Improvement of fiberboard manufacture through statistical process analytics

Master Thesis

Submitted by

Gerald Maschl, Bakk. techn.

Carried out at:

Institute of Wood Science and Technology
Department of Material Sciences and Process Engineering
University of Natural Resources and Applied Life Sciences
BOKU – Vienna/Austria

Supervisor:

Univ.Prof. Dipl.-Ing. Dr.nat.techn. Rupert Wimmer
Wood Technology & Wood-based Composites Unit
Faculty of Forest Sciences and Forest Ecology
Georg-August-Universität Göttingen
Büsgenweg 4, D-37077 Göttingen

Co-Supervision:

Dipl.-Ing. Dr. Bernhard Spangl
Department of Landscape, Spatial and Infrastructure Sciences
Institute of Applied Statistics and Computing
University of Natural Resources and Applied Life Sciences
Peter Jordan Str. 82, A-1190 Wien

Dipl.-Ing. Martin Weigl
Wood K plus / Kompetenzzentrum Holz GmbH
St.-Peter-Str. 25, A-4021 Linz

Vienna, December 2009

ABSTRACT

High-Density Fiberboard (HDF) has been produced globally in vast quantities over the last few years. Rejects in production are common due to high variation in board properties. This work statistically analyzes an HDF plant, with the aim of finding the key sources of variation. Process data on all steps of the process is used, but also data from the raw material (wood types, acidic groups measured by ion chromatography, hemicelluloses/extractives determination by methanolysis, pH, buffer capacities), and subjective variables (process evaluation, formaldehyde perceptibility). As response variables, the board properties Internal Bond Strength (IB), Surface Soundness (SS), Modulus of Rupture (MOR), Modulus of Elasticity (MOR) and Thickness Swelling 24 h (TS), and furthermore the press factor and resin fraction are analyzed. In total, the dataset consisted of 251 observations and 245 variables. Lag times were taken into account in data collection. Partial least squares regression (PLSR) was used to create 45 models. The main key sources of variation were determined 1) by the frequency in which variables occur in models and 2) by weighting the regression coefficients according to the technological relevance of the board properties. The models show that board properties were influenced by the raw material variables at an average of 21%, and the remainder by process variables. Furthermore, the appropriateness of (multivariate) control charting as a tool of Statistical Process Control (SPC) was shown on the data.

ZUSAMMENFASSUNG

Hochdichte Faserplatten (HDF) wurden in den letzten Jahren in großen Mengen hergestellt. Wegen hoher Streuung der Platteneigenschaften kommt es immer wieder zu Fehlproduktionen. Mit dem Ziel der Detektierung von Quellen der Streuung wird in dieser Arbeit eine statistische Analyse einer HDF-Produktion durchgeführt. Prozessdaten von allen Abschnitten des Prozesses wurden verwendet, und weiters auch Daten zum Rohmaterial (Holzarten; Säuregruppen, ermittelt mittels Ionenchromatographie; Hemizellulosen und Extraktstoffe, ermittelt durch Methanolyse; pH-Wert und Pufferkapazität), und subjektive Variablen (Prozessbeurteilung, Wahrnehmbarkeit von Formaldehyd). Als Zielvariablen wurden die Querzugfestigkeit, Abhebefestigkeit, Biegefestigkeit, E-Modul und Dickenquellung nach 24 h ausgewertet. Weitere Zielvariablen sind der Presszeitfaktor und der Festharzanteil. In Summe enthält der Datensatz 251 Beobachtungen und 245 Variablen. Die Prozessdurchlaufzeiten wurden bei der Datenerhebung berücksichtigt. Partial Least Squares-Regression (PLSR) wurde zur Modellierung von 45 Modellen verwendet. Die Hauptquellen der Streuung wurden ermittelt mittels 1) der Häufigkeit, in der die Variablen in den Modellen vorkommen und 2) einer Gewichtung der Regressionskoeffizienten, abhängig von der technologischen Bedeutung der Platteneigenschaft. Die Modelle zeigten, die Variablen des Rohmaterials zu durchschnittlich 21% in die Streuung der Platteneigenschaften einfließt, der Rest sind Prozessvariablen. Weiters wurde an diesem Datensatz die Anwendbarkeit von (multivariaten) Qualitätsregelkarten als Werkzeug der Statistischen Prozesskontrolle (SPC) gezeigt.

KEYWORDS

Process analytics, fiberboard, HDF, MDF, statistical modeling, PLS, Statistical Process Control, SPC

☼ Heidille ☼

Etäisyydellä ei ole merkitystä.

Olla lähellä on sydänasia.

Olet aurinko elämässäni. Minä rakastan sinua.

TABLE OF CONTENTS

1 Introduction

1.1 HDF manufacturing

1.1.1 Medium- and High-density fiberboard

Fiberboards are important engineered wood products. Medium-density fiberboards (MDF) are basically produced by the addition of usually synthetic resin to lingo-cellulosic fibers, followed by the application of temperature and pressure (CEN, 1999).

A schematic process flow diagram about the main steps of fiberboard production is shown in Figure 1. The logs are debarked and chipped, and the chips are broken down in a defibrator. Resin is applied to the fibers, which are dried afterwards. Fibers are discharged on a form band and pressed. The panels develop its final appearance during the conditioning and finishing stages (Deppe and Ernst, 1996).

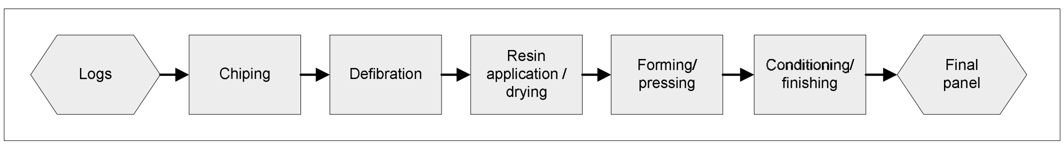


Figure 1. Main process steps of fiberboard production.

Development of the first fiberboards started around 1900. Over the years, boards with different properties and process technologies were created. Medium-density fiberboard (MDF) became a substitution product to particle boards. As it is a very homogenous product, it can be used for applications such as furniture and floorings. MDF is generally produced in a dry-process (Deppe and Ernst, 1996; Walker, 2006).

MDF is commercially available with thicknesses in a range from about 3 to 100 mm, with raw densities from 450 to 900 kg/m³. For MDF with a raw density ≥ 800 kg/m³, the name High-density fiberboard (HDF) has become common (Deppe and Ernst, 1996; CEN, 1999).

Another definition of distinction between MDF and HDF is by board thickness. MDF boards less than 12 mm in thickness are HDF with a density of 500-1,450 kg/m³ (André et al., 2008).

HDF is usually harder than MDF, because HDF fibers are produced under more severe conditions.

1.1.2 Standards in MDF/HDF manufacturing

The European Standard EN 622-5 (CEN, 2006) defines requirements of MDF/HDF according to the dry-process of EN 316 (CEN, 1999). The standard defines specification limits, which depend on the general use of the board. The limits have to be fulfilled by the 5% quantile values (for thickness swelling 95% quantile) calculated by board mean values according to EN 326-1 (CEN, 1994). Table 1 contains a listing of further standards for the production of MDF/HDF.

However, for use of MDF/HDF in certain applications such as laminate floorings etc; specific standards apply, where requirements for properties such as surface soundness and distance between joints are set for the final laminate flooring product. Therefore, for the HDF carrier boards, higher requirements apply indirectly (CEN, 2000b). Furthermore, modern click systems of laminate floorings demand an Internal Bond Strength (IB) of at least 1.2 N/mm².

Table 1. Selection of standards in testing MDF/HDF properties in dry use (Kuss, 2000).

Quality property	**Standard**	**Requirement of test**
Formaldehyde potential	EN 120 (CEN, 1992)	Obligatory
Modulus of rupture, bending strength (MOR)	EN 310 (CEN, 1993a)	Obligatory
Modulus of elasticity in bending (MOE)	EN 310 (CEN, 1993a)	Obligatory
Surface soundness (SS)	EN 311 (CEN, 2002a)	Not required, but values have to be published on demand
Thickness Swelling 24 h (TS)	EN 317 (CEN, 2002b)	Obligatory
Dimensional changes	EN 318 (CEN, 2002c)	Not required, but values have to be published on demand
Tensile strength perpendicular to the plane = Internal Bond Strength (IB)	EN 319 (CEN, 1993b)	Obligatory
Resistance to axial withdrawal of screws	EN 320 (CEN, 1993c)	Not required, but values have to be published on demand
Medium density (MD)	EN 323 (CEN, 1993d)	Obligatory
Surface absorption	EN 382-1 (CEN, 1993e)	Not required, but values have to be published on demand
Sand content	ISO 3340 (ISO, 1976)	Not required, but values have to be published on demand

According to EN 326-2 (CEN, 2000a), a production plant may use alternative tests methods for the quality properties, as long as the correlation coefficient between the result of the alternative method and the actual property tested in conformity with the standard (sample size is 32 boards per product type) is at least 0.70.

1.2 Concepts in Quality Management

1.2.1 Less variation resulting in less rejects

Due to different determining factors, boards are produced with variation in properties. If the variations are too high, it exceeds limits, and furthermore results in a waste of raw material and financial resources.

One major target in Quality Management (QM) and Six Sigma is to minimize variation in product properties, and to minimize deviations from the mean to the target. Thus, an aim is to increase precision in the production process, and to produce a smaller probability of products falling below or exceeding quality (specification) limits (Evans and Lindsay, 2002).

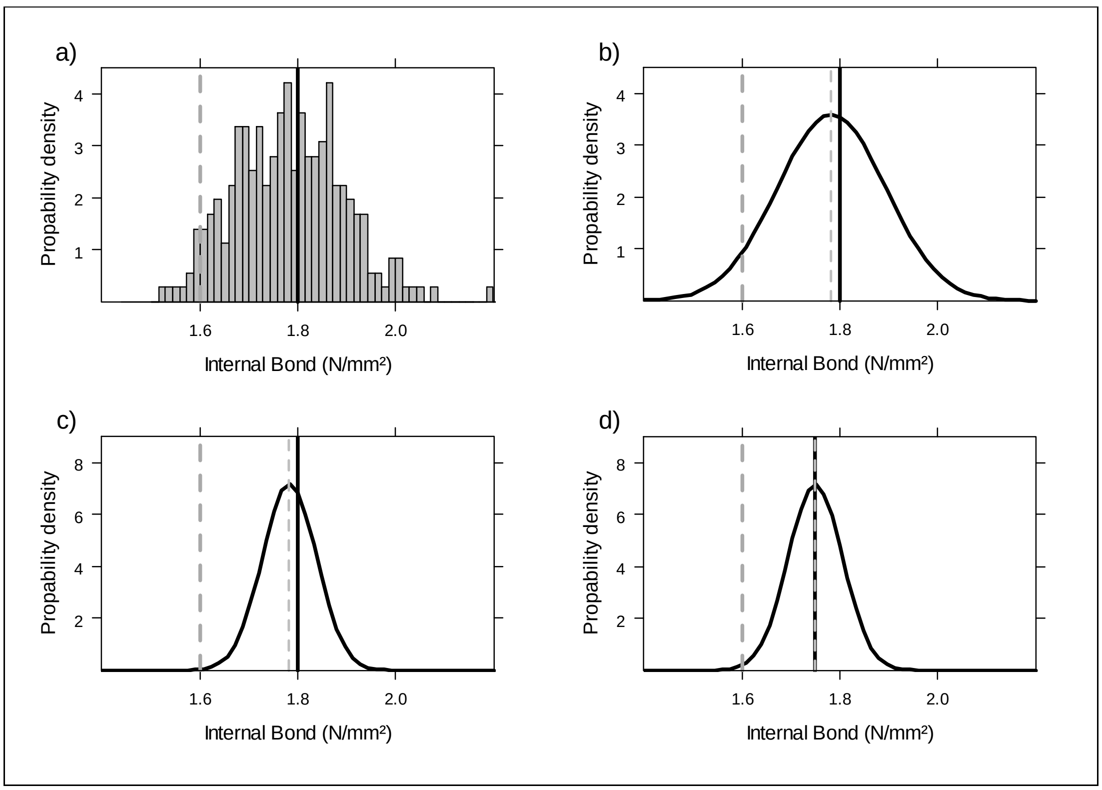


Figure 2. Reducing variability of IB results in fewer defect boards. a) Histogram of IB values of real plant data, b) Normal distribution of IB, 5% are below 1.6 N/mm², c) Transformed data with half standard deviation, d) Only less than 1% of values fall below 1.6 N/mm².

The principle of decreasing the variability is shown in Figure 2. Real factory data for the board property Internal Bond Strength (IB) was used for creating this figure. The current lower specification limit (and 5% quantile of the normal distributed board property) is 1.6 N/mm², and the target value is 1.8 N/mm².

Figure 2a shows a histogram of IB values. A normal distribution curve was fitted to Figure 2a. The normal distribution is represented in Figure 2b. About 5% of values fall below the lower specification limits. Figure 2c: The normal distribution was halved in order to simulate less variation with IB. Now, less than 1% of output falls below quality limits. Thus, the target value could be decreased to 1.75 N/mm², and still less than 1% would fall below the lower specification limit (Figure 2d).

1.2.2 Exploratory Data Analysis and Data Mining

This research is based on Exploratory Data Analysis (EDA). EDA aims to identify systematic relations between variables if there are no reasonable expectations regarding the nature of the relation. Hypotheses generation rather than hypotheses testing (as done in Confirmatory Data Analysis - CDA) is emphasized in EDA. EDA was established by Tukey (Mosteller and Tukey, 1977; Tukey, 1977). In an exploratory data analysis, many variables are taken into account and compared, and different techniques in searching systematic patterns are applied.

The idea of EDA is continued in the more recent approach of data mining. In data mining, statistical regression modeling is used extensively amongst classification, clustering and setting association rules (Witten, 2005; Han and Kamber, 2006).

Myers (1986) describes 4 objectives for statistical modeling. Usually these cannot be considered independently, but a mixture applies.

1. Prediction
2. Variable selection
3. Model specification (explanation of the system)
4. Parameter estimation

Prediction of technological properties can be done in different increments. Models should be validated internally and with external data. In a next step, long term prediction precision can be investigated. Model actualization using new data can improve the long term prediction precision.

This work does not consider long term prediction precision. However, internal model validation by cross validation is done as a part of the model development.

Variable selection in partly non-stationary technological processes includes the risk of omitting variables that could gain higher impact with time. Processes in plants are usually modified to some extent from time to time; therefore the significance of different variables may change as well. Variable selection has to consider this fact for robust long term prediction models. On the other hand, a selection of variables can improve the short term prediction precision considerably (Hasener, 2004).

In this work, variable selection has had a higher importance than long term prediction precision since the main objective was to find process and raw material variables of the plant that have significant influences on the produced HDF boards under the given production settings.

This work furthermore concentrates on an *explanation of the system*, whereas linear relations between variables can be assumed (Lobenhoffer, 1990).

Parameter estimation is strongly linked to prediction and variable selection. In this work, the focus in the model creation is mainly on the sign and relative significance of parameters, and only secondary on the true regression coefficients.

1.3 Models for MDF manufacture (literature review)

1.3.1 Fiber processing

The influence of pulping conditions of wood and fiber drying on the properties of MDF was studied by Schneider (1999). The effect of high-temperature defibration on the chemical structure of hard- and softwoods was shown by Widsten et al. (2001; 2002). These authors pointed out that the production of MDF and TMP fibers is characterized mainly by the process temperature, which is lower than with the pulping process. With the higher pulping temperature the fracture zone shifts from the secondary wall to the middle lamella, thus forming a crust of lignin. Studies on Water-soluble components from MDF fibers were studied by McDonald et al. (2006). It was shown, that the extracts are predominantly from the hemicelluloses, and not from cellulose or lignin.

Industrial applications of near-infrared (NIR) spectroscopy in combination with multivariate analysis methods (MVA) for controlling material properties of a MDF production process (e.g. moisture content, particle size, wood quality) as well as prediction of board properties are reviewed by Carlsson et al. (2002) and So et al. (2004). Reduced wavelength models of NIR data showed only little loss in predictive ability in comparison to full models, and thus demonstrating the potential of NIR spectroscopy for rapid process monitoring (Rials et al., 2002).

Changes in chemical composition, crystallinity and refining pressure during MDF processing were studied by Kelley (2005). A designed experiment was carried out, where wood chips were refined at different pressures. Resulting fiber material was analyzed using wet chemical methods, NIR spectroscopy as well as X-ray diffraction (XRD). Calibration models were developed using Partial Least Squares-Regression (PLSR, Chapter 2.3.4). Glucose and extractive contents increased, while mannose, xylose and galactose concentrations decreased with increased refining pressure.

The performance of IB affected by wood fiber acidity, size distribution and bulk density was studied by Xing et al. (2006). This study showed that IB was strongly related to the pH value of fibers. The mechanical properties also increased as buffering capacity went up. High bulk densities of fibers resulted in increased IB, MOR and MOE. Higher proportions of coarse fibers influenced the board properties negatively.

Models of laboratory produced MDF panels with special focus on the wood and fiber characteristics were developed by Li Shi et al. (2006). Various wood and fiber characteristics, i.e. wood density, fiber pH, buffering capacity and fiber coarseness, were used as predictor variables. Multiple linear regression (MLR) was used as regression method. Dummy variables were incorporated into the analysis to examine the effect of wood species. Because of the designed experiment, high coefficients of determination could be obtained, e.g. R^2=0.92 for the model with IB with 4 predictor variables only.

1.3.2 MDF/HDF process

The process control system for online quality control SPOC (Statistical Process Optimization and Control) was introduced by Bernardy and Scherff (1997). Online-quality prediction, simulation and process-/cost optimization are the main features. SPOC aims to predict Internal Bond Strength (IB) and Modulus of Rupture (MOR) using all available process parameters. Multiple Linear Regression (MLR) is used with the estimator 3SLS, which build an extension to the often used least square estimator (Amemiya, 1977; Fahrmeir et al., 2007). Bernardy and Scherff (1997) developed models using 72 observations on boards with a thickness of 16 mm. Without validation of the models with external data, coefficients of determination (R^2) of 0.91 for IB, and 0.86 for MOR were reached. Models with 19 mm thickness brought similar results. Important variables that indirectly carry information of the raw material include the *refiner power drain*, *vibration of refiner discs*, and *temperature of the chip silo*.

Models for the MDF process were developed using 45 process variables of industrial data including product properties (Greubel, 1999a; b). Nominal thicknesses of 8 and 19 mm were considered. The data set had 61 observations in total. Variable selection was done using knowledge of technological interrelationships and visual impression in time series plots. Validation of the data was not done; and furthermore no statistical indices about the quality of the model were given. Important process variables include the *resin fraction*, *refiner energy input*, *bulk density*, *grammage*, *press factor*, *moisture content band scale*, and *pressures in the press*.

A US patent on a method for controlling the production process of a cellulose fiber containing product such as MDF/HDF was submitted by Nordin et al. (2002) and assigned to Akzo Nobel N.V.. This patent includes methods for controlling production processes using mathematical models. Thus, this patent contains process control of MDF/HDF production with the use of models such as multiple linear regression (MLR) and partial least squares regression (PLSR).

The use of NIR spectroscopy for MDF, and the forest products industry in general, is summarized by So et al. (2004). NIR spectroscopy can be used to characterize properties such as density, mechanical properties, moisture contents or fiber length distribution.

Hasener (2004) developed several models for MDF using industrial data, where the focus was on the long term predicting quality of the models. Results for the Root Mean Square Error of Prediction (RMSEP) are summarized in Table 2. MLR and PLSR were used as regression methods. Models were validated with an external dataset that was obtained with a distance of several months to the calibration data set in order to evaluate long term prediction capabilities. Model results were similar for both modeling methods. The author, however, mentioned the advantages of PLSR for model building and interpretation of collinear data. Modeling is vastly simplified using PLSR, since no laborious variable selection step has to be performed as it is usually done in MLR.

Table 2. RMSEP for models on IB, validation data set was 3 month ahead (Hasener, 2004).

Modeling method	**RMSEP for IB (N/mm^2)**	**RMSEP for SS (N/mm^2)**	**RMSEP for TS (%)**
MLR	0.10 – 0.11	0.18 – 0.20	0.69 – 0.79
PLSR	0.09 – 0.11	0.19 – 0.22	0.63 – 0.73

Online UV-vis-NIR (ultraviolet-visual-near-infrared) spectroscopy was installed on full-scale industrial hardboard production by Dolezel-Horwarth et al. (2005). Spectra data was collected from fiber materials, intermediate fibermats and final hardboards by conducting a designed experiment. PCA and PLSR were performed on the acquired spectra, to investigate linkages between spectral information and the board quality parameters MOR, IB and water uptake (WU). RMSEP of the most models are summarized in Table 3.

Table 3. RMSEP of PLSR models on spectroscopic data with MOR and IB as predictor variables (Dolezel-Horwath et al., 2005).

Measured sample in spectral range	**RMSEP of models on MOR (N/mm^2)**	**RMSEP of models on IB (N/mm^2)**
Fibers 500-800 nm	5.92	0.30
Fibers 1100-2200 nm	4.92	0.26
Fibermat 200-730 nm	1.57	0.13
Fiberboards 380-1130 nm	3.17-4.30	0.15

Multivariate calibration models for the prediction of Internal Bond Strength (IB) of MDF process variables were developed by André et al. (2008). Chronological real-time process data was used considering time-lags with respect to the product quality parameters obtained by destructive testing. As calibration methods, radial basic function (RBF) neural networks, partial least squares (PLS), orthogonal-PLS (O-PLS) and supervised probabilistic principal component analysis (SPPCA) were used. In addition, genetic algorithms (GA) as a previous variable selection step were applied as well. Models were validated by using an independent data set. Results of these models are summarized in Table 4, where next to RMSEP a mean normalized RMSEP (NRMSEP, Chapter 2.3.4.4) is mentioned, that allows a direct comparison of the models. Most models benefitted from the GA-based variable selection. Best results without variable selection were obtained by O-PLS, and SPPCA with GA variable selection. Variables frequently obtained through GA-based variable selection were related to *refiner's parameters*, *fibers moisture contents*, *resin percentages* and *line speed*.

Table 4. RMSEP of full and GA variable selected models of IB (André et al., 2008).

Model calibration method	**Full model**		**GA variable selected model**	
	RMSEP (N/mm^2)	**Mean normalized RMSEP (%)**	**RMSEP (N/mm^2)**	**Mean normalized RMSEP (%)**
RBF neural networks	0.07526	6.81	0.09377	8.20
Traditional PLS	0.07288	6.80	0.06674	6.18
Orthogonal PLS (O-PLS)	0.06791	6.37	0.06695	6.20
Supervised probabilistic PCA (SPPCA)	0.07770	7.23	0.06522	5.89

Multiple linear regression (MLR) and quantile regression (QR) models for IB of MDF were developed by Young et al. (2008). The data set contained 184 variables derived from online detection sensors. Models were not validated. Results for MLR models are summarized in Table 5. Variables related to the *application rate of scavenger resin*, *amount of humidity in the face fiber layer during mat formation*, and the *amount of water added during the defibration process of wood fiber* were frequently selected as regressors for the MLR models. The use of quantile regression analysis gave a deeper insight of the behavior of significant variables with regard to certain quantiles of the response variable IB.

Table 5. RMSE and adjusted R^2 on MLR models of IB (Young et al., 2008).

Nominal board thickness	**RMSE (N/mm²)**	**adjusted R^2**
12.70 mm (0.500 inches)	0.0453	0.75
15.88 mm (0.625 inches)	0.0417	0.72
17.46 mm (0.6875 inches)	0.0430	0.81
19.05 mm (0.750 inches)	0.0531	0.75

Clapp et al. (2008) used a modified principal component analysis (PCA) to develop models for four board thicknesses to predict the IB of MDF. To improve prediction capabilities, PCA was combined with a heuristic algorithm which basically selects variables in several steps by determining correlations to IB. Time lags of the real-time process data were considered. The models were not cross validated using an independent data set. The root mean square errors (RMSE) between observed and predicted IB values and an adjusted R^2 are summarized in Table 6. Variables that occurred most frequently in the models were *core refiner total steam flow*, *face plate position of refiner* and *face plug feeder screw speed*.

Table 6. RMSE for models on IB (Clapp et al., 2008).

Nominal board thickness	**RMSE (N/mm²)**	**Mean normalized RMSE (%)**
12.70 mm (0.500 inches)	0.0883	9.3
15.88 mm (0.625 inches)	0.0951	10.1
17.46 mm (0.6875 inches)	0.1050	11.2

Prediction models for the MDF process are generally difficult to compare as they are only valid suited for specific production sites, with the models differently validated and characterized. A mean normalized RMSEP (NRMSEP) would be most desirable as a comparable characteristic, since it implies the validation of models, and also a standardization in the dimension of the response variable. The literature furthermore highlights the suitability of projection methods such as PCA, PLSR, and its extensions, in order to predict product quality parameters of MDF/HDF (Hasener, 2004; Dolezel-Horwath et al., 2005; André et al., 2008).

1.4 Research question

The main focus of this work is to determine interactions between the raw material, the process and board quality parameters.

In the course of this study, the following hypotheses are stated:

H1: Key sources of variation for HDF board quality properties (i.e. *IB*, *SS*, *MOE*, *MOR* and *TS;* besides *press factor* and *resin fraction*) can be determined using multivariate data analysis (MVA) of industrial data.

H1.1: Sources of variation can be determined through statistical models applied to board mean values of quality parameters as response variables.

H1.2: Sources of variation can be determined through statistical models applied to single values of quality parameters (which were obtained across the width of the production line after the press) as response variables.

H1.3: Sources of variation can be determined through statistical models applied to dispersion parameters of single values (which were obtained across the width of the production line after the press) as response variables.

H2: Influence of raw materials on HDF board quality properties can be elucidated through MVA, using industrial raw material and process data.

H3: Control charts on HDF production data can identify events that cause variation in the board quality properties.

2 Materials and Methods

2.1 Industrial production plant

Data is taken from an industrial HDF production. The site has been producing HDF carrier boards for laminate floorings. Data for different thicknesses are available; however the most frequent thicknesses are 7.4 mm and 6.4 mm. The plants' daily production capacity of HDF panels is about 1,000 m³.

2.1.1 Production process

A process flow diagram for the specific HDF production is shown in Figure 3.

Basically, there are two separate streams/flows of raw materials, one contains wood chips (S1/2), and the second one is sawdust (S3). The incoming wood chips can be further decomposed into chips that are debarked, washed, and chipped on site (S1), while another fraction are incoming wood chips bought from another supplier (S2). Flow S1 and S2 contain a mixture of hard- and softwood, whereas S3 is assumed to contain mostly softwood.

The two material flows of chips (S1/2) are merged into one flow before pre-cooking. Thus, there are two separate flows of material that are fiber-processed (P1 and P2).The S1/2 flows eventually become P1, while S3 supplied P2 as well as P1. P1 and P2 undergo both the same steps of cooking and refining. In the blowlines four different substances are added to the fiber material: 1) urea-melamine-formaldehyde (UMF) resin, 2) additional urea as formaldehyde scavenger, 3) paraffin emulsion, and 4) dye (depending on the recipe). The two material flows merge at the jet-stream dryer, where the fiber material is dried to about 2% moisture content.

Resinated and dried fiber material is stored in the fiber bin, from where it is applied to the forming head onto the band. The steel band with the furnish is discharged into a continuous pre-press and main press. The press consists of numerous press frames, which are bundled into 21 independently controllable *press systems*.

After pressing, boards are trimmed immediately and cooled by using star driers. Final sanding occurs after a conditioning step of several days.

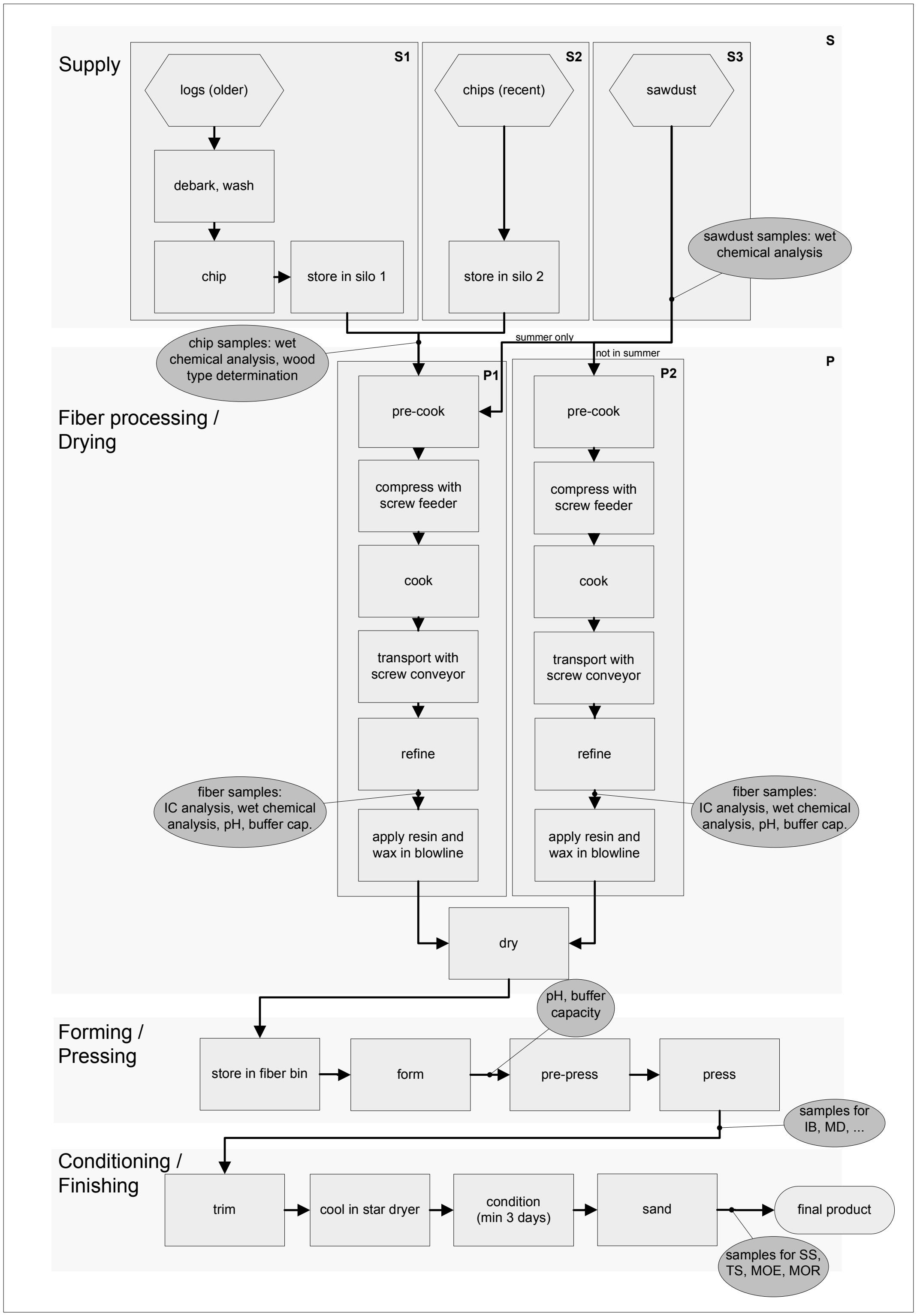


Figure 3.Flow diagram of the process. Oval cycles indicate sampling points for the raw material and final product.

2.1.2 Data

2.1.2.1 Sampling routines

Samples were taken in 2008 from calendar weeks 2 to 51. Basically, the samples were taken daily during the early shift. Data was acquired from boards produced with the two most common recipes, which only differed in the use of dyes. The plant produced several board thicknesses. Main thicknesses were 6.4 mm and 7.4 mm, respectively, besides rarely processed other thicknesses. Thicknesses were all kept in the dataset. In total, 251 sampling runs (observations) were included in the analysis.

2.1.2.2 Process and raw material data sources

Data was made available from four different sources (Table 7). The inhomogeneity of the underlying data required the development of a combined database to simplify data cleanup and selection. As database management systems MS Access ® (Microsoft, 2007) and MySQL ® (MySQL_AB, 2009) were used. Program codes in Visual Basic for Applications (VBA) and Java were written in order to enable data import from different file types and structures.

Data tables were brought into relation using unique keys, i.e. lot ID's and timestamps. SQL queries were created, which selected required data for further analysis by using statistical software packages.

The datasets had some missing values. This was caused mainly by problems during transmission of data from online sensors and test equipments, immeasurable laboratory samples, and problems that occurred in the mapping of database rows.

Table 7. Data sources for the process variables.

No	Data source description
1	Data from online-sensors at the plant, collected by the process control system (Siempelkamp, 2009). Additionally, subjective variables from evaluating the production and formaldehyde perceptibility.
2	Results of ion chromatography (IC) analysis on anions of the fiber material
3	Wet chemistry data obtained from raw material and processed fiber material; wood species / type identification data
4	Results of destructive testing of panels (single values across the board width, and board mean values)

Online-sensors

The process control system Prod-IQ® of the plant's production management system continuously collected data from all online-sensors in an integrated database using Programmable Logic Controllers (PLC) and bus systems (Siempelkamp, 2009). For the presented thesis a data subset was used, i.e. pressures of the press, temperatures in the dryer, and outside conditions.

Subjective evaluations

Subjective evaluation data, i.e. "*production in general"* and "*perceptibility of formaldehyde"* were provided by the shift supervisors. These parameters were defined as ordinal scales, using three scale levels. In the case of "*evaluation production"* the levels 1=bad, 2=average, and 3=good were used. For "*perceptibility of formaldehyde"* the levels were: 1=low, 2=medium and 3=high.

Raw and fiber material analysis

Data from sources number 2 and 3 (Table 7) was additionally collected within an observation period that was part of a greater research project[1]. Data was collected to determine the impact of the raw material on the HDF production process. Raw material samples were taken as chips, and sawdust, prior to the cooking stage. Furthermore, fiber samples were taken after refining the chips and the sawdust. Sampling points are indicated in Figure 3. Table 8 summarizes the analyses performed on the raw and fiber material samples for both processing lines.

Table 8. Overview of lab analysis on raw and fiber material samples.

Type of analysis	**S1/S2 wood chips (RHS[2])**	**S3 sawdust (RSS)**	**P1 fibers of mainly wood chips (FHS)**	**P2 fibers of sawdust (FSS)**	**Form band**
Wood types	X				
Hemicelluloses, extractives	X	X	X	X	
Anions (IC)			X	X	
Surface anions (sorption methylene blue)			X	X	
Buffer capacity, pH			X	X	X

[1] Wood K plus: „Egger MDF 4.2 / Produkt- und Prozessoptimierung in der MDF Produktion"; duration from 1/1/2008 to 12/31/2009
[2] Abbreviations are explained in Appendix A.1

The macroscopic and microscopic *wood type* (wood species) determination was carried out on the chip samples as described by Hoadley (1990).

Hemicelluloses and extractive percentages were determined by applying the methanolysis procedure; the content of single sugars was then analyzed via gas chromatography (GC) (Sundberg et al., 1996). The amount on different hemicelluloses can also be used as fingerprint information for identifying softwood or hardwood, or even the individual species. The amount of hydrophilic extractable substances increases due to the decomposition of cell wall polymers, predominantly by hemicelluloses (McDonald et al., 2006).

Ion (exchange) chromatography analyses on anions were conducted on the liquid extracts of the fiber material in order to quantify the amount of organic and inorganic anions. Formiate and acetate are anions of organic acids with relatively high pK_a-values (negative decadic logarithm of the acid dissociation constant), and products of decomposition from mainly hemicelluloses (Sundqvist et al., 2006). They can be found in high concentrations in wood. Inorganic anions are e.g. chlorides and nitrates that come from strong acids. Anions are expected to influence the resin curing.

The amount of anionic groups on the fiber surface was determined by *sorption of methylene blue*. Methylene blue sorption is an easy and straightforward method for the determination of ionisable anionic groups on the surface and in the fiber. Methylene blue, being a cationic colorant, shows a high affinity to replace the counter ion of the anionic group stoichiometrically. Details and the comparison with other methods are described by Fardim et al. (2003).

Measurements of *pH* were performed by using a conventional pH meter. *Buffer capacity* was measured on the fiber material by acid-based titration. Buffer capacity was expected to be more relevant than the pH values.

Results of destructive tests

After the press, samples from the continuous board are taken for destructive tests of IB (EN 319, CEN, 1993b) and MD (EN 323, CEN, 1993d). Finally, after conditioning, samples were taken from boards for the product properties MOE, MOR (EN 310, CEN, 1993a), SS (EN 311, CEN, 2002a) and TS (EN 317, CEN, 2002b).

For each property ten samples were taken across the board width. Samples had equal distances from each other. Dataset from the destructive testing consisted of ten single values and the derived board mean values for each board property, as calculated by EN 326-1 (CEN, 1994). All sampling points are indicated as oval cycles in the process flow chart (Figure 3).

2.1.2.3 Lag time

Real-time process variable were used with the data aligned in time-order. This means time-lags of all variables determined through destructive testing were considered. For the process variables this was done automatically by the plant's process control system. For the raw material the time-lags were estimated and considered during taking samples for the laboratory analysis.

2.1.2.4 Missing values

The used data contained variables and rows with missing values. Thus, only those variables were used that had the minimum of 50% valid data.

2.1.3 Variables

Table 9 lists the number of variables grouped by section across the production process. A distinction is made between primary and derived variables, where the latter contain calculated variables as well as dummy variables. Calculated variables describe variable interactions, i.e. variables weighted by material flow fractions. Other derived variables are dummies, which were extracted from primary variables. Dummy variables were created for each production month (12 variables), season (4 variables), shift (5 variables), and resin tank (7 variables). Derived variables were added to the dataset on the assumption of improving the strength of the statistical models.

For the analysis, some variables were omitted, since the required information of these variables was already incorporated in the dummy variables. Thus, the dataset for analysis consisted of 245 variables in total.

A listing of all variables including descriptions, categories and descriptive statistics can be found in Appendix B. Descriptive statistics are mentioned separately for the two main thicknesses. If relevant, the tables also contain information to which partial material flow the variable is assigned.

Table 9. Number of variables by process section

Process section / variable category	Primary variables	Derived variables	Total	Total for analysis
Process general	12	21	33	29[3]
Input requirement	2		2	2
Raw material	34	9	43	43
Fiber processing	9		9	9
Fiber	33	14	47	47
Bonding	10	7	17	16[4]
Drying	3		3	3
Form band	7		7	7
Hot press	23		23	23
Product property	60	6	66	66
Total	*193*	*57*	*250*	***245***

2.1.3.1 Response variables

Variables of special interest are board product properties, i.e. IB, SS, MOE, MOR, and TS.

Further, economically relevant variables are *press factor* (PF) and *resin fraction* (RF). To obtain detailed information about the dependencies to other process and raw material factors, PF and RF were both used as response variables in the modeling as well.

The press factor is calculated as shown in Equation 1 (Chapman, 2006).

$$Press\ factor\ (s\ mm^{-1}) = \frac{press\ time\ (s)}{thickness\ (mm)} \qquad (1)$$

The variable *resin fraction* summarized the fractions of resin used by both blowlines.

2.1.3.2 Predictor variables

Generally, all variables with exception of the product properties build the predictor variables sub-matrix. For the modeling, the predictor variables were further distinguished between raw material variables sensu lato (*raw*), and process variables (*proc*). “*Raw”* contained the

[3] LOTID, DATE_PRODUCTION, TIME_SAMLING and SHIFT_NO were omitted
[4] RESIN_TANK_NO was omitted

categories "*raw material*" sensu stricto and "*fiber*" of Table 9, and "*proc*" included the remaining categories, with exception of category "*product property*".

2.1.3.3 Special events

From 5/19/2008 to 5/23/2008 the plant underwent the annually scheduled general maintenance. Production lines were shut down completely. The continuous press had a complete service, along with an accompanied recalibration.

During summer, between 7/7/2008 and 9/11/2008, raw material from the sawdust flow (S3) was merged with the material chip flow (P1). In about the same period, the pressures on the press systems 13 and 14 were reduced, to be able to run the line with at a higher velocity.

From 8/17/2008 the nominal board density was decreased from 900 kg/m^3 to 875 kg/m^3. This resulted in slight changes of other process variables (e.g. lower pressures in press systems 17-22).

2.2 Statistical fundamentals

The notation in this work is as follows: Matrices are written as bold upper-case letters, (e.g. $\mathbf{X}$), vectors are written as bold lower-case letters (e.g. $\mathbf{x}$), and scalars are written in italics (e.g. a).

2.2.1 Sample Standard Deviation (*s*)

The sample standard deviation s is a measure of how variables spread around their mean value. The units are the same as with the original values. The calculation of s is described in Equation 2.

$$s = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} \tag{2}$$

Where:

$\bar{y}$	Mean value of y
y_i	Individual values of y
n	Number of observations

2.2.2 Coefficient of variation (CV)

The coefficient of variation (Equation 3) is defined as a relative standard deviation. It is a scaled measure of dispersion, which is the standard deviation divided by the mean, and multiplied by one hundred if indicated as percentage (Rosner, 2009).

$$CV = 100\% \times \frac{s}{\bar{x}} \tag{3}$$

Where:

CV	Coefficient of variation in %
s	Sample standard deviation
$\bar{x}$	Mean value of x

This value is useful when comparing dispersion statistics across sets of data with varying scales or measure and means, e.g., product types, etc.

However, limits apply in certain cases, such as with those that have an arbitrary definition of the zero point (e.g. pH values, temperatures), and where values tend to spread around zero (e.g. difference between nominal and actual values).

2.2.3 Pearson correlation coefficient (*r*)

The Pearson correlation coefficient r , also known as sample correlation coefficient, can be used to estimate the linear association of x and y. The sign of r indicates the direction of the relationship. By applying a t-test, the significance of r can be determined (Milton and Arnold, 2003). The calculation of r is described in Equation 4.

$$r_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{4}$$

Where:

$r_{x,y}$	Pearson coefficient, values from -1 to +1
s_x, s_y	Sample standard deviation of x and y respectively
$\bar{x}, \bar{y}$	Mean values of x and y
x_i, y_i	Individual values of x and y

2.2.4 Coefficient of determination (*R²*)

A general form of the coefficient of determination is shown by Equation 5.

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} = \frac{SS_{reg}}{SS_{tot}} \tag{5}$$

Where:

R^2	Coefficient of determination, values from 0 to 1
SS_{tot}	Total sum of squares
SS_{reg}	Regression sum of squares
SS_{err}	Residual sum of squares
y_i	Individual values of y

In the case of a multiple regression, the coefficient of determination can be seen as a squared correlation coefficient between the observations y and the estimations $\hat{y}$ (Equation 6).

$$R^2 = r_{y\hat{y}}^2 \tag{6}$$

Different models may be compared by the coefficient of determination if the response variable is the same and the number of parameters is equal (Fahrmeir et al., 2007).

2.3 Multivariate analysis (MVA)

Univariate methods, where only one variable at a time is studied, are very often limited when more complex datasets are present. To meet this problem, multivariate data analysis has to be performed. Introductions to multivariate data analysis are given e.g. by Esbensen (2002), or Eriksson et al. (2001a; b).

Generally, the objectives of multivariate data analysis are data exploration, classification and prediction. Data exploration can be performed by using principal component analysis (PCA), classification with discriminant and cluster analysis, and prediction with regression. Common methods for regression are multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR). PCR and PLSR are based on a projection of the original variables to latent variables, as it is done in PCA. Figure 4 gives an overview about how the methods are related.

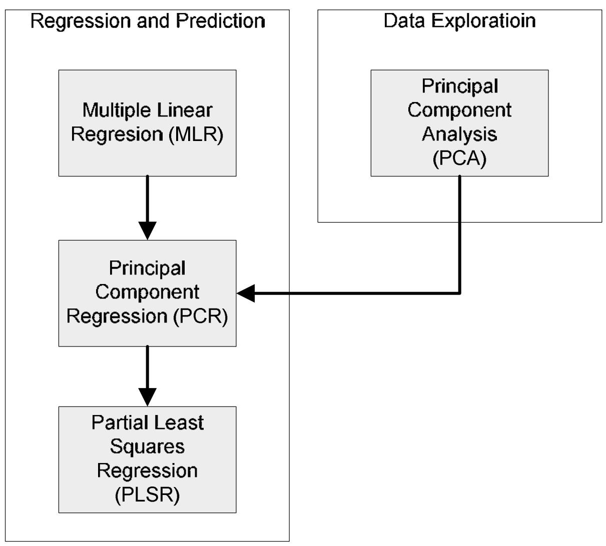


Figure 4. Relations among important multivariate analysis methods.

On the one hand, observations of the same object model, which build the “cheap” variables, are called predictors or X-variables. On the other hand, there are also response variables. The measurements of these Y-variables may be expensive, difficult, time consuming, labor intensive, dangerous, etc. These characteristics have in common that it would be desirable to replace them with measurements of X-variables. Appendix A.2 gives an overview about other terms used for predictor and response variables, depending on the context of use.

2.3.1 Multiple Linear Regression (MLR)

MLR is the classical method to combine a set of X-variables to a corresponding single vector $\mathbf{y}$. The matrix form of the MLR model is shown in Equation 7.

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f} \tag{7}$$

Where:

$\mathbf{y}$	Column vector of response (dependent) variable
$\mathbf{X}$	Matrix of predictor (independent) variables
$\mathbf{b}$	Vector of regression coefficients
$\mathbf{f}$	Error term

Estimation of $\mathbf{b}$ involves a matrix inversion, which may cause severe problems with MLR. In case of multicollinearities in $\mathbf{X}$, which is the case when variables are highly intercorrelated; matrix inversion may become difficult or even impossible. MLR could lead to misinterpretations of the results if multicollinearities are not handles satisfactory.

To avoid multicollinearity it is often practice to select X-variables by e.g. the variance inflation factor (VIF). With this factor some of the variables can be selected to be excluded from the model. A drawback is however that the omission of variables causes some loss of information.

Furthermore, MLR requires more samples than variables. Ignorance of this requirement results in over-fitted models. Moreover, noise and errors in $\mathbf{X}$ (e.g. caused by measurement or sampling), and interference among variables in $\mathbf{X}$ may also cause MLR to fail. In addition, MLR assumes normal distributed residuals (Esbensen, 2002; Kutner et al., 2004).

2.3.2 Principal Component Analysis (PCA)

Principal Component Analysis is a modeling method that extracts the main information in a multidimensional data table. The basic assumption is that the variables can be described by a small number of components, which assumes that the variables are intercorrelated. The information carried by the original variables is projected to a smaller number of latent variables called principal components (PC). The first principal component indicates the direction of maximum variation in the data, thus covering as much of variation as possible. Further principal components are orthogonal to the previous ones and cover as much as possible of the remaining variation.

Interrelationships between different variables can be viewed by plotting the principal components. In this way, patterns and groupings can be detected and interpreted.

An advantage of the projection onto components is that a separation between useful (information-carrying) components and components with noise can be performed. Therefore, the selection of the optimum number of components is crucial. Furthermore, the problem of multicollinearity can be handled with this projection since correlated variables abstracted into latent variables.

PCA results in decomposed smaller matrices, which are called loading and score matrices. Information about the variables is contained in the loading matrix $\mathbf{P}$. It is composed of a few vectors (principal components, PCs), which are linear combinations of the original X-variables. The score matrix $\mathbf{T}$ contains information about the observations. Each observation is described in terms of its projections onto the PCs, instead of the original variables. Equation 8 shows the relation. An additional error term indicates unexplained X-variance.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \tag{8}$$

Where:

$\mathbf{X}$	Matrix of predictor variables
$\mathbf{T}$	Score matrix as result of PCA
$\mathbf{P}^{\mathbf{T}}$	Transposed loading matrix
$\mathbf{E}$	Error matrix containing unexplained variance in $\mathbf{X}$ (zero if all components are used)

2.3.3 Principal Component Regression (PCR)

Principal Component Regression can be thought of as a two-step procedure: (1) transformation of $\mathbf{X}$ into principal components (PCA). This results in a matrix $\mathbf{T}$ as output; which contains scores for each sample, and loadings for each variable. As in PCA, they can be plotted and analyzed for patterns. (2) With $\mathbf{T}$ a MLR model is performed, the modified equation is shown in Equation 9.

$$\mathbf{y} = \mathbf{T}\mathbf{b} + \mathbf{f} \tag{9}$$

Where:

$\mathbf{y}$	Column vector of the response variable
$\mathbf{T}$	Score matrix as result of PCR
$\mathbf{b}$	Vector of regression coefficients
$\mathbf{f}$	Error term containing unexplained variance in $\mathbf{y}$ (zero if all components are used)

A problem with PCR is that there is no guarantee that the principal component decomposition of $\mathbf{X}$ necessarily produces exactly what is required. There is no certainty that the principal components only contain the information that is correlated to the Y-variable of interest. Because of a possible dominance of irrelevant structure parts in $\mathbf{X}$, remaining variance proportions correlated in $\mathbf{y}$ might be in higher order principal components, which never get into the regression stage.

2.3.4 Partial Least Squares Regression (PLSR)

2.3.4.1 Principle

Partial least squares regression (PLSR) was introduced by Wold (1966) and is a generalization of the Multiple Linear Regression. In the special case of a diagonal matrix $\mathbf{X}$ (i.e. X-variables do not show any correlation) and one single Y-variable, PLSR and MLR regression coefficients are identical. PLS is a relatively simple but powerful approach for the analysis of complex problems (Wold et al., 2001).

PLS allows the data structure of $\mathbf{y}$ to influence the decomposition of $\mathbf{X}$ directly, in order to incorporate the variance in $\mathbf{y}$ equally. Thus, the variance in $\mathbf{y}$ is directly used as a support for decomposing matrix $\mathbf{X}$. PLS can be seen as two simultaneous principal component analyses of $\mathbf{X}$ and $\mathbf{Y}$. Through this it is possible to obtain the same prediction results as with PCR, but only with a smaller number of components. As in PCA, a MLR is performed with the resulting components. Instead of the term *principal component (PC)* in PCA/PCR, the term *PLS component* is used for the latent variable.

The principle of PCR and PLSR is shown in Figure 5. The figure shows that response variables depend on prediction variables, with some are and others not incorporated in the model. The unincorporated variables have been omitted because some of them are “lurking” variables (Box, 1966).

Figure 6 shows a simplified overview of matrices and vectors involved in PLS, Equations 10 and 11 show the mathematical relations between the matrices. For $\mathbf{X}$, the scores are in matrix $\mathbf{T}$, and loadings in matrix $\mathbf{P}$. Additionally, there are alternative $\mathbf{W}$-loadings. Respectively for $\mathbf{Y}$, the scores are in matrix $\mathbf{U}$, and loadings in matrix $\mathbf{Q}$.

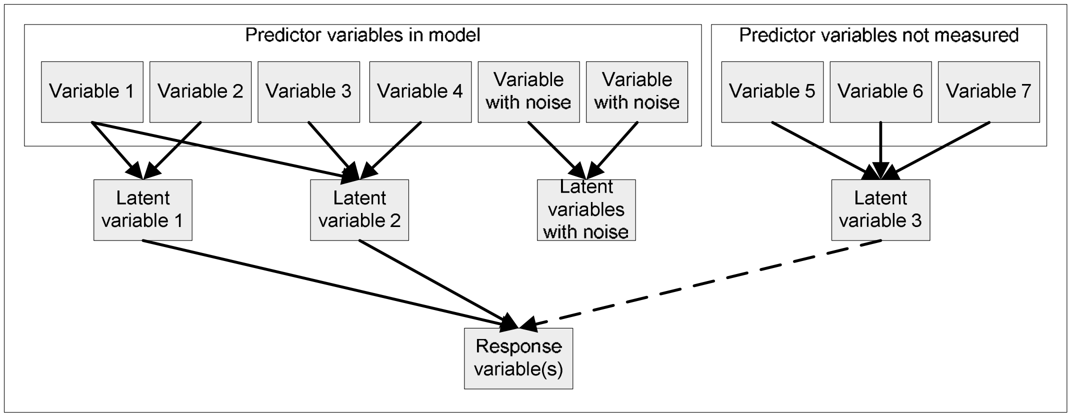


Figure 5. Principle of projection and regression.

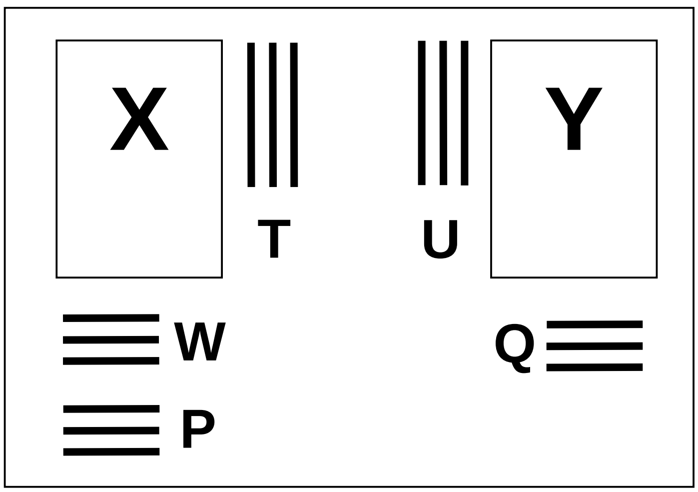


Figure 6. Schematic overview of PLSR (Esbensen, 2002).

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{10}$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{11}$$

Where:

$\mathbf{T}, \mathbf{U}$	Score matrices
$\mathbf{P}$	Loading matrix for $\mathbf{X}$ (with matrix $\mathbf{W}$ acting as precursor)
$\mathbf{Q}$	Loading matrix for $\mathbf{Y}$
$\mathbf{E}, \mathbf{F}$	Error matrices, contain unexplained X and Y-variance respectively

The two simultaneous PCA-analyses are not performed independently. PLS connects the $\mathbf{X}$- and $\mathbf{Y}$-matrices by specifying the $\mathbf{u}$-score vector(s) as starting points of the $\mathbf{t}$-score vectors for decomposing $\mathbf{X}$. Thus, the starting vector $\mathbf{t}_1$ is actually exchanged by $\mathbf{u}_1$. Subsequently $\mathbf{u}_1$ is later substituted by $\mathbf{t}_1$.

The vector $\mathbf{u}_1$ that first influenced the decomposition of $\mathbf{X}$ leads to the calculation of the loadings of $\mathbf{X}$, which are called loading-weights in vector $\mathbf{w}$. As in PCA, the $\mathbf{t}$-vectors are then calculated, but based on the newly calculated $\mathbf{w}$-vector. This $\mathbf{t}$-vector is now used as the starting vector instead of $\mathbf{u}_1$.

The interdependent substitutions of the $\mathbf{u}_1$- and $\mathbf{t}_1$-vectors are done iteratively until convergence, where a final set of $\mathbf{t}$- and $\mathbf{w}$-, and corresponding $\mathbf{u}$- and $\mathbf{q}$-vectors are calculated for the current PLS component. At convergence, a criterion is used, which is composed of the product of a modeling optimization term and a prediction error minimization term.

Thus, not only the data structure in $\mathbf{Y}$ influences the decomposition of matrix $\mathbf{X}$, but vice versa, also the data structure in $\mathbf{X}$ influences the decomposition of $\mathbf{Y}$. By balancing the information in $\mathbf{X}$ and $\mathbf{Y}$, PLS reduces the influence of large variations in $\mathbf{X}$ that do not correlate with $\mathbf{Y}$.

Besides the NIPALS (Nonlinear Estimation by Iterative Partial Least Squares) algorithm introduced by Wold (1966), other algorithms optimized for different purposes exist as well. Examples are an algorithm for missing data (Tenenhaus, 1998), a maximation of covariances when finding the PLS components by SIMPLS (de Jong, 1993), and an algorithm for large number of predictor variables (Ränner et al., 1995; Bhupinder and John, 1997).

2.3.4.2 Regression output

Components

Relevant information in $\mathbf{Y}$ is usually already expected in early components, since PLS focuses on $\mathbf{Y}$. Later components usually contain mostly noise only.

*Loadings (**P**)*

$\mathbf{P}$-loadings express the relationships between the raw data matrix $\mathbf{X}$ and the $\mathbf{T}$-scores.

*Loading-weights (**W**)*

The loading-weight matrix $\mathbf{W}$ represents the effective loadings directly connected to build the relationship between $\mathbf{X}$ and $\mathbf{Y}$. Vector $\mathbf{w}_1$ characterizes the first PLS-component direction in X-space. In general, this direction is not identical to the $\mathbf{p}_1$ direction.

$\mathbf{P}$ and $\mathbf{W}$ are quite similar in many applications. This means, the dominant structures in $\mathbf{X}$ are directed along the similar directions as those with maximum correlation in $\mathbf{Y}$.

*Y-loadings (**Q**)*

The Y-loadings $\mathbf{Q}$ are the regression coefficients from the Y-variables onto the $\mathbf{U}$-scores. $\mathbf{Q}$ and $\mathbf{W}$ may be used for interpreting relationships between X- and Y-variables.

*X-scores (**T**) and Y-scores (**U**)*

$\mathbf{T}$ and $\mathbf{U}$ can be used for interpretation and further exploration. For example, groups of data can be identified by opposing two different column vectors of $\mathbf{T}$ (of two different components), what may indicate that models should be made for each subgroup.

B coefficients

The formal regression equation is depicted in Equation 12.

$$\mathbf{Y} \cong \mathbf{XB} \tag{12}$$

Where:

$\mathbf{Y}$	Response variables
$\mathbf{X}$	Predictor variables
$\mathbf{B}$	Regression coefficients
$\cong$	Approximately equal, since error term is omitted

Vector $\mathbf{b}$ is a column vector of $\mathbf{B}$. Vector $\mathbf{b}$ contains the regression coefficients for one complete model with the number of components used for modeling only. The values of $\mathbf{b}$ can be used to predict new Y-values when only the prediction results are the point of interest. Calculation of $\mathbf{B}$ is shown in Equation 13.

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-\mathbf{1}}\mathbf{Q}^{\mathbf{T}} \tag{13}$$

Where:

$\mathbf{B}$	Matrix of regression coefficients
$\mathbf{W}$	Loading matrix
$\mathbf{P}^{\mathbf{T}}, \mathbf{Q}^{\mathbf{T}}$	Transposed loading matrices

However, vector $\mathbf{b}$ may be difficult to interpret with respect to important and unimportant variables. Therefore, $\mathbf{b}$ should be evaluated with the appropriate loading-weights.

B_w coefficients

B_w-coefficients are weighted regression coefficient. In case of a previous standardization of data, the B_w coefficients differ from the B coefficients. Large absolute values of coefficient indicate importance of the variable, but the sign might be wrong in case of interactions. In order to obtain significances of variables, the more reliable resampling method of jackknifing can be used (see paragraph 2.3.4.5).

The B_w-coefficients can be used to predict new Y-values from new weighted X-values. Large values for B_w indicate important X-variables. The formal equation for predicting weighted data is shown in Equation 14.

$$\mathbf{Y} \cong \mathbf{X_w B_w} \tag{14}$$

Where:

$\mathbf{B_w}$ Matrix of regression coefficients

$\mathbf{X_w}$ Standardized/weighted predictor variables

$\mathbf{Y}$ Response variables

2.3.4.3 Calibration and validation

Models are created in two steps, a calibration and a validation step. The validation step is required in order to test the capability of prediction.

Calibration

Calibration models require a training (also so-called calibration) set. It consists of the matrices $\mathbf{X}$ and the corresponding data of matrix $\mathbf{Y}$. The training set must meet some requirements. Most important is that it is representative of the future population, and the measuring conditions should be as similar as possible. The training set should cover all aspects of possible variation. In practical terms, this means modeling should include all variables that cause variations in the response variable.

Furthermore, the training set should span the X and the Y-space as widely and representatively as possible. Designed experiments play an important role in this. However, in cases of industrial process data, high variation is often not feasible.

The range of the training data set should also cover the range of the data to be predicted. Only in rare situations does extensive extrapolation beyond the range of the calibration set lead to reasonable results.

Validation

Firstly, the purpose of model validation is to avoid overfitting or underfitting of a model by finding the optimal number of components to be used in the calibration stage. Secondly, validation is an instrument for the evaluation of the prediction error, and thus the prediction strength of a future use of the model. Instead of the explained Y-variance R^2, for a validated models Q^2 is used.

Validation is based on a comparison between the model-based prediction results and reference values. Different approaches exist; the use of them depends mainly on the amount of available sampling data.

Test Set Validation

The concept of test set validation requires the availability of at least a second data set, the so-called test set. It should be drawn from the parent population as closely comparable to the calibration data set as possible. This means, the number of observations, the sampling conditions and the sampling time should be comparable to the parent population. This requires twice as many samples for the optimum as would be necessary for the training set only. At least 25% of the size of the training set is sufficient in practice. The difference in variance between the calibration set and an ideal test set is called the sampling variance (Esbensen, 2002).

Cross Validation

In practical situations it is often impossible to obtain separate test sets. In this case cross validation is the most common alternative. An advantage of cross validation is that it can be combined with jackknifing in order to find significant model variables (Martens and Martens, 2000).

An often used version is *full cross validation*. In this leave-one-out cross validation, each sample will be taken out individually from the model calibration set. The remaining $n-1$ data rows are used in the model, and the Y-value of the temporary left out sampled is predicted. This is carried out exactly n times, and for each time one specific Y-value is predicted and the prediction error computed. Finally, all n prediction results are averaged (Hastie et al., 2009).

This case is similar to the separate test set validation. Full cross validation often leads to over-optimistic validation results. This validation method is furthermore suggested for situations where only a few samples are available. For the purpose of finding significant variables, a full cross validation should be best since only model stability is targeted and not the future prediction capability (Esbensen, 2002).

A slightly modified approach is to create segments for cross validation, what is called *K-fold cross validation*. This comes closer to the ideal situation of an independent test set. The number of segments and thus the number of samples per segment is problem dependent. However, Hastie et al. (2009) suggested five- to tenfold cross-validation as a good compromise.

2.3.4.4 Important statistics

Root Means Square Error Prediction (RMSEP)

An important measure for the average prediction error and the modeling error in Y is the Root Mean Square Error of Prediction (RMSEP, Equation 15). The error of prediction is expressed as the Y-variable residual variance by validation. The unit of RMSEP is the original measurement unit of the Y-variable.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_{i,ref})^2}{n}} \qquad (15)$$

Where:

$RMSEP$	Root Mean Square Error of Prediction
$\hat{y}_i$	Predicted value of y_i by validation
$y_{i,ref}$	Reference value of y
n	Number of observations

If the residual are expected to be normally distributed then twice the RMSEP can be regarded as an estimate of the 95% confidence interval (Esbensen, 2002).

Mean Normalized Root Means Square Error Prediction (NRMSEP)

Models of Y-variables with different units and ranges can be compared easily with a mean normalized RMSEP (Feten et al., 2005). Regardless of how the models were created, the models with different preprocessing of X-variables, weightings, number of components, etc. can be compared with each other (NRMSEP, Equation 16).

$$NRMSEP = \frac{RMSEP}{\bar{y}} \qquad (16)$$

Where:

$NRMSEP$	Mean normalized RMSEP
$\bar{y}$	Mean value of y

Q² (cross-validated R²)

Similar to R^2 (Chapter 2.2.4), the Q^2 - being the cross-validated R^2 - can be calculated as shown in Equation 17. In models with several Y-variables, R_m^2 and Q_m^2 for each Y-variable can be obtained (Wold et al., 2001).

$$Q^2 = 1 - \frac{PRESS_A}{SS} \tag{17}$$

Where:

Q^2 Cross-validated R^2

SS Sum of squares of $\mathbf{Y}$

$PRESS_A$ Predictive residual sum of squares for the final model (using A PLS-components)

2.3.4.5 Miscellaneous issues

Standardization

Standardization of the data consists of centering and scaling of the X and Y variables. Other terms are auto-scaling or z-transform.

Scaling assures that all variables included in the analysis have a variance of 1 and thus have equal chance to influence the model regardless of the original variance. The projection methods PCA, PCR and PLS depend on the *relative* variance of the variables, hence scaling is often used if variables are measured with different units, which are of different range and type, respectively. Furthermore, higher or lower weight can be put on certain variables. Weighting modifies the relative influence of variables on a model. If there is no previous knowledge about the relative relevance, the variables are scaled to standardized variance by dividing the values by the standard deviation.

Centering is done by subtracting the mean value of each variable. It assures that variables have a mean value of 0 and results can be interpreted in terms of variation around the mean value.

Equation 18 shows the standardization using the standard deviation (Kutner et al., 2004; CAMO, 2006b).

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{18}$$

Where:

i Number of observations

j Number of variables

x_{ij} Measurement values of variable j of observation i

$\bar{x}_{.j}$ Mean value of all measurement values of variable j

s_j Standard deviation of variable j

z_{ij} Standardized measurement value of variable j of observation i

Missing data

In contrast to Multiple Linear Regression (MLR), PLS with the *NIPALS algorithm for missing data* can handle missing data in $\mathbf{X}$ and $\mathbf{Y}$ to some extent (Tenenhaus, 1998). This algorithm uses a simple imputation method.

Other more sophisticated strategies use the expectation-maximation (EM) algorithm and its extension multiple imputation (MI). EM and MI can result in better statistical models if data is completed in a previous step (Dempster et al., 1977; Rubin, 1987; Schafer, 1997).

Outliers

Observations that do not fit well with the bulk of the data or lie far away from other observations are possible outliers. On the other hand, extreme observations help to span the model over a reasonable range. Thus, it does not necessarily mean that outliers should be removed from the data set.

Data should first be checked for measurement errors. Plots (e.g. T vs. U-score plots, influence plots, leverage plots) and other statistical tools (such as Hotelling T^2 ellipse in score plots, based on a multivariate t-statistic) can help to identify groupings and potential outliers (CAMO, 2006a).

Removal of outliers should be done with only a few outliers at once, in order to evaluate significant changes in the model.

Residuals

Residuals on Y should show a normal distribution. Thus, a normal distribution probability plot on Y-residuals can reveal possible outliers. Furthermore, residual on Y plotted against predicted Y-values should be homoscedastic, i.e. show random scatter and no systematic deviation.

Category variables

Dichotomous, binary category variables (or dummy variables) can be used in models (e.g. simply by transforming it into the values 0 and 1).

If category variables have more than two levels, a re-coding into one dummy variable for each occurring level should be performed.

Optimal number of components

The optimal number of components is essential for multivariate regression methods. The overall prediction ability is best when the prediction error is minimal. The prediction error depends strongly on the number of components in use. Inclusion of more components may improve a modeling fit, but reduces the prediction ability, because the RMSEP may increase again with additional components.

Thus, an aid for finding the optimal number is minimizing the prediction error (also expressed as the residual Y-variance based on the validation). Therefore, it is usually relatively easy to obtain the correct number of components by choosing the number of PLS-components with the minimum Y-variable residual variance, or the maximum explained Y-variance obtained by validation (Q^2).

Significance of variables with jackknifing

Cross validation gives a number of individual sub-models. As extension to cross validation, perturbed regression coefficients, but also loadings, loading-weights and scores can be compared with the full model. The variance between the individual models and the full model reflects the stability towards removing one or more of the samples. The sum of these variances can be used for estimating uncertainties and significances of the model parameters. This so-called jackknifing (as a form of resampling) is described in detail by Martens and Martens (2000).

PLS1 and PLS2

PLS can be distinguished by how many Y-variables can be modeled on it. PLS1 models only one Y-variable, while PLS2 allows several Y-variables simultaneously. PLS2 gives one set of X- and Y-scores, and one set of X- and Y-loadings. They are valid for all Y-variables simultaneously. In practice however, only in rare cases PLS2 models produce better prediction models than a series of PLS1 models on the set of Y-variables (Kessler, 2006).

Linearity

PLS does not handle non-linearity automatically, but several techniques (e.g. transformation, neural networks) may help to overcome this problem (Wold et al., 2001). As for other regression methods, a manual approach is to add calculated variables that contain arbitrary interactions and polynomial effects. However, an addition of those variables may lack objectivity.

Criteria for good models

Good models have a small RMSEP, which should be at least smaller than the standard deviation of the data of the response variable. Furthermore, the correlation between the predicted and the measured Y values, thus being R^2, should be significant with α=0.05. A table for significant values of R^2 is given by Wakeling and Morris (1993). Moreover, the slope of the regression line between predicted versus measured Y values should be close to one; the intercept close to zero; and the bias (as difference between average values of the training set and validation set) should be close to zero (Esbensen, 2002).

Software implementing PLS

Several software packages exist that implement PLS. Examples are *MATLAB*® (The_MathWorks, 2009) in combination with *PLS_Toolbox* (Eigenvector_Research, 2009), the procedure PROC PLS of *SAS*® (SAS_Instutute, 2008), SPSS® PLS (SPSS_Inc, 2008), *SIMCA*® (Umetrics, 2009) and *The Unscrambler*® (CAMO, 2008).

2.3.5 Variation of data

The creation of meaningful models depends on the existence of variation in the underlying data space. If predictor and response variables contain constant values or only low variation, the mutual influence between the variables cannot be determined.

As measure for variation, the coefficient of variation (CV) can be used. The CV is a measure of a normalized standard deviation. Lobenhofer (1990) states that variables below a CV of 2% are more difficult to treat when developing models.

Low variation in the response variable causes models to fail if the worst comes to the worst. The problematic of low variation in predictor variables is depicted in Table 10. Interpretation of models should consider the possibility of too little variation in certain variables. Significant predictor variables in models are considered to have an influence on the response variable(s). However, only predictor variables with at least some variation and no significance in models can be considered to have no influence on response variable(s). Thus, Design of Experiment (DoE) tries to create variation within experiments (Montgomery, 2000).

Table 10. Detection of significant variables.

	X almost constant	**X with at least some variation**
Influence to Y	not detectable	detectable
No influence to Y	not detectable	detectable

2.3.6 Summary of regression modeling methods

Table 11 summaries some important features of the introduced regression methods, and Table 12 the possible output for each method respectively. Generally, PCR has the same limitations as MLR, except that PCR additionally can handle multicollinearities and more variables than observations. PLSR mainly differs from PCR that in PLSR components are built using predictor and response variables simultaneously.

Table 11. Summary of regression modeling methods (Wold et al., 2001; Esbensen, 2002; Kutner et al., 2004; Kessler, 2006).

Feature	**MLR**	**PCR**	**PLSR**
Projection to latent variables	No	Yes	Yes
Can handle multicollinearities	No (variable selection required)	Yes	Yes
Can handle missing data	Yes, with preceding data completion step	Yes (with most algorithms)	Yes (with most algorithms)
Number of required observations	Must be greater than number of variables	Can be less than number of variables	Can be less than number of variables
More than one response variable possible	Yes (multivariate regression)	Yes	Yes (PLS2)
Components are built using	-	Predictor variables	Predictor and response variable(s)
Sophisticated statistics available (e.g. for hypothesis testing)	Yes	Rudimentary	Rudimentary
Noise and errors allowed in $\boldsymbol{X}$	Yes (type 2 with random $\boldsymbol{X}$)	Yes	Yes
Normal distribution required	Yes	No, but better	No, but better[5]

[5] A critical view on the presumed immunity of the normal distribution assumption in PLS is given by Marcoulides (2009).

Table 12. Output of regression methods (CAMO, 2006b, modified; Kessler, 2006).

Result	**MLR**	**PCR**	**PLSR**
Regression coefficients B	X	X	X
Predicted Y-values	X	X	X
Residuals	X	X	X
Error Measures	X	X	X
Significance of model: ANOVA	X		
Scores and Loadings		X	X
Loading-weights			X
Significance of regression coefficient:: Student’s t test	X		
Significance of regression coefficients: Jackknife, Martens uncertainty		X	X

2.4 Statistical Models

To analyze board mean values of each product quality property, *press factor* and *resin fraction*, models were developed for each parameter. Models were also developed for each of the two dominating board thicknesses. Another separation was made by distinguishing between raw material and process variables. Thus, for each response variable five models were developed. The principle of this segmentation is shown in Figure 7.

In models of the *press factor* as Y, the *press velocity* was removed from $\boldsymbol{X}$, since the *press velocity* is closely linked to the *press factor*, which would mask important information. Similarity existed with *resin fraction, resin uses* of the chip line (P1) and sawdust line (P2), which were also removed from the $\boldsymbol{X}$-matrix. In the models of product quality properties, the *press factor* and *resin fraction* were used as predictor variables.

For IB, additional models for the ten single values of IB across the board width were developed as well.. This results in 45 models, as listed in Appendix C.

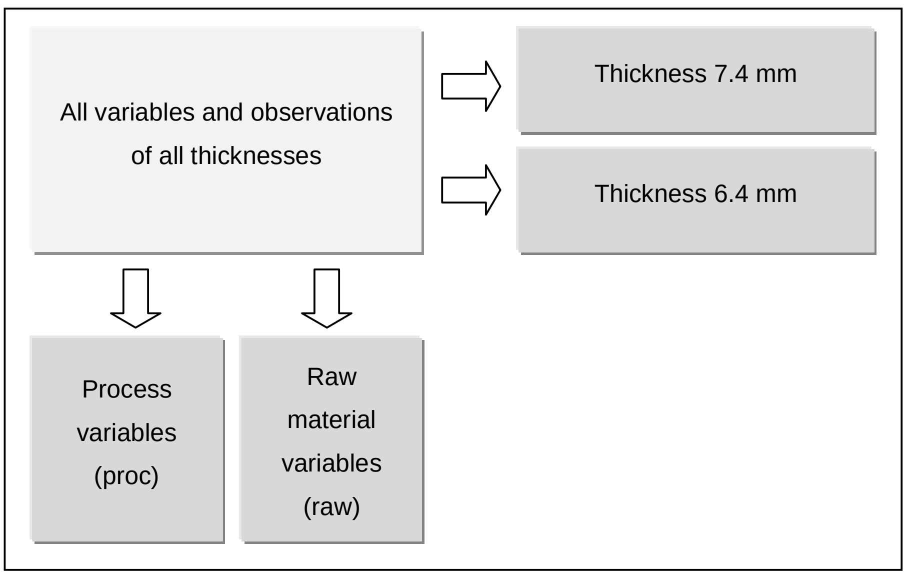


Figure 7. Principle of segmenting the data matrix.

2.4.1 Outlier detection

Outlier detection is done in terms of a technological plausibility check. Removal of data points was done conservatively, as these points could contain important information.

2.4.2 Outline for Modeling

The software package *The Unscrambler* (CAMO, 2008) was used for developing the models. Linear relations between the variables are assumed, as argued by Lobenhoffer (1990) in the case of particleboards, which is similar to HDF. Thus, no further variables were added to the dataset.

All data were standardized before modeling. Ten PLS components were calculated in a first run. Full cross validation was applied that gave deterministic and reproducible results, since no random parameter was required for a selection of segments. Full cross validation is applicable because internal model stability is of main interest. Additionally to the previous outlier detection and removal, further outliers are detected using different plots. Removal is done in a conservative way too. Using score plots, potential groupings were determined and tried to be interpreted.

In subsequent runs, new models were run iteratively with significant variables only (jackknifing with $\alpha=0.10$). The model with the lowest RMSEP and highest Q^2 of cross validation was chosen, wherefrom regression coefficients and further statistics were obtained.

2.5 Correlation and causality

Correlation is a statistical concept for relationship of data. However, a correlation does not automatically mean a chain of causation. Cause and effect deal with interpretation of deterministic relationships.

In order to prove causality, a design of experiment and path analysis (e.g. as application of PLS) can be used (Montgomery, 2000; Ringle et al., 2005).

2.6 Statistical Process Control (SPC)

Statistical Process Control (SPC) is a monitoring method that aims to improve an underlying process. It requires understanding of the process. SPC can be seen as both a learning and optimization framework for a process. The main tool is the control chart which is used for detecting if a process is under control. It can also be used to help to bring it back under control if the process steers off-course (Wheeler and Chambers, 1992).

2.6.1 Variation

The importance of management's obligation to learn about the sources of variation affecting a product and to take steps for the reduction of the variation is emphasized by Wheeler and Chambers (1992).

Variation can be separated into *controlled* and *uncontrolled variation*. *Controlled variation* is characterized by a stable and consistent pattern of variation over time, whereas *uncontrolled variation* contains a pattern of variation that changes over time. As a consequence of this classification, there are two different ways to improve any production process.

2.6.1.1 Controlled variation

When a process displays *controlled variation*, the process should be thought as stable and consistent. The variation present in the process consists only of that which belongs to the process itself. Therefore, to reduce the variation, the process itself has to be changed.

2.6.1.2 Uncontrolled variation

On the other hand, if a process displays *uncontrolled variation*, it is changing from time to time i.e. it is both inconsistent and unstable. The first step for improving the behavior of the process output is the identification of the *assignable causes* of large variation. If this *assignable cause* is unfavorable, then it should be removed.

Thus, it is important to determine whether or not the process displays uncontrolled variation. Shewhart’s control chart is a tool for the detection of uncontrolled variation.

2.6.1.3 Possible states of a process

In SPC, the concept of “zero defects” is insufficient and more efforts need to be done than meeting certain specifications only. Management should learn to understand this process.

In engineering the concept of variation has the objective of meeting specifications. Everything within specification limits is considered as sufficiently good. The concept of SPC is however a continuous process of improvement, and is aimed at products that are as consistent as possible.

Possible states of a process are depicted in Figure 8. The black bar indicates the border between a process showing lack of control, and a controlled process. The states of *total chaos* and *brink of chaos* below the line contain assignable causes of uncontrolled variation. Removal of the causes using control charts as identification tool can shift the process to the *threshold* and further to the *ideal state*. Thus, control charts can be used for continuous improvement, in order to reach the *ideal state*.

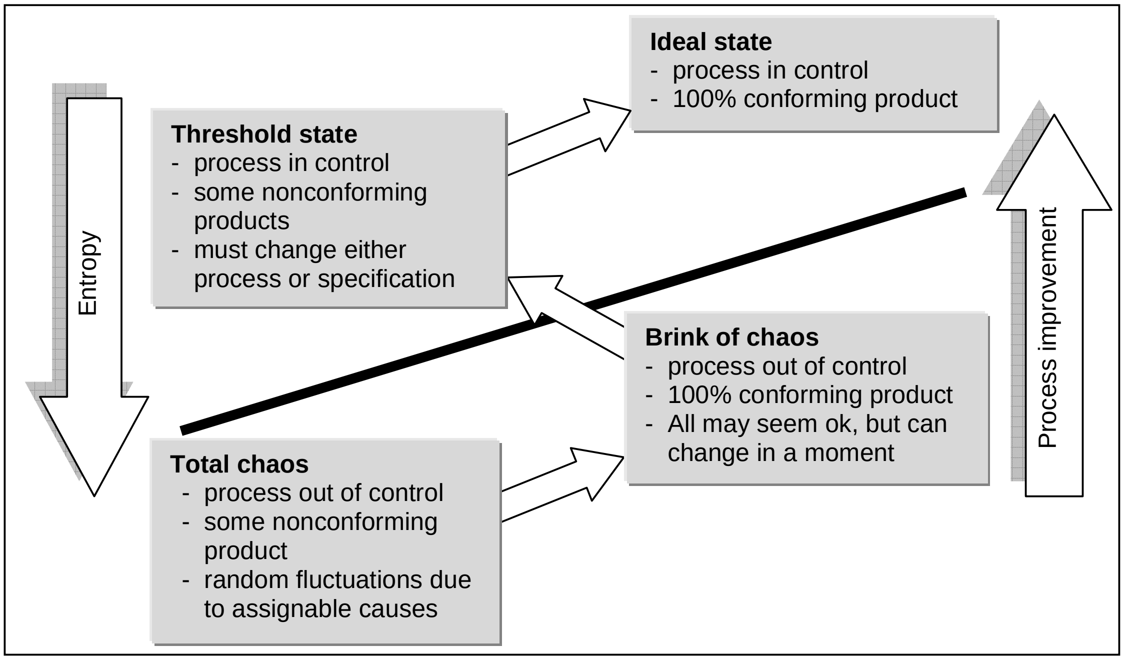


Figure 8. Possible states of a process (Wheeler and Chambers, 1992, modified).

The process that builds the base for this work can be considered to be in the *threshold state*. Thus, efforts have to be done in order to bring it to the *ideal state*. This can be done by changing the specification, or by changing the process.

2.6.2 Control Charts

Control charts provide a systematic way to detect and reduce process variation. Control charts basically display observed data in a time sequence horizontally, with additional horizontal lines indicating the mean value and control limits. So-called *run rules* for detection of unnatural patterns can be applied to control charts. Often used run rules are the Western Electric rules (Nelson, 1984). Different control charts exist that are applied depending on the type of underlying data, e.g. measurement or attribute data, auto-correlated data, data with subgroups, and multivariate data (Wheeler and Chambers, 1992; Wheeler, 1993; Fuchs and Kenett, 1998; Wheeler, 2004). Some important univariate and multivariate control charts are shown in Figure 9.

2.6.2.1 Univariate Control Charts

Shewhart created the control chart in order to detect the presence of uncontrolled variation (Shewhart and Deming, 1986). He provided both, a simple and efficient tool for the presentation of data, and an operational definition for a process in trouble. In the Shewhart control charts, the Upper Control Limit (UCL) is defined as the mean value of the data plus 3 times the standard deviation. The Lower Control Limit (LCL) is defined as mean value minus 3 times standard deviation.

2.6.2.2 *Multivariate Control Charts*

Multivariate control charts, such as Hotelling's T^2 control charts and control charts on principal components can be applied to process variables simultaneously (Fuchs and Kenett, 1998). Principal component regression residuals and scores can be plotted on multivariate SPC charts, which are basically univariate Shewhart charts, yet more powerful at detecting abnormalities (Nomikos and MacGregor, 1995). A software package that combines PCA/PLS and multivariate control charting is SIMCA-P+® (Umetrics, 2009).

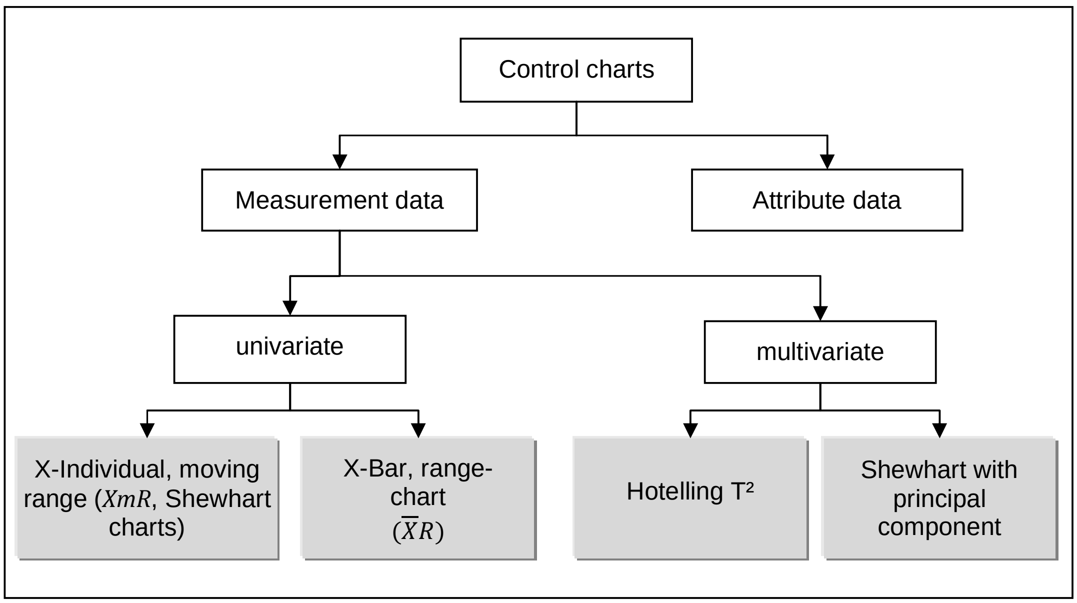


Figure 9. Overview of some important control charts.

3 Results and Discussion

3.1 Correlation of board quality parameters

With a board thickness of 7.4 mm all quality parameter mean values show significant correlations (p<0.05), with exception for TS and IB (Figure 10, Table 13).

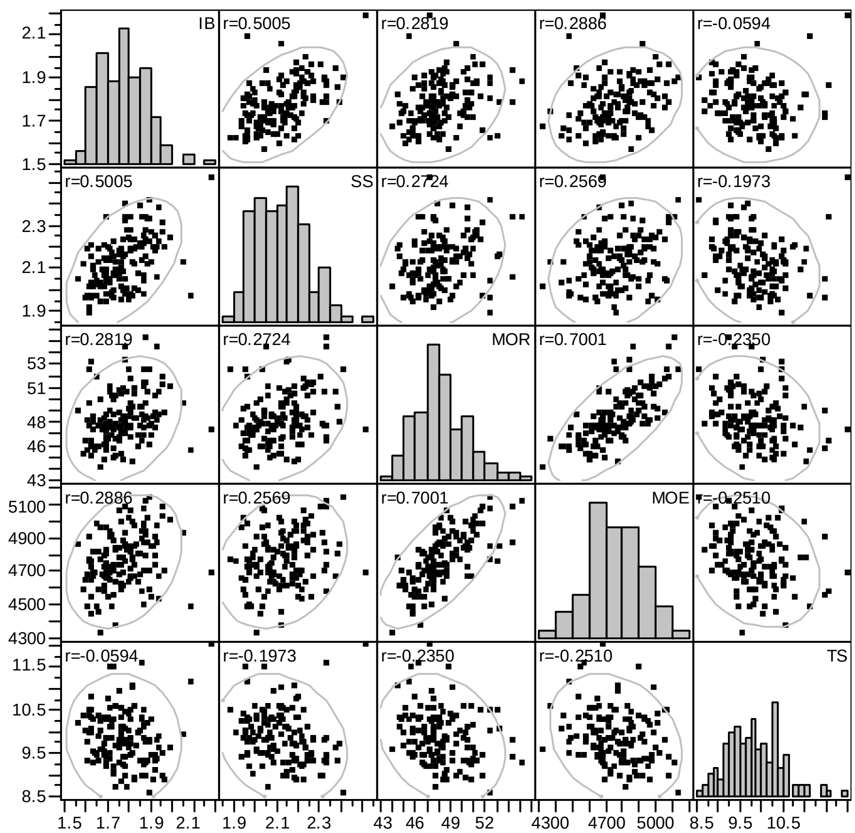


Figure 10. Scatterplot matrix of board properties for thickness of 7.4 mm.

Table 13. Property correlations for the board thickness 7.4 mm.

Variable[6]	by Variable	Pearson correlation	Count	Significance of correlation
IB	SS	0.5005	138	< 0.0001*
IB	MOR	0.2819	138	0.0008*
IB	MOE	0.2886	138	0.0006*
IB	TS	-0.0590	140	0.4884
SS	MOR	0.2724	138	0.0012*
SS	MOE	0.2569	138	0.0023*
SS	TS	-0.1973	138	0.0204*
MOR	MOE	0.7001	138	< 0.0001*
MOR	TS	-0.2350	138	0.0055*
MOE	TS	-0.2510	138	0.0030*

3.2 Number of rejects

In a first step those events were determined, where product properties failed to meet minimum specification limits (lower specification limits, LSL), or exceeding upper specification limits (UCL). The analysis showed that in 3.2 % of all cases IB exceeded LSL, thus producing boards that had to be rejected. This is in accordance with the fact that the LSL equals the 5% quantile value of the board mean IB. Table 14 summarizes the results.

Table 14. Number of rejects by property.

Property	Number of rejects	Rejects in %
IB	8	3.2
SS	0	0.0
MOR	1	0.4
MOE	0	0.0
TS	2	0.8

[6] See Appendix A.1 for abbreviations

3.3 In plane gradient of properties perpendicular to the production direction

The gradient of the product quality properties across the width of the continuous press, thus perpendicular to the production direction, was studied. Values from 10 specimens per board were considered. Observations were taken for a nominal thickness of 7.4 mm only. Single values are determined by destructive laboratory test. The mean of these single values was used to indicate the relevant product quality property.

The means from all positions were compared. Homogenous groups based on a Tukey-Kramer HSD test were significant at the level of α=0.05. Groups are indicated by letters on the top of the subsequent figures 11-15. Positions not assigned by the same letter are significantly different. The Tukey-Kramer HSD test ignores the fact that the data points of the positions are also intercorrelated since each of the 10 specimens came from the same board length position.

3.3.1 Gradient of IB across the board width

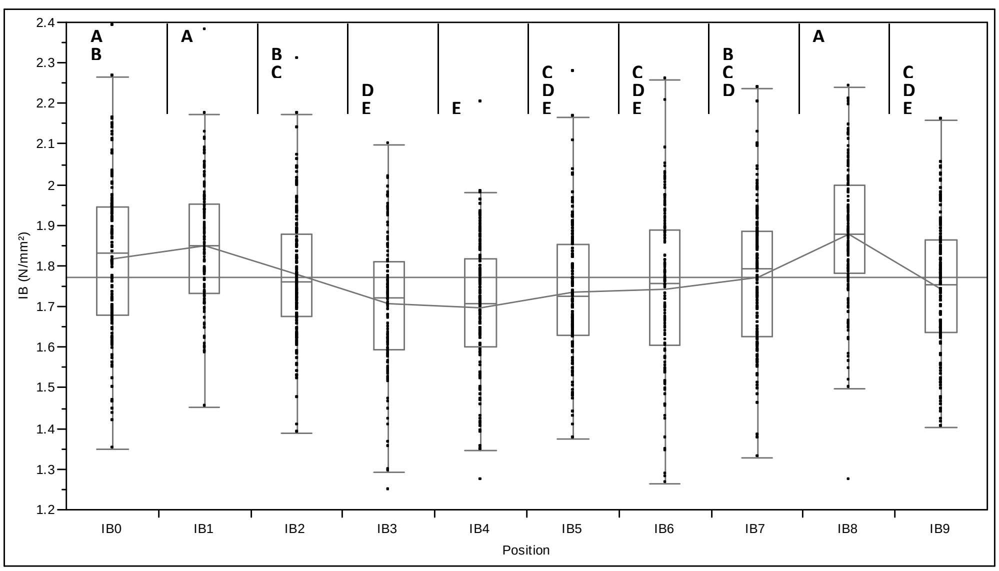


Figure 11. Gradient of IB across the board width, 7.4 mm.

The gradient of IB is shown in Figure 11. A similar gradient for IB can be seen in Hasener (2004). Five homogenous groups were identified. There is an IB minimum at the center, and maxima two points near the edges. At the very edges the IB dropped off again.

3.3.2 Gradient of SS across the board width

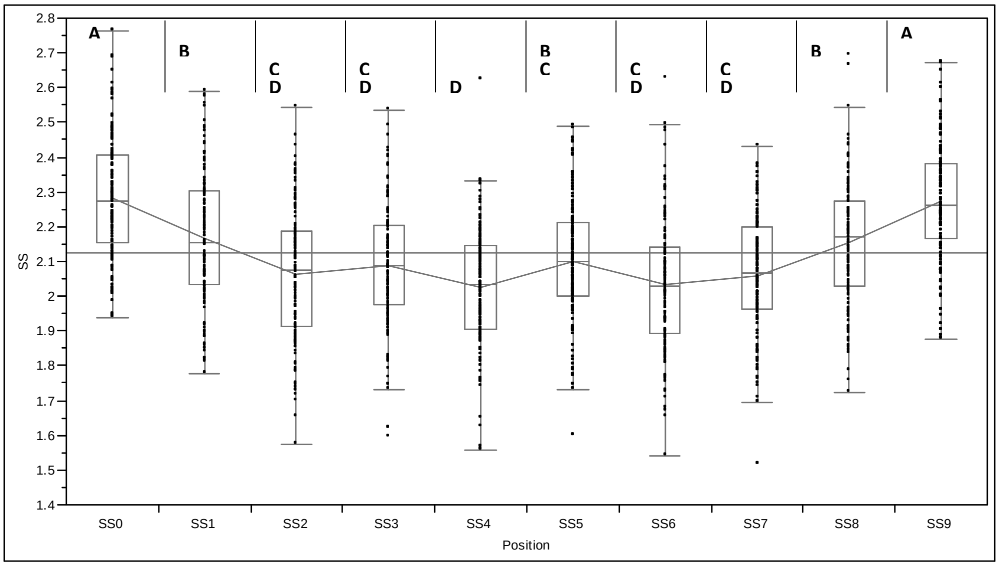


Figure 12. Gradient of SS across the board width, 7.4 mm.

The Tukey-Kramer HSD test reveals four homogenous groups in the gradient of SS (Figure 12). The drop in the center is similar to IB, but the maxima are at the very edges. A similar gradient can be found for SS by Hasener (2004).

3.3.3 Gradient of MOE across the board width

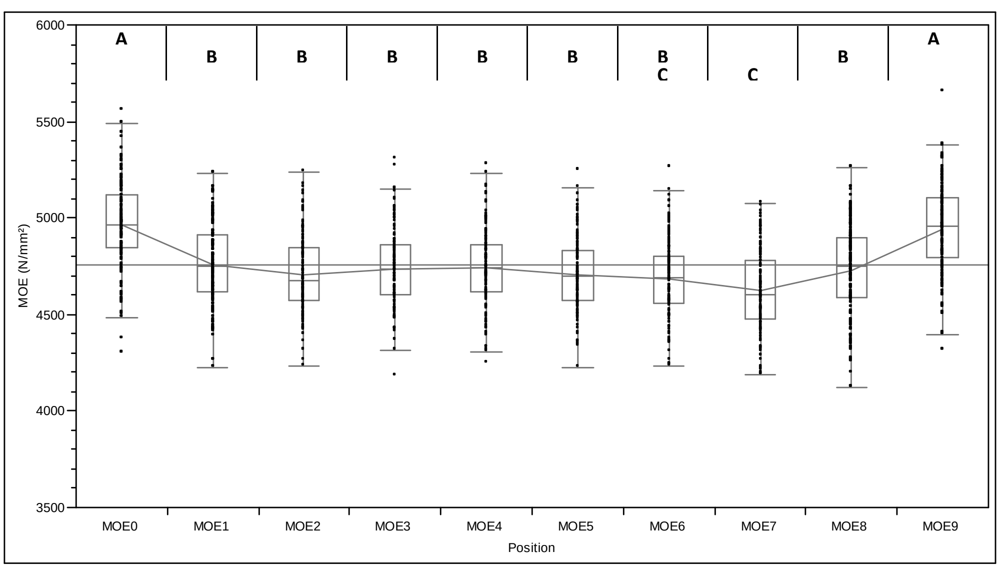


Figure 13. Gradient of MOE across the board width, 7.4 mm.

The gradient of MOE is shown in Figure 13. As for SS, the edges have higher mean values, and even build their own homogenous group. However, in contrast to IB and SS, after the maxima found at the edges values have dropped to a relatively uniform level of one homogenous groups only. An exception is position 7, which builds together with position 8 a third homogenous group (C).

3.3.4 Gradient of MOR across the board width

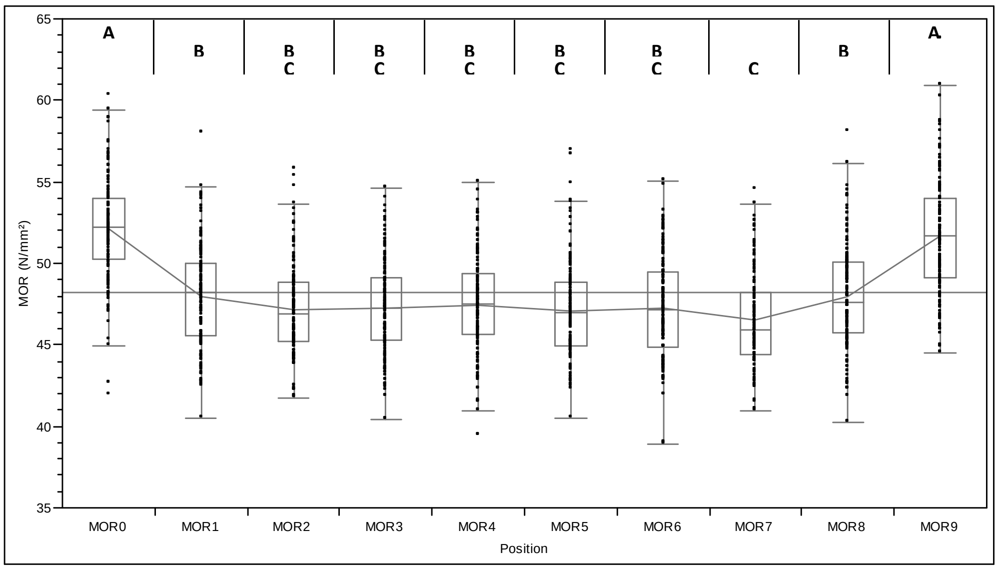


Figure 14. Gradient of MOR across the board width, 7.4 mm.

Figure 14 shows the MOR gradient, which has similarities with MOE. It is also noted that position 7 shows a slight drop, which could be due to a local flaw in the press configuration.

3.3.5 Gradient of thickness swelling across the board

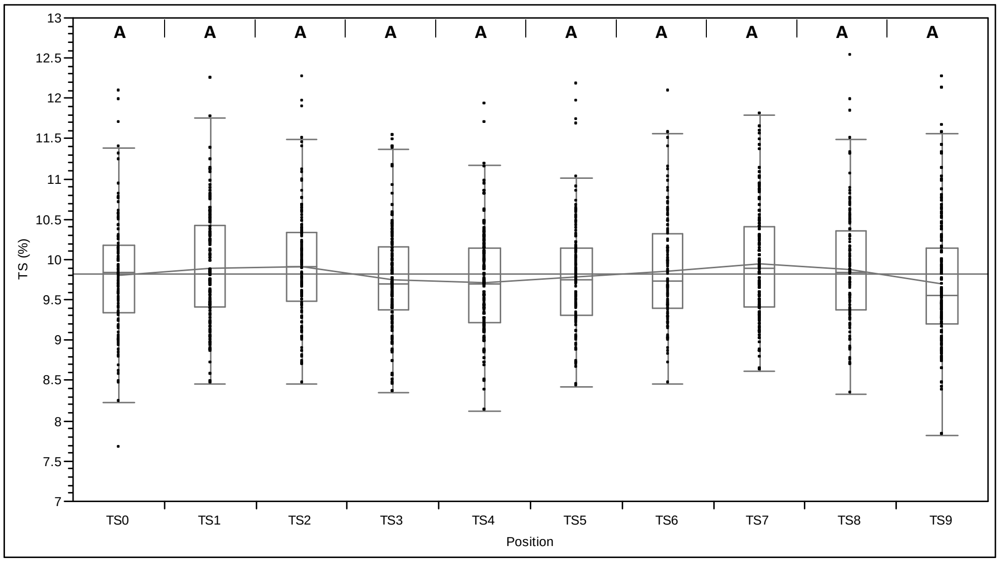


Figure 15. Gradient of TS across the board width, 7.4 mm.

The TS profile is shown in Figure 15. All positions belong statistically to one population, thus showing optimized process conditions for TS with 7.4 mm thick boards. Hasener (2004) presented a gradient with a slight drop in thickness swell at the edge positions. This effect was not seen in this study.

3.3.6 Summery for the gradient across the board width

Overall, quality parameters show minima in the center of the board, thus significantly better values were found closer to the edges. This can be explained by the sealing of the edges to avoid outgassing, what prevents a loss of moisture and heat. For IB, the technologically most important property, the effect was compensated by an altered press program at the edges only, in order to keep the IB uniform predominantly. Sealing has as side-effect that different density profiles are created on the edges that in the center of the board, what was however not studied in this work.

TS's behavior is different however, as there is no drop in the center of the board. This leads to the assumption, that the effect of outgassing has no or only minor effect on TS.

For MOE and MOR, the drop at position 7 can be explained by a flaw in the configuration of the press.

3.4 Results of PLSR models

In total, 45 models were developed. Models were made for all of the product properties. Furthermore, separate models were created for 1) taking into account the main thicknesses and 2) considering processes and raw material variables separately. In addition, models with single IB values were made as well.

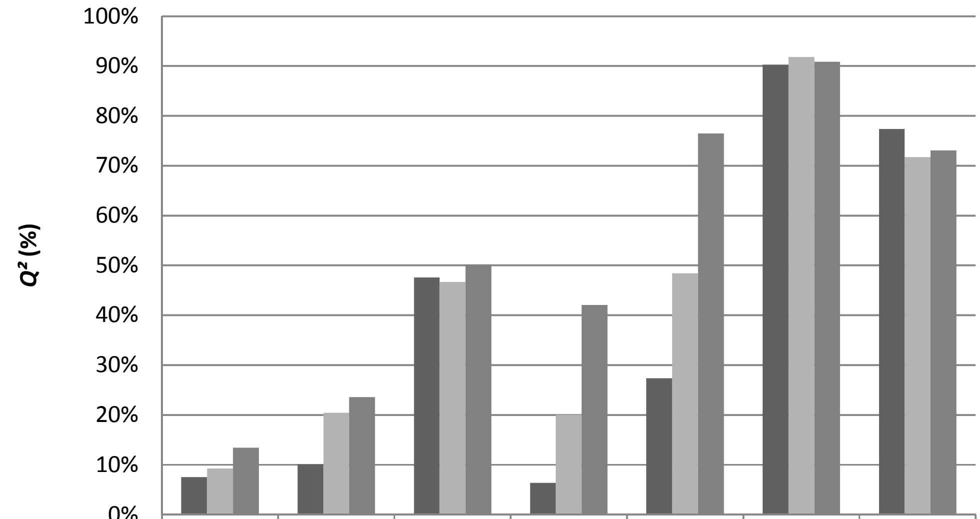
Appendix C lists summary statistics of all the models, and the weighted regression coefficients B_w are shown in Appendix D. Q^2 and the NRMSEP usually show a negative correlation, meaning that a high Q^2 causes a low prediction error.

3.4.1 Models by major board thicknesses

Models with IB and SS have generally low Q^2, see Figure 16. This is not necessarily surprising, since models with a high coefficient of variation (CV) of the response variable generally also have a high R^2. IB is the most important product property, thus it is attempted to keep its variation constant. SS has a high correlation to IB, thus the variation is generally low too. Models with certain thicknesses (for IB, SS, MOE and TS) show a lower Q^2 in comparison to models using all observations. However, the NRMSEP (Figure 17) is similar for models having the same response variable.

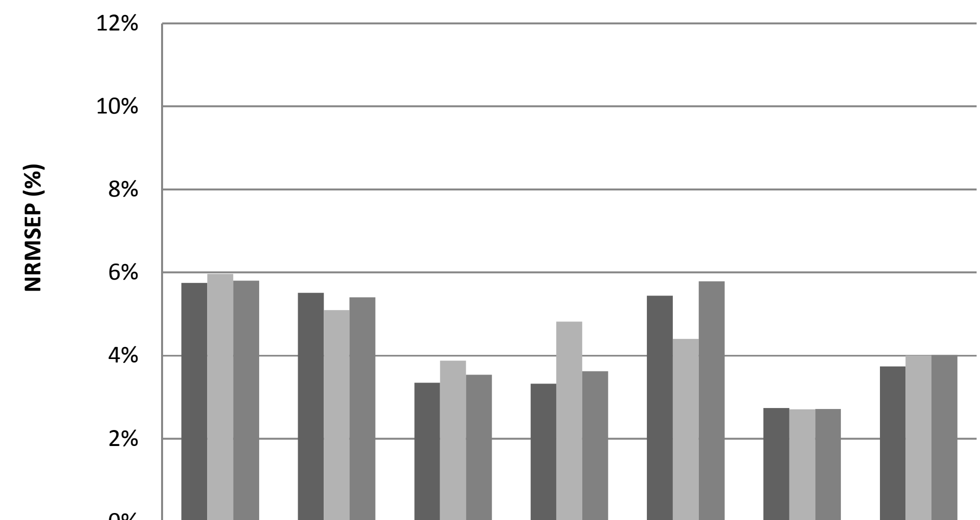
Models on MOR have similar values for Q^2 and RRMSEP respectively. Models for MOE and TS (with certain thicknesses) show huge differences in R^2 in comparison to models with all thicknesses. This can be explained by a strong effect due to the board thickness, especially with MOE and TS. Surprisingly, the thickness effect does not apply for MOR (which is expected, since it generally has strong correlation with MOE). A different picture can be seen with NRMSEP of MOE and TS, where no negative correlation to Q^2 can be seen. For example, the model for MOE with 7.4 mm shows a low NRMSEP, even if the corresponding Q^2 is very low.

Models with PF and RF are generally satisfying. They benefit from the variables expressing the amount of material processed. Q^2 and NRMSEP correlate as expected.



	IB	SS	MOR	MOE	TS	PF	RF
7.4 mm	7.6%	10.1%	47.6%	6.4%	27.4%	90.3%	77.3%
6.4 mm	9.3%	20.5%	46.7%	20.0%	48.4%	91.9%	71.8%
all thicknesses	13.4%	23.6%	50.1%	42.1%	76.5%	90.9%	73.1%

Figure 16. Q^2 of models with board thickness 7.4 and 6.4 mm, compared to models of all thicknesses.



	IB	SS	MOR	MOE	TS	PF	RF
7.4 mm	5.75%	5.51%	3.34%	3.32%	5.44%	2.74%	3.73%
6.4 mm	5.96%	5.09%	3.88%	4.81%	4.40%	2.71%	4.00%
all thicknesses	5.81%	5.40%	3.54%	3.62%	5.78%	2.72%	4.02%

Figure 17. NRMSEP of models with board thicknesses 7.4 and 6.4, compared to models with all thicknesses.

3.4.2 Models for raw material and process variables

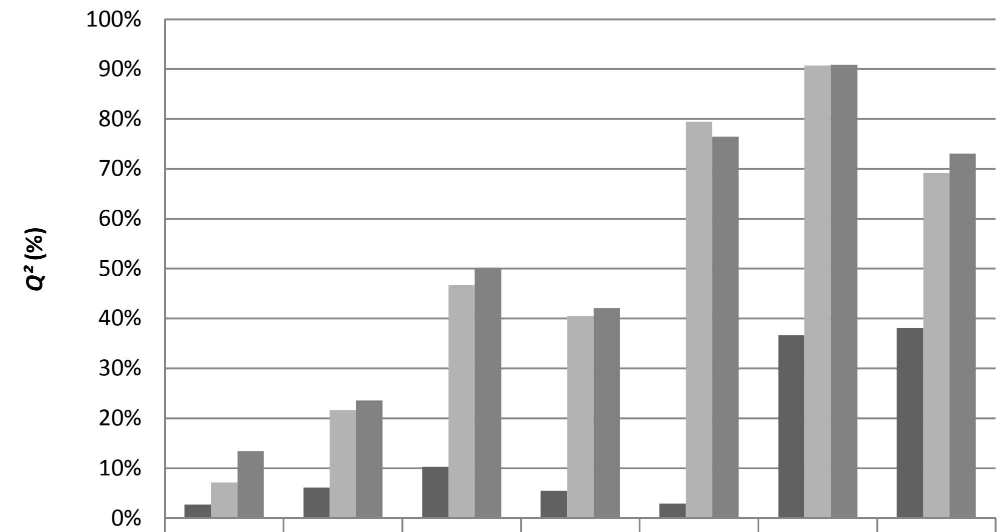
Generally, the process variables accounted for the largest fraction of cross-validated explained Y-variance (Q^2). In some cases the models with only process variables showed even better results than those with all variables. With this information the influence of the raw material variables is estimated at an average of 21% on Q^2. More detailed results are listed in Table 15.

Figure 18 shows the Q^2, and Figure 19 the NRMSEP for the models with all thicknesses, split by raw material and process variables. A low Q^2 on raw material models is generally reflected in a higher NRMSEP. However, this is not true for IB where the model with raw material have a lower NRMSEP (5.90%) than the model with the process parameters only (6.02%). It is noticed that TS and PF have a relatively high NRMSEP in the models with the raw material variables only. This is reflected by a low Q^2 in these models.

Table 15. Estimated fraction of Q^2 by raw material variables only.

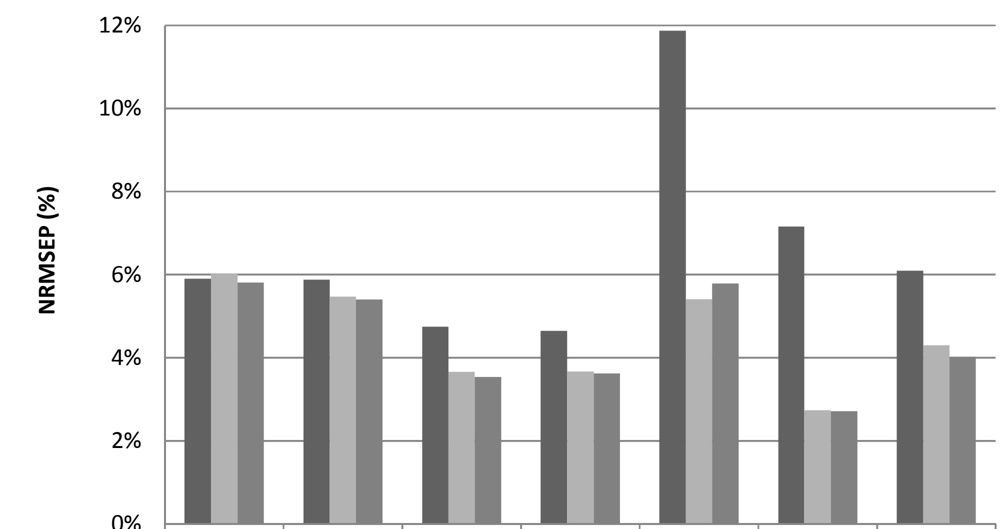
Property	**Fraction of raw material variables**[7]
IB	28%
SS	22%
MOR	18%
MOE	12%
TS	4%
press factor	29%
resin fraction	36%
average	***21%***

[7] Calculated as fraction of Q^2 from models with raw material on the total Q^2



	IB	SS	MOR	MOE	TS	PF	RF
raw material	2.7%	6.1%	10.3%	5.5%	2.9%	36.7%	38.2%
process	7.1%	21.6%	46.7%	40.5%	79.4%	90.7%	69.1%
all variables	13.4%	23.6%	50.1%	42.1%	76.5%	90.9%	73.1%

Figure 18. Q^2 of models with all thicknesses, variables split by raw material and process.



	IB	SS	MOR	MOE	TS	PF	RF
raw material	5.90%	5.88%	4.74%	4.65%	11.87%	7.16%	6.09%
process	6.02%	5.47%	3.66%	3.67%	5.41%	2.74%	4.30%
all variables	5.81%	5.40%	3.54%	3.62%	5.78%	2.72%	4.02%

Figure 19. NRMSEP of models with all thicknesses, variables split by raw material and process.

3.4.3 Models for IB on positions across the board width

Single values for IB of boards with a thickness of 7.4 mm were studies using separate models. Figure 20 shows the Q^2, and Figure 21 shows the NRMSEP of the models. The Q^2 is 10.2% in average, which is higher than the 7.6% of Q^2 in the model of IB with 7.4 mm Figure 16. For position 2, no validation of the model was possible. Where there was a large coefficient of variation (CV) (e.g. especially on position 0 and 6), good results are obtained. Surprisingly, in all models of the single values, the NRMSEP remains higher than the corresponding model with the board mean value of IB. This even applied to the models where a very high Q^2 is obtained.

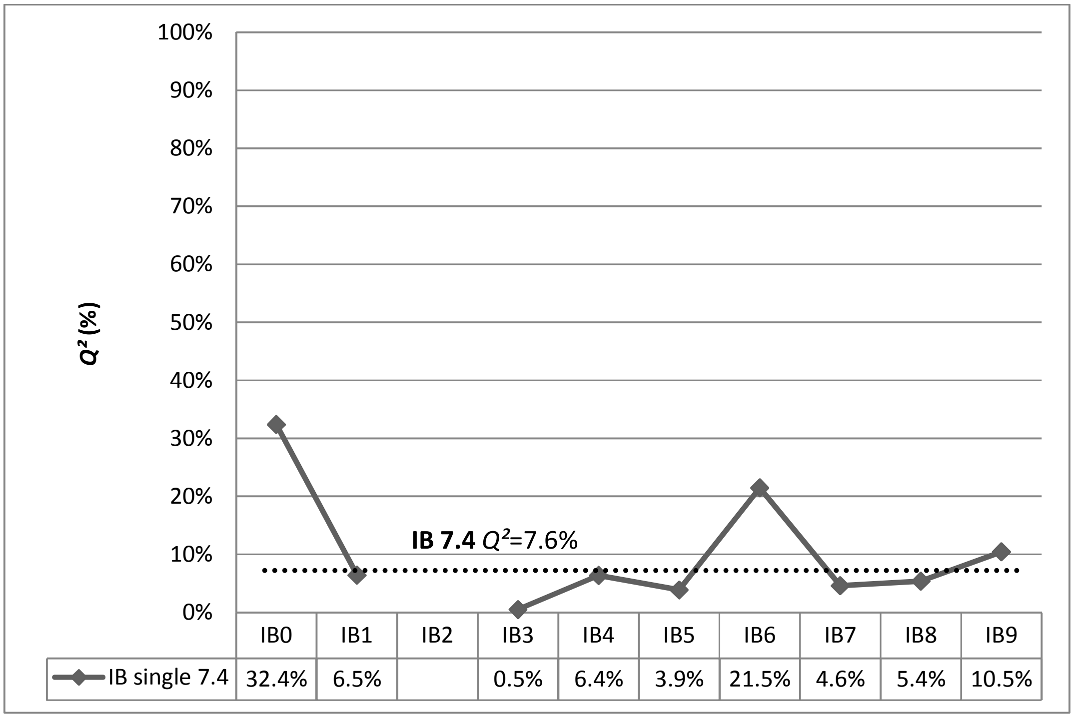


Figure 20. Q^2 of single values of IB, thickness 7.4 mm. Q^2 of IB 7.4 with board mean value is shown as dashed line.

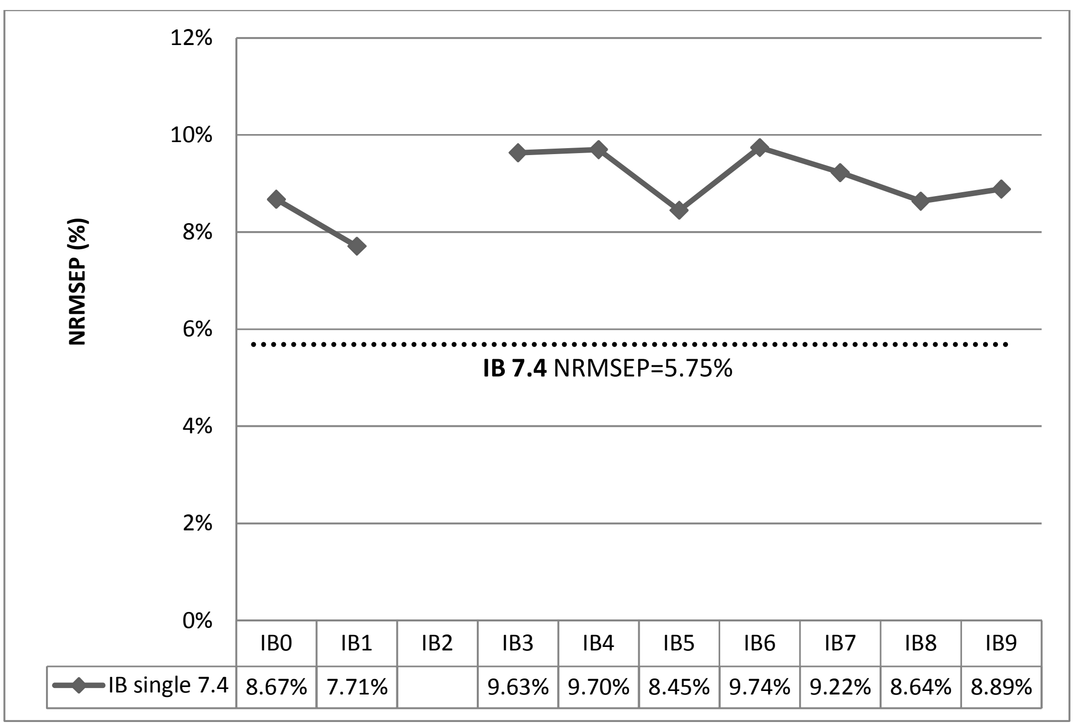


Figure 21. NRMSEP of models with single values of IB, thickness 7.4 mm. NRMSEP of IB 7.4 with board mean value is shown as dashed line.

3.4.4 Detailed results of the models

The weighted regression coefficients B_w for all models are listed in Appendix D. The most important variables were determined by the frequency of occurrence in the 45 models; a table for the number of occurrences of regression variables is listed in Appendix E. Unless otherwise noted, 5 models were used for each property (*7.4 mm, 6.4 mm, all thicknesses, process only, raw material only*) and the tables contain variables with a minimum of two occurrences. Five occurrences of variables are not possible because modes with type *process only* and *raw material only* are mutually exclusive in variable use.

3.4.4.1 Most important variables for internal bond strength (IB)

Models for IB are relatively low in Q^2 and high in (N)RMSEP. This is caused because the process is optimized for consistent IB, thus containing little variation regarding IB. Variation in data is however required to obtain meaningful models. Thus, models were made with the single values of IB across the board width as well.

Results of models with board mean value

The models based on the board mean value of IB showed minor seasonal fluctuation regarding IB (Table 16). Additionally, there seems to be a positive influence with IB with sawdust, what is indicated by the material flows S3 and P2. The perceptibility of formaldehyde can be linked to higher resin use, which improved IB.

Xylose (predominantly found in hardwood), and extractives in the raw material seem to have negative effects. Especially the first press systems seem to be more important. High pressures in these systems have a positive influence on IB.

Table 16. Summary of most frequent variables for IB.

Group of variables	**Variables**	**Frequency**	**Sign** [8]
Season	MONTH_07	2	+
Perceptibility formaldehyde	PERCEP_FORMALDEHYDE	2	+
Sawdust fraction	HE_SS_FRACT	2	+
Raw material chemical properties	R_XYLOSE_WGT R_TOTAL_EXTRACT_WGT	each with 2	-
Press	DI_AV_PRES_SYS02 DI_AV_PRES_SYS03 DI_AV_PRES_SYS04 DI_AV_PRES_SYS05	each with 2	+/-

[8] +/- indicates alternating sign of the regression coefficient in the models

Results of models with the single values

To obtain additional information modeling on single values across the board width was performed as well. Positions 0 and 6 have very high Q^2 compared to all the other positions. This correlates with the high variability of data on these positions. Results for models of single values for IB are summarized in Table 17. This table and the tables in the subsequent paragraphs contain only variables that occur at least twice in the models for the same response variable.

This reflects the results of the models with average IB, but contains additionally details. Table 17 contains variables that are significant at least twice. Generally, the significant variables of models with the board mean value are also included, and some additional variables emerge. The seasonal influence becomes more noticeable, especially in the summer when process modifications were applied.

It seems that wood chips have some negative influence (negative sign of variables WOOD_INPUT_TOTAL and RHS_DISCHARGE_SCREW), and sawdust positive (FIB_PROC_LINE2_FRACT and RSS_DISCHARGE_SCREW).

Some sugars of hemicelluloses show influence with inconclusive signs. As for the models with the board mean values of IB, xylose (which is predominantly in hardwood) generally show some negative influence.

A high temperature in the dryer has positive influence to IB. Sprinkling on the formband has negative influence.

The same press systems as the other models of IB have influence, and additionally systems 8 and 14 emerge. Moreover, high press velocities (meaning a low press factor) have some negative impact.

Table 17. Summary of the most frequent variables for models with single values of IB.

Group of variables	Variables	Frequency	Sign [9]
Season	SEASON_SPRING	2	+
	MONTH_05	2	+/-
	MONTH_07	5	+
	MONTH_08	3	-
Material flow	WOOD_INPUT_TOTAL	2	-
	FIB_PROC_LINE2_FRACT	2	+/-
	RHS_DISCHARGE_SCREW	2	-
	RSS_DISCHARGE_SCREW	2	+
Hemicelluloses	FHS_GALACTOSE	2	-
	RHS_XYLOSE	2	-
	F_XYLOSE_WGT	2	+/-
	F_RHAMNOSE_WGT	2	+/-
Dryer	A241_TEMP_AV_T	3	+
	FS_DRYER_TEMP_IN FS_DRYER_TEMP_OUT	each with 2	+
Sprinkling	FS_SPRINKLING	2	-
Press	DI_AV_PRES_SYS02 DI_AV_PRES_SYS03 DI_AV_PRES_SYS04 DI_AV_PRES_SYS05	each with 2	+
	DI_AV_PRES_SYS08	2	+/-
	DI_AV_PRES_SYS14	2	+
	PRESS_VELOCITY	2	-
	PRESSFACTOR	2	+

[9] +/- indicates alternating sign of the regression coefficient in the models

3.4.4.2 Most important variables for surface soundness (SS)

The models for SS are slightly better than those for IB. The most frequent variables for the models of SS are summarized in Table 18. SS does not show any obvious seasonal influence. Negative impact in SS is shown with material flow P1, which generally processed wood chips. This is indicated by the negative signs for variables on discharge screws and from the refiner power drain of line P1.

For fibers, high formiate content, methylene blue sorption, and the buffer capacity of the glued fiber indicate a good SS. Formiate indicates acidity, and the methylene blue sorption high polarity. This causes, together with a high buffer capacity, good conditions for curing the resin, which explains the better SS.

Furthermore, the latter systems of the press are also significant. Low pressures in these systems have a positive influence on SS. A possible explanation for this could be that the top layers, which are crucial for SS, are already formed at the beginning of the press and a high pressure at the end of the press has a harmful influence (in terms of breaking up adhesive bonds that have been already cured) on the top layer. It is assumed, that this harmful influence is much more effective than the positive influence of the first systems.

Table 18. Summary of most frequent variables for SS.

Group of variables	**Variables**	**Frequency**	**Sign**
Material flow	RHS_DISCHARGE_SCREW RHS_REFINER_POWER_DRAIN RSS_FILLLEVEL_DIGESTER	each with 3	-
Fiber properties	FHS_IC_FORMIATE	4	+
	F_IC_FORMIATE_WGT	3	
	F_METH_SORP_ANION_WGT_EX PUFFER_CAPACITY_FIB	each with 3	+
Press	DI_AV_PRES_SYS17 DI_AV_PRES_SYS18 DI_AV_PRES_SYS19	each with 3	-

3.4.4.3 Most important variables for modulus of rupture (MOR)

The most frequent variables for the models of MOR are summarized in Table 19.

Generally, there is a slight seasonal influence in MOR. There is a positive influence on MOR with the partial material flow P1, which primarily processes wood chips. This is indirectly indicated by the refiner usage (a positive sign means that chip refiner is only preferred to both refiners), the speed of the discharge from the chip refiner (RHS_DISCHARGE_SCREW), and the fraction of the material flow of P2 (FIB_PROC_LINE2_FRACT). However, there is an unexplained contradiction with the total wood input for material flow P1, which results in a decreased MOR.

A high bulk density results in reduced MOR properties. The finding is contradictory to the study of Xing et al. (2006). The cause for this could be matured wood or hi-density hardwoods with a high fraction of dust (e.g. beech). However, this cannot be proved by the models.

High mannose content (mannose is predominantly found in softwood) affects the MOR positively. A high resin use has positive influence to MOR. High pressures in press systems 1, 10, 11 and 16 are also positive for MOR. High press velocities (thus low press factors) show negative influence.

Table 19. Summary of the most frequent variables for MOR.

Group of variables	Variables	Frequency	Sign
Season	MONTH_04 MONTH_07	3	+
Material flow	REF_IN_USE RHS_DISCHARGE_SCREW	each with 3	+
	FIB_PROC_LINE2_FRACT WOOD_INPUT_TOTAL		-
Bulk density	FS_BULK_DENSITY	4	-
Resin	RSS_RESIN_USE	3	+
Hemicelluloses	RSS_MANNOSE F_MANNOSE_WGT	each with 3	+
Press	DI_AV_PRES_SYS01 DI_AV_PRES_SYS10 DI_AV_PRES_SYS11 DI_AV_PRES_SYS16	each with 3	+
	PRESS_VELOCITY		-
	PRESSFACTOR		+

3.4.4.4 Most important variables for modulus of elasticity (MOE)

Table 20 summarizes the most frequent variables for the models of MOE. Models show somewhat worse results in comparison to MOR in terms of the amount of significant variables.

No obvious seasonal effect for MOE was determined. There is a positive influence in MOE with material flow S3, which is dominated by sawdust. This is indicated by the positive impact of the sawdust fraction (HE_SS_FRACT). This finding is however partly contradictory to the models of MOR.

As with MOR, a high bulk density influences the MOE negatively. A high resin use affects the MOE positively. High pressures in press systems 1 and 11 are also positive for MOE, what is consistent with the results of MOR.

Table 20. Summary of the most frequent variables for MOE.

Group of variables	**Variables**	**Frequency**	**Sign**
Material flow	HE_SS_FRACT	4	+
Bulk density	FS_BULK_DENSITY	4	-
Resin	RSS_RESIN_USE	3	+
Press	DI_AV_PRES_SYS01	3	+
	DI_AV_PRES_SYS11	4	+

3.4.4.5 Most important variables for thickness swelling (TS)

The most frequent variables for the models of TS are summarized in Table 21. A positive sign on the variables indicates increased TS with higher values on the variables. In the case of TS higher values are not desirable. Thus the signs have to be interpreted inversely.

Generally, during winter time the TS tends to be higher compared to the rest of the year. This can be explained by generally higher moisture content (MC) of the raw material in winter. This led to a higher fiber MC on the formband (FS_MC_FORMBAND), and to higher thicknesses. On the other hand a high MC (measured by microwave in the continuous press) indicates lower TS.

Grammage frequently occurs in the models, yet the sign is unclear. However, for models of only certain thicknesses, high grammage means high thickness swelling. This finding correlates with the hardwood ash (*Fraxinus* spp.) and this corresponded with higher TS. This could be caused by insufficient ability to compress. Furthermore, a higher use of resin results in lower TS.

The pressure on system 10 of the press is relevant as high pressures in this system result in a low thickness swelling. This system has appeared important in models for MOR already. However, in subsequent systems, especially system 17 and 21, higher pressures resulted in higher TS. Thus, lower pressure should be targeted in these systems. System 17 has been already shown as important for SS. High press velocities have significant, yet unclear influences on the sign.

Although TS is relatively independent of other board quality parameters in terms of correlation (Figure 10), there are important common variables, such as the pressure on system 10 to MOR, the pressure on system 17 to SS, and the press velocity to MOR and IB.

Table 21. Summary of the most frequent variables for TS.

Group of variables	**Variables**	**Frequency**	**Sign**[10]
Season	SEASON_WINTER	4	+
	MONTH_01	3	+
	TEMP_OUTSIDE	3	-
Grammage	FS_GRAMMAGE	4	+/-
Wood type	RHS_NC_ASH	3	+
Resin	RESIN_FRACT	3	-
Moisture content	FS_MC_MICROWAVE_CONTI	each with 3	-
	FS_MC_FORMBAND		+
Press	DI_AV_PRES_SYS10	each with 3	-
	DI_AV_PRES_SYS17 DI_AV_PRES_SYS21		+
	PRESS_VELOCITY		+/-

[10] +/- indicates alternating sign of the regression coefficient in the models

3.4.4.6 Most important variables for press factor (PF)

Principally, very good models are obtained for the press factor (PF). Table 22 gives an overview of the most frequent variables. Similar to TS, a positive sign on the variables indicates a higher PF with higher values on the variables. However, a higher PF is not desirable in terms of the economically important production speed, thus the signs have to be interpreted inversely.

A seasonal influence is shown in winter and in June/July. The high PF in winter could be explained by the higher fiber moisture content, which has to be compensated during the process. This would explain the positive effect of higher temperatures in the dryer in terms of lower PF.

High material flow rates (expressed by variables of the partial flows P1 and P2, and the preference of both refiners in use), result in a higher press velocity and hence a lower PF. Thus, the sign of variables on P1 and P2 are no oppositional.

High grammage and a high sawdust fraction (S3) are seen as bad in terms of a high PF; meaning that a high PF can be reached by a high total material flow, but with a slight preference on the type of wood chips.

A high arabinose content results in a low PF. Arabinose is shown to be degraded significantly by fiber processing (Weigl et al., 2009). A high buffer capacity of the fiber on the formband, generally has a positive effect on curing, and results in a low PF. The positive effect has already been shown for SS.

There is high correlation in the occurrence of variables with the pressures of systems 2, 3 and 4 and IB; 11 with MOR and MOE; and 17 with SS and TS.

Table 22. Summary of most frequent variables for PF.

Group of variables	**Variables**	**Frequency**	**Sign**
Season	SEASON_WINTER	3	+
	MONTH_01	4	+
	MONTH_06	4	-
	MONTH_07	4	+
Grammage	FS_GRAMMAGE	4	+
Material flow	REF_IN_USE	4	+
	HE_SS_FRACT	3	+
	WOOD_INPUT_TOTAL FIB_PROC_LINE2_FRACT RHS_REFINER_POWER_DRAIN RHS_DISCHARGE_SCREW	each with 4	-
	RSS_DISCHARGE_SCREW	3	-
Fiber properties	F_ARABINOSE_WGT PUFFER_CAPACITY_FIB	4	-
Dryer	FS_DRYER_TEMP_IN A241_TEMP_AV_T FS_DRYER_TEMP_OUT	each with 4	-
Press	DI_AV_PRES_SYS02 DI_AV_PRES_SYS03	each with 3	+
	DI_AV_PRES_SYS04	4	+
	DI_AV_PRES_SYS11	3	+
	DI_AV_PRES_SYS13 DI_AV_PRES_SYS14	4	+
	DI_AV_PRES_SYS17	4	-
	DI_AV_PRES_SYS18	3	-

3.4.4.7 Most important variables for resin fraction (RF)

As with TS and press factor the signs in Table 23 have to be interpreted inversely, since the target is a low resin fraction due to economical reasons.

Table 23. Summary of most frequent variables for RF.

Group of variables	**Variables**	**Frequency**	**Sign**
Season	SEASON_SPRING	4	+
	SEASON_SUMMER	3	+
	SEASON_FALL	4	-
	SEASON_WINTER	3	-
	MONTH_01	3	-
	MONTH_03 MONTH_06	each with 4	+
	MONTH_09	4	-
	MONTH_10	4	-
Formaldehyde	PERCEP_FORMALDEHYDE	3	-
Material flow	REF_IN_USE	3	-
	HE_SS_FRACT	3	-
	WOOD_INPUT_TOTAL	4	+
Digester	RSS_FILLLEVEL_DIGESTER	4	-
	RHS_REFINER_DIGESTER_PRESSURE	4	-
	RHS_DWELL_TIME_DIGESTER	3	-
Raw material and fiber properties	RHS_C_PINE	3	-
	FHS_XYLOSE	3	-
	FHS_GLUCOSE	3	-
	F_RHAMNOSE_WGT	4	-
	F_XYLOSE_WGT	3	-
	F_GLUCOSE_WGT	3	-
	PUFFER_CAPACITY_SS	3	+
Dryer	A241_TEMP_AV_T FS_DRYER_TEMP_OUT	each with 3	+
Sprinkling	FS_SPRINKLING	4	-
Moisture content	FS_MC_MICROWAVE_CONTI FS_MC_FORMBAND	each with 3	-
Press	DI_AV_PRES_SYS01	3	+
	DI_AV_PRES_SYS06 DI_AV_PRES_SYS07 DI_AV_PRES_SYS08 DI_AV_PRES_SYS13 DI_AV_PRES_SYS14	each with 3	-

Generally, the resin fraction altered strongly during the year: In spring and summer there is a higher, and in fall and winter a lower amount of required resin. In spring and summer more resin has to be added in order to compensate for the faster pre-curing processes during these periods.

In the model, a higher perceptibility of formaldehyde indicates a low resin fraction. This is contradictory to the logical expectation that higher resin use results in a high perceptibility in formaldehyde. A possible explanation is that with higher resin use, the evaporating formaldehyde has a higher probability of getting associated with surrounding resin drops due to a larger surface. Thus, a lower perceptibility of formaldehyde occurred.

The more sawdust in use, the less resin is required. This is indicated by the negative coefficients of variables HE_SS_FRACT and REF_IN_USE (*both refiners* are preferred in contrast to *chip refiner only*). Furthermore, it is indicated by the positive coefficient of the wood input of partial material flow P1.

High pressure, dwell times and fill levels of the digester results in less resin consumption. This is generally the effect of good fiber processing, which could be the reason for this lower demand, since lignin is deposited on the fiber surface (Widsten et al., 2002).

A high use of pine wood (*Pinus* sp.) results in a lower demand of resin. The process in the plant is designed and optimized for pine as raw material, thus this finding was not surprising. Furthermore, a high content of monomer sugars of hemicelluloses (especially xylose, which mainly exist in hardwoods) and glucose results in a low demand of resin.

High buffer capacities of sawdust (P2) demand more resin, what is contracting the finding that high buffer capacity was shown to have a good impact on SS and PF.

High temperatures in the dryer demand more resin, that could be caused by pre-curing. A similar effect was observed by Xing et al. (2004).

Sprinkling (which is required for fibers that were dried) seems to decrease the need for resin. This explanation is supported by the low resin demand with higher fiber moisture contents on the form band.

The high pressure on press system 1 results in a higher demand in resin. Pressure on this system also causes better MOE and MOR.

Furthermore, high pressures on system 6-8, and 13-14 result in low resin use. This confirms that high pressure generally results in a lower demand in resin, since the performance is obtained by compression.

3.4.4.8 Most important variables on all models by frequency of occurrence

Table 24 shows the most frequent variables, occurring in all models. The listing contains only variables that have a minimum amount of 10/38 maximum possible occurrences. The listing summarizes the most important variables; however it ignores the relevance of quality parameters. This is true except in the case of IB, where the models of the single values of IB and with a nominal thickness of 7.4 mm are included in the frequency column as well.

Table 24. Most frequent variables for all response variables.

Group of variables	**Variables**	**Frequency**	**IB**	**IB sin**	**SS**	**MOR**	**MOE**	**TS**	**PF**	**RF**
			positive sign desirable					**negative sign desirable**		
Thickness	NOMINAL_THICKNESS	12	-		+	+	+	-	+	
Season	MONTH_01	11		+				+	+	-
	MONTH_07	17	+	+	+	+	+		+	
	MONTH_08	10		-	-	-	-			-
	SEASON_SPRING	10		+			+		-	+
	SEASON_WINTER	10						+	+	-
Material flow	HE_SS_FRACT	12	+				+		+	-
	WOOD_INPUT_TOTAL	18		-	-	-	-	-	-	+
Refiner	REF_IN_USE	11				+			+	-
	RHS_DISCHARGE_SCREW	14	-	-	-	+			-	+
	RHS_REFINER_POWER_DRAIN	12		-	-		-	+	-	
	RSS_DISCHARGE_SCREW	10	+	+			+	-	-	-
Resin	RSS_RESIN_USE	10		+	+	+	+	-	-	
Dryer	A241_TEMP_AV_T	10		+					-	+
Form band	FS_GRAMMAGE	15		+	+	+	+	+/-	+	
	FS_BULK_DENSITY	12			-	-	-	+		+
Press	DI_AV_PRES_SYS01	11				+	+	+	-	+
	DI_AV_PRES_SYS04	10	-	+			+		+	
	DI_AV_PRES_SYS10	10			+	+	+	-		
	DI_AV_PRES_SYS11	11		+		+	+		+	
	DI_AV_PRES_SYS14	10		+				+	+	-
	DI_AV_PRES_SYS17	11			-		+	+	-	
	PRESS_VELOCITY	12		-	-	-	-	+/-		

Thick boards usually resulted in better mechanical properties, which was not the case for IB. Seasonality could be observed: For IB, SS, MOR and MOE, July was good, while August was less favorable in the data used for this study. This can be most likely explained by the use of sawdust during summer, which influenced most board properties positively. However, this effect was masked by a change in the nominal density in August, which required a slight reconfiguration of the plant with some short term instability for the production process.

The positive effect of sawdust is furthermore visible by variables in the material flow and refiner data. For MOR, PF and RF, this effect is not as clear as it was for the other board quality parameters. Resin used on sawdust showed a consistent result and was evaluated positively. High drying temperatures cause better IB and lower PF. However, the use of resin is higher, as higher dryer temperatures cause faster pre-curing. In general, high grammage and low bulk density creates the best results for all properties. The pressure of some press systems, and the low press velocity creates positive results. The results of the press are discussed in more detail in Chapter 3.4.5.

Noticeably, only process variables, and no variables regarding the raw material are contained in the listing of Table 24.

3.4.4.9 Most important variables on all models by B_w, weighted by the technological relevance of the board properties

A further weighting of B_w was performed for each of the variables on selected models, to reflect the importance of certain board properties according to industrial requirements. In this approach, the five board properties with only three models for each were used: 1) models with observations of a thickness 7.4 mm only 2) 6.4 mm only, and 3) all observations. Weighting is done using the proportions in Equation 19. These proportions were determined by plant personnel. IB is the most important technical characteristic, and SS is as important as all remaining properties. MOR and MOE are overall as important as TS.

$$\mathrm{IB} : \mathrm{SS} : \mathrm{TS} : \mathrm{MOE} : \mathrm{MOR} = 50 : 25 : 12.5 : 6.25 : 6.25 \tag{19}$$

Equation 20 shows the calculation of the weighted B_w for each predictor variable.

$$\mathbf{b}_{\mathrm{wx_i}} = \sum \frac{\mathbf{B}_{\mathrm{w}_{ijk}} \times \mathbf{p}_{\cdot j \cdot}}{\mathbf{n}_{i \cdot\cdot}} \tag{20}$$

Where:

$\mathbf{b}_{\mathrm{wx}_i}$ Weighted B_w for predictor variable i

$\mathbf{B}_{\mathrm{w}_{ijk}}$ Weighted regression coefficient of the model with board property j as response variable and sub-model k, for predictor variable i (for TS, $\mathbf{B}_{\mathrm{w}_{ijk}}$ was used inversely)

i Index of the predictor variable

j Index of the board property (IB | SS | TS | MOE | MOR)

k Index of a model of a certain board property
(Observations with boards of 6.4 mm | 7.4 mm | and all thicknesses)

$\mathbf{n}_{i \cdot\cdot}$ Amount of regression coefficients used
(usually n=3; except if i indicates the variable NOMINAL_THICKNESS, where n=1)

$\mathbf{p}_{\cdot j \cdot}$ Weighting proportion for property j as shown in Equation 19

Then, the absolute values of $\mathbf{b}_{\mathrm{wx_i}}$ were sorted in descending order, and the top fifty most important variables were identified (Table 25)

Table 25. Most important variables determined by weighted B_w of selected models.

Rank	Variable	Weighted B_w for variables with positive influence	Weighted B_w for variables with negative influence
1	MONTH_07	0.08327	
2	NOMINAL_THICKNESS	0.07442	
3	FS_BULK_DENSITY		-0.03748
4	PERCEP_FORMALDEHYDE	0.03324	
5	MONTH_10	0.02850	
6	FHS_IC_FORMIATE	0.02732	
7	PUFFER_CAPACITY_FIB	0.02642	
8	FSS_EXTRACT_TOTAL		-0.02583
9	RSS_FILLLEVEL_DIGESTER		-0.02309
10	MONTH_04	0.02158	
11	RSS_RESIN_USE	0.02070	
12	DI_AV_PRES_SYS10	0.01798	
13	RHS_NC_MAPLE		-0.01771
14	RSS_DISCHARGE_SCREW	0.01684	
15	HE_SS_FRACT	0.01679	
16	RSS_DYE_FRACT		-0.01664
17	DI_AV_PRES_SYS21		-0.01661
18	SEASON_FALL	0.01653	
19	R_TOTAL_EXTRACT_WGT		-0.01651
20	DI_AV_PRES_SYS17		-0.01649
21	SEASON_WINTER		-0.01604
22	R_XYLOSE_WGT		-0.01528
23	FS_GRAMMAGE	0.01513	
24	SHIFT5		-0.01508
25	FSS_MANNOSE	0.01505	
26	F_METH_SORP_ANION_WGT_EX	0.01470	
27	RHS_REFINER_DIGESTER_PRESSURE	0.01410	
28	PRESS_VELOCITY		-0.01385
29	PUFFER_CAPACITY	0.01308	
30	RHS_REFINER_POWER_DRAIN		-0.01214
31	RHS_DISCHARGE_SCREW		-0.01204
32	DI_AV_PRES_SYS19		-0.01199
33	RHS_RESIN_USE	0.01159	
34	DI_AV_PRES_SYS18		-0.01065
35	F_IC_FORMIATE_WGT	0.01028	
36	DI_AV_PRES_SYS08		-0.01018

Rank	Variable	Weighted B_w for variables with positive influence	Weighted B_w for variables with negative influence
37	DI_AV_PRES_SYS11	0.00916	
38	F_MANNOSE_WGT	0.00915	
39	FS_MC_MICROWAVE_CONTI	0.00913	
40	MONTH_08		-0.00851
41	SHIFT2	0.00850	
42	DI_AV_PRES_SYS16	0.00849	
43	DI_AV_PRES_SYS01	0.00839	
44	RHS_C_FIR	0.00825	
45	RHS_NC_ASPEN	0.00802	
46	TEMP_OUTSIDE	0.00776	
47	RSS_MANNOSE	0.00746	
48	FS_MC_FORMBAND		-0.00733
49	RESIN_TANK05		-0.00683
50	REF_IN_USE	0.00656	

3.4.5 Summary of the press

Table 26 summarizes the influence of the average pressure on the press systems by the quality properties, PF and RF. Signs in bold letters have the highest frequency of occurrence within the models (minimum 2 out of 4). All other signs only occur once. The table additionally contains the PF and press velocity (PV) as predictor variables.

The models reveal for IB that the first press systems have the most influence, the sign is however not distinct. For SS, the focus is on latter systems. MOR and MOE are influenced positively by a similar press system setting; whereas MOE is somewhat more sensitive, with an additional focus on the final systems of the press. Remarkably, the very first system has positive influence on MOR and MOE, but a negative one on TS at the same time. Furthermore, pressure in the final systems caused higher TS. The signs for correlation between PF/RF and the press systems are ambiguous.

The influence of press system 1 leads to the assumption that this system causes a pre-compression, which can be seen as somewhat detached from the subsequent press system. In this specific press, i.e. system 11, the vertical density profiles (VDP) are shaped. This is reflected in the positive influence on pressure in this system to IB, MOR and MOE.

The modification of the press program in the summer with systems 13 and 14 is reflected in the models. It can be seen, that lower pressures in these systems actually result in a lower press factor and thus a higher press velocity. However, this modification had a slightly negative influence on IB, causing a higher demand in resin. TS was however slightly lower during this period, since for a low TS, pressure should be generally avoided with the last systems.

Table 26. Significance of pressure on press systems by the quality properties.

System	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	21	22	PF	PV
IB		+/-	+/-	+/-	+/-	-		+/-	+		+		+	+						-		+	-
SS										+							-	-	-				-
MOR	+									+	+					+						+	-
MOE	+	+	+	+	+				+	+	+	+					+	+	+	+	+	+	-
TS[11]	+									-		+	+	+	+		+	+	+	+	+	+	+/-
PF[12]	-	+	+	+	+			-	-		+	-	+	+	+		-	-				n/a	n/a
RF[13]	+					-	-	-					-	-						+		-	

3.5 Example control charts for thickness swelling

As an example for a possible control chart, Figure 22 shows a multivariate control chart (Figure 9) for the first PLS component of the models for thickness swelling with a thickness of 6.4 mm. The small numbers next to the data points indicate run rules (Nelson, 1984). For example, number 1 indicates a shift in the mean or an increase in the standard deviation, and is defined as an excess of the upper control limit (UCL) or falling below the lower control limit (LCL). This correlates with the presence of higher rates of swelling during the winter season. The rule number 2 occurred several times, which indicates a shift in the process mean. The shift correlates with the time period of the reduction of nominal density.

In comparison, Figure 23 shows a univariate X-Individual control chart for the variable TS of 6.4 mm thick boards only. It can be however seen, that the multivariate control chart in Figure 22 shows a clearer picture about changes in the process. The reason is that a multivariate control chart with the score of PLS components combines all existing variables of multivariate models that lie predominantly on the plotted PLS component.

Thus, when implementing SPC in the plant, the selection of meaningful control charts is crucial.

[11] Signs have to be interpreted inversely as high TS are undesirable.
[12] Signs have to be interpreted inversely as high press factor is undesirable.
[13] Signs have to be interpreted inversely as high resin fraction is undesirable.

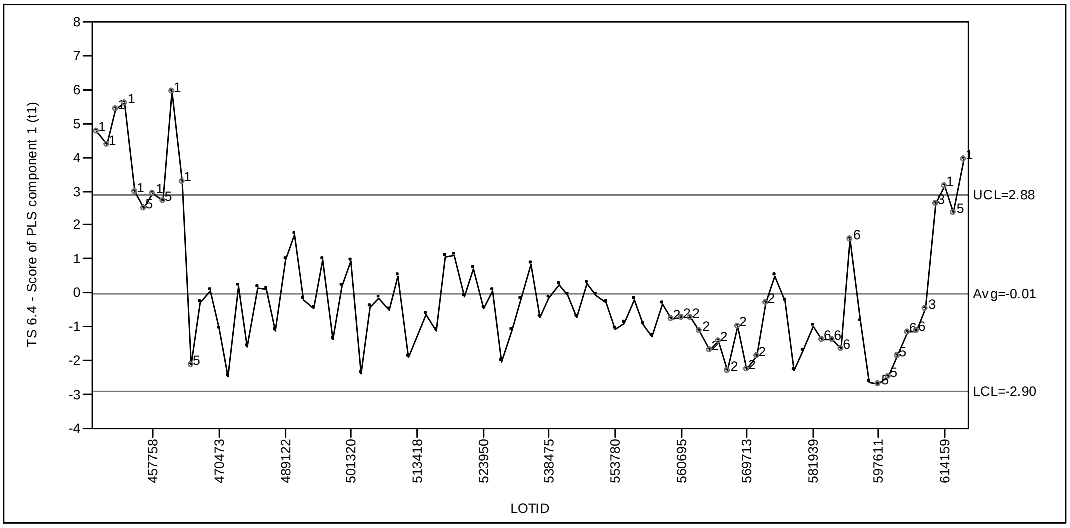


Figure 22. Multivariate control chart with score of first PLS component of models TS 6.4.

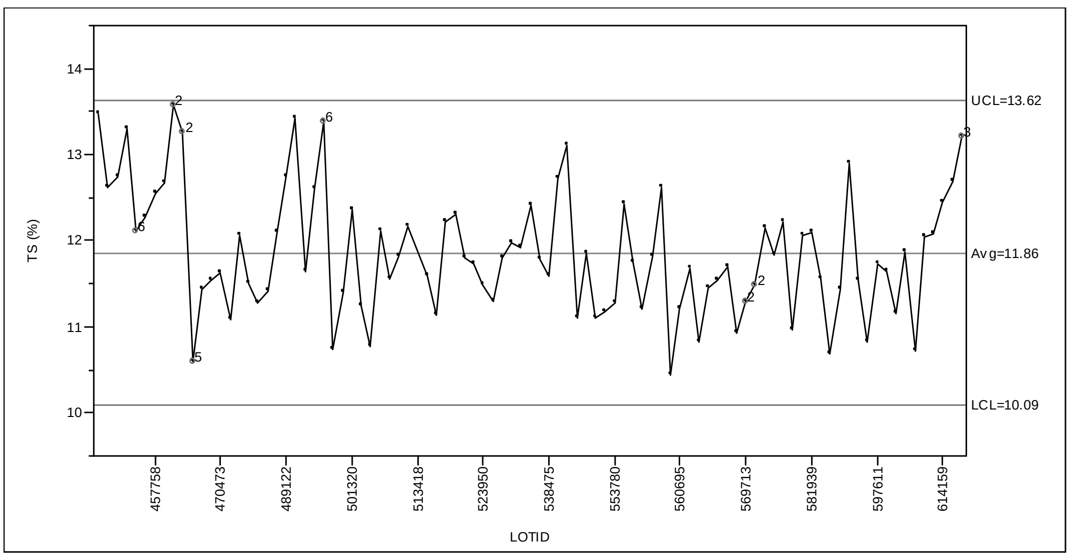


Figure 23. X-Individual control chart on thickness swelling 6.4 mm.

3.6 Further possible analysis

Improvement in the analysis can be done by further analysis of existing data, better statistical methods, and generation of new data.

3.6.1 Improvements with existing data

With improved processing knowledge additional models can be adopted. Potential is seen in a more detailed examination of the material flows, for example the partial material flows of S1 and S2.

Modes can be improved by considering polynomial effects and other interactions between existing variables. The subjective selection is problematic when it comes to understanding which variables have certain effects. However, these efforts are expected to lead to minor improvements in the models only, since variables behave approximately linear in the small observation range (Lobenhoffer, 1990). Computationally intensive algorithms could be developed and applied to try to find optimum combinations of polynomial effect of different orders and time laggings of existing variables.

Furthermore, transformations of skewed data (e.g. forth root transformation, log transformation) could improve the results. Exclusion of variables and observations with too much missing data, or variables with too little values different from the median (especially applies for the wood type fractions) could bring improvements too (Esbensen, 2002).

Models for each single position across the board width revealed that they were slightly better than models for mean values only (even with fewer significant variables only). Furthermore, not documenting attempts of modeling with measures of dispersion (e.g. variation) of the single values of a board showed very promising results.

3.6.2 Better statistical methods

Better calibration methods with the existing data base could be used, e.g. improved PLS algorithms (or extensions such as O-PLS) or possibly Structural Equation Modeling (SEM) in combination with path modeling and discriminant analysis (Montgomery, 2000; Ringle et al., 2005; André et al., 2008). Quantile regression (QR) could add insight into causes that lead to values in low or high quantiles of response variables (Young et al., 2008). Neural network analysis (NNA) and genetic algorithms (GA) could be used for variable selection as a step before actual modeling (André et al., 2008). Furthermore, multivariate time series analysis could be performed on the data set. Completion of missing data can be improved by using expectation-maximation (EM) algorithms (Dempster et al., 1977).

3.6.3 Use and collection of further data

Use of so far unused data (e.g. press temperatures) and additional existing data from the process control system may bring significant improvements as well. Certain issues are fiber processing, or better use of this existing information.

Gradients of properties across the board width could be analyzed for other board quality properties too, such as the medium density (MD). Furthermore, a comparison of the vertical density profile (VDP) on different positions across the board width might give further interesting results.

Another approach would be the collection of additional data of the raw material and the process, while focusing on certain questions. These could be fiber characteristics such as coarseness, fiber length or fines content. Good prediction models were developed and shown by Li Shi et al. (2006). Further studies on penetration of resin into the fiber material, stimulated by pressure, steam and heat, could gain further insight in resination, and improve the prediction quality of models based on this information (Cyr et al., 2008).

Information on the change of chemical composition as effect of age and processing conditions of the raw material could be linked to final board properties (Kelley et al., 2005). NIR spectroscopy can be used for MDF production to characterize properties such as density, mechanical properties, moisture contents, fiber length distribution (So et al., 2004). Another approach is intentional experimental design with the plant, where certain parameters are varied on purpose. However, this approach is cost intensive.

4 Conclusion

4.1 Validity of the hypotheses

Hypothesis H1 is accepted. Multivariate data analysis, especially PLSR, was shown to be a suitable statistical tool for analyzing industrial process data from a HDF production. When using more than one board property as response variable, care has to be taken with the interpretation and weighing of the results according industrial importance of the board properties.

Using board mean values (H1.1) has shown to give good results. Board mean values are also generally used for developing MDF/HDF models in literature available.

Commonly, models showed better results when the response variable exhibited high variation. Single values across the board width often show higher variation, and thus improved models, supporting H1.2. Developing models with dispersion parameters on single values, (which although showed plausible results) was only tried experimentally; thus slightly supporting H1.3.

Hypothesis H2 is accepted. The influence of the raw material on the models was detectable and estimated by 21% in average for all developed models. The fractions of raw material influence in detail were IB 28%, SS 22%, MOR 18%, MOE 12 %, TS 4%, PF 29% and RF 36 %.

Hypothesis H3 is accepted either. For TS with a board thickness of 6.4 mm, it was shown, that using multivariate control charts on all variables, events of special causes can be identified successfully. However, it has to be stated that selection of appropriate control charts and variables to chart is a crucial task and requires some intuition.

4.2 Further analysis

Based on the results obtained, the following improvements in the analysis could be performed:

- Minor improvements in modeling with the existing data could be performed considering polynomial effects of the variables, and transformation of skewed data.
- Models for each single position not only for IB, but also for all product properties could add additional insight into the process, with special regard on its position within the board.
- Improved statistical methods, e.g. combining PLSR with a previous step of variable selection using genetic algorithms (GA) could be performed.
- The plant in use recorded more data than was used for the analysis. Using all further existing data, e.g. data collected by the plants process system, vertical density profiles (VDP) would certainly bring further knowledge about the process.
- Collecting new data (e.g. more exact information about partial material flows, raw material data by online spectroscopy) could be performed if a completely new data set was to be used.

5 Summary

Medium-density fiberboards (MDF) are basically produced by addition of usually synthetic resin to lingo-cellulosic fibers, and subsequent application of temperature and pressure. For MDF with a raw density ≥ 800 kg/m³, the term High-density fiberboard (HDF) has become common.

Boards are produced with a certain variation in properties. Quality control is done by taking samples from the produced boards and analyzing them in the laboratory. Thus, this property data is available only with some delay of time. This may cause unnecessary waste of raw material and unwanted financial losses. Major targets in Quality Management (QM) and Six Sigma are to minimize product property variation, and also minimize deviations from targeted means. The main focus of this work is to study the interaction of the raw material, the process and the board quality parameters, with the target to determine the key sources of variation.

In this work, samples were taken in 2008 starting in calendar week 2, completing in week 51. Samples were routinely taken each day during the early shift. Only data from boards manufactured with the two most frequent recipes, which differed only in the use of dyes, were taken for analysis. The plant produces several HDF thicknesses. The main thicknesses are 6.4 mm and 7.4 mm, respectively, besides some other thicknesses. All thicknesses were kept in the dataset. In total, 251 samples (observations) were used for the analysis. Data was available from different inhomogeneous sources, which required the development of a combined database in order to simplify data selection. All data was aligned in time-order as accurately as possible, thus taking time-lags into consideration. In total, the dataset consists of 245 variables. The following types of variables are contained in the dataset:

- Board quality parameters, determined by destructive lab testing. They consist of Internal Bond Strength (IB), Surface Soundness (SS), Modulus of Rupture (MOR), Modulus of Elasticity (MOR) and Thickness Swelling 24 h (TS), determined by industrial test standards.
- A multiplicity of variables measured by online-sensors at each stage of the process, and collected by the plant's process control system
- Subjective variables on evaluation of the process and formaldehyde perception, recorded by the plant's staff.
- Ion chromatograph analyses on the fiber material
- Wet chemistry data of the raw material and processed fiber, containing amount of hemicelluloses/extractives determined by methanolysis.
- Wood types of the chips
- Buffer capacity and pH values of the fiber material

With the thickness of 7.4 mm, the board property mean values of the quality parameters show significant correlations ($p<0.05$), except between TS and IB.

The in-plane gradients for board properties perpendicular to the production direction (along board width) show that the quality drops in the center of the board, thus having significant better quality properties at the edges. For IB, the technologically most important property from an industrial view, the effect was compensated by an altered press program at the edges only, in order to keep the IB uniform predominantly. TS tends to behave differently however and does not drop in the center of the board.

Statistical models using PLSR were developed for board quality parameters (IB, SS, MOR, MOE, TS), resin fraction, and press factor as response variables. Since there are two dominating board thicknesses, specific models are developed for each major thickness as well. Another separation was made by distinguishing between *raw material* and *process* variables, in order to focus on these variable groups only. Thus, for one response variable, at least five models were developed. Models were validated using full cross validation. The most important variables were determined by two methods: (1) determining the frequency of occurrence with variables from the models, and (2) by weighting the regression coefficients B_w of selected models by industrial importance of the board properties.

Models for IB and SS show low cross-validated coefficients of determination (Q^2). This is not necessarily surprising, because the plant is quite well optimized for keeping IB and SS low, thus resulting in a low coefficients of variation (CV) if these two quality parameters. Models for the IB with single values show Q^2 up to 32%, the NRMSEP (mean normalized root mean square error of prediction) however did not improve in comparison to the models of IB. Models for MOR and MOE have a Q^2 of up to 50% and 42%, respectively. Models for MOE and TS for certain thicknesses show huge differences in Q^2 compared to models including all thicknesses. This can be explained by a strong thickness effect for MOE and TS. Surprisingly, despite the strong correlation between MOE and MOR, this thickness effect did not apply to MOR. Models with PF and RF are generally satisfying. They benefit from variables that indicated the material flows, which more or less directly influence the response variables. Table 27 summarizes the range of performance statistics on all multivariate models.

Table 27. Summary of performance statistics of all models.

Property	**RMSEP**	**Mean normalized RMSEP (=NRMSEP) (%)**	**Cross-validated R^2 (=Q^2) (%)**
Internal Bond Strength (IB)	0.1019 – 0.1703 N/mm²	5.75 – 6.02	2.73 – 13.44
IB single values	0.1425 – 0.1700 N/mm²	7.71 – 9.74	0.52 – 32.38
Surface Soundness (SS)	0.1038 – 0.1229 N/mm²	5.09 – 5.88	6.11 – 23.62
Modulus of Rupture (MOR)	1.6142 – 2.2619 N/mm²	3.34 – 4.74	10.33 – 50.14
Modulus of Elasticity (MOE)	158.03 – 219.74 N/mm²	3.32 – 4.81	5.47 – 42.06
Thickness Swelling (TS)	0.5222 – 1.2757 %	4.40 – 11.87	2.88 – 79.42
Press factor (PF)	0.1992 – 0.5360 s/mm	2.71 – 7.16	36.67 – 91.85
Resin fraction (RF)	3.832 – 6.2736 %	3.73 – 6.09	38.16 – 77.33

Generally, with models for *raw material* and *process* variables separately, the process variables accounted for the largest fraction of Q^2. With the information of the models, the influence of *raw material* variables to the board quality parameters is estimated at an average of 21% on Q^2.

The models for the continuous press reveal for IB that the first press systems have the most influence. For SS, the focus is on latter systems. MOR and MOE are influenced positively by a similar press system setting; whereas MOE has an additional focus on the final systems of the press. Remarkably, the very first system has positive influence on MOR and MOE, but a negative one on TS at the same time. Furthermore, pressure in the final systems caused higher TS.

A weighting of the regression coefficients of selected models according to technological relevance was performed. The relevance of the five main product quality parameters were assumed as follows: IB is the most important technical characteristic, and SS is as important as all remaining properties. MOR and MOE are overall as important as TS. The ten most important variables contain information of 1) dummy variables of certain months, favoring April, July and October, 2) *positive* impact on thicker boards, high perceptibility of formaldehyde, high formiate content of chips, and high buffer capacity of the glued fiber 3) *negative* impact of high bulk density; high fill level of the digester, high and amount of extractives of sawdust.

Example control charts for TS reveal that multivariate control charts on all available variables have better capabilities to detect changes in the process than control charting data of the board property TS only.

Acknowledgement

The findings and conclusions in this thesis are for a HDF plant of Fritz Egger GmbH & Co. The author would like to thank the company for offering him the special opportunity of working on the industrial data set.

The author would like to thank Rupert Wimmer (Georg-August-Universität Göttingen, Germany) for initiating a research project on process analytics in MDF production[14], and for supervising this work. The author would like to acknowledge all direct and indirect project members involved, especially Martin Weigl (Wood K plus, Austria) who supported with regular discussions, the input of ideas, and the analysis of wood types. Furthermore, acknowledgements go to Bernhard Spangl for regular help in questions about statistics (University of Applied Life Sciences, Vienna, Austria), Thomas Ruhnau (EGGER Holzwerkstoffe Wismar, Germany), Roland Mitter and Thomas Kuncinger (Fritz Egger GmbH & Co., Austria), Ralf Borchers (EGGER Holzwerkstoffe Wismar, Germany), Ulrich Müller (Wood K plus, Austria) and the Wood K plus chemistry team with Susanne Vrchoticky, Thomas Ters, Sonja Leder, Philip Heissl, Leopold Mahn that carried out most chemistry lab analysis. The analysis of samples was partly supported by the Austrian Government and the federal Government of Upper Austria, Lower Austria and Carinthia.

Furthermore, grateful acknowledgements go to Tim Young and Nicolas André (University of Tennessee, Forest Products Center, Knoxville/TN, USA) who have made large contributions to the statistical methods used in this work.

Thanks to Gerry McDermott and John Hodgshon for linguistically proofreading this work.

The author would also like to thank all the friends, especially those from *The Wickenburg*, who gave me an environment of great joy during the last some years, which was absolutely crucial for accomplishing this work. Very special thanks go to my girlfriend Heidi Rinta-Lusa, for giving me energy and at the same time freedom within the last years. Without her, my work would have never been carried out.

[14] Wood K plus: „Egger MDF 4.2 / Produkt- und Prozessoptimierung in der MDF Produktion“;

List of Tables, Figures and References

List of Tables

List of Figures

List of References

AMEMIYA, T. (1977) The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model. *Econometrica,* 45**,** 955-968.

ANDRÉ, N., CHO, H. W., BAEK, S. H., JEONG, M. K. & YOUNG, T. M. (2008) Prediction of internal bond strength in a medium density fiberboard process using multivariate statistical methods and variable selection. *Wood Science and Technology,* 42**,** 521-534, English.

BERNARDY, G. & SCHERFF, B. (1997) Prozessmodellierung führt zur Online-Qualitätskontrolle und Prozessoptimierung bei der Span- und Faserplattenproduktion. *Holz als Roh- und Werkstoff,* 55**,** 133-140, German.

BHUPINDER, S. D. & JOHN, F. M. (1997) Improved PLS algorithms. *Journal of Chemometrics,* 11**,** 73-85.

BOX, G. E. P. (1966) Use and Abuse of Regression. *Technometrics,* 8**,** 625-629.

CAMO (2006a) The Unscrambler - Method References, CAMO Inc., CAMO Software AS, http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Method%20References.pdf, 07/18/2009.

CAMO (2006b) The Unscrambler - Methods, CAMO Inc., CAMO Software AS, http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Methods.pdf, 07/22/2009.

CAMO (2008) The Unscrambler. 9.8 ed., http://www.camo.com.

CARLSSON, J., MATHIASSON, M. & GOEMINNE, J. (2002) On-line monitoring of fiber quality in the MDF process. *6th European Panel Products Symposium.* Llandudno, Wales.

CEN (1992) EN 120: Wood based panels. Determination of formaldehyde content. Extraction method called the perforator method.

CEN (1993a) EN 310: Wood-based panels - Determination of modulus of elasticity in bending and of bending strength.

CEN (1993b) EN 319: Particleboards and fibreboards - Determination of tensile strength perpendicular to the plane of the board.

CEN (1993c) EN 320: Fibreboards - Determination of resistance to axial withdrawal of screws.

CEN (1993d) EN 323: Wood-based panels - Determination of density.

CEN (1993e) EN 382-1: Fibreboards - Determination of surface absorption - Part 1: Test method for dry process fibreboards.

CEN (1994) EN 326-1: Wood-based panels - Sampling, cutting and inspection - Part 1: Sampling and cutting of test pieces and expression of test results.

CEN (1999) EN 316: Wood fibreboards - Definition, classification and symbols.

CEN (2000a) EN 326-2: Wood-based panels - Sampling, cutting and inspection - Part 2: Initial type testing and factory production control.

CEN (2000b) EN 13329: Laminate floor coverings. Specifications, requirements and test methods.

CEN (2002a) EN 311: Wood-based panels - Surface soundness - Test method.

CEN (2002b) EN 317: Particleboards and fibreboards - Determination of swelling in thickness after immersion in water.

CEN (2002c) EN 318: Wood based panels - Determination of dimensional changes associated with changes in relative humidity.

CEN (2006) EN 622-5: Fibreboards. Specifications. Requirements for dry process boards (MDF).

CHAPMAN, K. (2006) Wood-based panels: particleboard, fibreboards and oriented strand board. *Primary Wood Processing.*

CLAPP, N. E., YOUNG, T. M. & GUESS, F. M. (2008) Predictive modeling the internal bond of medium density fiberboard using a modified principal component analysis. *Forest Products Journal,* 58**,** 49-55.

CYR, P.-L., RIEDL, B. & WANG, X.-M. (2008) Investigation of Urea-Melamine-Formaldehyde (UMF) resin penetration in Medium-Density Fiberboard (MDF) by High Resolution Confocal Laser Scanning Microscopy. *Holz als Roh- und Werkstoff,* 66**,** 129-134.

DE JONG, S. (1993) SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems,* 18**,** 251-263.

DEMPSTER, A. P., M., L. N. & B., R. D. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society,* 39**,** 1-38.

DEPPE, H.-J. & ERNST, K. (1996) *MDF - Mitteldichte Faserplatten,* Leinfelden-Echterdingen, DRW-Verlag, German.

DOLEZEL-HORWATH, E., HUTTER, T., KESSLER, R. & WIMMER, R. (2005) Feedback and feedforward control of wet-processed hardboard production using spectroscopy and chemometric modelling. *Analytica Chimica Acta,* 544**,** 47-59.

EIGENVECTOR_RESEARCH (2009) PLS_Toolbox. http://software.eigenvector.com/.

ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., TRYGG, J., WIKSTRÖM, C. & WOLD, S. (2001a) *Multi- and Megavariate Data Analysis Part I: Basic Principles and Applications*.

ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., TRYGG, J., WIKSTRÖM, C. & WOLD, S. (2001b) *Multi- and Megavariate Data Analysis Part II: Advanced Applications and Method Extensions*.

ESBENSEN, K. H., GUYOT, D., WESTAD, F., HOUMØLLER L. P. (2002) *Multivariate Data Analysis - in practice*.

EVANS, J. R. & LINDSAY, W. M. (2002) *The Management and Control of Quality,* Cincinnati, Ohio, South-Western.

FAHRMEIR, L., KNEIB, T. & LANG, S. (2007) *Regression: Modelle, Methoden und Anwendungen,* Berlin Heidelberg, Springer-Verlag, German.

FARDIM, P., HOLMBOM, B., IVASKA, A., KARHU, J., MORTHA, G. & LAINE, J. E. (2003) Critical comparison and validation of methods for determination of anionic groups in pulp fibres. *Nord. Pulp Pap. Res. J.,* 17**,** 346-351.

FETEN, G., ALMØY, T. & AASTVEIT, A. H. (2005) Prediction of Missing Values in Microarray and Use of Mixed Models to Evaluate the Predictors. *Statistical Applications in Genetics and Molecular Biology,* 4.

FUCHS, C. & KENETT, R. S. (1998) *Multivariate Quality Control - Theory and Applications,* New York, Marcel Dekker.

GREUBEL, D. (1999a) Practical experiences with a process simulation models in particleboard and MDF production. *2nd European Wood-Based Panel Symp.* Hanover, Germany.

GREUBEL, D. (1999b) Untersuchungen von Methoden zur Qualitätssicherung durch Prozessmodelle. Abschlussbericht AiF 10508 N. Braunschweig, Germany, Internationaler Verein für Technische Holzfragen, German.

HAN, J. & KAMBER, M. (2006) *Data Mining. Concepts and Techniques*, Elsevier.

HASENER, J. (2004) Statistische Methoden der industriellen Prozessmodellierung zur Echtzeitqualitätskontrolle am Beispiel einer kontinuierlichen Produktion von Faserplatten. *Univsersität Hamburg, Fachbereich Biologie.* Hamburg, Universität Hamburg, German.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.

HOADLEY, R. B. (1990) *Identifying Wood: Accurate Results with Simple Tools,* Newtown, CT, Taunton Press.

ISO (1976) ISO 3340: Fibre building boards -- Determination of sand content.

KELLEY, S., ELDER, T. & GROOM, L. (2005) Changes in the Chemical Composition and Spectroscopy of Loblolly Pine Medium Density Fiberboard Furnish as a Function of age and Refining Pressure. *Wood and Fiber Science,* 37**,** 14-22.

KESSLER, W. (2006) *Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik* Wiley-VCH, German.

KUSS, H. (2000) Holzwerkstoffplatten - Leitfaden für die Anwendung der Europäischen Normen, Zusatz zu EN 312. Vienna, Österreichisches Normungsinstitut, German.

KUTNER, M. H., NACHTSHEIM, C. & NETER, J. (2004) *Applied Linear Regression Models*.

LI SHI, J., ZHANG, S. Y. & RIEDL, B. (2006) Multivariate modeling of MDF panel properties in relation to wood fiber characteristics. *Holzforschung,* 60**,** 285-293.

LOBENHOFFER (1990) Qualitätsbedingte Regelung eines Spanplattenformstrangs. Gött, Georg-August-Universität Göttingen, German.

MARCOULIDES, G. A., CHIN, W. W. & SAUNDERS, C. (2009) A Critical Look at Partial Least Squares Modeling. *MIS Quaterly,* 33**,** 171-175.

MARTENS, H. & MARTENS, M. (2000) Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference,* 11**,** 5-16.

MCDONALD, A. G., CLARE, A. B. & MEDER, A. R. (2006) Characterization of Cellulosic Cell Wall. IN STOKKE, D. & GROOM, L. (Eds.). Blackwell Pub Professional.

MICROSOFT (2007) Microsoft Access 2007. http://office.microsoft.com/access.

MILTON, J. S. & ARNOLD, J. C. (2003) *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*, McGraw-Hill Science Engineering.

MONTGOMERY, D. C. (2000) *Design and Analysis of Experiments*, Wiley.

MOSTELLER, F. & TUKEY, J. W. (1977) *Data Analysis and Regression - a second course in statistics*, Addison-Wesley.

MYERS, R. H. (1986) *Classical and Modern Regression with Applications,* Boston, Mass., Duxbury Press.

MYSQL_AB (2009) MySQL. 5.1 ed.

NELSON, L. S. (1984) The Shewhart Control Chart - Tests for Special Causes. *Journal of Quality Technology,* 16**,** 237-239.

NOMIKOS, P. & MACGREGOR, J. F. (1995) Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics,* 37**,** 41-59.

NORDIN, S., JOHNSSON, B. & ENGSTRÖM, B. (2002) Method for controlling a process for the production of a cellulose fiber containing product. IN PATENT, U. S. (Ed.). USA, Akzo Nobel N.V.

RÄNNER, S., GELADI, P., LINDGREN, F. & WOLD, S. (1995) A PLS Kernel Algorithm for data sets with many variables and few objects. Part II : cross-validation, missing data and examples. *Journal of Chemometrics,* 9**,** 459-470.

RIALS, T., KELLEY, S. & SO, C.-L. (2002) Use of Advanced Spectroscopic Techniques for Predicting the Mechanical Properties of Wood Composites. *Wood and Fiber Science,* 34**,** 398-407.

RINGLE, C. M., S., W. & WILL, S. (2005) SmartPLS 2.0 (M3) Beta. Hamburg.

ROSNER, B. A. (2009) *Fundamentals of biostatistics*, Duxbury Press.

RUBIN, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys,* New York, J. Wiley & Sons.

SAS_INSTUTUTE (2008) SAS. 9.2 ed., http://www.sas.com.

SCHAFER, J. L. (1997) *Analysis of Incomplete Multivariate Data,* London, Chapman & Hall.

SCHNEIDER, T. (1999) Untersuchungen über den Einfluß von Aufschlußbedingungen des Holzes und der Faserstofftrocknung auf die Eigenschaften von mitteldichten Faserplatten (MDF). *Fakultät für Forstwissenscahften und Waldökologie.* Göttingen, Georg-August-Universität Göttingen.

SHEWHART, W. & DEMING, W. (1986) *Statistical method from the viewpoint of quality control*, Dover Publications.

SIEMPELKAMP (2009) Siempelkamp Prod-IQ/SPOC. http://www.siempelkamp.de/Prod-IQ.584.0.html, 8/14/2009.

SO, C.-L., VIA, B. K., GROOM, L. H., SCHIMLECK, L. R., SHUPE, T. F., KELLEY, S. S. & RIALS, T. G. (2004) Near Infrared Spectroscopy in the Forest Products Industry. *Forest Products Journal,* 54**,** 6-16.

SPSS_INC (2008) SPSS. 17.0 ed., http://www.spss.com.

SUNDBERG, A., SUNDBERG, K., LILLANDT, C. & HOLMBOM, B. (1996) Determination of hemicelluloses and pectins in wood and pulp fibres by acid methanolysis and gas chromatography. *Nordic Pulp and Paper Research Journal,* 11**,** 216-219.

SUNDQVIST, B., KARLSSON, O. & WESTERMARK, U. (2006) Determination of formic-acid and acetic acid concentrations formed during hydrothermal treatment of birch wood and its relation to colour, strength and hardness. *Wood Science and Technology,* 40**,** 549-561.

TENENHAUS, M. (1998) *La régression PLS,* Paris, French.

THE_MATHWORKS (2009) MATLAB. R2009a ed., http://www.mathworks.com.

TUKEY, J. W. (1977) *Exploratory data analysis,* Massachusetts, Addision-Wesley.

UMETRICS (2009) SIMCA. 12.0.1 ed., www.umetrics.com.

WAKELING, I. N. & MORRIS, J. J. (1993) A test of significance for partial least squares regression. *Journal of Chemometrics,* 7**,** 291-304.

WALKER, J. C. F. (2006) *Primary wood processing. Principles and Practice*, Springer.

WEIGL, M., VRCHOTICKY, S. & TERS, T. (2009) Chemical Degradation of Cell Wall Polymers due to Industrial HDF Fibre Production. *IPPS.* Nantes, France.

WHEELER, D. J. (1993) *Understanding Variation - The Key to Managing Chaos,* Knoxville, Tennessee, SPC Press.

WHEELER, D. J. (2004) *Advanced Topics in Statistical Process Control - The Power of Shewharts's Charts,* Knoxville, Tennessee, SPC Press.

WHEELER, D. J. & CHAMBERS, D. S. (1992) *Understanding Statistical Process Control,* Knoxville, Tennessee, SPC Press.

WIDSTEN, P., LAINE, J. E., QVINTUS-LEINO, P. & TUOMINEN, S. (2001) Effect of High-Temperature Fiberization on the Chemical Structure of Softwood. *Journal of Wood Chemistry and Technology,* 21**,** 227-245.

WIDSTEN, P., LAINE, J. E., QVINTUS-LEINO, P. & TUOMINEN, S. (2002) Effect of High-Temperature Defibration on the Chemical Structure of Hardwood. *Holzforschung,* 56**,** 51-59.

WITTEN, I. H. (2005) *Data Mining. Practical Machine Learning Tools and Techniques*, Elsevier.

WOLD, H. (1966) *Estimation of principal components and related models by iterative least squares,* New York, Academic Press.

WOLD, S., SJÖSTRÖM, M. & ERIKSSON, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems,* 58**,** 109-130.

XING, C., RIEDL, B., CLOUTIER, A. & HE, G. (2004) The effect of urea-formaldehyde resin pre-cure on the internal bond of medium density fiberboard. *Holz als Roh-und Werkstoff,* 62**,** 439-444.

XING, C., ZHANG, S., DENG, J., RIEDL, B. & CLOUTIER, A. (2006) Medium-density fiberboard performance as affected by wood fiber acidity, bulk density, and size distribution. *Wood Science and Technology,* 40**,** 637-646.

YOUNG, T. M., SHAFFER, L. B., GUESS, F. M., BENSMAIL, H. & LEON, R. V. (2008) A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard. *Forest Products Journal,* 58**,** 39-48, English.

Appendix

A. Abbreviations and Terms

A.1 Abbreviations used in this document and within variable names

Abbreviation	Description
3SLS	Three-Stage Least Squares
B	Regression coefficient
B_w	Weighted regression coefficient
C	Conifer (softwood)
CDA	Confirmatory Data Analysis
CEN	European Committee for Standardization (French: *Comité Européen de Normalisation*)
CI	Confidence Interval
CV	Cross validation
CV	Coefficient of variation
DoE	Design of Experiment
EDA	Exploratory Data Analysis
EM	Expectation Maximation
EN	European Norm
F	Fiber
FTIR	Fourier transform infrared (spectroscopy)
GA	Genetic Algorithm
GC	Gas chromatography
HDF	High-density fiberboard
HS	Wood chips (from German: *Hackschnitzel*)
IB	Internal Bond Strength
IC	Ion (exchange) chromatography
ISO	International Organization for Standardization
LCL	Lower Control Limit
LSL	Lower Specification Limit
MC	Moisture content
MD	Medium density
MDF	Medium-density fiberboard
MI	Multiple Imputation
MLR	Multiple Linear Regression
MOE	Modulus of Elasticity
MOR	Modulus of Rupture, bending strength
NC	Non-conifer (hardwood)
NIPALS	Nonlinear Estimation by Iterative Partial Least Squares
NN	Neural Networks
NRMSEP	Mean normalized RMSEP
O-PLS	Orthogonal Partial Least Squares
P	Prefix for partial material flow in the fiber processing stage (Figure 3)

PCA	Principal Component Analysis
PCR	Principal Component Regression
PF	Press factor
PLC	Programmable Logic Controller
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PRESS	Predictive residual sum of squares
PV	Press velocity
Q^2	Cross-validated R^2
QM	Quality Management
QR	Quantile Regression
R	Raw material
r	Pearson correlation coefficient
R^2	Coefficient of determination, explained Y-variance
RBF	Radial basic function
RF	Resin fraction
RMSE	Root Mean Square Error (of Calibration)
RMSEP	Root Mean Square Error of Prediction
S	Prefix for partial material flow in the supply stage (Figure 3)
SEM	Structural Equation Modeling
SPOC	Statistical Process Optimization and Control; solution, that is a module of the production management system Prod-IQ® of Siempelkamp
SPPCA	Supervised probabilistic principal component analysis
SQL	Structured Query Language
SS	Surface soundness
SS	Sawdust (from German: *Sägespäne*)
SS	Sum of squares
TS	Thickness swelling 24 hours
UCL	Upper Control Limit
UMF	Urea-melamine-formaldehyde resin
USL	Upper Specification Limit
VBA	Visual Basic for Applications
VDP	Vertical Density Profile
WGT	Postfix used in variables that are weighted by the partial material flow fraction
WU	Water uptake
XRD	X-ray diffraction

A.2 Synonyms of terms in regression

Term[15]	**Synonym**
Observation	Object, sample (general); row, data row (for databases)
Variable	field (for databases); factor, parameter
X	Predictor variable, exogenous variable (general); Independent variable (for MLR)
Y	Response variable, endogenous variable (general); Dependent variable (for MLR)
Component	Principal Component, PC (for PCA/PCR); PLS Component (for PLS); latent variable, latent dimension, dimension (general)

B. Descriptive Statistics of the Variables

[15] The table is based on Esbensen (2002) and CAMO (2006b)

B.1 Descriptive Statistics of all observations

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign.[19]
-	LOTID	-	lot ID			process general	-				251	0			
-	DATE_PRODUCTION	-	date of production, incorporated into variable 8-23			process general	uncontrolled				251	0			
-	TIME_SAMPLING	-	sampling time			process general	uncontrolled				251	0			
-	SHIFT_NO	-	shift number, incorporated into variables 31-35			process general	controlled				251	0			
-	RESIN_TANK_NO	-	resin tank number, incorporated into variables 24-30			bonding	controlled				251	0			
1	NOMINAL_THICKNESS	mm	nominal thickness		proc	input requirement	controlled				251	0			
2	RECIPE_NO	-	recipe number (600 or 621)		proc	input requirement	controlled				251	0			
3	MOIST_OUTSIDE	°C	humidity/moisture outside		proc	process general	uncontrolled	10.38	74.03	14.03	217	34	0.37	-0.47	
4	TEMP_OUTSIDE	°C	temperature outside		proc	process general	uncontrolled	5.20	11.72		217	34	-0.98	0.04	
5	REF_IN_USE	-	refiner in use (0=both on, 1=sawdust refiner on P2 off)		proc	process general	controlled				250	1			
6	PERCEP_FORMALDEHYDE	-	perceptibility formaldehyde (1=low, 2=medium, 3=high)		proc	process general	intermediate	0.55	1.41	38.60	222	29	-0.35	0.85	
7	EVAL_PRODUCTION	-	evaluation production (1=bad, 2=average, 3=good)		proc	process general	intermediate	0.28	1.97	14.16	217	34	9.75	-1.16	
8	MONTH_01	-	Production in January (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
9	MONTH_02	-	Production in February (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
10	MONTH_03	-	Production in March (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
11	MONTH_04	-	Production in April (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
12	MONTH_05	-	Production in May (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
13	MONTH_06	-	Production in June (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
14	MONTH_07	-	Production in July (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
15	MONTH_08	-	Production in August (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
16	MONTH_09	-	Production in September (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
17	MONTH_10	-	Production in October (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
18	MONTH_11	-	Production in November (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
19	MONTH_12	-	Production in December (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
20	SEASON_SPRING	-	Production in spring (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
21	SEASON_SUMMER	-	Production in summer (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
22	SEASON_FALL	-	Production in fall (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
23	SEASON_WINTER	-	Production in winter (1 if applicable, else 0)		proc	process general	uncontrolled				251	0			
24	RESIN_TANK01	-	Resin tank 1 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
25	RESIN_TANK02	-	Resin tank 2 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			

[16] Indicates mapping of the variable to the partial material flows as given in the process flow diagram
[17] Distinction of variables as proposed by SIMCA-P+® (Umetrics, 2009)
[18] The CV is given for variables where this calculation is applicable, which is not the case for pH-values and temperatures
[19] Indicates if variable follow a normal distribution using a Shapiro-Wilk test (p>0.05 indicated by **, 0.01≤p≤0.05 by *)

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign. [19]
26	RESIN_TANK03	-	Resin tank 3 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
27	RESIN_TANK04	-	Resin tank 4 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
28	RESIN_TANK05	-	Resin tank 5 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
29	RESIN_RES01	-	Resin tank reserve 1 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
30	RESIN_RES02	-	Resin tank reserve 2 in use (1 if applicable, else 0)		proc	bonding	controlled				251	0			
31	SHIFT1	-	Production by shift 1 (1 if applicable, else 0)		proc	process general	controlled				251	0			
32	SHIFT2	-	Production by shift 2 (1 if applicable, else 0)		proc	process general	controlled				251	0			
33	SHIFT3	-	Production by shift 3 (1 if applicable, else 0)		proc	process general	controlled				251	0			
34	SHIFT4	-	Production by shift 4 (1 if applicable, else 0)		proc	process general	controlled				251	0			
35	SHIFT5	-	Production by shift 5 (1 if applicable, else 0)		proc	process general	controlled				251	0			
36	HE_SS_FRACT	%	raw material fraction sawdust	S3	proc	process general	controlled	3.56	26.46	13.45	237	14	2.94	-1.07	
37	WOOD_INPUT_TOTAL	t/h	total wood input (usually without sawdust)	P1	proc	process general	controlled	4.26	41.33	10.30	250	1	-1.54	-0.11	
38	FIB_PROC_LINE2_FRACT	%	material flow fiber processing line 2	P2	proc	process general	controlled	11.23	19.57	57.39	247	4	-0.65	-1.06	
39	RHS_C_FIR	%	percentage raw material chips conifer fir	S1/2	raw	raw material	raw material	2.95	0.48	612.86	239	12	138.76	10.88	
40	RHS_C_SPRUCE	%	percentage raw material chips conifer spruce	S1/2	raw	raw material	raw material	19.21	37.02	51.89	239	12	-0.58	0.24	
41	RHS_C_PINE	%	percentage raw material chips conifer pine	S1/2	raw	raw material	raw material	21.05	47.92	43.93	239	12	-0.40	0.25	
42	RHS_C_LARCH	%	percentage raw material chips conifer larch	S1/2	raw	raw material	raw material	4.56	0.97	469.96	239	12	127.01	10.12	
43	RHS_C_TOTAL	%	percentage raw material chips conifer total	S1/2	raw	raw material	raw material	17.38	86.39	20.11	239	12	2.88	-1.71	
44	RHS_NC_ALDER	%	percentage raw material chips non conifer alder	S1/2	raw	raw material	raw material	0.96	0.15	656.34	239	12	58.56	7.29	
45	RHS_NC_ASH	%	percentage raw material chips non conifer ash	S1/2	raw	raw material	raw material	8.50	3.02	281.29	239	12	13.52	3.53	
46	RHS_NC_BEECH	%	percentage raw material chips non conifer beech	S1/2	raw	raw material	raw material	7.73	2.22	348.23	239	12	32.40	5.30	
47	RHS_NC_CHERRY	%	percentage raw material chips non conifer cherry	S1/2	raw	raw material	raw material	0.91	0.13	723.40	239	12	71.54	8.07	
48	RHS_NC_ELM	%	percentage raw material chips non conifer elm	S1/2	raw	raw material	raw material	0.46	0.04	1090.8	239	12	116.97	10.86	
49	RHS_NC_HORNBEAM	%	percentage raw material chips non conifer hornbeam	S1/2	raw	raw material	raw material	0.68	0.04	1545.9	239	12	239.00	15.46	
50	RHS_NC_LOCUST	%	percentage raw material chips non conifer locust	S1/2	raw	raw material	raw material	0.46	0.04	1090.8	239	12	116.97	10.86	
51	RHS_NC_MAPLE	%	percentage raw material chips non conifer maple	S1/2	raw	raw material	raw material	2.11	0.63	335.60	239	12	20.63	4.17	
52	RHS_NC_OAK	%	percentage raw material chips non conifer oak	S1/2	raw	raw material	raw material	0.90	0.17	538.48	239	12	25.46	5.22	
53	RHS_NC_ASPEN	%	percentage raw material chips non conifer aspen	S1/2	raw	raw material	raw material	9.54	7.22	132.28	239	12	1.76	1.52	
54	RHS_NC_TOTAL	%	percentage raw material chips non conifer total	S1/2	raw	raw material	raw material	17.40	13.65	127.48	239	12	2.83	1.70	
55	RHS_ARABINOSE	mg/g DM	raw material chips arabinose	S1/2	raw	raw material	raw material	3.48	14.28	24.36	242	9	0.32	0.18	**
56	RHS_XYLOSE	mg/g DM	raw material chips xylose	S1/2	raw	raw material	raw material	20.08	56.06	35.83	242	9	2.45	1.32	
57	RHS_RHAMNOSE	mg/g DM	raw material chips rhamnose	S1/2	raw	raw material	raw material	0.80	3.39	23.75	242	9	0.35	0.23	**
58	RHS_MANNOSE	mg/g DM	raw material chips mannose	S1/2	raw	raw material	raw material	21.98	89.24	24.63	242	9	0.74	0.06	
59	RHS_GALACTOSE	mg/g DM	raw material chips galactose	S1/2	raw	raw material	raw material	5.37	22.42	23.95	242	9	0.95	0.32	**
60	RHS_GLUCOSE	mg/g DM	raw material chips glucose	S1/2	raw	raw material	raw material	7.75	31.14	24.89	242	9	2.74	1.04	
61	RHS_GALACT_ACID	mg/g DM	raw material chips galactoronic acid	S1/2	raw	raw material	raw material	3.93	13.91	28.25	242	9	0.48	0.41	
62	RHS_HEMI_TOTAL	mg/g DM	raw material chips hemicelluloses total	S1/2	raw	raw material	raw material	47.08	233.60	20.16	242	9	0.92	-0.05	
63	RHS_EXTRACT_TOTAL	%	raw material chips total extractives	S1/2	raw	raw material	raw material	1.10	2.57	42.69	244	7	3.51	1.29	

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign. [19]
64	RSS_ARABINOSE	mg/g DM	raw material sawdust chips arabinose	S3	raw	raw material	raw material	3.39	13.71	24.74	225	26	1.45	0.50	
65	RSS_XYLOSE	mg/g DM	raw material sawdust chips xylose	S3	raw	raw material	raw material	9.70	44.63	21.73	225	26	1.34	0.42	
66	RSS_RHAMNOSE	mg/g DM	raw material sawdust chips rhamnose	S3	raw	raw material	raw material	0.64	2.98	21.48	225	26	0.15	0.04	**
67	RSS_MANNOSE	mg/g DM	raw material sawdust chips mannose	S3	raw	raw material	raw material	22.54	96.48	23.36	225	26	1.73	-0.08	
68	RSS_GALACTOSE	mg/g DM	raw material sawdust chips galactose	S3	raw	raw material	raw material	5.58	24.32	22.94	225	26	0.11	0.10	**
69	RSS_GLUCOSE	mg/g DM	raw material sawdust chips glucose	S3	raw	raw material	raw material	7.18	29.43	24.39	225	26	1.62	0.67	
70	RSS_GALACT_ACID	mg/g DM	raw material sawdust chips galactoronic acid	S3	raw	raw material	raw material	4.03	12.42	32.44	225	26	2.63	0.99	
71	RSS_HEMI_TOTAL	mg/g DM	raw material sawdust chips hemicelluloses total	S3	raw	raw material	raw material	49.48	226.58	21.84	225	26	0.88	0.13	
72	RSS_EXTRACT_TOTAL	%	raw material sawdust total extractives	S3	raw	raw material	raw material	0.59	2.56	23.11	227	24	0.36	-0.21	**
73	R_ARABINOSE_WGT	mg/g DM	raw material arabinose weighted		raw	raw material	raw material	2.93	14.08	20.81	216	35	0.05	0.28	**
74	R_XYLOSE_WGT	mg/g DM	raw material xylose weighted		raw	raw material	raw material	15.14	53.43	28.33	216	35	1.34	1.01	
75	R_RHAMNOSE_WGT	mg/g DM	raw material rhamnose weighted		raw	raw material	raw material	0.65	3.27	19.81	216	35	-0.25	0.09	**
76	R_MANNOSE_WGT	mg/g DM	raw material mannose weighted		raw	raw material	raw material	18.34	91.16	20.12	216	35	1.46	0.41	
77	R_GALACTOSE_WGT	mg/g DM	raw material galactose weighted		raw	raw material	raw material	4.61	22.91	20.14	216	35	0.89	0.44	*
78	R_GLUCOSE_WGT	mg/g DM	raw material glucose weighted		raw	raw material	raw material	6.47	30.73	21.06	216	35	2.62	1.08	
79	R_GALACT_ACID_WGT	mg/g DM	raw material galactoronic acid weighted		raw	raw material	raw material	3.38	13.48	25.04	216	35	0.65	0.40	
80	R_HEMI_TOTAL_WGT	mg/g DM	raw material hemicelluloses total weighted		raw	raw material	raw material	39.72	232.07	17.12	216	35	0.88	0.11	
81	R_TOTAL_EXTRACT_WGT	%	raw material total extractives weighted		raw	raw material	raw material	0.82	2.55	31.95	220	31	4.01	1.26	
82	RHS_FILLLEVEL_DIGESTER	m	raw material chips digester fill level	P1	proc	fiber processing	controlled	0.44	3.78	11.55	251	0	2.30	-1.88	
83	RHS_DWELL_TIME_DIGESTER	min	raw material chips digester dwell time	P1	proc	fiber processing	controlled	0.31	3.67	8.43	120	131	10.03	2.58	
84	RHS_REFINER_DIGESTER_PRESSUR	bar	raw material chips refiner digester pressure	P1	proc	fiber processing	controlled	0.43	7.79	5.50	251	0	-0.93	0.93	
85	RHS_DISCHARGE_SCREW	rot/min	raw material chips discharge screw	P1	proc	fiber processing	controlled	9.05	72.19	12.54	251	0	-0.61	-0.37	
86	RHS_REFINER_POWER_DRAIN	kW	raw material chips refiner power drain	P1	proc	fiber processing	uncontrolled	541.10	4326.01	12.51	251	0	0.47	-0.57	
87	RSS_FILLLEVEL_DIGESTER	m	raw material sawdust digester fill level	P2	proc	fiber processing	controlled	0.21	0.76	27.81	184	67	3.12	1.48	
88	RSS_REFINER_DIGESTER_PRESSURE	bar	raw material sawdust refiner digester pressure	P2	proc	fiber processing	controlled	0.11	7.51	1.40	188	63	47.35	5.63	
89	RSS_DISCHARGE_SCREW	rot/min	raw material sawdust discharge screw	P2	proc	fiber processing	controlled	5.48	34.14	16.06	188	63	-0.84	-0.20	
90	RSS_REFINER_POWER_DRAIN	kW	raw material sawdust refiner power drain	P2	proc	fiber processing	uncontrolled	201.71	1433.68	14.07	188	63	-0.39	-0.11	**
91	FHS_ARABINOSE	mg/g DM	fiber chips arabinose	P1	raw	fiber	raw material	2.84	10.70	26.55	237	14	-0.34	0.22	*
92	FHS_XYLOSE	mg/g DM	fiber chips xylose	P1	raw	fiber	raw material	18.62	51.88	35.88	237	14	4.99	1.71	
93	FHS_RHAMNOSE	mg/g DM	fiber chips rhamnose	P1	raw	fiber	raw material	0.73	2.87	25.61	237	14	0.51	0.54	
94	FHS_MANNOSE	mg/g DM	fiber chips mannose	P1	raw	fiber	raw material	24.90	94.71	26.29	237	14	0.56	0.41	*
95	FHS_GALACTOSE	mg/g DM	fiber chips galactose	P1	raw	fiber	raw material	5.03	23.32	21.58	237	14	-0.26	-0.10	**
96	FHS_GLUCOSE	mg/g DM	fiber chips glucose	P1	raw	fiber	raw material	10.01	37.42	26.75	237	14	2.25	0.90	
97	FHS_GALACT_ACID	mg/g DM	fiber chips galactoronic acid	P1	raw	fiber	raw material	3.52	12.20	28.82	237	14	0.25	0.42	
98	FHS_HEMI_TOTAL	mg/g DM	fiber chips hemicelluloses total	P1	raw	fiber	raw material	55.15	236.10	23.36	237	14	0.82	0.19	
99	FHS_IC_ACETATE	mg/kg DM	fiber chips IC acetate	P1	raw	fiber	raw material	2156.4	10367.7	20.80	223	28	3.61	0.45	
100	FHS_IC_FORMIATE	mg/kg DM	fiber chips IC formiate	P1	raw	fiber	raw material	137.92	463.20	29.78	224	27	0.03	0.70	
101	FHS_IC_CHLORIDE	mg/kg DM	fiber chips IC chloride	P1	raw	fiber	raw material	78.93	135.94	58.06	248	3	12.87	2.58	

No	Variable	Unit	Description	material flow[16]	proc/raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign.[19]
102	FHS_IC_NITRATE	mg/kg DM	fiber chips IC nitrate	P1	raw	fiber	raw material	44.19	58.98	74.92	170	81	27.89	4.95	
103	FHS_EXTRACT_TOTAL	%	fiber chips total extractives	P1	raw	fiber	raw material	0.99	4.38	22.56	241	10	8.54	1.15	
104	FHS_METH_SORP_ANION_EX	µmol/g	fiber chips methylene blue sorption of anionic groups extr.	P1	raw	fiber	raw material	10.13	97.49	10.40	238	13	0.66	-0.31	**
105	FHS_PH_UNEX	-	fiber chips pH unextracted	P1	raw	fiber	raw material	0.25	5.07	n/a	241	10	24.81	-3.56	
106	FSS_ARABINOSE	mg/g DM	fiber sawdust arabinose	P2	raw	fiber	raw material	2.59	9.04	28.68	180	71	0.09	0.42	*
107	FSS_XYLOSE	mg/g DM	fiber sawdust xylose	P2	raw	fiber	raw material	12.18	43.71	27.87	180	71	6.93	1.64	
108	FSS_RHAMNOSE	mg/g DM	fiber sawdust rhamnose	P2	raw	fiber	raw material	0.61	2.66	22.96	180	71	2.35	0.86	
109	FSS_MANNOSE	mg/g DM	fiber sawdust mannose	P2	raw	fiber	raw material	24.81	100.71	24.63	180	71	1.03	-0.06	*
110	FSS_GALACTOSE	mg/g DM	fiber sawdust galactose	P2	raw	fiber	raw material	5.41	26.15	20.69	180	71	0.06	-0.04	**
111	FSS_GLUCOSE	mg/g DM	fiber sawdust glucose	P2	raw	fiber	raw material	10.53	38.14	27.60	180	71	1.64	0.95	
112	FSS_GALACT_ACID	mg/g DM	fiber sawdust galactoronic acid	P2	raw	fiber	raw material	3.31	10.77	30.74	180	71	0.39	0.71	
113	FSS_HEMI_TOTAL	mg/g DM	fiber sawdust hemicelluloses total	P2	raw	fiber	raw material	53.76	233.99	22.98	180	71	0.12	0.21	*
114	FSS_IC_ACETATE	mg/kg DM	fiber sawdust IC acetate	P2	raw	fiber	raw material	2222.4	12455.2	17.84	185	66	0.94	0.11	**
115	FSS_IC_FORMIATE	mg/kg DM	fiber sawdust IC formiate	P2	raw	fiber	raw material	212.77	785.50	27.09	184	67	4.43	1.38	
116	FSS_IC_CHLORIDE	mg/kg DM	fiber sawdust IC chloride	P2	raw	fiber	raw material	84.70	160.53	52.76	184	67	4.24	1.62	
117	FSS_IC_NITRATE	mg/kg DM	fiber sawdust IC nitrate	P2	raw	fiber	raw material	38.91	67.39	57.74	127	124	19.46	3.55	
118	FSS_EXTRACT_TOTAL	%	fiber sawdust total extractives	P2	raw	fiber	raw material	0.93	4.07	22.85	184	67	15.22	2.00	
119	FSS_METH_SORP_ANION_EX	µmol/g	fiber sawdust methylene blue sorption of anionic groups	P2	raw	fiber	raw material	11.20	95.27	11.76	182	69	0.14	0.11	**
120	FSS_PH_UNEX	-	fiber sawdust pH unextracted	P2	raw	fiber	raw material	0.25	4.82	n/a	184	67	10.49	-1.27	
121	F_ARABINOSE_WGT	mg/g DM	fiber arabinose weighted		raw	fiber	raw material	2.42	10.20	23.72	177	74	-0.32	0.31	*
122	F_XYLOSE_WGT	mg/g DM	fiber xylose weighted		raw	fiber	raw material	15.09	50.35	29.96	177	74	3.25	1.39	
123	F_RHAMNOSE_WGT	mg/g DM	fiber rhamnose weighted		raw	fiber	raw material	0.60	2.85	21.18	177	74	-0.02	0.32	**
124	F_MANNOSE_WGT	mg/g DM	fiber mannose weighted		raw	fiber	raw material	19.95	94.16	21.19	177	74	1.17	0.50	*
125	F_GALACTOSE_WGT	mg/g DM	fiber galactose weighted		raw	fiber	raw material	4.23	23.99	17.63	177	74	-0.29	-0.14	**
126	F_GLUCOSE_WGT	mg/g DM	fiber glucose weighted		raw	fiber	raw material	8.50	37.17	22.87	177	74	2.95	1.07	
127	F_GALACT_ACID_WGT	mg/g DM	fiber galactoronic acid weighted		raw	fiber	raw material	2.89	11.66	24.78	177	74	0.36	0.50	
128	F_HEMI_TOTAL_WGT	mg/g DM	fiber hemicelluloses total weighted		raw	fiber	raw material	45.11	233.66	19.31	177	74	0.85	0.20	
129	F_IC_ACETATE_WGT	mg/kg DM	fiber sawdust IC acetate weighted		raw	fiber	raw material	1903.7	10876.7	17.50	184	67	3.06	0.67	
130	F_IC_FORMIATE_WGT	mg/kg DM	fiber sawdust IC formiate weighted		raw	fiber	raw material	133.13	547.33	24.32	184	67	-0.22	0.49	
131	F_IC_CHLORIDE_WGT	mg/kg DM	fiber sawdust IC chloride weighted		raw	fiber	raw material	57.32	137.76	41.61	184	67	2.25	1.24	
132	F_TOTAL_EXTRACT_WGT	%	fiber total extractives weighted		raw	fiber	raw material	0.79	4.39	18.07	181	70	5.96	1.11	
133	F_METH_SORP_ANION_WGT_EX	µmol/g	fiber methylene blue sorption of anionic groups extracted weighted		raw	fiber	raw material	8.69	96.12	9.04	178	73	0.47	-0.09	**
134	F_PH_WGT_UNEX	-	fiber pH unextracted weighted		raw	fiber	raw material	0.22	4.98	n/a	181	70	20.66	-3.29	
135	PUFFER_CAPACITY	ml/g	buffer capacity of fiber chips, ml 0.1 N NaOH/2g fiber	P1	raw	fiber	raw material	0.14	0.91	15.57	248	3	2.33	1.01	
136	PUFFER_CAPACITY_FIB	ml/g	buffer capacity of glued fiber on form band, ml 0.1 N NaOH/2g fiber		raw	fiber	raw material	0.13	1.09	12.05	249	2	1.48	0.50	
137	PUFFER_CAPACITY_SS	ml/g	buffer capacity of fiber sawdust, ml 0.1 N NaOH/2g fiber	P2	raw	fiber	raw material	0.14	0.96	14.88	184	67	-0.16	0.25	**
138	RESIN_FRACT	kg/m³ MDF	resin fraction (RF)		proc	bonding	controlled	7.95	102.95	7.72	250	1	-0.74	0.02	
139	RHS_DYE_FRACT	% pure	chips dye fraction		proc	bonding	controlled	0.32	0.63	51.01	248	3	0.69	-0.39	

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign.[19]
140	RHS_EMULSION_FRACT	% dry fiber	chips paraffin emulsion fraction		proc	bonding	controlled	0.06	1.60	4.01	251	0	28.77	3.34	
141	RHS_RESIN_USE	% dry fiber	chips resin fraction dry		proc	bonding	controlled	0.94	12.73	7.36	251	0	-0.74	-0.01	
142	RHS_UREA_FRACT	% pure	chips urea fraction		proc	bonding	controlled	0.41	4.06	10.20	251	0	42.08	6.55	
143	RSS_DYE_FRACT	% pure	sawdust dye fraction		proc	bonding	controlled	0.33	0.68	48.73	186	65	0.67	-0.71	
144	RSS_EMULSION_FRACT	% dry fiber	sawdust paraffin emulsion fraction		proc	bonding	controlled	0.10	1.59	6.04	188	63	46.45	-2.91	
145	RSS_RESIN_USE	% dry fiber	sawdust resin fraction		proc	bonding	controlled	0.98	12.79	7.69	188	63	-0.92	-0.18	
146	RSS_UREA_FRACT	% pure	sawdust urea fraction		proc	bonding	controlled	0.48	4.09	11.69	188	63	30.48	5.61	
147	FS_DRYER_TEMP_IN	°C	dryer temperature inflow		proc	drying	controlled	23.01	158.18	n/a	251	0	0.19	0.52	
148	A241_TEMP_AV_T	°C	dryer cyclone discharge temperature mean		proc	drying	controlled	3.04	57.84	n/a	217	34	-0.32	0.10	*
149	FS_DRYER_TEMP_OUT	°C	dryer temperature discharge		proc	drying	uncontrolled	3.10	57.88	n/a	251	0	-0.42	0.15	
150	FS_BULK_DENSITY	kg/m³	bulk density		proc	form band	controlled	3.61	35.73	10.09	251	0	-0.41	-0.28	
151	FS_GRAMMAGE	kg/m²	grammage		proc	form band	uncontrolled	0.59	7.55	7.83	251	0	-0.83	-0.15	
152	FS_SPRINKLING	ml/m²	mat sprinkling top		proc	form band	controlled	1.04	3.65	28.60	217	34	-0.76	0.15	
153	FS_MAT_TEMP	°C	mat temperature		proc	form band	controlled	1.35	38.60	n/a	250	1	0.00	-0.51	
154	FS_MC_BANDSCALE	% dry fiber	fiber moisture content bandscale		proc	form band	intermediate	0.40	12.86	3.10	251	0	9.68	2.17	
155	FS_MC_FORMBAND	% dry fiber	fiber moisture content formband		proc	form band	intermediate	0.37	10.85	3.41	251	0	1.23	-0.44	
156	FS_MC_MICROWAVE_CONTI	%	fiber moisture content microwave continuous press		proc	form band	intermediate	0.24	6.17	3.95	247	4	0.43	0.02	
157	DI_AV_PRES_SYS01	bar	average pressure system 01		proc	hot press	controlled	10.58	69.38	15.24	217	34	-0.98	-0.04	
158	DI_AV_PRES_SYS02	bar	average pressure system 02		proc	hot press	controlled	6.20	113.93	5.44	217	34	0.25	1.04	
159	DI_AV_PRES_SYS03	bar	average pressure system 03		proc	hot press	controlled	6.06	110.31	5.49	217	34	0.37	1.10	
160	DI_AV_PRES_SYS04	bar	average pressure system 04		proc	hot press	controlled	6.30	106.02	5.94	217	34	0.32	0.97	
161	DI_AV_PRES_SYS05	bar	average pressure system 05		proc	hot press	controlled	5.61	51.77	10.83	217	34	0.50	1.14	
162	DI_AV_PRES_SYS06	bar	average pressure system 06		proc	hot press	controlled	5.47	88.66	6.17	217	34	3.36	1.50	
163	DI_AV_PRES_SYS07	bar	average pressure system 07		proc	hot press	controlled	4.46	84.00	5.30	217	34	2.07	1.16	
164	DI_AV_PRES_SYS08	bar	average pressure system 08		proc	hot press	controlled	4.09	43.31	9.44	217	34	0.11	0.52	
165	DI_AV_PRES_SYS09	bar	average pressure system 09		proc	hot press	controlled	3.10	36.90	8.39	217	34	-0.17	0.34	
166	DI_AV_PRES_SYS10	bar	average pressure system 10		proc	hot press	controlled	2.89	37.08	7.80	217	34	0.63	0.39	
167	DI_AV_PRES_SYS11	bar	average pressure system 11		proc	hot press	controlled	3.01	41.78	7.20	217	34	0.96	1.11	
168	DI_AV_PRES_SYS12	bar	average pressure system 12		proc	hot press	controlled	2.25	19.87	11.34	217	34	3.88	-1.04	
169	DI_AV_PRES_SYS13	bar	average pressure system 13		proc	hot press	controlled	8.55	11.88	71.95	217	34	-1.35	0.61	
170	DI_AV_PRES_SYS14	bar	average pressure system 14		proc	hot press	controlled	9.38	13.80	67.97	217	34	-1.32	0.64	
171	DI_AV_PRES_SYS15	bar	average pressure system 15		proc	hot press	controlled	2.61	26.37	9.90	217	34	6.89	-1.44	
172	DI_AV_PRES_SYS16	bar	average pressure system 16		proc	hot press	controlled	11.93	52.69	22.65	217	34	45.82	5.69	
173	DI_AV_PRES_SYS17	bar	average pressure system 17		proc	hot press	controlled	15.71	131.66	11.93	217	34	-0.58	-0.36	
174	DI_AV_PRES_SYS18	bar	average pressure system 18		proc	hot press	controlled	16.80	126.53	13.27	217	34	-0.72	-0.29	
175	DI_AV_PRES_SYS19	bar	average pressure system 19		proc	hot press	controlled	16.51	126.39	13.06	217	34	-0.39	-0.27	**
176	DI_AV_PRES_SYS21	bar	average pressure system 21		proc	hot press	controlled	18.71	124.95	14.98	217	34	-0.62	-0.34	
177	DI_AV_PRES_SYS22	bar	average pressure system 22		proc	hot press	controlled	11.76	99.59	11.80	217	34	-0.52	-0.26	*

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign.[19]
178	PRESS_VELOCITY	mm/s	press velocity (PV)		proc	hot press	controlled	96.63	780.38	12.38	251	0	-0.70	0.37	
179	PRESSFACTOR	s/mm	press factor (PF)		proc	hot press	intermediate	0.67	7.48	8.98	251	0	-1.41	0.12	
180	IB	N/mm²	actual internal bond			product property	result	0.11	1.78	6.23	251	0	0.03	0.27	**
181	IB0	N/mm²	internal bond, single value position 0			product property	result	0.19	1.83	10.19	251	0	0.09	-0.08	**
182	IB1	N/mm²	internal bond, single value position 1			product property	result	0.16	1.85	8.45	251	0	-0.05	-0.05	**
183	IB2	N/mm²	internal bond, single value position 2			product property	result	0.17	1.79	9.54	251	0	0.11	0.08	**
184	IB3	N/mm²	internal bond, single value position 3			product property	result	0.18	1.72	10.15	251	0	-0.17	0.04	**
185	IB4	N/mm²	internal bond, single value position 4			product property	result	0.17	1.70	10.00	251	0	-0.08	-0.22	**
186	IB5	N/mm²	internal bond, single value position 5			product property	result	0.17	1.75	9.75	251	0	0.16	0.19	**
187	IB6	N/mm²	internal bond, single value position 6			product property	result	0.19	1.76	10.63	251	0	0.44	-0.38	*
188	IB7	N/mm²	internal bond, single value position 7			product property	result	0.16	1.79	9.10	251	0	-0.23	-0.03	**
189	IB8	N/mm²	internal bond, single value position 8			product property	result	0.17	1.88	9.29	251	0	0.42	-0.33	**
190	IB9	N/mm²	internal bond, single value position 9			product property	result	0.16	1.75	9.26	251	0	-0.53	-0.14	**
191	SS	N/mm²	actual surface soundness			product property	result	0.13	2.09	6.17	244	7	-0.16	0.33	
192	SS0	N/mm²	surface soundness, single value position 0			product property	result	0.18	2.28	8.11	244	7	0.02	0.06	**
193	SS1	N/mm²	surface soundness, single value position 1			product property	result	0.19	2.14	8.91	244	7	0.02	0.23	**
194	SS2	N/mm²	surface soundness, single value position 2			product property	result	0.18	2.02	9.06	244	7	-0.52	0.21	**
195	SS3	N/mm²	surface soundness, single value position 3			product property	result	0.19	2.05	9.05	244	7	-0.07	-0.11	**
196	SS4	N/mm²	surface soundness, single value position 4			product property	result	0.18	1.98	9.31	244	7	0.31	-0.06	**
197	SS5	N/mm²	surface soundness, single value position 5			product property	result	0.18	2.06	8.81	244	7	0.61	-0.18	**
198	SS6	N/mm²	surface soundness, single value position 6			product property	result	0.21	1.99	10.30	244	7	0.18	0.22	**
199	SS7	N/mm²	surface soundness, single value position 7			product property	result	0.18	2.02	8.70	244	7	-0.36	0.00	**
200	SS8	N/mm²	surface soundness, single value position 8			product property	result	0.18	2.13	8.35	244	7	0.15	0.03	**
201	SS9	N/mm²	surface soundness, single value position 9			product property	result	0.17	2.24	7.64	244	7	-0.36	0.00	**
202	MOR	N/mm²	actual modulus of rupture			product property	result	2.38	47.68	5.00	244	7	0.46	0.42	*
203	MOR0	N/mm²	modulus of rupture, single value position 0			product property	result	3.47	51.83	6.70	244	7	0.12	-0.21	**
204	MOR1	N/mm²	modulus of rupture, single value position 1			product property	result	3.26	47.42	6.87	244	7	-0.23	0.08	**
205	MOR2	N/mm²	modulus of rupture, single value position 2			product property	result	3.05	46.59	6.55	244	7	0.77	0.09	**
206	MOR3	N/mm²	modulus of rupture, single value position 3			product property	result	3.15	46.63	6.76	244	7	0.16	0.21	**
207	MOR4	N/mm²	modulus of rupture, single value position 4			product property	result	3.24	46.76	6.93	244	7	0.08	0.05	**
208	MOR5	N/mm²	modulus of rupture, single value position 5			product property	result	3.19	46.59	6.86	244	7	0.31	0.37	*
209	MOR6	N/mm²	modulus of rupture, single value position 6			product property	result	3.26	46.44	7.03	244	7	0.04	0.06	**
210	MOR7	N/mm²	modulus of rupture, single value position 7			product property	result	2.99	46.08	6.49	244	7	-0.06	0.45	
211	MOR8	N/mm²	modulus of rupture, single value position 8			product property	result	3.50	47.30	7.40	244	7	-0.13	0.22	**
212	MOR9	N/mm²	modulus of rupture, single value position 9			product property	result	3.74	51.17	7.30	244	7	0.01	0.24	**
213	MOE	N/mm²	actual modulus of elasticity			product property	result	222.11	4677.07	4.75	244	7	1.71	-0.66	
214	MOE0	N/mm²	modulus of elasticity, single value position 0			product property	result	302.81	4910.36	6.17	244	7	1.69	-0.71	
215	MOE1	N/mm²	modulus of elasticity, single value position 1			product property	result	263.46	4670.86	5.64	244	7	0.22	-0.26	**

No	Variable	Unit	Description	material flow[16]	proc/ raw	section	type of variable[17]	std dev	mean	CV[18]	n	n missing	Kurtosis	Skeweness	p sign. [19]
216	MOE2	N/mm²	modulus of elasticity, single value position 2			product property	result	254.61	4617.81	5.51	244	7	0.80	-0.43	
217	MOE3	N/mm²	modulus of elasticity, single value position 3			product property	result	254.06	4648.16	5.47	244	7	1.05	-0.23	*
218	MOE4	N/mm²	modulus of elasticity, single value position 4			product property	result	263.38	4647.42	5.67	244	7	1.54	-0.56	
219	MOE5	N/mm²	modulus of elasticity, single value position 5			product property	result	250.13	4630.86	5.40	244	7	1.29	-0.54	
220	MOE6	N/mm²	modulus of elasticity, single value position 6			product property	result	250.30	4590.54	5.45	244	7	1.28	-0.55	
221	MOE7	N/mm²	modulus of elasticity, single value position 7			product property	result	254.61	4539.48	5.61	244	7	0.36	-0.26	**
222	MOE8	N/mm²	modulus of elasticity, single value position 8			product property	result	282.04	4639.20	6.08	244	7	-0.08	-0.37	*
223	MOE9	N/mm²	modulus of elasticity, single value position 9			product property	result	271.25	4876.04	5.56	244	7	0.25	-0.43	*
224	TS	%	actual thickness swelling			product property	result	1.28	10.74	11.89	247	4	-0.44	0.51	
225	TS0	%	thickness swelling, single value position 0			product property	result	1.24	10.65	11.62	247	4	-0.25	0.44	
226	TS1	%	thickness swelling, single value position 1			product property	result	1.32	10.80	12.21	247	4	-0.16	0.53	
227	TS2	%	thickness swelling, single value position 2			product property	result	1.32	10.84	12.19	247	4	-0.42	0.48	
228	TS3	%	thickness swelling, single value position 3			product property	result	1.33	10.71	12.46	247	4	-0.39	0.51	
229	TS4	%	thickness swelling, single value position 4			product property	result	1.33	10.66	12.45	247	4	-0.01	0.58	
230	TS5	%	thickness swelling, single value position 5			product property	result	1.30	10.68	12.14	247	4	-0.29	0.58	
231	TS6	%	thickness swelling, single value position 6			product property	result	1.33	10.80	12.36	247	4	-0.41	0.57	
232	TS7	%	thickness swelling, single value position 7			product property	result	1.38	10.91	12.64	247	4	-0.42	0.55	
233	TS8	%	thickness swelling, single value position 8			product property	result	1.34	10.79	12.40	247	4	-0.29	0.57	
234	TS9	%	thickness swelling, single value position 9			product property	result	1.32	10.59	12.47	247	4	-0.73	0.36	
235	MD	kg/m³	actual medium density			product property	result	12.90	898.33	1.44	251	0	-0.71	-0.28	
236	MD0	kg/m³	medium density, single value position 0			product property	result	21.59	905.01	2.39	251	0	-0.45	0.10	**
237	MD1	kg/m³	medium density, single value position 1			product property	result	19.14	897.73	2.13	251	0	0.31	-0.19	**
238	MD2	kg/m³	medium density, single value position 2			product property	result	18.21	892.36	2.04	251	0	-0.26	-0.12	**
239	MD3	kg/m³	medium density, single value position 3			product property	result	18.27	897.76	2.03	251	0	-0.21	0.02	**
240	MD4	kg/m³	medium density, single value position 4			product property	result	18.74	893.06	2.10	251	0	-0.19	-0.34	*
241	MD5	kg/m³	medium density, single value position 5			product property	result	19.25	899.41	2.14	251	0	-0.29	-0.31	**
242	MD6	kg/m³	medium density, single value position 6			product property	result	19.02	899.91	2.11	251	0	-0.36	-0.28	**
243	MD7	kg/m³	medium density, single value position 7			product property	result	19.30	895.27	2.16	251	0	-0.34	-0.29	**
244	MD8	kg/m³	medium density, single value position 8			product property	result	19.27	904.55	2.13	251	0	0.01	0.13	**
245	MD9	kg/m³	medium density, single value position 9			product property	result	17.20	898.27	1.91	251	0	-0.10	-0.22	**

B.2 Descriptive statistics for properties with thickness 7.4 mm

Variable	std dev	Mean	CV	n	n missing	Kurtosis	Skeweness	p sign.[20]
IB	0.11	1.77	6.26	143	0	0.62	0.52	*
IB0	0.19	1.82	10.51	143	0	-0.16	-0.01	**
IB1	0.15	1.85	7.94	143	0	0.29	0.23	**
IB2	0.15	1.78	8.67	143	0	0.48	0.35	**
IB3	0.16	1.71	9.62	143	0	0.01	-0.23	**
IB4	0.17	1.70	9.99	143	0	0.30	-0.36	**
IB5	0.16	1.74	9.08	143	0	0.34	0.29	**
IB6	0.19	1.74	10.95	143	0	-0.15	-0.19	**
IB7	0.17	1.77	9.65	143	0	-0.18	0.01	**
IB8	0.17	1.88	8.85	143	0	0.50	-0.30	**
IB9	0.16	1.74	9.36	143	0	-0.63	-0.07	**
MD	12.39	899.03	1.38	143	0	-0.68	-0.43	
MD0	21.83	903.55	2.42	143	0	-0.57	0.17	**
MD1	17.91	897.52	2.00	143	0	-0.15	-0.28	**
MD2	17.19	892.95	1.92	143	0	-0.42	-0.11	**
MD3	16.45	898.58	1.83	143	0	-0.04	0.08	**
MD4	17.48	894.09	1.95	143	0	0.13	-0.27	**
MD5	18.68	900.09	2.07	143	0	-0.34	-0.39	**
MD6	17.16	901.62	1.90	143	0	-0.22	-0.39	**
MD7	18.65	897.56	2.08	143	0	-0.49	-0.21	**
MD8	17.34	906.65	1.91	143	0	-0.33	-0.11	**
MD9	16.70	897.67	1.86	143	0	-0.27	-0.19	**
MOE	162.75	4759.78	3.42	138	5	-0.26	-0.01	**
MOE0	225.16	4963.72	4.54	138	5	0.35	-0.23	**
MOE1	211.92	4757.29	4.45	138	5	-0.45	-0.04	**
MOE2	199.87	4708.81	4.24	138	5	-0.20	0.32	**
MOE3	200.36	4736.48	4.23	138	5	0.06	0.23	**
MOE4	200.87	4742.69	4.24	138	5	0.02	-0.01	**
MOE5	189.85	4709.04	4.03	138	5	-0.10	0.22	**
MOE6	201.44	4683.11	4.30	138	5	-0.06	0.10	**
MOE7	207.82	4621.14	4.50	138	5	-0.53	0.07	**
MOE8	235.38	4729.63	4.98	138	5	-0.49	-0.28	**
MOE9	232.38	4945.91	4.70	138	5	0.03	-0.14	**
MOR	2.22	48.26	4.60	138	5	0.32	0.66	
MOR0	3.24	52.15	6.21	138	5	0.58	-0.16	**
MOR1	3.12	47.96	6.51	138	5	-0.06	0.19	**
MOR2	2.81	47.18	5.95	138	5	0.41	0.67	
MOR3	2.88	47.27	6.09	138	5	-0.35	0.20	**
MOR4	3.01	47.45	6.34	138	5	-0.10	0.03	**
MOR5	2.98	47.11	6.32	138	5	0.77	0.75	
MOR6	3.15	47.27	6.66	138	5	-0.33	0.09	**
MOR7	2.90	46.54	6.23	138	5	-0.07	0.62	
MOR8	3.23	47.98	6.72	138	5	0.06	0.29	**
MOR9	3.60	51.71	6.96	138	5	0.20	0.41	**
SS	0.12	2.12	5.79	138	5	-0.22	0.41	*
SS0	0.17	2.28	7.65	138	5	-0.35	0.19	**
SS1	0.19	2.17	8.90	138	5	-0.59	0.14	**
SS2	0.19	2.06	9.22	138	5	-0.46	-0.08	**
SS3	0.18	2.09	8.69	138	5	0.24	-0.09	**
SS4	0.17	2.02	8.21	138	5	0.79	-0.10	**
SS5	0.17	2.10	8.07	138	5	0.12	-0.07	**
SS6	0.19	2.03	9.45	138	5	0.21	0.35	**
SS7	0.17	2.06	8.20	138	5	-0.08	-0.30	**
SS8	0.18	2.15	8.19	138	5	0.21	0.07	**
SS9	0.17	2.27	7.33	138	5	-0.13	-0.02	**
TS	0.62	9.83	6.36	140	3	0.65	0.57	*
TS0	0.70	9.81	7.16	140	3	1.17	0.27	**
TS1	0.69	9.89	7.01	140	3	0.12	0.39	**
TS2	0.69	9.92	6.94	140	3	0.86	0.57	*
TS3	0.65	9.76	6.67	140	3	0.37	0.34	**
TS4	0.66	9.72	6.79	140	3	0.45	0.42	**
TS5	0.67	9.78	6.83	140	3	1.37	0.70	
TS6	0.66	9.87	6.73	140	3	0.33	0.68	
TS7	0.70	9.96	7.02	140	3	-0.30	0.41	*
TS8	0.71	9.88	7.17	140	3	1.09	0.67	
TS9	0.78	9.70	8.06	140	3	0.68	0.73	

[20] Indicates if variable follow a normal distribution using a Shapiro-Wilk test (p>0.05 indicated by **, 0.01≤p≤0.05 by *)

B.3 Descriptive statistics for properties with thickness 6.4 mm

Variable	std dev	Mean	CV	n	n missing	Kurtosis	Skeweness	p sign.[21]
IB	0.11	1.79	6.23	93	0	-0.50	-0.03	**
IB0	0.17	1.84	9.33	93	0	0.56	-0.42	**
IB1	0.17	1.84	9.08	93	0	-0.30	-0.32	**
IB2	0.19	1.79	10.38	93	0	0.08	-0.18	**
IB3	0.19	1.75	10.62	93	0	-0.50	0.24	**
IB4	0.17	1.69	10.20	93	0	-0.52	0.02	**
IB5	0.18	1.76	10.48	93	0	-0.09	-0.04	**
IB6	0.19	1.76	10.54	93	0	1.55	-0.60	*
IB7	0.15	1.82	8.14	93	0	-0.41	0.07	**
IB8	0.18	1.88	9.71	93	0	0.50	-0.48	**
IB9	0.16	1.76	9.36	93	0	-0.31	-0.18	**
MD	12.94	895.71	1.44	93	0	-0.46	0.02	**
MD0	19.59	904.04	2.17	93	0	-0.50	-0.22	**
MD1	19.24	895.70	2.15	93	0	0.93	-0.37	**
MD2	19.17	889.57	2.16	93	0	-0.03	-0.02	**
MD3	19.68	895.24	2.20	93	0	-0.77	0.02	**
MD4	19.95	890.42	2.24	93	0	-0.53	-0.37	**
MD5	19.30	896.14	2.15	93	0	-0.26	-0.24	**
MD6	20.53	896.36	2.29	93	0	-0.55	-0.15	**
MD7	19.26	891.36	2.16	93	0	-0.26	-0.32	**
MD8	20.21	900.04	2.25	93	0	0.13	0.33	**
MD9	18.06	898.24	2.01	93	0	0.17	-0.27	**
MOE	244.37	4564.07	5.35	92	1	1.82	-0.45	
MOE0	376.76	4819.12	7.82	92	1	0.61	-0.46	**
MOE1	282.87	4550.42	6.22	92	1	0.57	0.07	**
MOE2	266.65	4494.97	5.93	92	1	0.50	-0.44	**
MOE3	275.62	4535.99	6.08	92	1	1.58	0.08	**
MOE4	281.48	4515.58	6.23	92	1	1.94	-0.36	
MOE5	287.49	4541.39	6.33	92	1	0.94	-0.43	**
MOE6	260.18	4468.87	5.82	92	1	1.32	-0.74	
MOE7	270.84	4429.00	6.12	92	1	0.78	-0.03	**
MOE8	293.46	4516.31	6.50	92	1	0.13	-0.09	**
MOE9	287.01	4769.08	6.02	92	1	-0.07	-0.37	**
MOR	2.48	46.90	5.28	92	1	0.53	0.45	**
MOR0	3.84	51.27	7.49	92	1	-0.38	-0.11	**
MOR1	3.39	46.69	7.27	92	1	-0.60	0.00	**
MOR2	3.34	45.89	7.27	92	1	0.34	-0.31	**
MOR3	3.38	45.83	7.37	92	1	0.77	0.36	**
MOR4	3.35	45.79	7.32	92	1	0.45	0.38	**
MOR5	3.41	46.07	7.41	92	1	-0.62	0.25	**
MOR6	3.22	45.28	7.12	92	1	0.46	0.09	**
MOR7	3.08	45.45	6.77	92	1	-0.13	0.40	**
MOR8	3.68	46.49	7.92	92	1	-0.13	0.37	**
MOR9	3.91	50.28	7.77	92	1	-0.32	0.23	**
SS	0.12	2.04	5.68	91	2	-0.27	0.22	**
SS0	0.19	2.27	8.41	91	2	-0.32	-0.43	**
SS1	0.17	2.08	8.11	91	2	-0.08	-0.14	**
SS2	0.15	1.94	7.79	91	2	-0.45	0.41	**
SS3	0.18	2.00	9.22	91	2	-0.77	0.04	**
SS4	0.18	1.92	9.33	91	2	0.57	0.01	**
SS5	0.18	2.01	8.91	91	2	1.56	-0.30	*
SS6	0.21	1.93	10.66	91	2	-0.25	0.08	**
SS7	0.17	1.96	8.76	91	2	0.21	0.41	**
SS8	0.17	2.08	8.37	91	2	-0.01	-0.08	**
SS9	0.16	2.19	7.39	91	2	-0.61	-0.05	**
TS	0.72	11.86	6.10	92	1	-0.26	0.44	**
TS0	0.77	11.67	6.60	92	1	-0.60	0.45	
TS1	0.77	11.91	6.48	92	1	-0.41	0.38	*
TS2	0.75	11.98	6.27	92	1	-0.18	0.06	**
TS3	0.79	11.89	6.61	92	1	0.56	0.54	**
TS4	0.74	11.78	6.25	92	1	0.66	0.71	*
TS5	0.80	11.76	6.79	92	1	-0.23	0.58	*
TS6	0.86	11.94	7.18	92	1	0.15	0.50	**
TS7	0.90	12.06	7.45	92	1	-0.17	0.37	**
TS8	0.87	11.86	7.37	92	1	0.41	0.40	**
TS9	0.84	11.70	7.20	92	1	-0.58	-0.16	**

[21] Indicates if variable follow a normal distribution using a Shapiro-Wilk test ($p>0.05$ indicated by **, $0.01 \le p \le 0.05$ by *)

C. Summary statistics of all models

Model no.	Model name	Objects with thickness	X variables[22]	Y variable	A_{opt}[23]	no. obser-vationss	RMSEP[24]	Q^2 [25]	NRMSEP[26]	Number of variables[27]
1	IB 7.4	7.4	all	IB	2	142	0.1019	7.57%	5.75%	5
2	IB 6.4	6.4	all	IB	2	93	0.1067	9.29%	5.96%	12
3	IB all	all	all	IB	1	251	0.1036	13.44%	5.81%	11
4	IB proc	all	proc	IB	1	251	0.1073	7.14%	6.02%	4
5	IB raw	all	raw	IB	1	221	0.1053	2.73%	5.90%	2
6	SS 7.4	7.4	all	SS	1	138	0.1171	10.06%	5.51%	15
7	SS 6.4	6.4	all	SS	1	91	0.1038	20.47%	5.09%	8
8	SS all	all	all	SS	2	244	0.1129	23.62%	5.40%	24
9	SS proc	all	proc	SS	2	244	0.1143	21.65%	5.47%	19
10	SS raw	all	raw	SS	1	220	0.1229	6.11%	5.88%	3
11	MOR 7.4	7.4	all	MOR	3	138	1.6142	47.59%	3.34%	26
12	MOR 6.4	6.4	all	MOR	2	92	1.8198	46.68%	3.88%	6
13	MOR all	all	all	MOR	3	244	1.6867	50.14%	3.54%	24
14	MOR proc	all	proc	MOR	3	244	1.7436	46.72%	3.66%	20
15	MOR raw	all	raw	MOR	1	244	2.2619	10.33%	4.74%	21
16	MOE 7.4	7.4	all	MOE	1	138	158.02	6.41%	3.32%	12
17	MOE 6.4	6.4	all	MOE	1	92	219.74	20.02%	4.81%	32
18	MOE all	all	all	MOE	2	244	169.41	42.06%	3.62%	14
19	MOE proc	all	proc	MOE	3	244	171.72	40.47%	3.67%	18
20	MOE raw	all	raw	MOE	1	223	217.45	5.47%	4.65%	4
21	TS 7.4	7.4	all	TS	1	140	0.5345	27.37%	5.44%	35
22	TS 6.4	6.4	all	TS	2	92	0.5222	48.44%	4.40%	20
23	TS all	all	all	TS	3	247	0.6211	76.47%	5.78%	13
24	TS proc	all	proc	TS	2	247	0.5808	79.42%	5.41%	19
25	TS raw	all	raw	TS	1	236	1.2757	2.88%	11.87%	4
26	PRESSFACTOR 7.4	7.4	all	PRESSFACTOR	3	143	0.2070	90.26%	2.74%	34
27	PRESSFACTOR 6.4	6.4	all	PRESSFACTOR	2	93	0.1992	91.85%	2.71%	32
28	PRESSFACTOR all	all	all	PRESSFACTOR	3	251	0.2034	90.88%	2.72%	31
29	PRESSFACTOR proc	all	proc	PRESSFACTOR	3	251	0.2049	90.74%	2.74%	30
30	PRESSFACTOR raw	all	raw	PRESSFACTOR	2	251	0.5360	36.67%	7.16%	15
31	RESIN_FRACT 7.4	7.4	all	RESIN_FRACT	3	143	3.8320	77.33%	3.73%	32
32	RESIN_FRACT 6.4	6.4	all	RESIN_FRACT	1	92	4.1386	71.76%	4.00%	22
33	RESIN_FRACT all	all	all	RESIN_FRACT	3	250	4.1363	73.05%	4.02%	52
34	RESIN_FRACT proc	all	proc	RESIN_FRACT	3	250	4.4274	69.13%	4.30%	36
35	RESIN_FRACT raw	all	raw	RESIN_FRACT	3	247	6.2736	38.16%	6.09%	15
36	IB0 7.4	7.4	all	IB0	2	143	0.1577	32.38%	8.67%	23
37	IB1 7.4	7.4	all	IB1	1	143	0.1425	6.46%	7.71%	16
38	IB2 7.4	7.4	all	IB2	n/a	n/a	n/a	n/a	n/a[28]	0
39	IB3 7.4	7.4	all	IB3	1	143	0.1646	0.52%	9.63%	2
40	IB4 7.4	7.4	all	IB4	1	143	0.1644	6.41%	9.70%	2
41	IB5 7.4	7.4	all	IB5	1	143	0.1466	3.90%	8.45%	4
42	IB6 7.4	7.4	all	IB6	2	143	0.1700	21.48%	9.74%	5
43	IB7 7.4	7.4	all	IB7	1	129	0.1633	4.64%	9.22%	6
44	IB8 7.4	7.4	all	IB8	1	143	0.1621	5.41%	8.64%	10
45	IB9 7.4	7.4	all	IB9	1	143	0.1551	10.46%	8.89%	12
									Total	**750**

[22] Category of variables used as predictor: *all*, *proc* and *raw* (as indicated in Appendix B.1)
[23] Number of PLS components used
[24] Units of RMSEP are the same as unit of the Y-variable of the model
[25] Q^2 obtained by full cross validation
[26] RMSEP normalized by the mean value of Y
[27] Indicating the number of remaining significant variables in the final model
[28] Validation was not possible due to a too bad model

D. Weighted regression coefficients (Bw) of all models

Empty cells indicate that the variable was not used for modeling to the specific model.

Cells with 0 indicate that the variable did not show any significance in modeling. Rows with variables that were not significant in any models are omitted in the listings.

D.1 B_w for IB

Var.no.	Variable	IB0 7.4	IB1 7.4	IB2 7.4	IB3 7.4	IB4 7.4	IB5 7.4	IB6 7.4	IB7 7.4	IB8 7.4	IB9 7.4	IB 7.4	IB 6.4	IB all	IB proc	IB raw
1	NOMINAL_THICKNESS													-0.072	-0.097	
6	PERCEP_FORMALDEHYDE	0	0	0	0	0	0	0	0	0	0	0	0	0.129	0.172	
8	MONTH_01	0.106	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	MONTH_04	0.164	0	0	0	0	0	0	0	0	0	0	0	0.087	0	
12	MONTH_05	0	-0.026	0	0	0	0	0	0	0	0.063	0	0	0	0	
13	MONTH_06	0	0	0	0	0	0	-0.293	0	0	0	0	0	0	0	
14	MONTH_07	0.147	0.032	0	0.100	0.201	0	0.136	0	0	0	0.351	0	0.077	0	
15	MONTH_08	-0.085	0	0	0	0	0	-0.212	0	-0.036	0	0	0	0	0	
17	MONTH_10	0	0	0	0	0	0	0	0	0	0	0	0.171	0	0	
20	SEASON_SPRING	0	0	0	0	0	0	0	0	0.036	0.050	0	0	0	0	
22	SEASON_FALL	0	0	0	0	0	0	0	0	0	0	0	0.087	0	0	
32	SHIFT2	0	0	0	0	0	0	0	0	0	0	0	0	0	0.123	
35	SHIFT5	0	0	0	0	0	0	0	0	0	-0.063	0	0	-0.091	0	
36	HE_SS_FRACT	0	0	0	0	0	0	0	0	0	0	0	0	0.079	0.106	
37	WOOD_INPUT_TOTAL	-0.071	-0.023	0	0	0	0	0	0	0	0	0	0	0	0	
38	FIB_PROC_LINE2_FRACT	0	0	0	-0.078	0	0	0	0	0.038	0	0	0	0	0	
45	RHS_NC_ASH	-0.139	0	0	0	0	0	0	0	0	0	0	0	0		0
53	RHS_NC_ASPEN	0	0	0	0	0	0	-0.289	0	0	0	0	0	0		0
56	RHS_XYLOSE	-0.084	0	0	0	-0.187	0	0	0	0	0	0	0	0		0
61	RHS_GALACT_ACID	-0.008	0	0	0	0	0	0	0	0	0	0	0	0		0
62	RHS_HEMI_TOTAL	-0.008	0	0	0	0	0	0	0	0	0	0	0	0		0
74	R_XYLOSE_WGT	-0.078	0	0	0	0	0	0	0	0	0	0	0	-0.086		-0.111
79	R_GALACT_ACID_WGT	0.005	0	0	0	0	0	0	0	0	0	0	0	0		0
81	R_TOTAL_EXTRACT_WGT	-0.113	0	0	0	0	0	0	0	0	0	0	0	-0.099		-0.149
84	RHS_REFINER_DIGESTER_PRESSURE	-0.191	0	0	0	0	0	0	0	0	0	0	0.089	0	0	
85	RHS_DISCHARGE_SCREW	-0.059	0	0	0	0	0	0	0	-0.035	0	0	-0.036	0	0	
86	RHS_REFINER_POWER_DRAIN	-0.106	0	0	0	0	0	0	0	0	0	0	0	0	0	
89	RSS_DISCHARGE_SCREW	0	0	0	0	0	0	0	0	0.049	0.071	0	0	0.097	0	
91	FHS_ARABINOSE	0	0	0	0	0	-0.058	0	0	0	0	0	0	0		0
95	FHS_GALACTOSE	0	0	0	0	0	-0.065	0	0	0	-0.047	0	0	0		0
97	FHS_GALACT_ACID	0	0	0	0	0	-0.068	0	0	0	0	0	0	0		0
98	FHS_HEMI_TOTAL	0	0	0	0	0	-0.063	0	0	0	0	0	0	0		0
109	FSS_MANNOSE	0	0	0	0	0	0	0	0	0	0	0	0	0.090		0
118	FSS_EXTRACT_TOTAL	0	0	0	0	0	0	0.194	0	0	0	0	-0.155	0		0
122	F_XYLOSE_WGT	0	0.035	0	0	0	0	0	0	0	-0.045	0	0	0		0
123	F_RHAMNOSE_WGT	0	0.033	0	0	0	0	0	0	0	-0.054	0	0	0		0
125	F_GALACTOSE_WGT	0	0	0	0	0	0	0	0	0	-0.053	0	0	0		0

Var.no.	Variable	IB0 7.4	IB1 7.4	IB2 7.4	IB3 7.4	IB4 7.4	IB5 7.4	IB6 7.4	IB7 7.4	IB8 7.4	IB9 7.4	IB 7.4	IB 6.4	IB all	IB proc	IB raw
138	RESIN_FRACT	0	0	0	0	0	0	0	0	0.036	0	0	0	0	0	
143	RSS_DYE_FRACT	0	0	0	0	0	0	0	0	0	0	0	0	-0.095	0	
145	RSS_RESIN_USE	0.158	0	0	0	0	0	0	0	0	0	0	0	0	0	
147	FS_DRYER_TEMP_IN	0	0	0	0	0	0	0	0	0.040	0.049	0	0	0	0	
148	A241_TEMP_AV_T	0.123	0	0	0	0	0	0	0	0.044	0.055	0	0	0	0	
149	FS_DRYER_TEMP_OUT	0	0	0	0	0	0	0	0	0.044	0.043	0	0	0	0	
151	FS_GRAMMAGE	0	0.024	0	0	0	0	0	0	0	0	0	0	0	0	
152	FS_SPRINKLING	0	0	0	0	0	0	0	0	-0.052	-0.047	0	0	0	0	
155	FS_MC_FORMBAND	0	0.024	0	0	0	0	0	0	0	0	0	0	0	0	
156	FS_MC_MICROWAVE_CONTI	-0.119	0	0	0	0	0	0	0	0	0	0	0	0	0	
158	DI_AV_PRES_SYS02	0	0.024	0	0	0	0	0	0.028	0	0	-0.042	0.033	0	0	
159	DI_AV_PRES_SYS03	0	0.024	0	0	0	0	0	0.031	0	0	0.002	0.029	0	0	
160	DI_AV_PRES_SYS04	0	0.024	0	0	0	0	0	0.029	0	0	-0.003	-0.007	0	0	
161	DI_AV_PRES_SYS05	0	0.022	0	0	0	0	0	0.024	0	0	-0.042	0.039	0	0	
162	DI_AV_PRES_SYS06	-0.058	0	0	0	0	0	0	0	0	0	0	0	0	0	
164	DI_AV_PRES_SYS08	-0.115	0	0	0	0	0	0	0.023	0	0	0	-0.061	0	0	
165	DI_AV_PRES_SYS09	0	0.020	0	0	0	0	0	0	0	0	0	0.000	0	0	
167	DI_AV_PRES_SYS11	0	0	0	0	0	0	0	0.032	0	0	0	0	0	0	
169	DI_AV_PRES_SYS13	0	0.033	0	0	0	0	0	0	0	0	0	0	0	0	
170	DI_AV_PRES_SYS14	0.075	0.032	0	0	0	0	0	0	0	0	0	0	0	0	
176	DI_AV_PRES_SYS21	0	0	0	0	0	0	0	0	0	0	0	-0.076	0	0	
178	PRESS_VELOCITY	-0.089	-0.021	0	0	0	0	0	0	0	0	0	0	0	0	
179	PRESSFACTOR	0.089	0.021	0	0	0	0	0	0	0	0	0	0	0	0	

D.2 B_w for SS

Var.no.	Variable	SS 7.4	SS 6.4	SS all	SS proc	SS raw
1	NOMINAL_THICKNESS			0.119	0.129	
6	PERCEP_FORMALDEHYDE	0	0	0.099	0.096	
14	MONTH_07	0	0	0.090	0.108	
15	MONTH_08	0	0	-0.052	-0.074	
25	RESIN_TANK02	0	0	0.067	0	
28	RESIN_TANK05	0	0	-0.082	0	
32	SHIFT2	0	0	0.102	0.102	
33	SHIFT3	0	0	-0.075	-0.081	
37	WOOD_INPUT_TOTAL	0	0	-0.069	-0.074	
39	RHS_C_FIR	0	0	0.099		0
51	RHS_NC_MAPLE	0	-0.140	0		0
56	RHS_XYLOSE	-0.044	0	0		0
57	RHS_RHAMNOSE	-0.041	0	0		0
60	RHS_GLUCOSE	-0.037	0	0		0
62	RHS_HEMI_TOTAL	-0.043	0	0		0
83	RHS_DWELL_TIME_DIGESTER	0	0	0	0.152	
85	RHS_DISCHARGE_SCREW	-0.029	0	-0.065	-0.062	
86	RHS_REFINER_POWER_DRAIN	-0.033	0	-0.070	-0.079	
87	RSS_FILLLEVEL_DIGESTER	0	-0.137	-0.128	-0.130	
100	FHS_IC_FORMIATE	0.062	0.148	0.067		0.109
106	FSS_ARABINOSE	-0.058	0	0		0
114	FSS_IC_ACETATE	0.055	0	0		0
115	FSS_IC_FORMIATE	0.048	0	0		0
130	F_IC_FORMIATE_WGT	0.075	0	0.057		0.107
133	F_METH_SORP_ANION_WGT_EX	0.060	0	0.116		0.101
135	PUFFER_CAPACITY	0	0.157	0		0
136	PUFFER_CAPACITY_FIB	0.032	0.183	0.102		0
141	RHS_RESIN_USE	0	0	0.089	0.092	
143	RSS_DYE_FRACT	0	0	0	-0.150	
145	RSS_RESIN_USE	0	0.096	0	0	
150	FS_BULK_DENSITY	0	-0.130	0	0	
151	FS_GRAMMAGE	0	0	0.070	0.069	
166	DI_AV_PRES_SYS10	0	0	0.109	0.120	
173	DI_AV_PRES_SYS17	0	-0.125	-0.020	-0.049	
174	DI_AV_PRES_SYS18	-0.052	0	-0.026	-0.058	
175	DI_AV_PRES_SYS19	-0.049	0	-0.038	-0.071	
178	PRESS_VELOCITY	0	0	-0.083	-0.088	

D.3 B_w for MOR

Var.no.	Variable	MOR 7.4	MOR 6.4	MOR all	MOR proc	MOR raw
1	NOMINAL_THICKNESS			0.101	0.114	
5	REF_IN_USE	0.022	0.232	0.061	0.052	
11	MONTH_04	0.202	0	0.091	0.105	
14	MONTH_07	0.118	0	0.061	0.084	
15	MONTH_08	0	0	-0.097	-0.170	
18	MONTH_11	0.133	0	0.134	0	
37	WOOD_INPUT_TOTAL	-0.006	0	-0.029	-0.049	
38	FIB_PROC_LINE2_FRACT	-0.015	0	-0.042	-0.042	
41	RHS_C_PINE	0	0	0		0.031
43	RHS_C_TOTAL	-0.027	0	0		0.040
45	RHS_NC_ASH	0	0	0		-0.024
46	RHS_NC_BEECH	0	0	0		-0.023
53	RHS_NC_ASPEN	0	0	0		-0.028
54	RHS_NC_TOTAL	0.020	0	0		-0.041
56	RHS_XYLOSE	-0.001	0	0		0
58	RHS_MANNOSE	0	0	0		0.043
60	RHS_GLUCOSE	0	0	0		0.022
61	RHS_GALACT_ACID	0	0	0		0.022
62	RHS_HEMI_TOTAL	0	0	0		0.021
65	RSS_XYLOSE	0	0	0		0.027
67	RSS_MANNOSE	0.142	0	0.050		0.033
70	RSS_GALACT_ACID	0	0	0		0.030
71	RSS_HEMI_TOTAL	0	0	0.015		0.027
74	R_XYLOSE_WGT	-0.044	0	0		0
76	R_MANNOSE_WGT	0	0	0.060		0.050
78	R_GLUCOSE_WGT	0	0	0		0.029
80	R_HEMI_TOTAL_WGT	0	0	0		0.028
85	RHS_DISCHARGE_SCREW	0.037	0	0.050	0.024	
92	FHS_XYLOSE	-0.045	0	0		0
93	FHS_RHAMNOSE	-0.058	0	0		0
94	FHS_MANNOSE	0	0	0		0.034
99	FHS_IC_ACETATE	0	0.178	0		0
100	FHS_IC_FORMIATE	0.112	0	0.093		0
105	FHS_PH_UNEX	0	0	0		0.024
107	FSS_XYLOSE	-0.042	0	0		0
108	FSS_RHAMNOSE	-0.054	0	0		0
112	FSS_GALACT_ACID	-0.020	0	0		0
124	F_MANNOSE_WGT	0.180	0	0.138		0.037
129	F_IC_ACETATE_WGT	0	0	0.096		0
136	PUFFER_CAPACITY_FIB	0	0	0		-0.026
139	RHS_DYE_FRACT	-0.047	0	0	-0.040	
140	RHS_EMULSION_FRACT	-0.098	0	0	-0.113	
145	RSS_RESIN_USE	0	0.131	0.123	0.067	
147	FS_DRYER_TEMP_IN	0	0	0	0.040	
150	FS_BULK_DENSITY	-0.305	-0.311	-0.367	-0.446	
151	FS_GRAMMAGE	0	0	0.093	0.086	
157	DI_AV_PRES_SYS01	0	0.159	0.091	0.020	
166	DI_AV_PRES_SYS10	0.062	0	0.109	0.124	
167	DI_AV_PRES_SYS11	0.163	0	0.068	0.048	
172	DI_AV_PRES_SYS16	0	0.334	0.074	0.120	
178	PRESS_VELOCITY	-0.028	0	-0.075	-0.113	
179	PRESSFACTOR	0.040	0	0.032	0.072	

D.4 B_w for MOE

Var.no.	Variable	MOE 7.4	MOE 6.4	MOE all	MOE proc	MOE raw
1	NOMINAL_THICKNESS			0.183	0.186	
4	TEMP_OUTSIDE	0	0.029	0	0.100	
14	MONTH_07	0.034	0	0	0	
15	MONTH_08	0	0	-0.102	-0.181	
19	MONTH_12	0	0	-0.123	-0.129	
20	SEASON_SPRING	0	0	0.096	0.050	
22	SEASON_FALL	0	-0.029	0	0	
26	RESIN_TANK03	0	0.020	0	0	
36	HE_SS_FRACT	0.031	0.032	0.110	0.093	
37	WOOD_INPUT_TOTAL	0	0	0	-0.059	
41	RHS_C_PINE	0.039	0.020	0		0
54	RHS_NC_TOTAL	-0.030	0	0		0
65	RSS_XYLOSE	0	0.026	0		0
67	RSS_MANNOSE	0	0.032	0.135		0
69	RSS_GLUCOSE	0	0.038	0		0.089
71	RSS_HEMI_TOTAL	0	0.030	0		0
84	RHS_REFINER_DIGESTER_PRESSURE	0	-0.032	0	0	
86	RHS_REFINER_POWER_DRAIN	0	0	-0.097	-0.209	
87	RSS_FILLLEVEL_DIGESTER	0	-0.048	0	-0.086	
89	RSS_DISCHARGE_SCREW	0	0.029	0	0	
115	FSS_IC_FORMIATE	0	-0.032	0		0
121	F_ARABINOSE_WGT	0	0	0		0.090
124	F_MANNOSE_WGT	0	0	0.121		0.074
127	F_GALACT_ACID_WGT	0	0	0		0.079
130	F_IC_FORMIATE_WGT	0	-0.036	0		0
132	F_TOTAL_EXTRACT_WGT	0	-0.030	0		0
139	RHS_DYE_FRACT	0	-0.032	0	0	
141	RHS_RESIN_USE	0	0.023	0	0	
143	RSS_DYE_FRACT	0	-0.039	0	0	
145	RSS_RESIN_USE	0	0.041	0.155	0.096	
147	FS_DRYER_TEMP_IN	0	0	0	0.090	
150	FS_BULK_DENSITY	-0.048	-0.028	-0.169	-0.243	
151	FS_GRAMMAGE	0	0	0.139	0.095	
157	DI_AV_PRES_SYS01	0	0.040	0.165	0.120	
158	DI_AV_PRES_SYS02	0.032	0.026	0	0	
159	DI_AV_PRES_SYS03	0.035	0.026	0	0	
160	DI_AV_PRES_SYS04	0.027	0.021	0	0	
161	DI_AV_PRES_SYS05	0.028	0.019	0	0	
165	DI_AV_PRES_SYS09	0.037	0	0	0	
166	DI_AV_PRES_SYS10	0.036	0	0	0.033	
167	DI_AV_PRES_SYS11	0.055	0.027	0.126	0.085	
168	DI_AV_PRES_SYS12	0	0.027	0	0	
173	DI_AV_PRES_SYS17	0	0.033	0	0	
174	DI_AV_PRES_SYS18	0	0.033	0	0	
175	DI_AV_PRES_SYS19	0	0.026	0	0	
176	DI_AV_PRES_SYS21	0	0.033	0	0	
177	DI_AV_PRES_SYS22	0	0.040	0	0	
178	PRESS_VELOCITY	0	0	-0.129	-0.141	
179	PRESSFACTOR	0	0	0	0.042	

D.5 B_w for TS

Var.no.	Variable	TS 7.4	TS 6.4	TS all	TS proc	TS raw
1	NOMINAL_THICKNESS			-0.504	-0.404	
4	TEMP_OUTSIDE	-0.031	-0.141	0	-0.058	
6	PERCEP_FORMALDEHYDE	0.027	-0.110	0	0	
8	MONTH_01	0.052	0.076	0	0.094	
9	MONTH_02	0	0.090	0	0	
10	MONTH_03	0	-0.086	0	0	
11	MONTH_04	-0.024	0	0	0	
16	MONTH_09	0	0	0	-0.023	
19	MONTH_12	0	0.055	0	0	
22	SEASON_FALL	0	-0.062	0	0	
23	SEASON_WINTER	0.038	0.136	0.211	0.114	
32	SHIFT2	0	0	0	-0.024	
37	WOOD_INPUT_TOTAL	-0.033	0	0	-0.086	
40	RHS_C_SPRUCE	0	0	0		-0.058
43	RHS_C_TOTAL	0	0	0		-0.054
45	RHS_NC_ASH	0.052	0	0.048		0.082
46	RHS_NC_BEECH	0	0.118	0		0
51	RHS_NC_MAPLE	0	0.145	0		0
53	RHS_NC_ASPEN	0	-0.087	-0.105		0
54	RHS_NC_TOTAL	0	0	0		0.054
58	RHS_MANNOSE	-0.031	0	0		0
59	RHS_GALACTOSE	-0.026	0	0		0
72	RSS_EXTRACT_TOTAL	-0.029	0	0		0
76	R_MANNOSE_WGT	-0.031	0	0		0
77	R_GALACTOSE_WGT	-0.026	0	0		0
86	RHS_REFINER_POWER_DRAIN	0	0	0.037	0.030	
89	RSS_DISCHARGE_SCREW	0	0	0	-0.045	
90	RSS_REFINER_POWER_DRAIN	0	0.100	0	0.053	
92	FHS_XYLOSE	0.039	0	0.001		0
93	FHS_RHAMNOSE	0.039	0	0		0
122	F_XYLOSE_WGT	0.043	0	0.009		0
123	F_RHAMNOSE_WGT	0.039	0	0		0
137	PUFFER_CAPACITY_SS	-0.034	0	0		0
138	RESIN_FRACT	-0.026	-0.117	0	-0.101	
141	RHS_RESIN_USE	0	-0.089	0	-0.063	
144	RSS_EMULSION_FRACT	0	-0.055	0	0	
145	RSS_RESIN_USE	0	-0.079	0	0	
150	FS_BULK_DENSITY	0.025	0	0	0	
151	FS_GRAMMAGE	0.045	0.087	-0.238	-0.272	
155	FS_MC_FORMBAND	0.036	0	0.140	0.129	
156	FS_MC_MICROWAVE_CONTI	-0.042	0	-0.177	-0.150	
157	DI_AV_PRES_SYS01	0.026	0	0	0	
166	DI_AV_PRES_SYS10	-0.030	0	-0.080	-0.110	
168	DI_AV_PRES_SYS12	0.033	0	0	0	
169	DI_AV_PRES_SYS13	0.028	0	0	0	
170	DI_AV_PRES_SYS14	0.029	0	0	0	
171	DI_AV_PRES_SYS15	0.033	0	0	0	
173	DI_AV_PRES_SYS17	0.038	0.084	0	0.044	
174	DI_AV_PRES_SYS18	0.038	0.078	0	0	
175	DI_AV_PRES_SYS19	0.044	0.085	0	0	
176	DI_AV_PRES_SYS21	0.045	0	0.067	0.088	
177	DI_AV_PRES_SYS22	0.031	0	0	0	
178	PRESS_VELOCITY	-0.026	0	0.076	0.211	
179	PRESSFACTOR	0.025	0	0	0	

D.6 B_w for press factor

Var.no.	Variable	PRESS-FACTOR 7.4	PRESS-FACTOR 6.4	PRESS-FACTOR all	PRESS-FACTOR proc	PRESS-FACTOR raw
1	NOMINAL_THICKNESS			0.058	0.051	
5	REF_IN_USE	0.067	0.050	0.070	0.093	
8	MONTH_01	0.049	0.070	0.066	0.054	
9	MONTH_02	0	0	0	-0.026	
12	MONTH_05	-0.021	0	0	0	
13	MONTH_06	-0.085	-0.060	-0.092	-0.076	
14	MONTH_07	0.068	0.025	0.069	0.087	
16	MONTH_09	0	0	0	-0.091	
19	MONTH_12	0	0	0.037	0	
20	SEASON_SPRING	0	-0.072	-0.042	0	
23	SEASON_WINTER	0.052	0	0.047	0.062	
30	RESIN_RES02	0.103	0	0	0.043	
35	SHIFT5	0	0.041	0	0	
36	HE_SS_FRACT	0.068	0	0.058	0.044	
37	WOOD_INPUT_TOTAL	-0.143	-0.117	-0.156	-0.189	
38	FIB_PROC_LINE2_FRACT	-0.050	-0.034	-0.047	-0.080	
41	RHS_C_PINE	0	0	0		0.086
52	RHS_NC_OAK	0	0	0		-0.058
56	RHS_XYLOSE	0	0	-0.057		0
57	RHS_RHAMNOSE	0	-0.045	0		0
58	RHS_MANNOSE	0	0	0		0.074
73	R_ARABINOSE_WGT	0	0	0		-0.067
75	R_RHAMNOSE_WGT	0	-0.045	0		0
76	R_MANNOSE_WGT	0	0.013	0		0
81	R_TOTAL_EXTRACT_WGT	0	-0.060	-0.063		0
85	RHS_DISCHARGE_SCREW	-0.062	-0.062	-0.080	-0.064	
86	RHS_REFINER_POWER_DRAIN	-0.057	-0.073	-0.078	-0.081	
89	RSS_DISCHARGE_SCREW	-0.038	-0.040	-0.059	0	
91	FHS_ARABINOSE	0	-0.056	0		-0.061
94	FHS_MANNOSE	0	0	0		0.078
97	FHS_GALACT_ACID	0	-0.046	0		0
100	FHS_IC_FORMIATE	0.032	0	0		0.082
105	FHS_PH_UNEX	0	0	0		0.076
106	FSS_ARABINOSE	0	0	0		-0.116
115	FSS_IC_FORMIATE	0	0.032	0		0
120	FSS_PH_UNEX	-0.057	0	0		0
121	F_ARABINOSE_WGT	-0.069	-0.064	-0.067		-0.161
123	F_RHAMNOSE_WGT	0	0	0		0.160
127	F_GALACT_ACID_WGT	0	0	0		-0.061
130	F_IC_FORMIATE_WGT	0	0	0		0.172
135	PUFFER_CAPACITY	-0.059	0	-0.050		0
136	PUFFER_CAPACITY_FIB	-0.072	-0.048	-0.072		-0.142
137	PUFFER_CAPACITY_SS	0	0	0		-0.156
138	RESIN_FRACT	0	0	0	0.045	
141	RHS_RESIN_USE	0	0	0	-0.014	
143	RSS_DYE_FRACT	0.039	0	0	0	
145	RSS_RESIN_USE	0	0	0	-0.035	
147	FS_DRYER_TEMP_IN	-0.080	-0.084	-0.076	-0.102	
148	A241_TEMP_AV_T	-0.066	-0.073	-0.066	-0.080	
149	FS_DRYER_TEMP_OUT	-0.070	-0.076	-0.068	-0.085	
151	FS_GRAMMAGE	0.074	0.046	0.082	0.091	
156	FS_MC_MICROWAVE_CONTI	-0.050	0	0	-0.061	
157	DI_AV_PRES_SYS01	0	-0.062	0	0	
158	DI_AV_PRES_SYS02	0.010	0.033	0.022	0	
159	DI_AV_PRES_SYS03	0.011	0.032	0.021	0	
160	DI_AV_PRES_SYS04	0.014	0.039	0.024	0.072	
161	DI_AV_PRES_SYS05	0.012	0	0	0	
164	DI_AV_PRES_SYS08	0	0	0	-0.020	
165	DI_AV_PRES_SYS09	0	0	0	-0.024	
167	DI_AV_PRES_SYS11	0.040	0.025	0.045	0	
168	DI_AV_PRES_SYS12	0	0	0	-0.037	
169	DI_AV_PRES_SYS13	0.118	0.110	0.119	0.150	
170	DI_AV_PRES_SYS14	0.115	0.106	0.115	0.146	
171	DI_AV_PRES_SYS15	0.016	0	0	0	
173	DI_AV_PRES_SYS17	-0.060	-0.058	-0.056	-0.059	
174	DI_AV_PRES_SYS18	-0.059	0	-0.045	-0.050	

D.7 B_w for resin fraction

Var.no.	Variable	RESIN_ FRACT 7.4	RESIN_ FRACT 6.4	RESIN_ FRACT all	RESIN_ FRACT proc	RESIN_ FRACT raw
3	MOIST_OUTSIDE	0	0	0	-0.053	
5	REF_IN_USE	0	-0.072	-0.029	-0.028	
6	PERCEP_FORMALDEHYDE	-0.064	0	-0.052	-0.061	
7	EVAL_PRODUCTION	0	0	0.054	0.077	
8	MONTH_01	-0.051	0	-0.038	-0.052	
9	MONTH_02	0	0	0	-0.026	
10	MONTH_03	0.149	0.057	0.100	0.113	
11	MONTH_04	0	0.057	0	0	
13	MONTH_06	0.093	0.084	0.076	0.105	
15	MONTH_08	-0.059	0	0	0	
16	MONTH_09	-0.063	-0.090	-0.074	-0.096	
17	MONTH_10	-0.178	-0.063	-0.107	-0.128	
18	MONTH_11	0	0	0.045	0.054	
20	SEASON_SPRING	0.104	0.101	0.088	0.094	
21	SEASON_SUMMER	0.057	0	0.040	0.057	
22	SEASON_FALL	-0.122	-0.102	-0.089	-0.112	
23	SEASON_WINTER	-0.049	0	-0.048	-0.050	
36	HE_SS_FRACT	0	-0.071	-0.071	-0.083	
37	WOOD_INPUT_TOTAL	0.077	0.108	0.062	0.070	
40	RHS_C_SPRUCE	0.053	0	0		0
41	RHS_C_PINE	-0.068	0	-0.038		-0.147
42	RHS_C_LARCH	0	0	0		0.092
52	RHS_NC_OAK	0	0	0.042		0.091
64	RSS_ARABINOSE	0	0.064	0.046		0
67	RSS_MANNOSE	0.057	0	0.039		0
71	RSS_HEMI_TOTAL	0	0	0.040		0
72	RSS_EXTRACT_TOTAL	0.042	0	0		0
73	R_ARABINOSE_WGT	0	0.057	0.061		0
83	RHS_DWELL_TIME_DIGESTER	-0.093	0	-0.070	-0.088	
84	RHS_REFINER_DIGESTER_PRESSURE	-0.124	-0.108	-0.103	-0.127	
85	RHS_DISCHARGE_SCREW	0	0	0	0.042	
87	RSS_FILLLEVEL_DIGESTER	-0.077	-0.109	-0.070	-0.072	
89	RSS_DISCHARGE_SCREW	0	0	-0.056	-0.063	
90	RSS_REFINER_POWER_DRAIN	0	0	-0.048	-0.068	
91	FHS_ARABINOSE	0	0	0.040		0.216
92	FHS_XYLOSE	0	-0.054	-0.039		-0.039
93	FHS_RHAMNOSE	0	0	-0.042		0
94	FHS_MANNOSE	0	0	0		-0.138
96	FHS_GLUCOSE	0	-0.061	-0.031		-0.098
98	FHS_HEMI_TOTAL	0	0	-0.023		-0.059
100	FHS_IC_FORMIATE	0	0	0		-0.129
103	FHS_EXTRACT_TOTAL	0.073	0	0		0
120	FSS_PH_UNEX	0	0	-0.042		0
121	F_ARABINOSE_WGT	0	0	0.042		0.246
122	F_XYLOSE_WGT	0	-0.078	-0.038		-0.115
123	F_RHAMNOSE_WGT	-0.083	-0.066	-0.050		-0.110
126	F_GLUCOSE_WGT	0	-0.076	-0.038		-0.147
127	F_GALACT_ACID_WGT	0	0	0		0.129
128	F_HEMI_TOTAL_WGT	0	0	-0.028		0
135	PUFFER_CAPACITY	-0.051	0	-0.038		0
137	PUFFER_CAPACITY_SS	0	0.091	0.065		0.208
148	A241_TEMP_AV_T	0.038	0	0.033	0.035	
149	FS_DRYER_TEMP_OUT	0.041	0	0.033	0.035	
150	FS_BULK_DENSITY	0	0	0.042	0.060	
152	FS_SPRINKLING	-0.114	-0.098	-0.095	-0.115	
155	FS_MC_FORMBAND	-0.063	0	-0.039	-0.035	
156	FS_MC_MICROWAVE_CONTI	-0.093	0	-0.057	-0.061	
157	DI_AV_PRES_SYS01	0	0.093	0.059	0.060	
162	DI_AV_PRES_SYS06	-0.045	0	-0.028	-0.038	
163	DI_AV_PRES_SYS07	-0.050	0	-0.027	-0.046	
164	DI_AV_PRES_SYS08	-0.053	0	-0.035	-0.057	
169	DI_AV_PRES_SYS13	-0.044	0	-0.043	-0.039	
170	DI_AV_PRES_SYS14	-0.039	0	-0.036	-0.030	
176	DI_AV_PRES_SYS21	0	0	0.037	0	
179	PRESSFACTOR	0	0	0	-0.022	

E. Frequency of occurrence of variables in models

Var.no.	Variable	IB	IB single	SS	MOR	MOE	TS	PF[29]	RF[30]	sum
1	NOMINAL_THICKNESS	2	0	2	2	2	2	2	0	12
2	RECIPE_NO	0	0	0	0	0	0	0	0	0
3	MOIST_OUTSIDE	0	0	0	0	0	0	0	1	1
4	TEMP_OUTSIDE	0	0	0	0	2	3	0	0	5
5	REF_IN_USE	0	0	0	4	0	0	4	3	11
6	PERCEP_FORMALDEHYDE	2	0	2	0	0	2	0	3	9
7	EVAL_PRODUCTION	0	0	0	0	0	0	0	2	2
8	MONTH_01	0	1	0	0	0	3	4	3	11
9	MONTH_02	0	0	0	0	0	1	1	1	3
10	MONTH_03	0	0	0	0	0	1	0	4	5
11	MONTH_04	1	1	0	3	0	1	0	1	7
12	MONTH_05	0	2	0	0	0	0	1	0	3
13	MONTH_06	0	1	0	0	0	0	4	4	9
14	MONTH_07	2	5	2	3	1	0	4	0	17
15	MONTH_08	0	3	2	2	2	0	0	1	10
16	MONTH_09	0	0	0	0	0	1	1	4	6
17	MONTH_10	1	0	0	0	0	0	0	4	5
18	MONTH_11	0	0	0	2	0	0	0	2	4
19	MONTH_12	0	0	0	0	2	1	1	0	4
20	SEASON_SPRING	0	2	0	0	2	0	2	4	10
21	SEASON_SUMMER	0	0	0	0	0	0	0	3	3
22	SEASON_FALL	1	0	0	0	1	1	0	4	7
23	SEASON_WINTER	0	0	0	0	0	4	3	3	10
24	RESIN_TANK01	0	0	0	0	0	0	0	0	0
25	RESIN_TANK02	0	0	1	0	0	0	0	0	1
26	RESIN_TANK03	0	0	0	0	1	0	0	0	1
27	RESIN_TANK04	0	0	0	0	0	0	0	0	0
28	RESIN_TANK05	0	0	1	0	0	0	0	0	1
29	RESIN_RES01	0	0	0	0	0	0	0	0	0
30	RESIN_RES02	0	0	0	0	0	0	2	0	2
31	SHIFT1	0	0	0	0	0	0	0	0	0
32	SHIFT2	1	0	2	0	0	1	0	0	4
33	SHIFT3	0	0	2	0	0	0	0	0	2
34	SHIFT4	0	0	0	0	0	0	0	0	0
35	SHIFT5	1	1	0	0	0	0	1	0	3
36	HE_SS_FRACT	2	0	0	0	4	0	3	3	12
37	WOOD_INPUT_TOTAL	0	2	2	3	1	2	4	4	18
38	FIB_PROC_LINE2_FRACT	0	2	0	3	0	0	4	0	9
39	RHS_C_FIR	0	0	1	0	0	0	0	0	1
40	RHS_C_SPRUCE	0	0	0	0	0	1	0	1	2
41	RHS_C_PINE	0	0	0	1	2	0	1	3	7
42	RHS_C_LARCH	0	0	0	0	0	0	0	1	1
43	RHS_C_TOTAL	0	0	0	2	0	1	0	0	3
44	RHS_NC_ALDER	0	0	0	0	0	0	0	0	0
45	RHS_NC_ASH	0	1	0	1	0	3	0	0	5
46	RHS_NC_BEECH	0	0	0	1	0	1	0	0	2
47	RHS_NC_CHERRY	0	0	0	0	0	0	0	0	0
48	RHS_NC_ELM	0	0	0	0	0	0	0	0	0
49	RHS_NC_HORNBEAM	0	0	0	0	0	0	0	0	0
50	RHS_NC_LOCUST	0	0	0	0	0	0	0	0	0
51	RHS_NC_MAPLE	0	0	1	0	0	1	0	0	2
52	RHS_NC_OAK	0	0	0	0	0	0	1	2	3
53	RHS_NC_ASPEN	0	1	0	1	0	2	0	0	4
54	RHS_NC_TOTAL	0	0	0	2	1	1	0	0	4
55	RHS_ARABINOSE	0	0	0	0	0	0	0	0	0
56	RHS_XYLOSE	0	2	1	1	0	0	1	0	5
57	RHS_RHAMNOSE	0	0	1	0	0	0	1	0	2
58	RHS_MANNOSE	0	0	0	1	0	1	1	0	3
59	RHS_GALACTOSE	0	0	0	0	0	1	0	0	1
60	RHS_GLUCOSE	0	0	1	1	0	0	0	0	2
61	RHS_GALACT_ACID	0	1	0	1	0	0	0	0	2
62	RHS_HEMI_TOTAL	0	1	1	1	0	0	0	0	3
63	RHS_EXTRACT_TOTAL	0	0	0	0	0	0	0	0	0
64	RSS_ARABINOSE	0	0	0	0	0	0	0	2	2

[29] Press factor (PRESSFACTOR)
[30] Resin fraction (RESIN_FRACT)

Var.no.	Variable	IB	IB single	SS	MOR	MOE	TS	PF[29]	RF[30]	sum
65	RSS_XYLOSE	0	0	0	1	1	0	0	0	2
66	RSS_RHAMNOSE	0	0	0	0	0	0	0	0	0
67	RSS_MANNOSE	0	0	0	3	2	0	0	2	7
68	RSS_GALACTOSE	0	0	0	0	0	0	0	0	0
69	RSS_GLUCOSE	0	0	0	0	2	0	0	0	2
70	RSS_GALACT_ACID	0	0	0	1	0	0	0	0	1
71	RSS_HEMI_TOTAL	0	0	0	2	1	0	0	1	4
72	RSS_EXTRACT_TOTAL	0	0	0	0	0	1	0	1	2
73	R_ARABINOSE_WGT	0	0	0	0	0	0	1	2	3
74	R_XYLOSE_WGT	2	1	0	1	0	0	0	0	4
75	R_RHAMNOSE_WGT	0	0	0	0	0	0	1	0	1
76	R_MANNOSE_WGT	0	0	0	2	0	1	1	0	4
77	R_GALACTOSE_WGT	0	0	0	0	0	1	0	0	1
78	R_GLUCOSE_WGT	0	0	0	1	0	0	0	0	1
79	R_GALACT_ACID_WGT	0	1	0	0	0	0	0	0	1
80	R_HEMI_TOTAL_WGT	0	0	0	1	0	0	0	0	1
81	R_TOTAL_EXTRACT_WGT	2	1	0	0	0	0	2	0	5
82	RHS_FILLLEVEL_DIGESTER	0	0	0	0	0	0	0	0	0
83	RHS_DWELL_TIME_DIGESTER	0	0	1	0	0	0	0	3	4
84	RHS_REFINER_DIGESTER_PRESSURE	1	1	0	0	1	0	0	4	7
85	RHS_DISCHARGE_SCREW	1	2	3	3	0	0	4	1	14
86	RHS_REFINER_POWER_DRAIN	0	1	3	0	2	2	4	0	12
87	RSS_FILLLEVEL_DIGESTER	0	0	3	0	2	0	0	4	9
88	RSS_REFINER_DIGESTER_PRESSURE	0	0	0	0	0	0	0	0	0
89	RSS_DISCHARGE_SCREW	1	2	0	0	1	1	3	2	10
90	RSS_REFINER_POWER_DRAIN	0	0	0	0	0	2	0	2	4
91	FHS_ARABINOSE	0	1	0	0	0	0	2	2	5
92	FHS_XYLOSE	0	0	0	1	0	2	0	3	6
93	FHS_RHAMNOSE	0	0	0	1	0	1	0	1	3
94	FHS_MANNOSE	0	0	0	1	0	0	1	1	3
95	FHS_GALACTOSE	0	2	0	0	0	0	0	0	2
96	FHS_GLUCOSE	0	0	0	0	0	0	0	3	3
97	FHS_GALACT_ACID	0	1	0	0	0	0	1	0	2
98	FHS_HEMI_TOTAL	0	1	0	0	0	0	0	2	3
99	FHS_IC_ACETATE	0	0	0	1	0	0	0	0	1
100	FHS_IC_FORMIATE	0	0	4	2	0	0	2	1	9
101	FHS_IC_CHLORIDE	0	0	0	0	0	0	0	0	0
102	FHS_IC_NITRATE	0	0	0	0	0	0	0	0	0
103	FHS_EXTRACT_TOTAL	0	0	0	0	0	0	0	1	1
104	FHS_METH_SORP_ANION_EX	0	0	0	0	0	0	0	0	0
105	FHS_PH_UNEX	0	0	0	1	0	0	1	0	2
106	FSS_ARABINOSE	0	0	1	0	0	0	1	0	2
107	FSS_XYLOSE	0	0	0	1	0	0	0	0	1
108	FSS_RHAMNOSE	0	0	0	1	0	0	0	0	1
109	FSS_MANNOSE	1	0	0	0	0	0	0	0	1
110	FSS_GALACTOSE	0	0	0	0	0	0	0	0	0
111	FSS_GLUCOSE	0	0	0	0	0	0	0	0	0
112	FSS_GALACT_ACID	0	0	0	1	0	0	0	0	1
113	FSS_HEMI_TOTAL	0	0	0	0	0	0	0	0	0
114	FSS_IC_ACETATE	0	0	1	0	0	0	0	0	1
115	FSS_IC_FORMIATE	0	0	1	0	1	0	1	0	3
116	FSS_IC_CHLORIDE	0	0	0	0	0	0	0	0	0
117	FSS_IC_NITRATE	0	0	0	0	0	0	0	0	0
118	FSS_EXTRACT_TOTAL	1	1	0	0	0	0	0	0	2
119	FSS_METH_SORP_ANION_EX	0	0	0	0	0	0	0	0	0
120	FSS_PH_UNEX	0	0	0	0	0	0	1	1	2
121	F_ARABINOSE_WGT	0	0	0	0	1	0	4	2	7
122	F_XYLOSE_WGT	0	2	0	0	0	2	0	3	7
123	F_RHAMNOSE_WGT	0	2	0	0	0	1	1	4	8
124	F_MANNOSE_WGT	0	0	0	3	2	0	0	0	5
125	F_GALACTOSE_WGT	0	1	0	0	0	0	0	0	1
126	F_GLUCOSE_WGT	0	0	0	0	0	0	0	3	3
127	F_GALACT_ACID_WGT	0	0	0	0	1	0	1	1	3
128	F_HEMI_TOTAL_WGT	0	0	0	0	0	0	0	1	1
129	F_IC_ACETATE_WGT	0	0	0	1	0	0	0	0	1
130	F_IC_FORMIATE_WGT	0	0	3	0	1	0	1	0	5
131	F_IC_CHLORIDE_WGT	0	0	0	0	0	0	0	0	0
132	F_TOTAL_EXTRACT_WGT	0	0	0	0	1	0	0	0	1
133	F_METH_SORP_ANION_WGT_EX	0	0	3	0	0	0	0	0	3
134	F_PH_WGT_UNEX	0	0	0	0	0	0	0	0	0
135	PUFFER_CAPACITY	0	0	1	0	0	0	2	2	5

Var.no.	Variable	IB	IB single	SS	MOR	MOE	TS	PF[29]	RF[30]	sum
136	PUFFER_CAPACITY_FIB	0	0	3	1	0	0	4	0	8
137	PUFFER_CAPACITY_SS	0	0	0	0	0	1	1	3	5
138	RESIN_FRACT	0	1	0	0	0	3	1	0	5
139	RHS_DYE_FRACT	0	0	0	2	1	0	0	0	3
140	RHS_EMULSION_FRACT	0	0	0	2	0	0	0	0	2
141	RHS_RESIN_USE	0	0	2	0	1	2	1	0	6
142	RHS_UREA_FRACT	0	0	0	0	0	0	0	0	0
143	RSS_DYE_FRACT	1	0	1	0	1	0	1	0	4
144	RSS_EMULSION_FRACT	0	0	0	0	0	1	0	0	1
145	RSS_RESIN_USE	0	1	1	3	3	1	1	0	10
146	RSS_UREA_FRACT	0	0	0	0	0	0	0	0	0
147	FS_DRYER_TEMP_IN	0	2	0	1	1	0	4	0	8
148	A241_TEMP_AV_T	0	3	0	0	0	0	4	3	10
149	FS_DRYER_TEMP_OUT	0	2	0	0	0	0	4	3	9
150	FS_BULK_DENSITY	0	0	1	4	4	1	0	2	12
151	FS_GRAMMAGE	0	1	2	2	2	4	4	0	15
152	FS_SPRINKLING	0	2	0	0	0	0	0	4	6
153	FS_MAT_TEMP	0	0	0	0	0	0	0	0	0
154	FS_MC_BANDSCALE	0	0	0	0	0	0	0	0	0
155	FS_MC_FORMBAND	0	1	0	0	0	3	0	3	7
156	FS_MC_MICROWAVE_CONTI	0	1	0	0	0	3	2	3	9
157	DI_AV_PRES_SYS01	0	0	0	3	3	1	1	3	11
158	DI_AV_PRES_SYS02	2	2	0	0	2	0	3	0	9
159	DI_AV_PRES_SYS03	2	2	0	0	2	0	3	0	9
160	DI_AV_PRES_SYS04	2	2	0	0	2	0	4	0	10
161	DI_AV_PRES_SYS05	2	2	0	0	2	0	1	0	7
162	DI_AV_PRES_SYS06	0	1	0	0	0	0	0	3	4
163	DI_AV_PRES_SYS07	0	0	0	0	0	0	0	3	3
164	DI_AV_PRES_SYS08	1	2	0	0	0	0	1	3	7
165	DI_AV_PRES_SYS09	1	1	0	0	1	0	1	0	4
166	DI_AV_PRES_SYS10	0	0	2	3	2	3	0	0	10
167	DI_AV_PRES_SYS11	0	1	0	3	4	0	3	0	11
168	DI_AV_PRES_SYS12	0	0	0	0	1	1	1	0	3
169	DI_AV_PRES_SYS13	0	1	0	0	0	1	4	3	9
170	DI_AV_PRES_SYS14	0	2	0	0	0	1	4	3	10
171	DI_AV_PRES_SYS15	0	0	0	0	0	1	1	0	2
172	DI_AV_PRES_SYS16	0	0	0	3	0	0	0	0	3
173	DI_AV_PRES_SYS17	0	0	3	0	1	3	4	0	11
174	DI_AV_PRES_SYS18	0	0	3	0	1	2	3	0	9
175	DI_AV_PRES_SYS19	0	0	3	0	1	2	0	0	6
176	DI_AV_PRES_SYS21	1	0	0	0	1	3	0	1	6
177	DI_AV_PRES_SYS22	0	0	0	0	1	1	0	0	2
178	PRESS_VELOCITY	0	2	2	3	2	3	0	0	12
179	PRESSFACTOR	0	2	0	3	1	1	0	1	8
									Total	**750**