

Simulation of a real-time process adaptation in the manufacture of high-density fibreboards using multivariate regression analysis and feedforward control

Martin Riegler · Bernhard Spangl · Martin Weigl ·
Rupert Wimmer · Ulrich Müller

Received: 8 November 2012 / Published online: 20 July 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The industrial manufacturing of wood-based panels has become a highly technological process, where all parameters have to be perfectly adjusted to manufacture products of high quality. However, variations caused by differing wood characteristics as well as variations of single process parameters can cause out-of-control events. These undesirable events can be diminished by monitoring and controlling the entire manufacturing process using multivariate statistical techniques. Hence, a real-time process adaptation of an industrial scale fibreboard manufacturing process was simulated. Regression results revealed a mean normalised root mean squared error of prediction of 4.6 %, when predicting the internal bond strength of fibreboards. The regression model is regularly validated and, if necessary, recalibrated using the offline determined board properties (feedback control). Consequently, the process can immediately be adapted as soon as the board is produced (feedforward control). The investigations resulted in reliable models and revealed high potential for permanent industrial implementation.

M. Riegler (✉) · M. Weigl · U. Müller
Wood K plus, Competence Centre for Wood Composites and Wood Chemistry,
Konrad Lorenzstrasse 24, 3430 Tulln, Austria
e-mail: m.riegler@kplus-wood.at

M. Riegler · R. Wimmer · U. Müller
Institute of Wood Technology and Renewable Resources, University of Natural Resources
and Life Sciences, Konrad Lorenzstrasse 24, 3430 Tulln, Austria

B. Spangl
Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences,
Peter-Jordan-Straße 82, 1190 Vienna, Austria

Introduction

The supply of wood as most important raw material for the wood-based panels industry has been increasingly challenging, due to the rising demand for wood used as an energy resource (Marutzky 2006). A more efficient utilisation of raw materials in combination with lower energy consumption is both seen as strategies to address this situation. In this respect, the quality control of industrially produced wood-based panels has become a major prerequisite in achieving higher efficiency levels. During the past decades, procedures for testing of physical, chemical and biological properties of wood-based panels under standardised conditions have been developed (EN 622-5 2006). Such procedures are normally used for evaluating the adherence of the required minimum product quality. Additionally, data from manufacturing process parameters as well as technological properties of raw materials are determined. These data can be used for building models, which are capable of predicting the properties of wood-based panels using multivariate regression analysis (Andre et al. 2010; Weigl et al. 2012). Once models of high predictive capability are developed, the process of manufacturing wood-based panels can be automatically adapted in real time to reduce process variability and out-of-control events.

While real-time adaptation is commonly used in chemometrics (Danzer et al. 2001) for sectors such as biotechnology (Vojinović et al. 2006) or chemical engineering, it is rarely used in wood industry. Bernardy and Scherff (1997) developed a software programme that is able to predict final properties of wood-based panels. Cook and Chiu (1997) predicted internal bond strength (IB) of particleboards using a radial basis function (RBF) neural network, resulting in an error of prediction of 12.5 %. Andre et al. (2008) concluded, after comparing four different algorithms [RBF neural network, partial least squares regression (PLSR), orthogonal PLSR and supervised probabilistic principal component analysis (SPPCA)] for predicting IB in medium density fibreboards (MDF), that SPPCA performed best (error of prediction = 5.9 %). PLSR performed second best with an error of prediction of 6.18 %. Esteban et al. (2011) used an artificial neural network with feedforward multilayer perceptron to predict the bonding quality of plywood with an overall average accuracy of 93 % (number of correctly classified testing samples divided by the total number of testing samples in per cent). Predicting strength properties of particleboards using a combination of algorithms (multiple linear regression (MLR), ridge regression, PLSR and neural network) resulted in errors of prediction of 9.0 % for predicting IB and 6.9 % for predicting the bending strength (MOR) of particleboards (Andre et al. 2010). Here, an error of prediction of 15 % was used as a threshold for recalibrating the PLSR model. Young et al. (2008) used MLR and quantile regression for modelling the IB of MDF. Dolezel-Horwath et al. (2005) obtained an error of prediction of about 11 % when predicting IB of hardboards using UV–Vis spectra of fibres. The database management was frequently realised by relational databases using SQL programming language (Andre et al. 2008; Bernardy and Scherff 1997; Young and Guess 2002).

The occurrence of missing values is a common problem when analysing data sets from industrial processes. As the deletion of entire observations would lead to

considerable data loss, algorithms such as the non-linear iterative partial least squares (NIPALS) algorithm (Nelson et al. 1996) or the expectation–maximisation algorithm (Zeng 2011) are appropriate for replacing missing values. Detailed information about treating missing values in statistical analysis can be found in Little and Rubin (2002).

Advantages of PLSR compared to MLR lie in the possibility to analyse highly multi-collinear, noisy and numerous predictor variables as well as to simultaneously model several predicted variables (Wold et al. 2001). Commonly used algorithms for PLSR are the simple partial least squares (SIMPLS) algorithm (introduced by De Jong 1993), which maximises the variance–covariance between predictor and predicted variables, and the NIPALS algorithm. Although the SIMPLS algorithm is faster than NIPALS, both will lead to the same result in the case of using a single predicted variable (Wise et al. 2006). Ideally, all relevant product properties should be optimised simultaneously (multivariate case). However, if properties of panels are not correlated, they should not be analysed together, as this would cause a higher number of selected significant variables, lower predictability and hence a more difficult interpretation of results (Wold et al. 2001).

The exclusion of variables that do not contribute to the explanation of the model improved interpretability and predictability of the PLSR model (Mehmood et al. 2012). When predicting properties of wood-based panels, the selection of significant process parameters has been carried out using F tests paired with the experience of an operator (Steffen et al. 2001), genetic algorithms (GA) in combination with PLSR (Andre et al. 2008) and MLR (Andre et al. 2010) as well as principal component analysis (PCA) (Clapp Jr et al. 2008). Mehmood et al. (2012) gave an overview of possible methods for variable selection with PLSR. Nørgaard et al. (2000) presented an iterative procedure for selecting intervals of spectra, called interval partial least squares regression (IPLS), from near infrared spectroscopy (NIR).

To improve the predictability and interpretability of established models, information on raw material properties should be recorded in more detail (Weigl et al. 2012). In particular, Weigl et al. (2012) detected significant influences of raw material parameters when modelling thickness swelling of high-density fibreboards (HDF) using PLSR. A major factor for the mechanical properties of wood-based panels is the distribution of resin on particles (Wilson and Kraemer 1976). The fluorescent dye brilliant sulphaflavine can serve as indicator for a semi-online measurement of the distribution and size of resin droplets (Riegler et al. 2012). Similarly, NIR only or in combination with UV–Vis spectroscopy can be used for determining wood characteristics. These characteristics are used to model strength properties of particleboards with PCA and PLSR (Dolezel-Horwath et al. 2005; Rials et al. 2002; Sjöblom et al. 2004). Sjöblom et al. (2004) detected cost savings of 2.8 euros/m³ due to a lower variability of particleboard properties.

As high product variability entails higher safety margins and thus higher energy and raw material input, product variability should be kept to a minimum. This can be achieved by defining optimisation problems and using feedback control to adapt process parameters (Chachuat et al. 2009). Hence, the objective of the present study was to evaluate the potential of adapting process parameters in real time to minimise product variability using multivariate statistical methods.

Materials and methods

Data logging

The industrial production of 7-mm-thick HDF with a target density of 875 kg/m³ was investigated. Therefore, 804 process and raw material parameters were recorded every 20 seconds. Figure 1 shows the approach of the applied process adaptation with the four main steps, i.e. calibration (I), validation (II), prediction (III) and feedforward adaptation (IV). Data records had to be time corrected by defining time-lags between consecutive process sections before storing in a “real-time database” (Fig. 1). As the internal bond strength (IB) is one of the most important board parameters in the industrial production of fibreboards (Dunky and Niemz 2002), it was used as predicted variable to exemplify the method used in the present study. At four-hour intervals, the IB of panels was determined offline, following EN 319 (1993) and EN 326-1 (1994). IB data were stored in an “offline database”. To simulate the adaptation of the manufacturing process of HDF, statistical models were calculated using PLSR analysis. Thus, 440 data records of

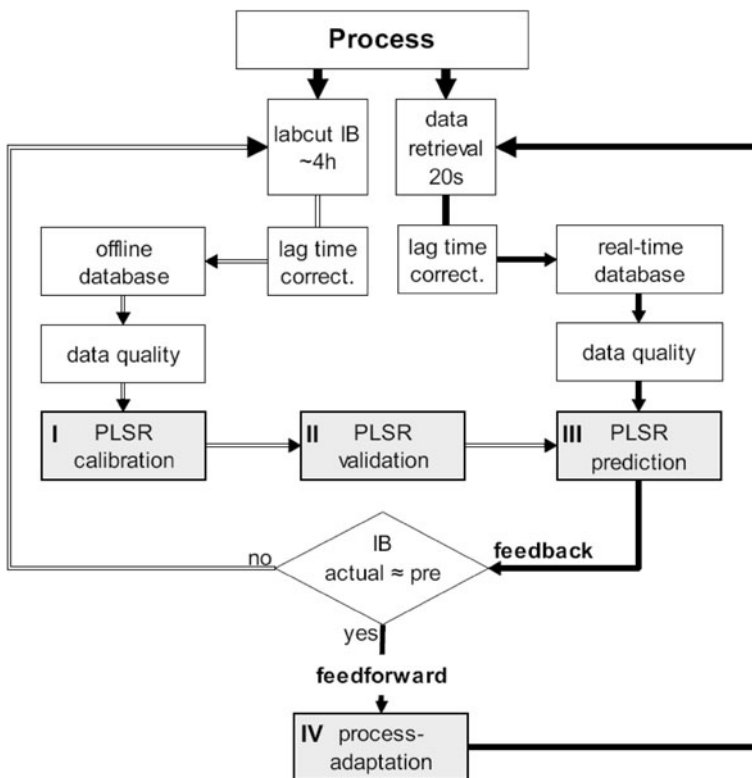


Fig. 1 Approach of process adaptation using feedback and feedforward control of offline and online determined parameters using 4 main steps: calibration (I), validation (II), prediction (III) and feedforward adaptation (IV)

offline data, representing the recent 6 months of production, were merged with the time-lag-corrected real-time database records (data fusion). These 440 data records were divided by alternately selecting every 11th data record for evaluating the error of prediction (40) and the rest for calibration (400). Statistical analyses were carried out with MATLAB (version R2010b) and PLS Toolbox from Eigenvector Research Inc.

Pre-processing of data

After the above-mentioned data fusion step, predictor variables were grouped according to their respective process sections: silo, digester, refiner, gluing, drying, forming, pressing and laboratory. Variables with values outside of a predefined realistic range were removed to exclude corrupt sensor signals. In addition, predictor variables with more than 50 % missing values were excluded as well. Predictor variables with less than 50 % missing values were imputed by iteratively fitting a PCA model to the data. Missing values were initially replaced by the arithmetic mean of their corresponding variables. From the obtained data matrix, the covariance matrix was calculated. This covariance matrix was factorised using the singular value decomposition to impute originally missing values with values that are most consistent with the loadings of the PCA model. This routine was successively carried out (maximum 100 times) until the change in the replaced values dropped below the threshold of 1E-6, using PLS Toolbox (Wise et al. 2012). Predictor variables that were technologically related were manually clustered into groups (refining, gluing, drying, etc.), to improve imputation of missing values. If single observations of the predicted variable (IB) were missing, the entire respective observation was removed from the data set. Predictor variables with a coefficient of variation below 0.2 % were excluded, as these variables did not contribute to the explanation of the predicted IB. Afterwards, the filtered data ($X^{(\text{orig})}$) were standardised to $X^{(\text{sc})}$ (scaled X -matrix) that comprises variables with an arithmetic mean ($\bar{x}_{.j}$) of 0 and a standard deviation ($s_{.j}$) of 1 by

$$X_{ij}^{(\text{sc})} = \frac{X_{ij}^{(\text{orig})} - \bar{x}_{.j}}{s_{.j}}, \quad i = 1 \dots n, j = 1 \dots m, \quad (1)$$

where m is the number of variables and n the number of observations.

Regression analysis with PLSR

Using PLSR and the SIMPLS algorithm (Wold et al. 2001) as basis for adapting the process of manufacturing fibreboards, the scaled X -matrix $X^{(\text{sc})}$ was transformed into X -scores T and X -loadings P aiming at minimising the residuals E by

$$X^{(\text{sc})} = TP' + E. \quad (2)$$

The scores matrix T was used to predict the scaled Y -matrix using Y -loadings C and minimised residuals F by

$$Y^{(c)} = TC' + F. \quad (3)$$

These transformations were calculated using the weights matrix W as linear combinations of $X^{(sc)}$ by

$$T = X^{(sc)}W(P'W)^{-1}, \quad (4)$$

resulting in regression coefficients B by calculating

$$B = W(P'W)^{-1}C'. \quad (5)$$

Calibration and validation

Calibration (1st PLSR model) was carried out with the offline database using 400 data records from the most recent 6 months (Fig. 1, I). To evaluate the predictability of PLSR models, the root mean squared error of calibration (RMSEC) (Formula 6) and the root mean squared error of cross validation (RMSECV) (Formula 7) were calculated, where the deviation of the predicted value ($\hat{Y}^{(c)}, \hat{Y}^{(cv)}$) from the actual value (Y) was of interest. The RMSECV was determined by applying a b -fold cross validation, where b is the number of contiguous blocks, and n/b is the number of observations per block. Here, n was 400 and b was set to 10. In addition, RMSECV values were standardised for better comparison by dividing them by the absolute arithmetic mean of the actual values (mean normalised root mean squared error of cross validation (MNRMSECV) in Formula 8).

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i^{(c)} - Y_i)^2}{n}} \quad (6)$$

$$\text{RMSECV} = b^{-1} * \sum_{j=1}^b \sqrt{\frac{b * \sum_{i=1}^{n/b} (\hat{Y}_i^{(cv)} - Y_i)^2}{n}} \quad (7)$$

$$\text{MNRMSECV} = \text{RMSECV} * \left(\frac{\sum_{i=1}^n Y_i}{n} \right)^{-1} * 100 \quad (8)$$

The 1st calculated PLSR model used all predictor variables ($X^{(sc)}$) and an optimum number of latent variables (LV) by searching for a significant change (see below) in the RMSEC and RMSECV using the function choosecomp (Wise et al. 2012). The function's principle is described in Formulas 9–11. The first significant drop in RMSEC is indicated by

$$\text{knee} = 1 + \min_{l=1, \dots, k-2} \left\{ l : \frac{(\text{RMSEC}_l)^{1+\alpha}}{\text{RMSEC}_{l+1} + \varepsilon} - \frac{(\text{RMSEC}_{l+1})^{1+\alpha}}{\text{RMSEC}_{l+2} + \varepsilon} > 0 \right\} \quad (9)$$

where knee is the minimum subscript of the positive integers, α is a sensitivity parameter (set to 0.2) for detecting a drop in RMSEC, k is the maximum number of LV and $\varepsilon > 0$ (small to avoid division by zero). The resulting suggestion of LV (knee) was altered if the addition or removal of further LV resulted in lower

RMSECV values, using the mean of the absolute difference of adjacent relative RMSECV values as threshold. In this case, RMSECV values were searched for drops up to the suggested number of LV by

$$pc = 1 + \max_{l=1, \dots, p} \left\{ l : \frac{\text{RMSECV}_{l+1} - \text{RMSECV}_l}{\max(\text{RMSECV})} < \frac{-\frac{1}{k-1} \sum_{i=1}^{k-1} |\text{RMSECV}_{i+1} - \text{RMSECV}_i|}{\max(\text{RMSECV})} \right\} \quad (10)$$

where p is equal to knee, and pc is the maximum subscript of the positive integers. Afterwards, the final suggestion was made by comparing pc with the change of the remaining number of LV by

$$\text{optlv} = pc - 1 + \min_{l=1, \dots, k-pc} \left\{ l : \frac{\text{RMSECV}_{pc+l} - \text{RMSECV}_{pc+l-1}}{\max(\text{RMSECV})} > \frac{-\frac{1}{k-1} \sum_{i=1}^{k-1} |\text{RMSECV}_{i+1} - \text{RMSECV}_i|}{\max(\text{RMSECV})} \right\} \quad (11)$$

where optlv is the minimum subscript of the positive integers. Finally, optlv was used as the optimum number of LV for the 1st PLSR model.

To improve the predictability of this 1st PLSR model, the most important predictor variables were selected using IPLS (Nørgaard et al. 2000). In this iterative process of variable selection, a 2nd PLSR model (Fig. 1, II—validation) was calculated by successively selecting predictor variables that minimised the RMSECV (determined by cross validation with 10 contiguous blocks). The number of LV for the 2nd PLSR model was equal to the number of LV that was used to calculate the final IPLS iteration.

Prediction

Estimated regression coefficients (B) from the calibrated PLSR model were used to simulate a real-time prediction of the IB. The simulation was carried out using collected real-time data of predictor variables obtained over 2 days of production ($X^{(r)}$), retrieved at 20-second intervals (Fig. 1, III). Missing values were replaced according to the algorithm described in section pre-processing of data, using the most recent 100 real-time records. Time-lag correction as well as data quality improvements were applied to the real-time data, analogous to the steps mentioned above. The prediction of the IB ($\hat{Y}^{(p)}$) with new real-time data was achieved by using the linear regression equation

$$\hat{Y}^{(p)} = X^{(r)} * B. \quad (12)$$

To evaluate the predictability of the PLSR model in real time, the predicted values $\hat{Y}^{(p)}$ were compared to the newly measured actual values ($Y^{(\text{new})}$) that were obtained from the offline database. Thus, the root mean squared error of prediction (RMSEP) was calculated using every 11th data record from the most recent offline database, resulting in d observations (Formula 13). Here, d was 40.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^d \left(\hat{Y}_i^{(p)} - Y_i^{(\text{new})} \right)^2}{d}} \quad (13)$$

Recalibration

For the evaluation of the feedback control (“actual \approx pre” in Fig. 1), two thresholds were defined to indicate whether a recalibration was necessary. If the average of three consecutive predictions of $\hat{Y}^{(p)}$ (to exclude short-term variations) was out of the range of ± 3 times the standard deviation of $\hat{Y}^{(c)}$, a negative decision was obtained from the feedback control and the PLSR model was recalibrated using the most recent offline data. For a definition of the confidence interval see Wold et al. (2001). Additionally, a recalibration was carried out if the mean normalised root mean squared error of prediction (MNRMSEP) values (Formula 14) were $>5\%$. Recalibration was neglected if the MNRMSEP value of the newly calibrated PLSR model was higher than the MNRMSEP value of the previous model.

$$\text{MNRMSEP} = \text{RMSEP} * \left(\left| \frac{\sum_{i=1}^d Y_i^{(\text{new})}}{d} \right| \right)^{-1} * 100 \quad (14)$$

Process adaptation

If the result of the feedback function was positive (Fig. 1), theoretical values for predictor variables of the current model were calculated, which should result in fibreboards with a defined target IB ($\hat{Y}^{(t)}$) of 1.70 N/mm² (Fig. 1, IV). In particular, the function

$$\tilde{X}^{(a)} = \arg \min_{X^{(a)}} \left(X^{(a)} * B^{(a)} + X^{(u)} * B^{(u)} - \hat{Y}^{(t)} \right)^2 \quad (15)$$

was minimised [using the trust-region-reflective algorithm (Coleman and Li 1994)] while searching for controllable variables ($X^{(a)}$) and using the predefined limits for each selected variable as constraints. Nonadjustable predictor variables and predictor variables that had been changed manually by the operator (uncontrollable variables $X^{(u)}$) were not allowed to be adapted by the function and were defined as absolute terms. $B^{(a)}$ and $B^{(u)}$ were the corresponding regression coefficients of the controllable and uncontrollable variables, respectively. As abrupt changes in specific processes could destabilise the overall manufacturing process, controllable variables in $\tilde{X}^{(a)}$ were not allowed to vary by more than \pm the standard deviation of the same variables in $X^{(\text{sc})}$.

Results and discussion

Model calibration and validation

The simulated process adaptation was an iterative process, i.e. one calibration (A) and one recalibration stage (B) (Fig. 2), each stage consisting of a calibrating

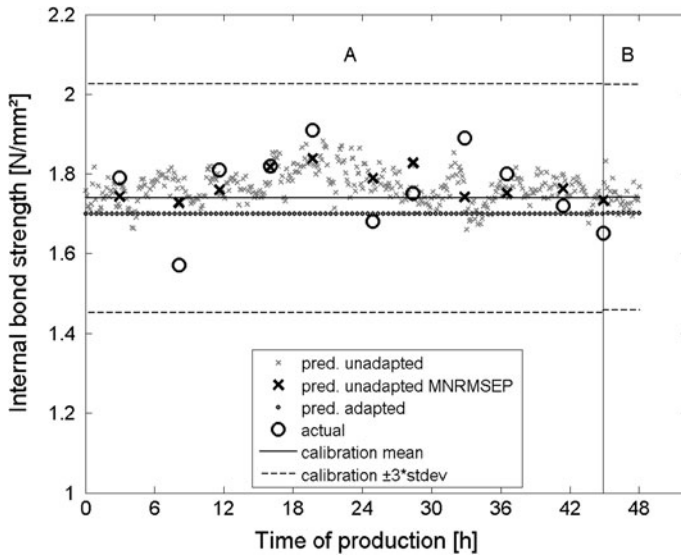


Fig. 2 Simulation of a real-time process adaptation with predictions of unadapted (Fig. 1, III) and adapted (Fig. 1, IV) predictor variables over 2 days of production using two calibrated PLSR models (A calibration, B recalibration). Actual IB values (circled points) and the calibrated range of IB values (solid and dashed lines) are used as criteria for recalibration

(“1st PLSR model”) and a validating (“2nd PLSR model”) step, were carried out. To exemplify the results of model calibration, the first calibration stage (A) is presented in detail. Offline determined IB data with an arithmetic mean of 1.74 N/mm^2 ($\text{CV} = 7.1 \%$) were used for calibrating the 1st PLSR model. From the original 804 predictor variables, 19 were excluded due to values deviating from the predefined range. 22 predictor variables were excluded because of unavailability in the real-time database, 156 predictor variables were excluded because of $>50 \%$ missing values, and 47 predictor variables were excluded due to low variation ($\text{CV} < 0.2 \%$). In total, 244 variables were filtered and the remaining 560 predictor variables were used to calibrate the 1st PLSR model. The number of remaining variables was almost three times higher than the 179 variables used by Clapp Jr et al. (2008), who suggested the collection of additional predictor variables to improve model performance. A high number of collected predictor variables, from which the most significant variables are selected, should diminish unobserved influences and consequently the RMSEP (Clapp Jr et al. 2008). Thus, it is assumed that the 560 predictor variables used for calibration in the present study can sufficiently explain the variation within the predicted variable.

The MNRMSECV of the 1st PLSR model using all 560 predictor variables and an optimum number of LV of 1 was 7.2% . Using IPLS, this error was reduced by selecting the variables that contributed most to the explanation of the variance of the predicted variable (IB). In addition, selecting the most important predictor variables allowed a technological interpretation of the current process and facilitated an adaptation of process parameters. Thus, 47 predictor variables were selected in the

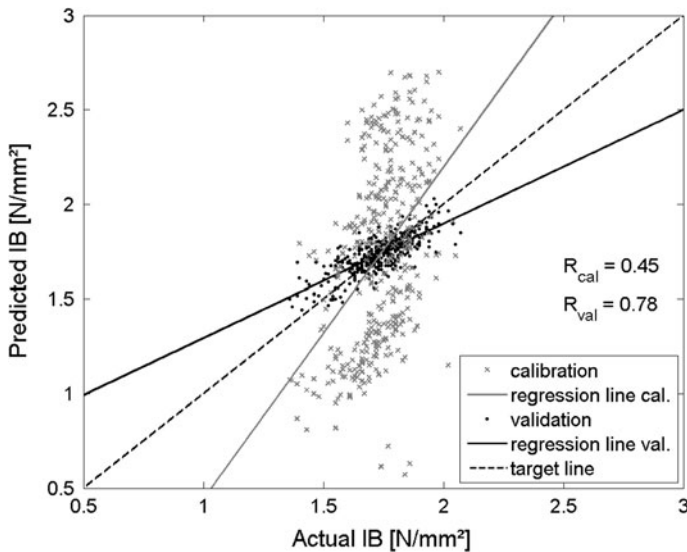


Fig. 3 Correlation between actual and predicted IB values obtained from the 1st PLSR model [calibration (Fig. 1, *I*)] and the 2nd PLSR model [validation (Fig. 1, *II*)] in stage A, showing the target line ($R = 1$) and regression lines

2nd PLSR model with 5 LV resulting in an MNRMSECV of 4.7 %. The coefficient of correlation of actual IB values versus predicted IB values from the 1st PLSR model was 0.45 (R_{cal} in Fig. 3). Using the 2nd PLSR model, the correlation between actual and predicted IB values could be improved to a coefficient of correlation (R_{val}) of 0.78.

Model prediction with recalibration (feedback)

Based on the calibrated PLSR model, the IB was predicted in a simulation using real-time process data (Fig. 2). The grey crosses in Fig. 2 are real-time predictions of the IB at 20-second intervals, with the dashed line (mean) and the dotted lines (± 3 times standard deviation) depicting the range of predicted values from the calibrated model. The circled points show the actual IB values determined by destructive offline measurements. The crosses in black depict predicted IB values at the time when actual IB values were measured to evaluate MNRMSEP values. Predicted IB values (black crosses) deviated with a CV of 2.2 % from their arithmetic mean of 1.77 N/mm² and were consistently within the range of ± 3 times the standard deviation of their respective $\hat{Y}^{(c)}$. The arithmetic mean of actual IB values was 1.76 N/mm² with a CV of 5.8 %. Thus, the model predictions overestimated the IB on average by 0.01 N/mm². The CV of predicted IB values was more than half lower than the CV of actual IB values, which indicates that short-term deviations between two subsequent actual IB values (see second actual value in Fig. 2) are neglected by the calibrated PLSR model. This can especially be

seen in the predicted IB values at 7.7 h after starting the simulation as they are not affected by the low actual IB value of 1.57 N/mm².

After 11.1, 35.0 and 43.0 h of simulating the production of HDF, MNRMSEP values were above the threshold of 5 % and PLSR models were recalibrated (Table 1). MNRMSEP values obtained after 11.1 h (7.1 %) and after 35.0 h (6.7 %) were higher than their corresponding MNRMSEP values obtained from calibration (5.5 % after 11.1 h and 5.1 % after 35.0 h). Thus, the calibrated PLSR model, which was used up to these points (11.1 and 35.0 h), was used as basis for further simulation instead of the recalibrated PLSR models. After 43.0 h, predicted IB values deviated from actual IB values by an MNRMSEP of 6.1 %. The corresponding MNRMSEP obtained from recalibration was 5.6 % (Table 1). As this recalibration stage improved the MNRMSEP, the recalibrated PLSR model was used for further simulation (Fig. 2, B). The mean of all MNRMSEP values from PLSR models used for predicting IB values was 4.6 %. The coefficient of correlation between actual and predicted IB values that were used for calculating MNRMSEP values was 0.74 (Fig. 4). Looking at the regression line in Fig. 4, low actual values were overestimated and high actual values were underestimated. This effect can be ascribed to the length of time periods chosen for calibration, which could be observed in preliminary investigations explained in the following paragraph.

In preliminary investigations, the approach presented in the present study was applied to different time periods. In this pre-examination, the usage of a time period of 2 months (100 data records for calibration) resulted in MNRMSEP values down to 2.9 % but also in a high number of recalibration stages (up to five stages within 2 days of production). A process modelling scheme that needs such a high number of recalibration stages could produce over-fitted PLSR models and is considered to be unstable for a process adaptation in real time. The higher susceptibility to over-fitted models could also be seen in the higher variation of predicted IB values (CV = 3.6 %). Models using time periods longer than 2 months resulted in more stable models (one or even no recalibration stage necessary within 2 days of production), but also in higher MNRMSEP values (increase from 2.9 up to 6.1 %).

Table 1 MNRMSEP values of calibration and recalibration stages

| Time of production (h) | MNRMSEP calibration (%) | MNRMSEP recalibration (%) |
|------------------------|-------------------------|---------------------------|
| 2.8 | 4.0 | – |
| 7.7 | 4.3 | – |
| 11.1 | 5.5 | 7.1 |
| 15.3 | 4.5 | – |
| 18.8 | 4.9 | – |
| 23.9 | 4.4 | – |
| 27.2 | 4.0 | – |
| 31.5 | 4.1 | – |
| 35.0 | 5.1 | 6.7 |
| 39.6 | 4.4 | – |
| 43.0 | 6.1 | 5.6 |

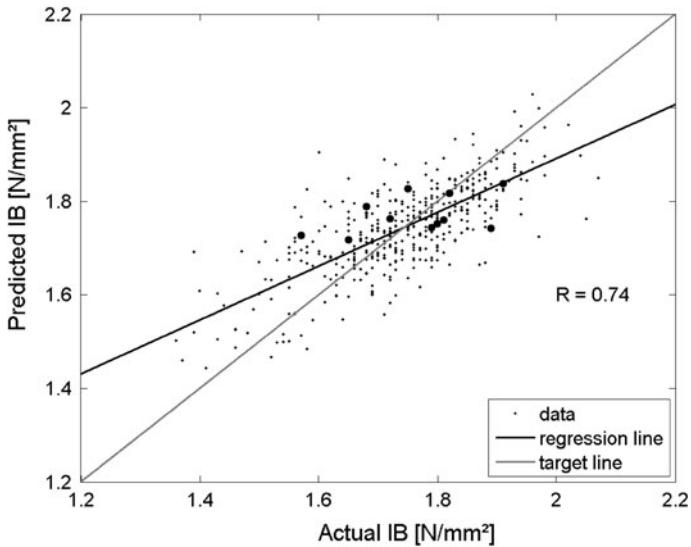


Fig. 4 Correlation between actual and predicted IB values obtained from PLSR models for calculating MNRMSEP values in all stages, showing the *target line* ($R = 1$) and the *regression line*. Larger symbols indicate the 11 IB values used in the simulation of process adaptation in Fig. 2

Thus, the periods for model calibration, validation and prediction have to be carefully chosen in practice, to obtain models that are both stable over time and able to predict short-term variations (e.g. in the range of a few hours). In the present study, the alternate selection of data for evaluating the MNRMSEP ensures the inclusion of recently generated data for calibration (to allow the analysis of most recent process variability) as well as considering data for a longer time period to obtain stable and statistically reliable models. The resulting MNRMSEP of 4.6 % seems to be an ideal prerequisite for the following adaptation.

Andre et al. (2008) obtained a MNRMSEP of 6.2 % while selecting 56 variables from a total of 164. Hasener (2004) obtained MNRMSEP values of 5.8 % when predicting IB over a one-year period using up to 281 variables. One explanation for the lower MNRMSEP in the present study could be the higher number of available predictor variables (560), of which the final variables ranging from 41 to 47 were selected (Table 2). Results of calibration and recalibration stages are shown in Table 2, indicating high predictability of all PLSR models with MNRMSECV values ranging from 4.7 to 4.8 %.

The most important selected predictor variables over all calibration stages are shown in Table 3 ranked by their scaled regression coefficients ($>|0.15|$) in descending order. Significant variables that were selected at multiple calibration stages showed the same trend of influence at all calibration stages. During model generation, the selection as well as the ranking of variables depended on the used number of LV and the inclusion of new observations. However, the three most important variables were selected consistently. Higher board densities as well as increased steam consumptions for preheating chips prior to refining would have

Table 2 Predictability and model parameters of calibration stages A and B

| Calibration stage | Number of observations for calibration | Selected predictor variables | Number of LV | MNRMSECV (%) |
|-------------------|--|------------------------------|--------------|--------------|
| A | 400 | 47 | 5 | 4.7 |
| B | 400 | 41 | 4 | 4.8 |

Table 3 Scaled regression coefficients and deviation of actual values from adapted values (RMS $_{\delta}$) of most important model variables (scaled regression coefficients >0.15) at two calibration stages A and B

| | Variables | Scaled regression coefficient | | RMS $_{\delta}$ of actual-adapted values | |
|----|---------------------------------------|-------------------------------|-------|--|------|
| | | A | B | A | B |
| 1 | Board density | 0.32 | 0.41 | 0.44 | 0.33 |
| 2 | Steam consumption preheating chips | 0.25 | 0.26 | 0.34 | 0.21 |
| 3 | Amount water-repellent | -0.21 | -0.17 | 0.28 | 0.13 |
| 4 | Distance press system 06 right | | 0.20 | | 0.16 |
| 5 | Moisture content sawdust ^a | | -0.18 | | - |
| 6 | Distance press system 12 left | | 0.18 | | 0.14 |
| 7 | Drying temperature | | -0.18 | | 0.14 |
| 8 | Sprinkling water fibre mat bottom | -0.17 | | 0.22 | |
| 9 | Cooking time chips | | 0.17 | | 0.13 |
| 10 | Pressure press system 19 | -0.16 | | 0.21 | |

^a Uncontrollable

increased the IB of boards (positive regression coefficients). Increased amounts of water-repellent led to deteriorated IB values (negative regression coefficient). An increase in moisture content of added sawdust decreased IB values. Similarly, the amount of water sprinkled onto the forming belt prior to forming should have been lower to increase IB values. The drying temperature of fibres, cooking time of chips as well as press parameters such as press platen distance and press pressure additionally influenced IB values of boards. Lower press pressures at the end of the press would have increased IB values. This could be ascribed to an undesired breaking of already cured bonds between the fibres. The inclusion of a three-dimensional model of the hot-pressing process of fibreboards, introduced by Thömen (2000), could further improve the technological interpretation of significant press parameters.

Adaptation (feedforward)

After PLSR models were obtained that fulfilled the feedback function (actual \approx pre in Fig. 1), significant model parameters were adapted to lower the variation of predicted IB values (Fig. 1, IV). Similarly, the variation of actual IB values should

be decreased as well when using the approach of process adaptation in an industrial environment. Due to this feedforward process control, the mean of predicted IB values obtained from models that used adapted predictor variables was in accordance with the target IB value of 1.70 N/mm^2 (adapted values depicted as grey dots with black edge colour in Fig. 2). The variation of predicted IB values was minimised to 0.2 % (CV). This was achieved by adapting controllable model variables using the minimisation function in Formula 15. Scaled mean values of actual and adapted predictor variables (Table 3) are depicted in Fig. 5. As the mean of actual IB values (1.76 N/mm^2) was above the target IB value of 1.70 N/mm^2 , adaptable predictor variables with positive regression coefficients were decreased and adaptable predictor variables with negative regression coefficients were increased to gain the target IB. The root mean square deviation (RMS_δ) of actual versus adapted values of predictor variables in Fig. 5 was 0.17. This deviation was clearly below the maximum deviation allowed in function 15 (\pm the standard deviation of variables in $X^{(sc)}$ which is ± 1). The single RMS_δ values of controllable predictor variables are shown in Table 3, ranging from 0.13 to 0.44.

The developed optimisation programme predicted IB values of fibreboards with high precision. The simulation of adapting controllable predictor variables to gain a specific target value revealed promising results, as all predictions over the 2 days of simulation had a value of 1.70 N/mm^2 . Using the presented adaptation technique, process variables that consume high energy, gain low yield or have low cost efficiency can be optimised. This can be achieved by defining uncontrollable variables (Formula 15) and assigning them a specific value. Adapting the controllable predictor variables of the calibrated PLSR model should allow a compensation of this measure up to a certain degree. The validation of this

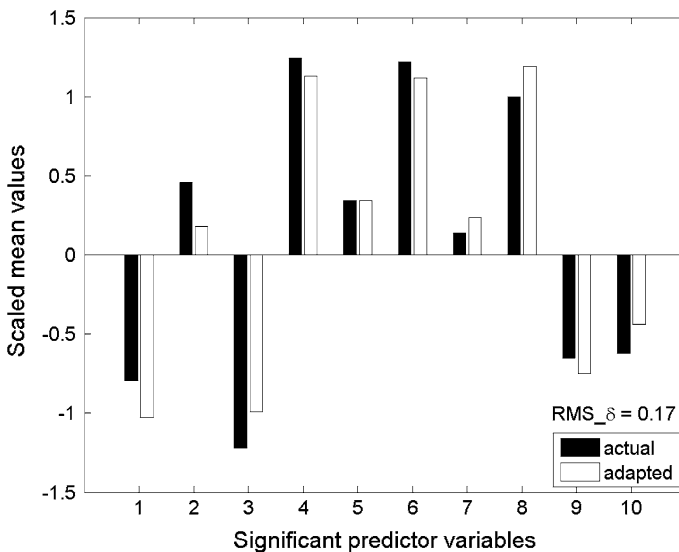


Fig. 5 Scaled mean values of most significant actual ($X^{(a)}$, $X^{(a)}$) and adapted ($\tilde{X}^{(a)}$) predictor variables

hypothesis should be carried out in an industrial environment by using simultaneous modelling of different product properties, which can be achieved with a combination of PCA and PLSR (Wold et al. 2001). The determination of additional process and raw material parameters, such as the gluing quality of wood particles, will lead to higher predictability of PLSR models and should be considered in future trials. The presented study serves as basis for a simultaneous optimisation of relevant board properties mentioned in EN 622-5 (2006).

Conclusion

The outcome of this study shows that process parameters can be adapted in a real-time simulation using the multivariate statistical tools presented. Particularly, IPLS seems to be appropriate as algorithm for selecting significant variables when modelling the manufacturing process of wood-based panels in real time. The first of the two main benefits of the developed process adaptation is the newly gained knowledge about interactions between properties of the final board and process as well as raw material parameters. In this respect, significant parameters such as press conditions, preheating conditions of chips, amounts of water-repellent or drying conditions could be detected and their influence on the IB was interpreted. The second benefit is the possibility to change process parameters in real time to gain a specific target value, minimise safety margins and save precious resources. The adaptation of controllable predictor variables can be limited by predefining upper and lower limits to avoid excessive changes in the variables.

To test the performance of the developed adaptation tool under real conditions, the simulation should be carried out in an industrial environment. Consequently, the influence of adapted process variables on the subsequent adaptation periods is subject to future researches. The approach presented in this study serves as basis for implementing a feedforward real-time process adaptation in industrial manufacturing processes.

Acknowledgments The authors gratefully acknowledge the financial support by the Competence Centre for Wood Composites and Wood Chemistry, Wood K plus, Austria. Special thanks go to Mr. Ralf Borchers and Mr. Thomas Kuncinger for their professional support throughout the study and Mrs. Cornelia Riegler for proofreading.

References

- Andre N, Cho HW, Baek SH, Jeong MK, Young TM (2008) Prediction of internal bond strength in a medium density fiberboard process using multivariate statistical methods and variable selection. *Wood Sci Technol* 42(7):521–534
- Andre N, Young TM, Zaretzki P (2010) Real-time process modeling of wood composite panels. In: Teischinger A, Barbu MC, Dunky M et al (eds) *Processing technologies for the forest and biobased products industries*, PTF BPI, Kuchl, Austria. Salzburg University of Applied Sciences, Salzburg, pp 28–33
- Bernardy G, Scherff B (1997) Process modelling provides online quality control and process optimization in particleboard and fiberboard production. *Holz Roh Werkst* 55(3):133–140

- Chachuat B, Srinivasan B, Bonvin D (2009) Adaptation strategies for real-time optimization. *Comput Chem Eng* 33(10):1557–1567
- Clapp NE Jr, Young TM, Guess FM (2008) Predictive modeling the internal bond of medium density fiberboard using a modified principal component analysis. *For Prod J* 58(4):49–55
- Coleman TF, Li Y (1994) On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Math Program* 67(1–3):189–224
- Cook DF, Chiu CC (1997) Predicting the internal bond strength of particleboard, utilizing a radial basis function neural network. *Eng Appl Artif Intell* 10(2):171–177
- Danzer K, Hobert H, Fischbacher C, Jagemann K-U (2001) *Chemometrik*. Springer, Berlin
- De Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. *Chemom Intell Lab Syst* 18(3):251–253
- Dolezel-Horwath E, Hutter T, Kessler R, Wimmer R (2005) Feedback and feedforward control of wet-processed hardboard production using spectroscopy and chemometric modelling. *Anal Chim Acta* 544(1–2):47–59
- Dunky M, Niemz P (2002) *Holzwerkstoffe und Leime*. Springer, Berlin
- EN 319 (1993) Particleboards and fibreboards—determination of tensile strength perpendicular to the plane of the board. European Committee for Standardization
- EN 326-1 (1994) Wood-based panels—sampling, cutting and inspection—part 1: sampling and cutting of test pieces and expression of test results. European Committee for Standardization
- EN 622-5 (2006) Fibreboards—specifications—part 5: requirements for dry process boards (MDF). European Committee for Standardization
- Esteban LG, Fernández FG, De Palacios P (2011) Prediction of plywood bonding quality using an artificial neural network. *Holzforschung* 65(2):209–214
- Hasener J (2004) Statistische Methoden der industriellen Prozessmodellierung zur Echtzeitqualitätskontrolle am Beispiel einer kontinuierlichen Produktion von Faserplatten. Univ, Hamburg
- Little RJA, Rubin DB (2002) *Statistical analysis with missing data*. Wiley series in probability and statistics, 2nd edn. Wiley, New York
- Marutzky R (2006) Energetische und stoffliche Verwertung von Holzresten und Altholz in der Holzwerkstoffindustrie—eine aktuelle Bestandsaufnahme. In: 3. Fachtagung Umweltschutz in der Holzwerkstoffindustrie Göttingen, 18–19. Mai 2006. Univ. Göttingen, Institut für Holzbiologie und Holztechnologie, pp 17–26
- Mehmood T, Liland KH, Snipen L, Sæbø S (2012) A review of variable selection methods in partial least squares regression. *Chemom Intell Lab Syst* 118:62–69
- Nelson PRC, Taylor PA, MacGregor JF (1996) Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemom Intell Lab Syst* 35(1):45–65
- Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB (2000) Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Appl Spectrosc* 54(3):413–419
- Rials TG, Kelley SS, So CL (2002) Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. *Wood Fiber Sci* 34(3):398–407
- Riegler M, Gindl-Altmutter W, Hauptmann M, Müller U (2012) Detection of UF resin on wood particles and in particleboards: potential of selected methods for practice-oriented offline detection. *Eur J Wood Prod* 70(6):829–837
- Sjöblom E, Johnsson B, Sundström H (2004) Optimization of particleboard production using NIR spectroscopy and multivariate techniques. *For Prod J* 54(6):71–75
- Steffen A, Janssen A, Kruse K (2001) Analysis of the MDF production process by means of statistical process modelling. *Holz Roh Werkst* 58(6):419–431
- Thömen H (2000) Modeling the physical processes in natural fiber composites during batch and continuous pressing. Oregon State University, Corvallis
- Vojinović V, Cabral JMS, Fonseca LP (2006) Real-time bioprocess monitoring: part I: in situ sensors. *Sens Actuators, B* 114(2):1083–1091
- Weigl M, Maschl G, Wimmer R, Mitter R (2012) Within-process and seasonal changes during industrial production of high-density fibreboard. Part 2: PLS modelling of chemical alterations, refining conditions and panel thickness swell. *Holzforschung* 66(5):673–678
- Wilson JB, Krahmer RL (1976) Particleboard: microscopic observations of resin distribution and board fracture. *For Prod J* 26(11):42–45
- Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS (2006) *Chemometrics tutorial—for PLS_Toolbox and Solo*. Wenatchee, Washington

- Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS (2012) Eigenvector Documentation Wiki. Eigenvector Research, Inc., <http://wiki.eigenvector.com>. Accessed 8 May 2012
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58(2):109–130
- Young TM, Guess FM (2002) Mining information in automated relational databases for improving reliability in forest products manufacturing. *Int J Reliab Appl* 3(4):155–164
- Young TM, Shaffer LB, Guess FM, Bensmail H, Leon RV (2008) A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard. *For Prod J* 58(4):39–48
- Zeng Y (2011) A study of missing data imputation and predictive modeling of strength properties of wood composites. University of Tennessee, Knoxville