	Speedratio
x106	Conductivity measurement headbox
x107	Redox measurement headbox
x108	Topformer consistency
x109	Topformer consistency (additives)
x110	Dewatering vacuumsection
x111	Vacuum suction roll
x112	Process value vacuumbox 7
x113	Valve position vacuumbox 7
x114	Process value vacuumbox 8
x115	Valve position vacuumbox 8
x116	Process value vacuumbox 9
x117	Valve position vacuumbox 9
x118	Process value vacuumbox 10
x119	Valve position vacuumbox 10
x120	Process value vacuumbox 1
x121	Valve position vacuumbox 1
x122	Process value vacuumbox 2
x123	Valve position vacuumbox 2
x124	Process value vacuumbox 3
x125	Valve position vacuumbox 3
x126	Process value vacuumbox 4
x127	Valve position vacuumbox 4
x128	Process value vacuumbox 5
x129	Valve position vacuumbox 5
x130	Process value vacuumbox 6
x131	Valve position vacuumbox 6
x132	Process value vacuumbox 11
x133	Valve position vacuumbox 11
x134	Press dewatering
x135	Press 4 dewatering
x136	Machine speed
x137	Press 1 dewatering
x138	Press 2 dewatering
x139	Press 3 dewatering

Appendix II: Analysis of the dataset 2 to 7

AII.1 Analysis dataset 2

In the second dataset no variables were detected which were offline during the production run. The variables x62, x71, x96, x117, x120, x121 and x129 had a constant value during the run period. These variables were however left as is, since they will be automatically ignored in the development of a predictive model because of a lack of variance. The principal component analysis on this dataset yielded a model with 15 principal components that explained 71.4% of the variation present in the dataset. The score scatter plot of the first two principal components is shown in figure AII.1.

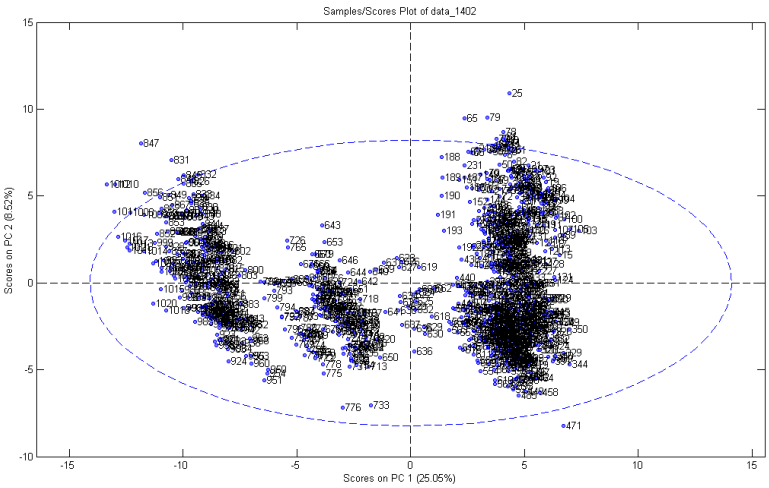

Figure AII.5.1 Score scatter plot of the first two principal components of the PCA model of dataset 2.

As can be observed from figure AII.1 most of the data points fall within the 95% confidence limit. A more detailed examination yielded that the number of outliers present, is in accordance with the five percent of acceptable outliers. Further it can be observed from figure AII.1, that the data is again distributed in cluster of consecutive data. An overview of the exact division of the data points in the clusters can be found in table AII.1.

Table AII.1 Distribution of the data points in the cluster in dataset 2.

Cluster	Data points	Period
1	1 – 617	14-02-2010 12:00 – 14-02-2010 22:16
2	618 – 640	14-02-2010 22:17 – 14-02-2010 22:39
3	641 – 799	14-02-2010 22:40 – 15-02-2010 01:18
4	800 – 1021	15-02-2010 01:19 – 15-02-2010 05:00

In order to determine the cause of the clustering the contribution of the variables in each cluster was determined. This yielded the plot as shown in AII.2.

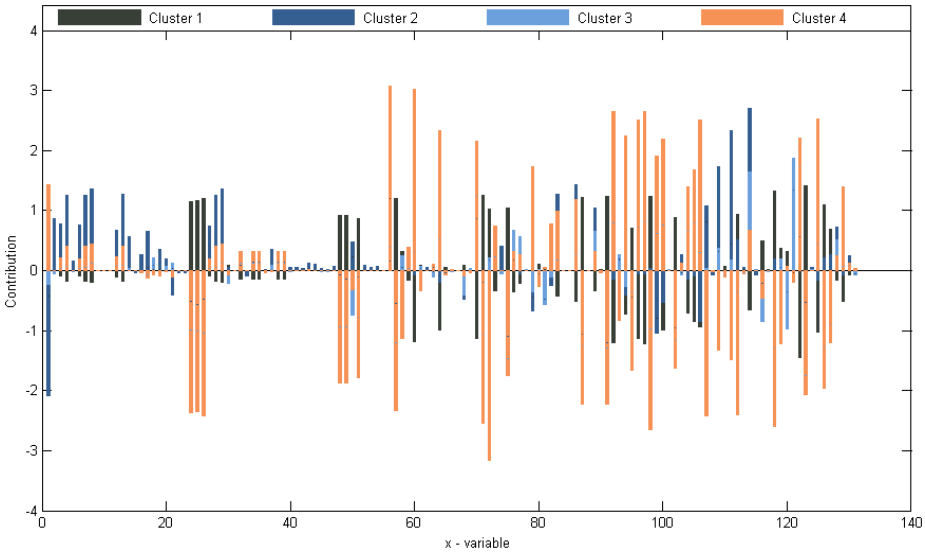

Figure AII.5.2 Variable contribution in each cluster in dataset 2.

From figure AII.2 it can be observed that the major deviations in the variable contributions can be found at variables x60 and higher. From Appendix I it can be found that these variables represent the additive section and paper machine section. A more detailed analysis showed that the deviations were mainly caused by levels of storage tanks, additive flows and vacuum boxes valve positions. As it is likely that these values will change over time, it can be assumed that the data in this dataset represents normal process operation. For this the sum of the cluster contributions was also determined. This yielded the results as shown in table AII.2.

Table AII.2 Cluster contributions for dataset 2.

Cluster	Sum of contribution
1	0,031
2	0,165
3	-0,482
4	0,242

From table AII.2 it can be observed that the results obtained are not completely in accordance with the results shown in figure AII.1. The sum of contribution for the second and forth cluster deviate from the placement in the score scatter plot. The cause of these deviations is not known and requires further investigation.

AII.2 Analysis dataset 3

A visual inspection of dataset 3 showed that the variables x95 and x108 were offline during the run period. Furthermore it was found that the variables x59, x62, x63, x71, x96, x117, x120, x121 and x129 had a constant value during the production run. Just as in dataset 1, the variables x95 and x 108 were removed to prevent difficulties in the development of a predictive model. The performed principal component analysis yielded a model with 12 principal components that explained 70.3% of the variation present in the dataset. The score scatter plot of the first two components is shown in figure AII.3.

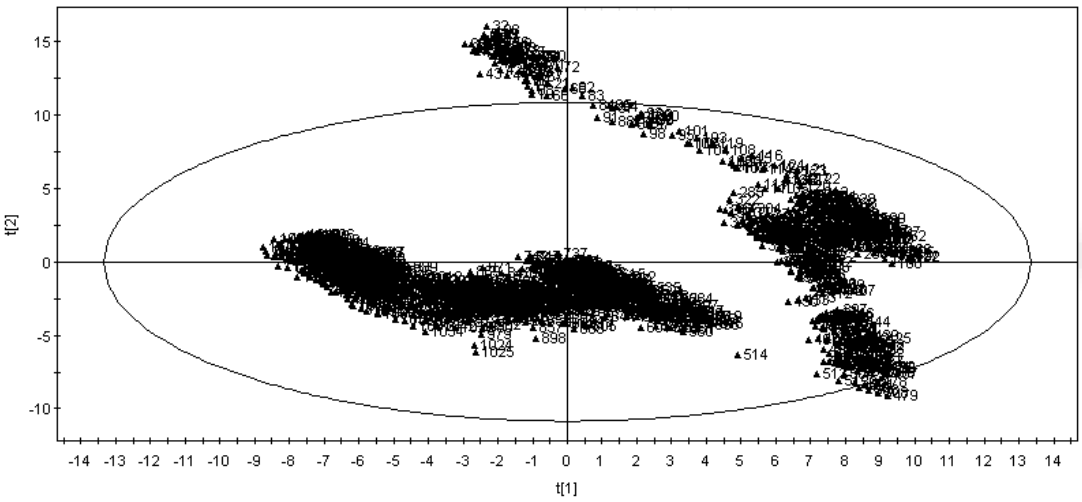

Figure AII.5.3 Score scatter plot of the first two principal components of the PCA model of dataset 3.

As can be observed from figure AII.3 most data points fall within the 95% confidence limit. A more detailed examination shows that the number of outliers is higher than the acceptable five percent of outliers. The latter is however not considered as being problematic, because of the relatively small excess. For this dataset it can also be observed that the data is arranged in cluster of consecutive data points. The exact cluster arrangement is shown in table AII.3.

Table AII.3 Distribution of the data points in the cluster in dataset 3.

Cluster	Data points	Period
1	1 – 121	20-02-2010 22:00 – 21-02-2010 00:00
2	122 – 414	21-02-2010 00:01 – 21-02-2010 04:53
3	415 – 513	21-02-2010 04:54 – 21-02-2010 06:32
4	514 – 820	21-02-2010 06:33 – 21-02-2010 11:39
5	821 – 1070	21-02-2010 11:40 – 21-02-2010 15:49
6	1071 – 1441	21-02-2010 15:50 – 21-02-2010 22:00

In order to find the cause of the clustering of the data points, the contribution of the variables in each cluster has been determined and is shown in figure AII.4.

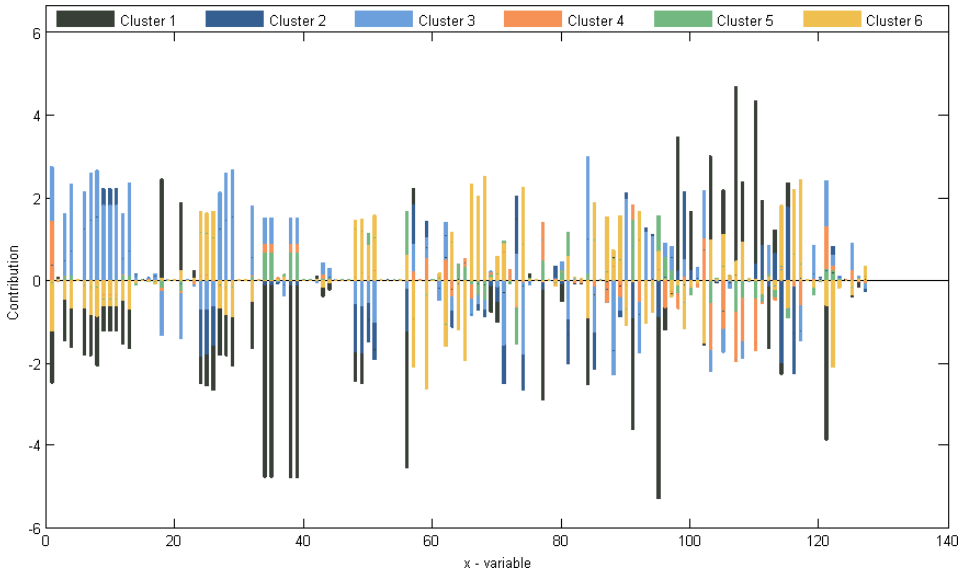

Figure AII.5.4 Variable contribution in each cluster in dataset 3.

From figure AII.4 it can be observed that most of the deviation in the contribution of the variables occurs between the clusters 1, 3 and 6. The major deviations are again found in the refiner section (variables x1 to x30), the additive section (variables x60 to x95) and the paper machine section (variables x100 to x121). As expected the deviations can be attributed to flows, temperatures and valve positions of the vacuum boxes. As it is expected that these variables show a variation over time, it can be concluded that the data in this dataset can be considered as normal process operation data. The results yielded for the sum of the cluster contributions are shown in table AII.4.

Table AII.4 Cluster contributions for dataset 3.

Cluster	Sum of contribution
1	-0,935
2	4,610
3	5,055
4	3,077
5	-6,239
6	-3,027

Comparing the result from figure AII.3 to the results from table AII.4 indicates that these results are in agreement. Thus it can be most likely assumed that the clusters in the data are formed by variations in the average values of the variables over time.

AII.3 Analysis dataset 4

For dataset 4 it was found that the variables x59, x63, x80, x96, x120, x121 and x129 had a constant value during the production run. The principal component analysis on this dataset yielded a model with 13 principal components that described 72.5% of the variation present in the dataset. The score scatter plot of the first two components is shown in figure AII.5.

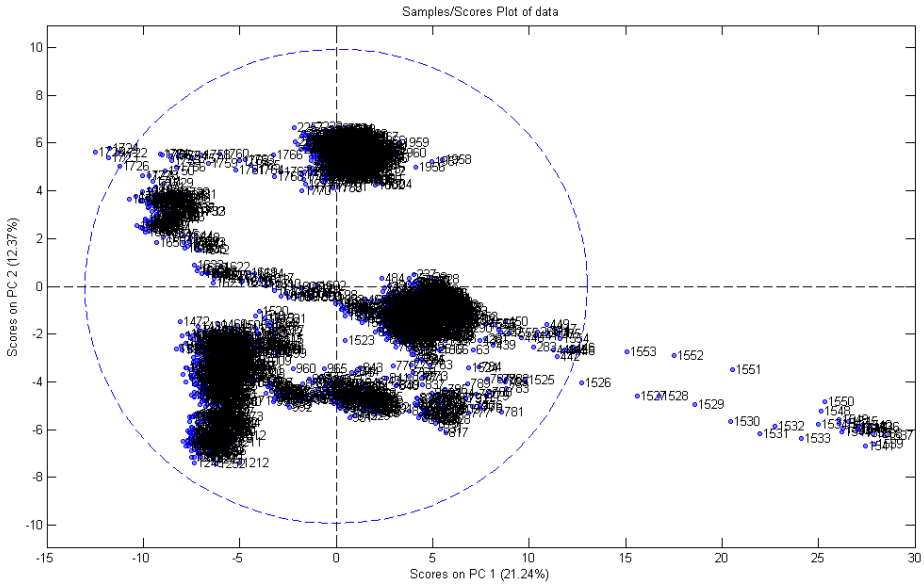

Figure AII.5.5 Score scatter plot of the first two principal components of the initial PCA model of dataset 4.

As can be observed from figure AII.5 there is a sequence of outliers present in the dataset. A more detailed examination of the contribution of the variables in this sequence of outliers is shown in figure AII.6.

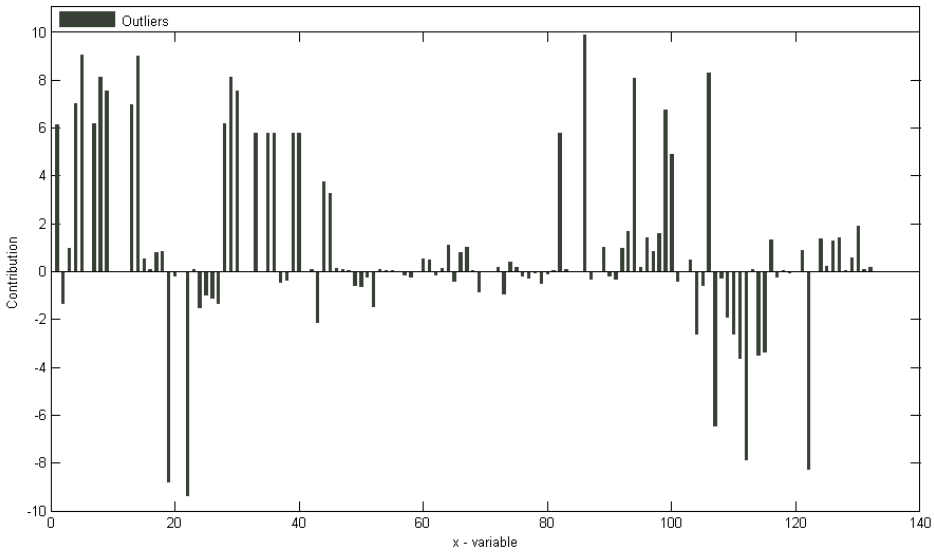

Figure AII.5.6 Variable contribution plot of the outlier sequence.

From figure AII.6 it can be observed that this sequence of outliers is mostly influenced by variables from the refiner section, as well as the additive section and vacuum section. When comparing figure AII.6 to figure 4.2, it can be concluded that the figures show strong similarities. Therefore it can be assumed that the sequence of outliers in this dataset is also caused by a sheet breakage. As this sequence of outliers will distort the predictive capacity of a predictive model, the sequence was removed from the dataset. Due to this change in dataset, a new principal component analysis had to be performed.

The new principal component analysis yielded a model with 13 principal components that explained 71.7% of the variation present in the modified dataset. The score scatter plot of the newly obtained model is shown in figure AII.7.

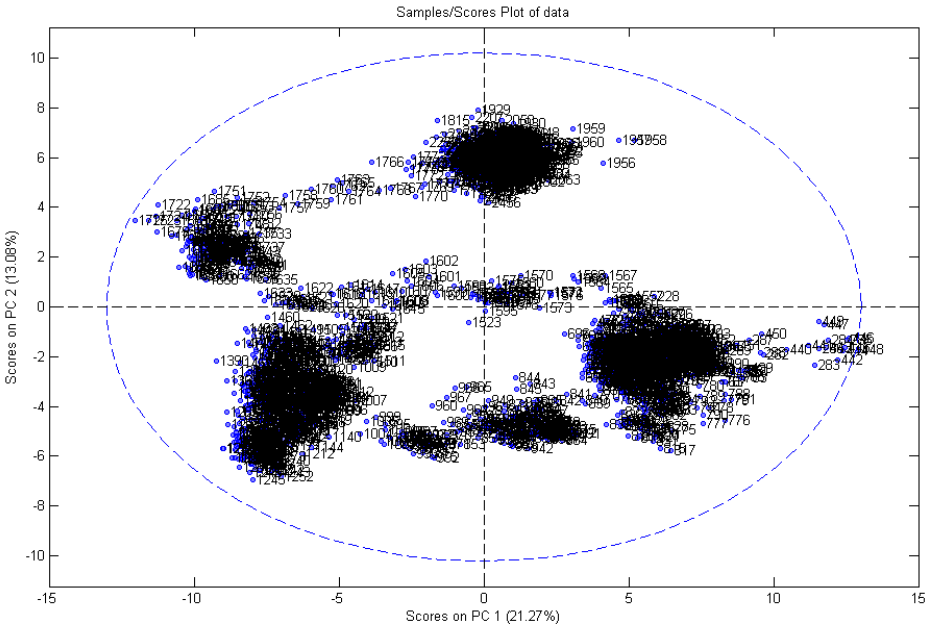

Figure AII.5.7 Score scatter plot of the first two principal components of the second PCA model of dataset 4.

From figure AII.7 it can be observed that all data points now fall within the 95% confidence limit. However it can also be observed that the data is again arranged in clusters through the confidence ellipse. The clusters present in this dataset are shown in table AII.5.

Table AII.5 Distribution of the data points in the cluster in dataset 4.

Cluster	Data points	Period
1	1 – 838	17-04-2010 14:00 – 18-04-2010 03:57
2	839 – 1005	18-04-2010 03:58 – 18-04-2010 06:44
3	1006 – 1481	18-04-2010 06:45 – 18-04-2010 14:40
4	1482 – 1522	18-04-2010 14:41 – 18-04-2010 15:21
5	1557 – 1633	18-04-2010 15:56 – 18-04-2010 17:12
6	1634 – 1768	18-04-2010 17:13 – 18-04-2010 19:27
7	1769 – 2461	18-04-2010 19:28 – 19-04-2010 07:00

Also for this dataset a plot has been made of the contribution of the variables in each cluster. The thus obtained figure is shown in figure AII.8.

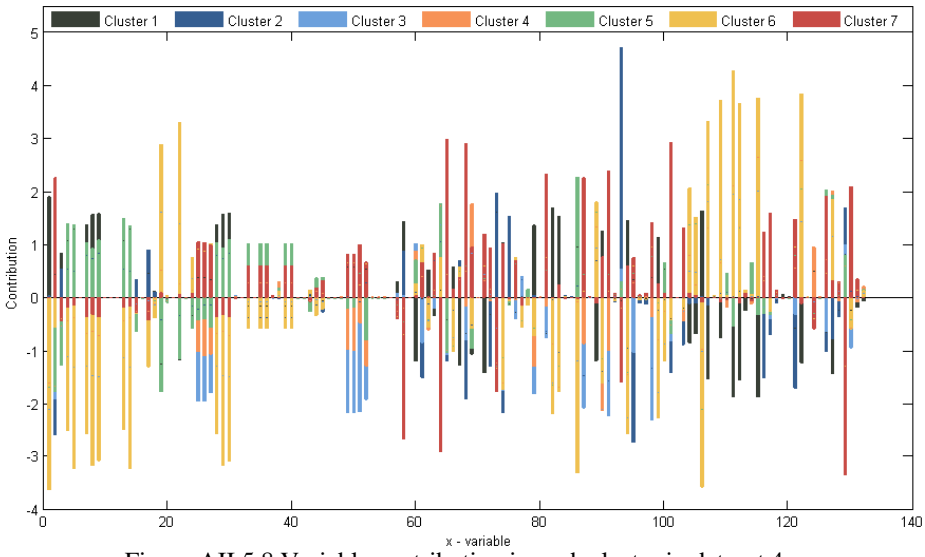

Figure AII.5.8 Variable contribution in each cluster in dataset 4.

It can be observed from figure AII.8 that most of the deviation in the contribution of the variables can be found in the refiner section (variables x1 to x 30), the additive section (variables x60 to x95) and the paper machine section (variables x100 to x121).
Just like the previous analyses, most of the deviations in contribution can be found in variables representing a flow, temperature, level of a storage tank or the valve position of the vacuum boxes. As it is likely that values of these types of variables change over time, it can be concluded that the data in this data represent normal process operation.
The results of the sum of the cluster contribution for this dataset are shown in table AII.6.

Table AII.6 Distribution of the data points in the cluster in dataset 4.

Cluster	Sum of contribution
1	1,949
2	-0,421
3	-0,960
4	0,456
5	-4,033
6	-1,608
7	1,949

The result of the contribution of the clusters is mostly in accordance with the result shown in figure AII.7. The only deviant value is obtained for the fifth cluster. The latter can possibly be assigned as an effect from the sheet breakage prior to the period of the fifth cluster. Due to the sheet breakage, set-points for some of the variables might still be different in the fifth cluster in order to prevent another sheet breakage.

AII.4 Analysis dataset 5

The inspection of dataset 5 showed that the variables x56, x96, x117, x120, x121 and x129 had a constant value during the production run. The principal component analysis on the dataset yielded a model with 12 principal components that explained 73.1% of the variation present in the dataset. The score scatter plot of the first two principal components of the model obtained is shown in figure AII.9.

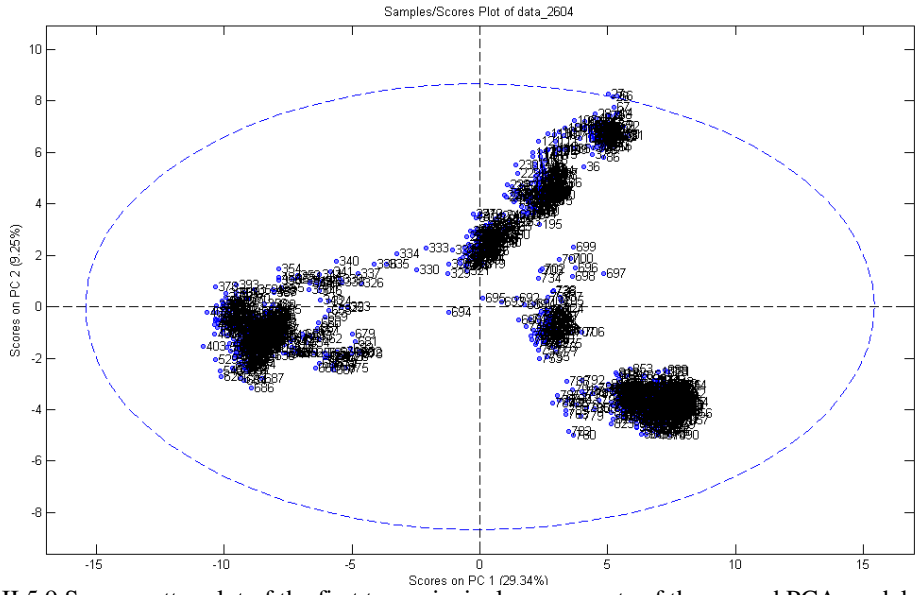

Figure AII.5.9 Score scatter plot of the first two principal components of the second PCA model of dataset 5.

From figure AII.9 it can be observed that almost all data points fall within the 95% confidence limit. A more detailed examination showed that the number of outliers present falls within the limit of 5% of acceptable outliers. Furthermore it can be observed from figure AII.9 that the data in this dataset is also arranged in clusters of consecutive data. The structure of each cluster in this dataset is shown in table AII.7.

Table AII.7 Distribution of the data points in the cluster in dataset 5.

Cluster	Data points	Period
1	1 – 108	26-04-2010 02:00 – 26-04-2010 03:47
2	109 – 233	26-04-2010 03:48 – 26-04-2010 05:52
3	234 – 327	26-04-2010 05:53 – 26-04-2010 07:26
4	328 – 430	26-04-2010 07:27 – 26-04-2010 09:09
5	431 – 640	26-04-2010 09:10 – 26-04-2010 12:39
6	641 – 687	26-04-2010 12:40 – 26-04-2010 13:26
7	688 – 777	26-04-2010 13:27 – 26-04-2010 14:56
8	778 – 1099	26-04-2010 14:57 – 26-04-2010 20:18

Figure AII.10 shows the result of the analysis of the contribution of the variables in each cluster.

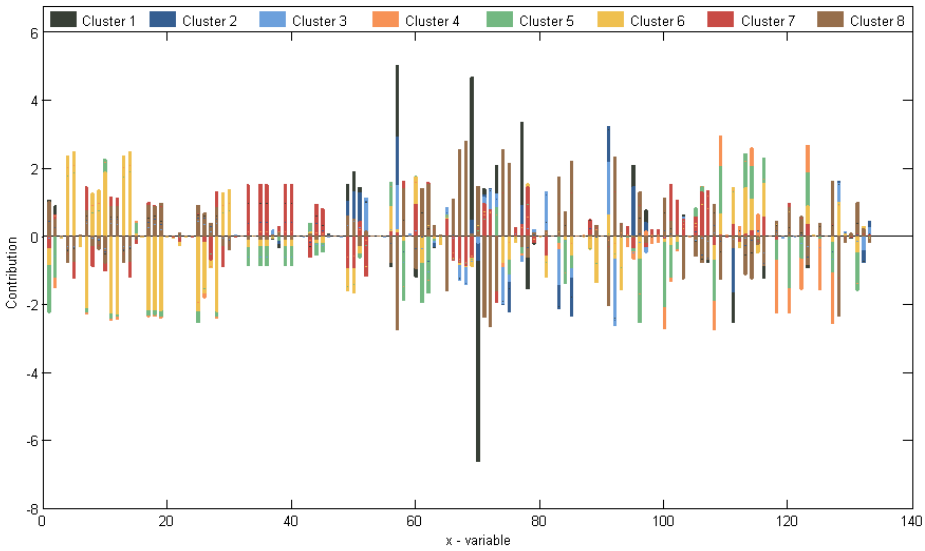

Figure AII.5.10 Variable contribution in each cluster in dataset 5.

It can be observed from figure AII.10 that for this dataset most of the deviation in the contribution of the variables can be found in the refiner section (variables x1 to x30), as well as the additive section (variables x60 to x 95) and the valve position of the vacuum boxes (variables x110 to x121). A more detailed examination yielded that most deviating variables represent flows or temperatures, it can therefore be concluded that the data in this dataset represents normal process operation.

Table AII.8 Distribution of the data points in the cluster in dataset 5.

Cluster	Sum of contribution
1	4,008
2	1,428
3	0,434
4	-5,019
5	-6,297
6	-3,718
7	5,174
8	3,249

From table AII.8 it can be observed that the sum of contribution is in accordance with the cluster arrangement as shown in figure AII.9. It can therefore be concluded that the different clusters are a result of changing averages in the dataset.

AII.5 Analysis dataset 6

The visual inspection of dataset 6 yielded that the variables x96, x117, x120, x121 and x129 had a constant value during the production run. As this will not distort the development of a predictive model, the variables were left in the same state. The principal component analysis for this dataset yielded a model with 16 principal components which explained 72.1% of the variance present in the dataset. The score scatter plot for the first two principal components of this model is shown in figure AII.11.

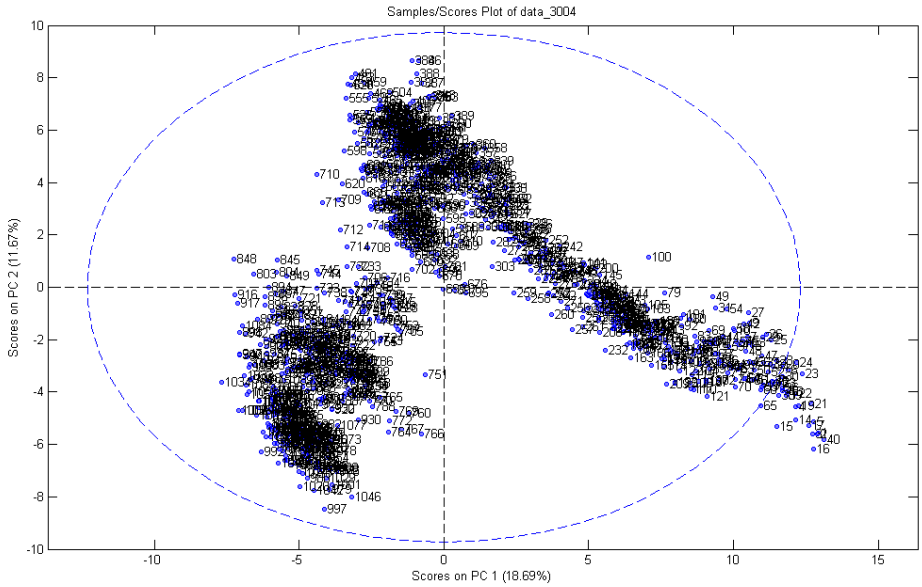

Figure AII.5.11 Score scatter plot of the first two principal components of the second PCA model of dataset 6.

As can be observed from figure AII.11 most of the data points fall within the confidence limit of 95%. The number of outliers present in the dataset is in accordance with the 5% of acceptable outliers, therefore no data points have to be removed.
From figure AII.11 it can furthermore be observed that in this dataset the data points are also arranged in clusters of consecutive data. The actual arrangement of the data points in the clusters is shown in table AII.9.

Table AII.9 Distribution of the data points in the cluster in dataset 6.

Cluster	Data points	Period
1	1 – 559	30-04-2010 14:00 – 30-04-2010 23:18
2	560 – 711	30-04-2010 23:19 – 01-05-2010 01:50
3	712 – 756	01-05-2010 01:51 – 30-04-2010 02:35
4	757 – 943	01-05-2010 02:36 – 01-05-2010 05:42
5	944 – 1099	01-05-2010 05:43 – 01-05-2010 08:18

For this dataset the contribution of the variables in each cluster was also determined. The latter yielded the figure as shown in figure AII.12.

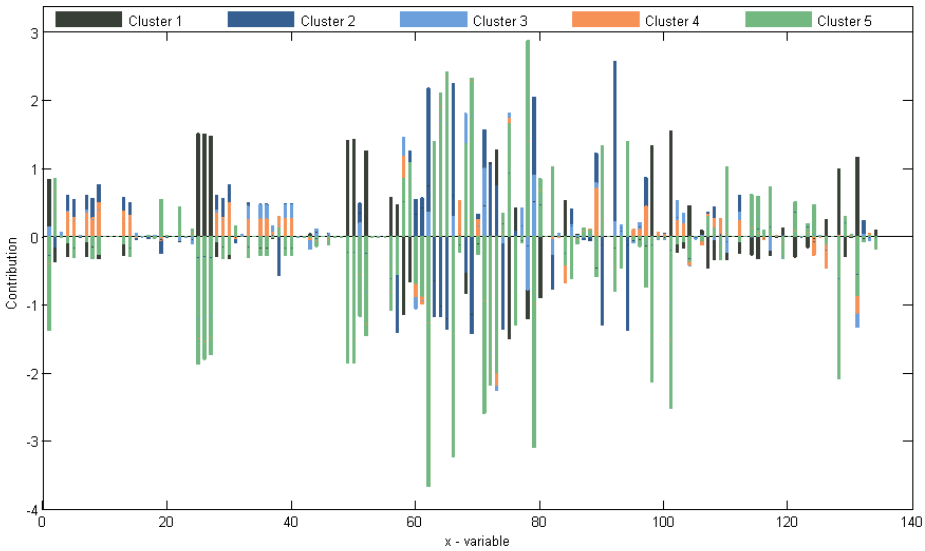

Figure AII.5.12 Variable contribution in each cluster in dataset 6.

As can be observed from figure AII.12 most of the deviation in contribution of the variables can be found in cluster five. As can be expected from the previous analyses the deviation is primarily found in the additive section and especially in variables that represent a flow. As the value of these variables is likely to change over time, it can be concluded that the data in this dataset represent normal process operation.

Table AII.10 Distribution of the data points in the cluster in dataset 6.

Cluster	Sum of contribution
1	-1,559
2	-0,147
3	3,426
4	2,215
5	2,086

As can be observed from table AII.10, the sum of contribution does not match with the cluster arrangement in figure AII.11. However a more detailed examination shows that the sum of contribution is actually mirrored with respect to the cluster arrangement in figure AII.11. Therefore it can assumed that the sum of contributions represents the clustering in figure AII.11 only the values have to have the opposite sign, which can be translated in a 180º counter-clockwise of figure AII.11.

AII.6 Analysis dataset 7

The inspection of dataset 7 showed that the variable x95 was offline during the production run, whereas the variables x59, x96, x106, x117, x120, x121 and x129 had a constant value. As the variable x95 will distort the development of a predictive model, the variable was removed from the dataset. The subsequent performed principal component analysis yielded a model with 13 principal component analysis which explained 73.9% of the variation present in the dataset. The score scatter plot for the first two principal components of the obtained model is shown in figure AII.13.

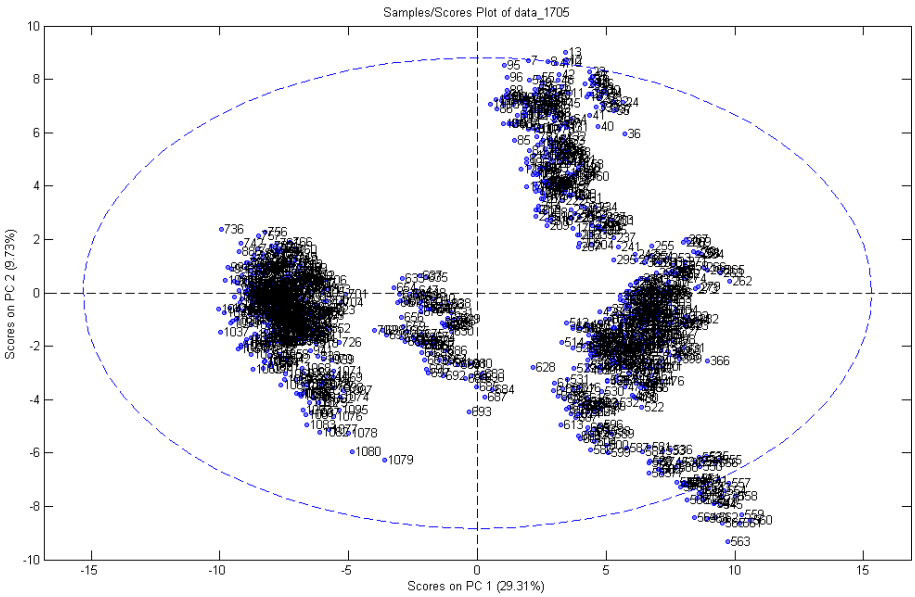

Figure AII.5.13 Score scatter plot of the first two principal components of the second PCA model of dataset 7.

From figure AII.13 it can be observed that most of the data points fall within the 95% confidence limit. A detailed examination yielded that the number of outliers present in the dataset is in accordance with the 5% limit of acceptable outliers. As expected, bearing in mind the results of the previous analyses, the data points in this dataset are arranged in clusters of consecutive data points. The actual arrangement of the data points for each cluster are shown in table AII.11.

Table AII.11 Distribution of the data points in the cluster in dataset 7.

Cluster	Data points	Period
1	1 – 241	17-05-2010 14:00 – 17-05-2010 18:00
2	242 – 529	17-05-2010 18:01 – 17-05-2010 22:48
3	530 – 628	17-05-2010 22:49 – 18-05-2010 00:27
4	629 – 699	18-05-2010 00:28 – 18-05-2010 01:38
5	700 – 1058	18-05-2010 01:39 – 18-05-2010 07:37
6	1059 – 1099	18-05-2010 07:38 – 18-05-2010 08:18

The result of the variable contribution in each cluster for this dataset is shown in figure AII.14.

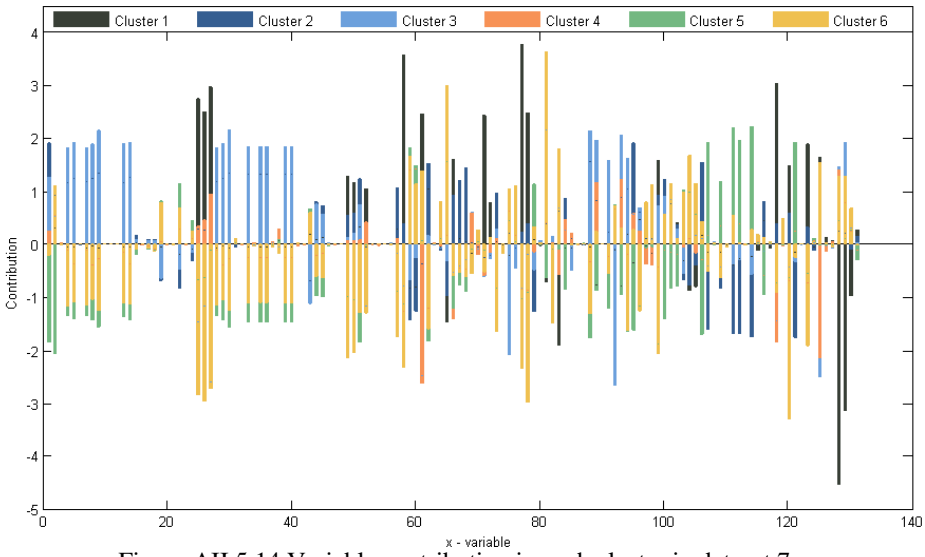

Figure AII.5.14 Variable contribution in each cluster in dataset 7.

As can be observed from figure AII.14 most deviations in variable contribution can be found between the clusters 1, 3, 5 and 6. As expected from the previous analyses, most deviations occur in the refiner section, the additive section and the paper machine section. A detailed analysis of the deviations, showed that these are caused by variables that represent a temperature, flow, or valve position of a vacuum box. As these variables are likely to change, it can be concluded that the data in this dataset represent normal operating conditions.

Table AII.12 Distribution of the data points in the cluster in dataset 6.

Cluster	Sum of contribution
1	4,008
2	1,428
3	0,434
4	-5,019
5	-6,297
6	-3,718

The sum of contribution of the clusters, as shown in table AII.12, is in accordance with the arrangement if the clusters in figure AII.13. It can therefore be concluded that the different clusters are the result of changing average values in the dataset.

Appendix III: Cross-correlation results grade specific PLS model

	Variable name	Time delay from cross-correlation analysis	Time delay from operating experience
y	Paper quality		
x5	Flow refiners	5	20
x8	Refiner 2 loading	5	20
x9	Refiner 3 loading	7	20
x14	Flow refiner	6	20
x29	Refiner 2 loading	5	20
x30	Refiner 3 loading	7	20
x33	Flow refiners	6	20
x35	Refiner 4 loading	4	20
x36	Refiner 5 loading	4	20
x39	Refiner 4 loading	4	20
x40	Refiner 5 loading	4	20
x57	Consistency tank 711	9	30
x58	Conductivity tank 711	0	…
x61	Consistency tank 712	0	30
x66	Conductivity tank 713	0	…
x67	Redox tank 713	0	…
x68	Flow tank 795	1	30
x69	Consistency tank 795	0	30
x70	Conductivity tank 795	5	…
x72	Flow tank 796	0	30
x73	Consistency tank 796	0	30
x74	Conductivity tank 796	3	…
x75	Redox tank 796	2	…
x76	Flow tank 723	1	20
x79	Redox tank 724	0	25
x82	Flow proptower	6	20
x85	Valve position tank 726	7	…
x86	Flow fixation aid	1	20
x89	Flow binder	3	20
x90	Flow retention aid	0	2
x92	Flow binder	0	2
x93	Flow flocculation aid	0	2
x94	Flow retention aid	4	2
x97	Consistency suspension	3	x
x98	Flow suspension	2	2
x99	Flow water	2	0
x100	Consistency	2	x
x101	Consistency	1	x
x102	Temperature tank 721	0	0
x103	Headbox consistency	1	x
x104	Headbox consistency	0	x
x106	Headbox conductivity	0	0

x107	Headbox redox	0	0
x109	Topformer consistency	0	x
x110	Autoslice water flow	3	0
x111	Vacuum suction roll	2	0
x113	Valve position vacuumbox 5	1	0
x115	Valve position vacuumbox 6	0	0
x116	Process value vacuumbox 7	2	0
x125	Valve position vacuumbox 2	2	0
x127	Valve position vacuumbox 3	2	0
x128	Process value vacuumbox 1	0	0
x131	Valve position vacuumbox 4	0	0
x133	Valve position vacuumbox 9	0	0
x134	Press dewatering	1	0
x137	Press dewatering	4	0

Appendix IV: Cross-correlation results multi-grade PLS model

	Variable name	Time delay from cross-correlation analysis
y	Paper quality	
x4	Flow refiner	5
x5	Flow refiner	4
x7	Refiner 1 loading	5
x8	Refiner 2 loading	4
x9	Refiner 3 loading	4
x14	Flow refiner	5
x28	Refiner 1 loading	5
x29	Refiner 2 loading	4
x30	Refiner 3 loading	4
x33	Flow refiner	5
x35	Refiner 4 loading	4
x36	Refiner 5 loading	4
x39	Refiner 4 loading	4
x40	Refiner 5 loading	4
x58	Conductivity tank 711	0
x61	Consistency tank 712	0
x63	Redox tank 712	0
x66	Conductivity tank 713	0
x68	Flow tank 795	1
x69	Consistency tank 795	1
x70	Conductivity tank 795	4
x73	Consistency tank 796	1
x74	Conductivity tank 796	3
x75	Conductivity tank 724	3
x79	Redox tank 724	0
x82	Flow proptower	2
x85	Valve position tank 726	4
x89	Flow cationic starch	2
x90	Flow retention aid	3
x92	Flow chalk	2
x94	Flow percol	4
x97	Consistency suspension	2
x98	Flow suspension	2
x99	Flow water	1
x102	Water temperature	0
x103	Headbox consistency	3
x108	Topformer consistency	0
x109	Topformer consistency	0
x110	Autoslice dewatering	5
x111	Vacuum suction roll	2
x113	Valve position vacuumbox 5	1
x115	Valve position vacuumbox 6	0
x116	Process value vacuumbox 7	3

x119	Valve position vacuumbox 8	0
x125	Valve position vacuumbox 2	2
x127	Valve position vacuumbox 3	3
x128	Process value vacuumbox 1	1
x131	Valve position vacuumbox 4	0
x133	Valve position vacuumbox 9	3

Appendix V: Variables used for the development of the cause effect PLS model

	Variable name
y	Paper quality
x1	Gram weight
x2	Grade
x3	Consistency flow refiners
x4	Flow refiner
x5	Flow refiner
x6	Level tank 703
x7	Refiner 1 loading
x8	Refiner 2 loading
x9	Refiner 3 loading
x10	Refiner 1 energy
x11	Refiner 2 energy
x12	Refiner 3 energy
x13	Flow refiner
x14	Flow refiner
x15	Refiner 1 loading
x16	Refiner 2 loading
x17	Refiner 3 loading
x18	Level tank 702
x19	Consistency flow refiners
x20	Flow refiners
x21	Level tank 704
x22	Refiner 4 loading
x23	Refiner 5 loading
x24	Refiner 4 loading
x25	Refiner 5 loading
x26	Pressure after refiner 5
x27	Pressure after refiner 4
x28	Pressure after refiners
x29	Flow tank 711
x30	Consistency tank 711
x31	Conductivity measurement tank 711
x32	Redox measurement tank 711
x33	Flow tank 712
x34	Consistency tank 712
x35	Conductivity measurement tank 712
x36	Redox measurement tank 712
x37	Flow tank 713
x38	Consistency tank 713
x39	Conductivity measurement tank 713
x40	Redox measurement tank 713
x41	Flow tank 795
x42	Consistency tank 795
x43	Conductivity measurement tank 795

x44	Redox measurement tank 795
x45	Flow tank 796
x46	Consistency tank 796
x47	Conductivity measurement tank 796
x48	Redox measurement tank 796
x49	Flow tank 723
x50	Consistency tank 723
x51	Conductivity measurement tank 724
x52	Redox measurement tank 724
x53	Charge measurement tank 724
x54	Flow material recovery
x55	Flow recovered material
x56	Consistency flow recovered material
x57	Flow recovered material (storage)
x58	Valve position tank 726
x59	Flow fixation additive
x60	Flow retention additive
x61	Flow retention additive
x62	Flow binder additive
x63	Flow dewatering additive
x64	Flow dilution water
x65	Flow binder additive
x66	Flow flocculation additive
x67	Flow retention additive
x68	Consistency natural dewatering
x69	Consistency natural dewatering (additives)
x70	Incoming consistency suspension on wire
x71	Flow suspension to wire
x72	Flow excessive water from wire
x73	Consistency natural dewatering (additives)
x74	Consistency natural dewatering
x75	Temperature tank 721
x76	Headbox consistency
x77	Headbox consistency (additives)
x78	Speedratio
x79	Conductivity measurement headbox
x80	Redox measurement headbox
x81	Topformer consistency
x82	Topformer consistency (additives)
x83	Dewatering vacuumsection
x84	Vacuum suction roll
x85	Process value vacuumbox 7
x86	Process value vacuumbox 8
x87	Process value vacuumbox 9
x88	Process value vacuumbox 10
x89	Process value vacuumbox 1
x90	Process value vacuumbox 2
x91	Process value vacuumbox 3

x92	Process value vacuumbox 4
x93	Process value vacuumbox 5
x94	Process value vacuumbox 6
x95	Process value vacuumbox 11
x96	Press dewatering
x97	Press 4 dewatering
x98	Machine speed
x99	Press 1 dewatering
x100	Press 2 dewatering
x101	Press 3 dewatering

Appendix VI: Cross-correlation results for the cause effect PLS model

	Variable name	Time delays from cross-correlation analysis
y	Paper quality	
x4	Flow refiner	4
x5	Flow refiner	3
x7	Refiner 1 loading	5
x8	Refiner 2 loading	3
x9	Refiner 3 loading	4
x13	Flow refiner	3
x14	Flow refiner	4
x15	Refiner 1 loading	5
x16	Refiner 2 loading	3
x17	Refiner 3 loading	4
x20	Flow refiners	5
x22	Refiner 4 loading	4
x23	Refiner 5 loading	4
x24	Refiner 4 loading	4
x25	Refiner 5 loading	4
x30	Consistency tank 711	9
x31	Conductivity measurement tank 711	0
x34	Consistency tank 712	0
x36	Redox measurement tank 712	0
x39	Conductivity measurement tank 713	0
x41	Flow tank 795	1
x42	Consistency tank 795	1
x43	Conductivity measurement tank 795	4
x46	Consistency tank 796	0
x47	Conductivity measurement tank 796	3
x48	Redox measurement tank 796	1
x49	Flow tank 723	3
x51	Conductivity measurement tank 712	3
x52	Redox measurement tank 712	0
x55	Flow recovered material	4
x62	Flow binder additive	2
x63	Flow dewatering additive	3
x65	Flow binder additive	1
x66	Flow flocculation additive	1
x67	Flow retention additive	3
x70	Incoming consistency suspension on wire	2
x71	Flow suspension to wire	2
x72	Flow excessive water from wire	1
x75	Temperature tank 721	0

x76	Headbox consistency	1
x77	Headbox consistency (additives)	0
x81	Topformer consistency	1
x82	Topformer consistency (additives)	2
x83	Dewatering vacuumsection	3
x84	Vacuum suction roll	2
x87	Process value vacuumbox 9	3
x93	Process value vacuumbox 5	1
x96	Press dewatering	1
x99	Press 1 dewatering	3