Holz als Roh- und Werkstoff 55 (1997) 377–382 © Springer-Verlag 1997

Multivariate prediction of wood surface features using an imaging spectrograph

O. Hagman

This paper presents a multispectral system for evolution of linear algorithms for prediction of wood surface features important for automatic inspection of lumber. The selection of training samples, the imaging spectrograph scanning method, raw data representation, evaluation of linear algorithms and testing of performance is discussed. A possible on line implementation for high speed wood scanning with a smart sensor is outpointed. An example, showing the evolution of linear algorithms for prediction of compression wood in softwood species (*Picea abies*, *Pinus sylvestris*), is reported, showing verified 92–94% correct classification. It is shown that compression wood classification could be reduced to an uncomplicated linear model using just a few spectral components where the most important one is around the limit for visible light going to the Ultraviolet spectra. This almost univariate behaviour for the model is not the common behaviour for other wood surface features (Brunner et al., 1996; Hagman, 1995; Hagman, 1996).

Multivariate Vorhersage der Eigenschaften von Holzoberflächen mit Hilfe eines Bild-Spektrographen

Diese Arbeit beschreibt ein Mehrkanalsystem zur Entwicklung eines linearen Algorithmus, der es gestattet, Eigenschaften von Holzoberflächen für eine automatische Gütesortierung von Schnittholz vorherzusagen. Die Auswahl von Testproben zum Kalibrieren des Systems, die Bilderzeugung durch Abrastern mit dem Bildspektrographen, die Darstellung der Rohdaten, die Beurteilung des linearen Algorithmus und die Eignung des Systems werden diskutiert. Die mögliche On-Line-Implementierung in ein Hochgeschwindigkeitsprüsystem für Schnittholz mit Hilfe eines schnellen Sensors wird aufgezeigt. Als Beispiel wird der lineare Algorithmus zur Vorhersage von Druckholzanteilen in Nadelholz (*Picea abies*, *Pinus sylvestris*) beschrieben, der eine 92–94% richtige Klassifizierung ermöglicht. Es konnte gezeigt werden, daß die Erkennung von Druckholz auf ein einfaches Modell reduziert werden kann, das nur wenige spektrale Komponenten erfordert, wobei die wichtigsten Wellenlängen im Bereich von der Grenze des sichtbaren Lichts bis zum UV-Bereich liegen. Dieses fast univariate Verhalten des Modells ist allerdings für die Bewertung anderer Oberflächeneigenschaften nicht der Fall.

O. Hagman
Department of Wood Technology, Skellefteå campus
S-931 87 Skellefteå, Sweden
Email: Olle.hagman@tt.luth.se

1 Introduction

Today every piece of wood is examined and classified several times in the production chain from the forest to the finished wood product. The classification has so far been based on human judgements after visual inspection. The classification is often combined with some kind of human "optimization" process.

The workers do often have to make up to 30 classification/"optimization" decisions every minute. The consequence of this situation is many sub-optimal decisions and a very high mental load on the workers. Furthermore an incorrect classification has a very big impact on the production economy. Consequently very big efforts have been made in order to replace man in this kind of classification- optimization work.

For some simple classification- optimization tasks replacement of man has been successful. However in many cases attempts to replace man have failed due to the fact that the implemented systems have not been robust enough to handle the variability of wood. The main reason for these failures is that we do not have basic knowledge about the relationship between reflected or transmitted electromagnetic waves and different wood features.

Today classification and analysis of wood and its features are important field for industrial applications and research. Methods used are often based on image analysis using defect related features within the spatial- and frequency domain. By introducing the RGB-technique some industrial applications has also started (Vogrig et al., 1993) to use spectral information for defect separation.

In research multispectral approaches have been tested using spectral information from more than three wavelengths using Multivariate Image Analysis (MIA) (Esbensen et al., 1989) as a tool to create more reliable and sharp classification algorithms. However there is so far no method that gives a perfect solution to the classification and prediction problems due to the fact that wood itself is a polymer material with a large variation within species, within population and within each object.

The measured variations are due to primary factors such as:

- chemical composition
- cellular structure and orientation
- surface roughness
- light and sensor arrangement

These primary factors are dependent on a lot of secondary factors such as wood species, growth conditions, inheritance, machining method, angle between surface and annular rings, drying conditions, storing conditions, light conditions, type of sensor etc. The measured value on a certain place of the object is also dependant on the wood features in the neighbourhood. This clearly shows that measurement, classification and analysis of wood and its features is a multivariate problem. It is not possible to develop robust scanning systems for automatic wood classification with trial and error methods. A systematic approach is needed.

In order to perform a systematic approach we need fundamental knowledge about the relationship between reflected, transmitted or absorbed electromagnetic waves and different wood features. The objective with this project is to investigate and describe these relations and the create well performing and stable prediction algorithms.

2
Materials and methods

A wood surface is characterized by the reflected or transmitted electromagnetic waves from the surface. The response, i.e. the number and energy of photons reaching the sensor element, is a combination of reflected or transmitted waves from the area of the scene that correspond to a certain sensor position and scattered waves from the surrounding scene areas. The scattering effect is a problem that must be taken into consideration.

The responses are mainly measured by a multispectral image scanner. This scanner is a linescanner with the possibility to generate spectral images within the wavelengths 400–720 nm. Spectral resolution and integration is varying due to the sensitivity of the used sensor in the examined spectral area. The imaging spectrometer uses a CCD- matrix camera and a PGP imaging spectrograph developed by VTT Finland (Hyvärinen, 1993) for evaluation of raw data. This gives a maximal spectral resolution of 768 discrete levels and a spatial resolution of 512pixels/ scanwidth. In addition to the multispectral image scanner we also use x-ray scanners, CT scanning and SEM micrographs.

The scene, in this case the piece of wood, is described and characterized with aid of the results from SEM-analysis, surface roughness measurements, lighting conditions, neighbourhood description and macroscopic description of different biological features such as type of knot, heart-sapwood, early-latewood, compression wood etc.

How the sensor response depend on the different wood characteristics will be analysed with aid of MIPLS – technique (Multivariate Image Projections to Latent Structures) (Hagman, 1996). The method based on the Kernel algorithm (Lindgren et al., 1993) is an application of the PLS (Martens, Naes, 1989) concept and gives a possibility to investigate several, parallel, interacting depending variables using images. MIPLS is used as a typical problem-dependent strategy for image decomposition guided by the nature of the dependent-variable and/or training data set delineation in the (independent, dependent) image domains. The method is introduced in chemometrics (Esbensen et al., 1992) and is also implemented in multispectral image analysis of wood (Hagman, Grundberg, 1993).

In order to analyse the problem systematically the experimental design is of great importance. It is, however, hard to find biological samples that vary according to the experimental design. Thus it is only the above described secondary factors that can be controlled by choosing samples that enable a large variation, regarding to relevant variables, to be obtained. The samples are chosen from an extremely well-substained population of trees from all over Sweden. There exist a well described transform between the wood samples (logs, boards, defects), the digital image stacks and the segmented and classified models for the log and some of it's features. These models will be used for simulation and testing of error estimations.

Evaluated models for feature classification could be used for developing of a wood surface scanner with selected chromatic filters optimized for certain features (Brunner et al., 1993) or as tested by Åstrand et al., 1995 using the MAPP-sensor with the PGP where the linear discriminant functions extracted by MIA and MIPLS or by using an unconstrained optimization strategy are calculated using variable exposure times and analog summation of pixel data. After A/D-conversion the sums are compared and classified pixels are output from the sensor chip. This method not fully evaluated for wood feature classification but demonstrated on more straight forward applications (Åstrand et al., 1995).

3
Prediction and classification of compression wood

A test and training set containing samples from five different pine (*Pinus sylvestris*) logs and three different spruce (*Picea abies*) logs, Fig. 1 was trained and classified to three different classes Clearwood (CW), Compression wood (Compr) and Black knot (knot). Images obtained by using a linear scanner, a Nikkor 20 mm lens, a PGP imaging spectrometer, a MTI matrix CCD-cameras and a Macintosh Quadra 950 are stored in a three dimensional (X,Y,λ) image stack (Fig. 2). Two similar methods (MIA/ MIPLS and PLS) is tested to obtain prediction models. The reason for using two methods is first to compare the ability and validate the image based algorithms to the more tested iterative method PLS and second to test if the performance operators (*sensitivity*) could be used with similar results for both methods. The calculation for the MIPLS prediction performance is somewhat more complicated, using image subtraction between predicted image and training image and the calculating areas belonging to false positive respective false negative predictions.

3.1
Image compression

The initial multispectral image stack with the size X (transverse the board) and Y (lengthwise the board) of 258 by 226 pixels and λ, 768 spectral levels representing the VIS spectra from 400 to 720 nm were reduced to a 258 (X) by 226 (Y) by 56 (λ) voxel image stack for MIA/MIPLS analysis and to a 557 (points on surface) by 56 (λ) matrix for PLS. In Fig. 3 the MIA/MIPLS stack is shown in three different cuts, XY, Xλ, and Yλ.

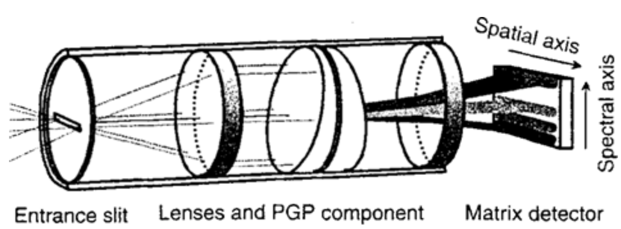


Fig. 1. Schematic layout of prism-grating-prism (PGP), from Hyverinän 1993, and matrix-sensor combined to image spectrometer

Bild 1. Schema des PGP-Systems (Prisma-Gitter-Prisma), nach Hyverinän 1993, sowie des Matrix-Sensors in Verbindung mit dem Bildspektrometer

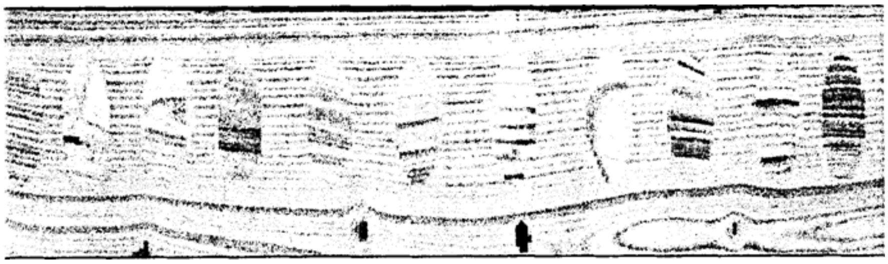


Fig. 2. The train and test set board scanned at 470 nm wavelength with the PGP imaging spectrometer, showing a pine board with black knots, down left, a hole down middle, some streaks of compression wood and 11 eleven samples containing more or less compression wood
Bild 2. Zusammengesetztes Probebrett zum Trainieren und Testen des Systems. Die Probe wurde bei 470 nm eingescannt. Das Bild zeigt ein Kiefernbrett mit Totast (links), einem Loch (mitte, unten) und einigen Streifen Druckholz sowie 11 Teile mit mehr oder weniger hohem Druckholzanteil

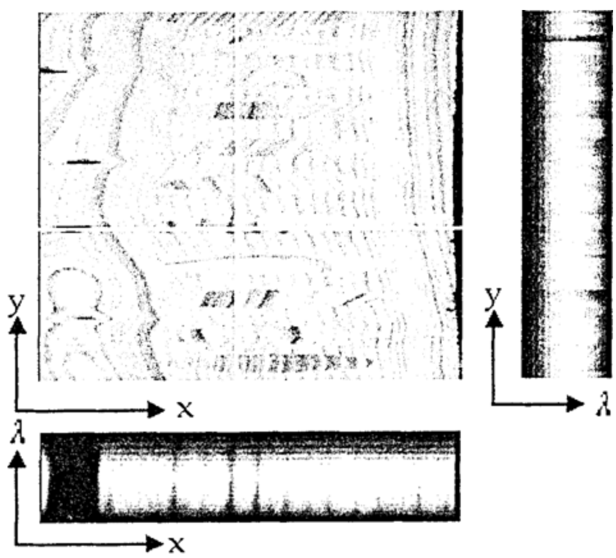


Fig. 3. MIA/MIR reduced spectral image stack represented by three cuts, XY (upper left), Xλ (down left), and Yλ,(upper right) respectively. The Xλ and Yλ planes are sliced as indicated by white lines in XY plane. XY plane is cut in the 470 nm spectra
Bild 3. Ein mittels MIA/MIR reduziertes spektrales Bild einer Holzoberfläche, dargestellt in drei Schnittebenen: XY (links oben), Xλ (links unten) und Yλ (rechts oben). Die beiden letzten Ebenen werden entsprechend den weißen Linien in der XY-Ebene geschnitten. Die XY-Ebene wird bei 470 nm gescannt

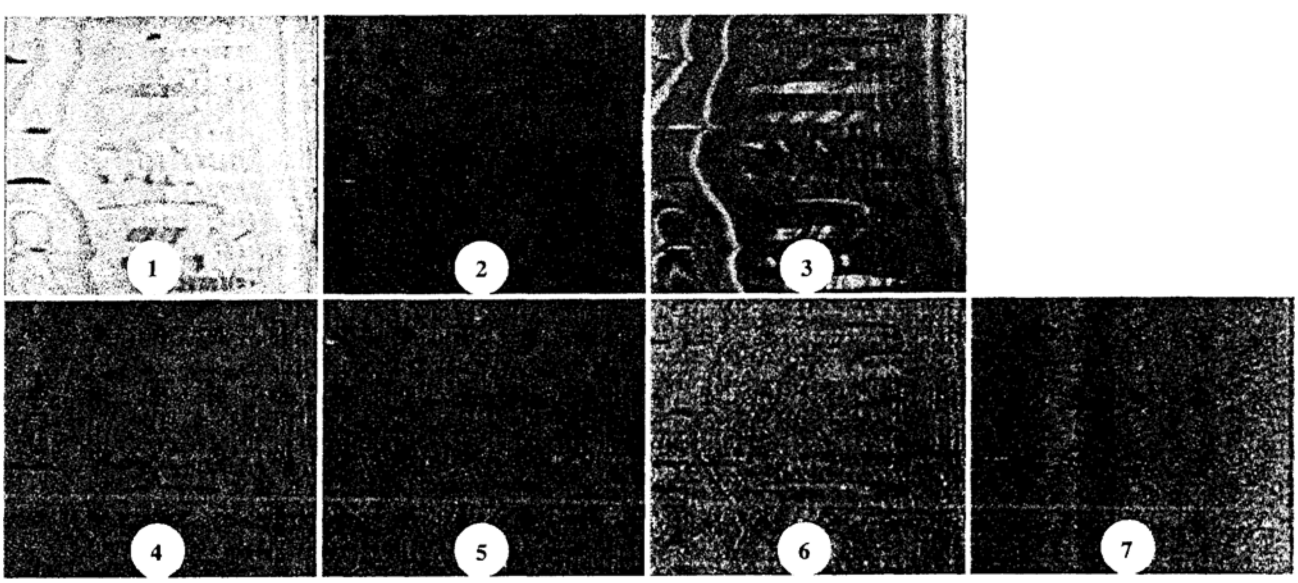
4 Results and discussion

A MIA analysis was done, pc-score plots shown in Fig. 4, indicating that the wanted classes are separable in the score scatter plots pc2 against pc3 (Fig. 5). A training set was obtained by mascing in the score plot pc2/pc3 resulting in the Y-dummy images shown in Fig. 6. Fig. 7 shows selected training areas containing clearwood areas, compression wood areas and knot areas representing the main features in the scene. The selected training areas covers approximately 12% of the scene.

The PLS training set was selected using a "sawing in the XY view" operation, where rectangular shaped pixel set was selected within the different feature areas (Fig. 7).

The selected training set was converted to text files, dummy variables added and then exported to Simca for PLS analysis.

Fig. 4. PC score plots showing principal component 1 to 7, pc1-6 contain information clearly correlated to the scanned board but pc7 is showing another mechanism, probably variation in the X-direction of the CCD-sensor
Bild 4. PC-Auswertung der Hauptkomponenten 1 bis 7 eines Oberflächenscans. Die Plots 1–6 zeigen deutlich Korrelationen mit Eigenschaften des Brettes. Plot 7 zeigt einen anderen Mechanismus, der möglicherweise durch Abweichung des CCD-Sensors con der X-Richtung herrührt
▼



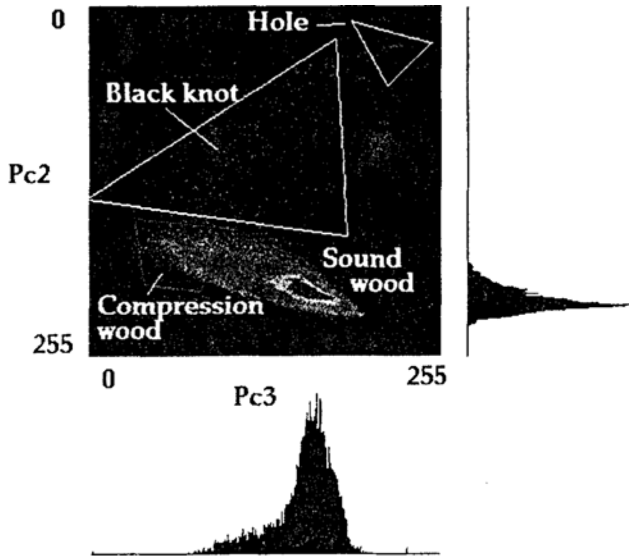


Fig. 5. PC-score scatter plot with areas masced for dummy segmentation marked with white borders. A score scatter plot is a two dimensional histogram (indicated by grey histogram outside the plot) where grey-scale (or preferably colour tones) levels indicate increased density of pixels with identical score values in pc2 and pc3
Bild 5. PC-Auswertung von Probeflächen, die mit weißen Grenzlinien markiert waren. Die Auswertung beruht auf einem zweidimensionalen Histogramm (s. Histogramme rechts und unten), wobei die Grauskalen (oder vorzugsweise Farbtöne) als erhöhte Pixeldichte angeben, wo die beiden Plots (ps2 und pc3) übereinstimmen

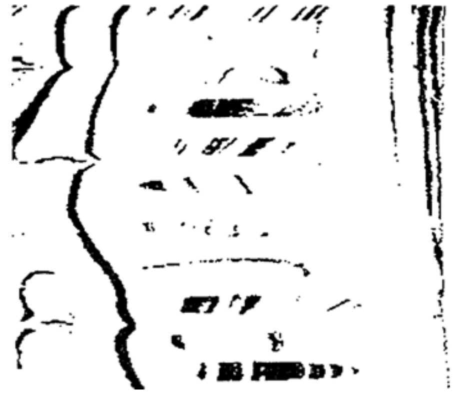


Fig. 6. Y-dummy image showing segmented compression wood areas in the training image
Bild 6. Y-Teilbild eines Testbildes mit Segmenten von Druckholzbereichen der Trainingsprobe

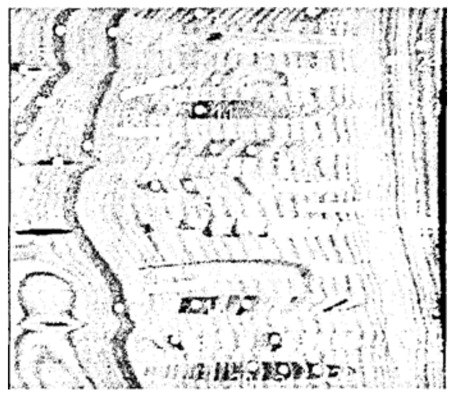


Fig. 7. Position of selected training areas. POs 1–24 compression wood areas, 25–26 knot areas and 27–33 clear wood areas. 12 object pixels are chosen from every position
Bild 7. Positionen ausgewählter Trainingsbereiche. Pos. 1–24: Druckholzflächen; Pos. 25–26: Astanteile; Pos. 27–33 fehlerfreie Oberfläche. Von jedem Testobjekt werden 12 Pixels ausgewählt

The MIPLS modelling result was that after 3 principal components the problem was solved i.e that 2 pc:s was needed the first is just a mean value transformation. The regression coefficients (Fig. 8) indicate that the separating information is within the blue spectra and the bluer the better, except when reaching near 400 nm where the sensor sensitivity decreases the signal/noise ratio. The PLS modell is somewhat different, a bit sharper, in selecting prediction coefficients and it is clear that the modell complexity could be reduced rather dramatically. Test with a reduction to 5 x-variables 429, 487, 575, 639 and 679 nm or even 3 x-variables, 429, 487 and 575 nm are showing almost equal results.

The performance of the model prediction can be shown in contingency tables (Table 1 and 2). A prediction is here considered correct for an prediction value (considered as a probability for that feature (Hagman, 1993)) for correct feature higher than 0.5.

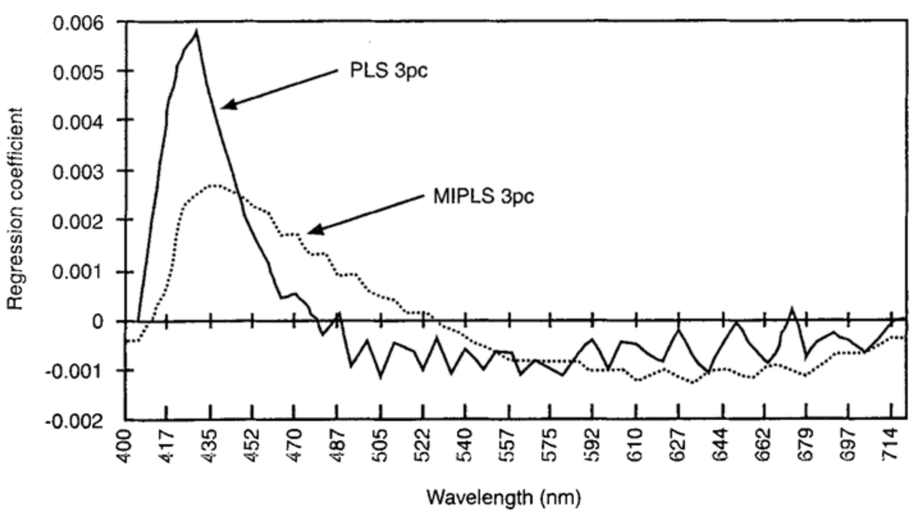


Fig. 8. Regression coefficients for predicting compression wood in training and testing area using PLS and MIPLS methods
Bild 8. Regressionskoeffizienten für die Vorhersage von Druckholzanteilen in Trainingsflächen Nach der PLS- und MIPLS-Methode

Table 1. Contingency table for compression wood feature prediction PLS model (%)
Tabelle 1. Trefferquoten beim Schätzen der Druckholzanteile nach dem PLS-Modell (%)

Status	Facit pos	Facit neg
Pred pos	93.8	5.6
Pred neg	6.2	94.4

Table 2. Contingency table for compression wood feature prediction MIR model (%)
Tabelle 2. Trefferquoten beim Schätzen der Druckholzanteils nach dem MIR-Modell (%)

Status	Facit pos	Facit neg
Pred pos	92	3
Pred neg	8	97

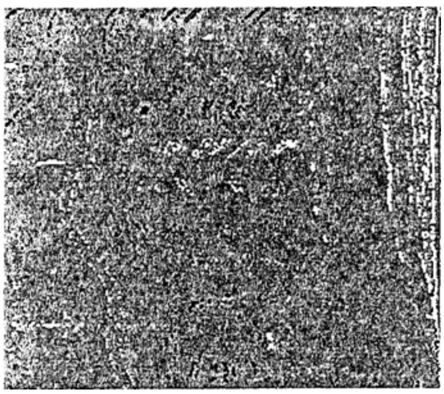


Fig. 9. Resulting image after adding the thresholded predicted image and the traning image. Grey areas are correctly predicted, white areas are false positive and black areas are false negative, respectively
Bild 9. Ergebnis nach Überlagerung des ausgewerteten Bildes mit dem Trainingsbild. Graue Flächen bedeuten korrekte Schätzung, weiße Flächen falsche Schäzung bei positiver Vorhersage, schwarze Flächen falsche Schätzung bei negativer Vorhersage

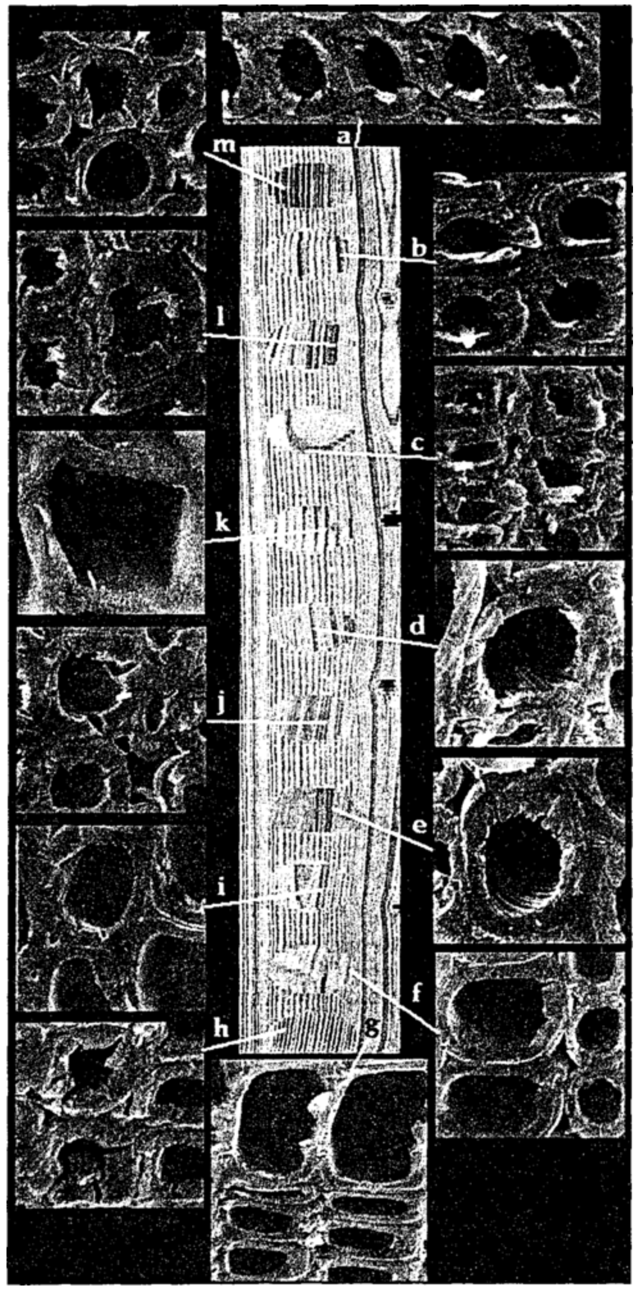


Fig. 10. SEM micrographs as indicators of compression wood areas within the training surface: **a** Compression wood (Cow) stripe on the frame board (Norwegian spruce (Ns)); **b** Cow stripe Ns; c Narrow Cow stripe (Scots pine (Sp)); **d** Cow stripe Ns; **e** Cow stripe Sp; **f** Narrow Cow stripe Ns; **g** Clear wood frame board; **h** Cow stripe Ns; **i** Narrow Cow stripe Sp; **j** Cow stripe Ns; **k** Late wood with high pitch content Sp; **l** Cow stripe Ns; **m** Cow stripe Ns.
Bild 10. REM-Aufnahmen als Indikatoren für Druckholzanteile in den Testflächen: **a** Streifen mit Druckholz (=Cow) im Hauptbrett (Fichte = Ns); **b** Druckholz in Ns; **c** schmaler Druckholzstreifen in Kiefer (Sp); **d** Druckholz in Ns; **e** Druckholz in Sp; **f** schmaler Druckholzstreifen in Ns; **g** fehlerfreies Holz im Brett; **h** Druckholz in Ns; **i** schmaler Druckholzstreifen in Sp; **j** Druckholz in Ns; **k** Spätholz mit hohem Harzgehalt; **l** Druckholz in Ns; **m** Druckholz in Ns

The resulting images from the performance calculation are indicating good performance for the prediction model, this is intuitively realised by a human being when knowing that grey areas are good behaviour.

The models where validated by comparing the predictions of compression wood to the SEM micrographs, Fig. 10, showing good model behaviour, in fact in some cases better than the initial visual classification, which were wrong for two areas.

The strong and stable almost univariate model obtained for compression wood model for is rather unique for wood surface features. Other models evaluated for knots, rot, and blue stain that are published in Hagman, 1996, are more multivariate in their nature and uses more variations in the spectrum. The strong dependency in deep blue spectra is probably due to the high contents of lignin in compression wood. Compression wood cells could be found not only in clearly visible compression wood streaks but also in other part of the stem and some of the miss classified areas in the validation test has after SEM analysis been judged as correct. The classification of compression wood could be essential for detecting light sound knots in spruce, because the large amount of that feature in the knot.

5
Conclusion

To summarise the project exemplified in this paper, the claimed contributions to the field are the following:

- The use of experimental design for wood surface object concentration on the test boards as a practical data compression method to reduce the amount of scanning needed. The designed variation is required to determine the variable space needed to get a proper description of the classes. Classes are used for describing the quality grade of a piece of softwood lumber.
- The introduction of PLS and MIPLS as soft modelling tools for calibration and prediction modelling of wood features.
- The evaluation of the imaging spectrograph/MIPLS concept to model the spectral behaviour of the interaction between visible light, the wood surface and the sensor.
- The power of a multivariate concept (PLS, MIA and MIPLS) to determine latent variables in low and high resoluted measurements of wood is identified. It is shown that models based on these compressed data can be used for prediction of objects appearing on the surface of lumber.
- The decomposition into latent variables also makes it possible to separate unwanted mechanisms that occur from the mechanisms caused by physical, chemical or optical features in the wood surface, thus making spectral classification possible. These external effects can be due to external influences such as moisture variations, drying, light exposure or angular effects.
- Latent variables based on principal component compression can be found for complex scenes that often are based on the dominating relationships in the measured (X) data or by correlation between X and dependent (Y) variables. These latent variables can be studied and give causal insight into complex mechanisms such as the interaction between light and wood surfaces, i.e. tracheid effects, dichromatic modelling etc.
- The necessity of decomposition of the data into information and noise before modelling is demonstrated.

Further examples of multispectral modelling of wood features are found in Hagman, 1996.

References

Åstrand E, Österberg M, Hagman, O Åström A (1995) Near-Sensor Real Time Spectral Time Classification for Industrial Applications. 9th Scandinavian Conference on Image Analysis. Uppsala. Sweden. 6–9 Jun.

Åström A (1993) Smart Image Sensors. Linköping Studies in Science and Technology. Dissertations. No. 319. pp. II 2–II 11

Brunner CC, Maristany AG, Butler DA, Funck JW (1993) Enhancing Color-image Data for Wood-Surface Feature Identification. Proc. 5th int. Conf. on Scanning Technology & process Control for the Wood Products Industry. 25.27 Oct. Atlanta, USA

Brunner CC, Maristani AG, Butler DA (1996) Wood species identification using spectral reflectance. Forest Prod. J. 46(2). pp. 82–85

Esbensen K, Geladi P, Grahn HF (1992) Strategies for Multivariate image regression. Chemometrics and Intelligent Laboratory Systems. 14, pp 357–374

Grundberg S, Hagman O (1993) Measuring active wood fracture surfaces. Holz Roh- Werkstoff 51, pp 49–54

Hagman POG (1993) Prediction of Wood Quality Features by Multivariate Models Based on Scanning Techniques. Thesis no. 20L. Luleå Technical University. Sweden

Hagman O, Grundberg S (1993) Multivariate Image Analysis Methods to Classify Features on Scots Pine: Evaluation of a Multisensor Approach. Proc. 5th int. Conf. on Scanning Technology & Process Control for the Wood Products Industry. 25.27 Oct. Atlanta, USA

Hagman O (1996) On Reflections of Wood – Wood Quality Features Modelled by Means of Multivariate Image Projections to Latent Structures in Multispectral Images Ph.D thesis no 1996: 198 D. Luleå Technical University

Hyvärinen T (1993) Towards Rugged and Miniaturized NIR instruments Proc. The Second European Symposium on Near Infrared (NIR) Spectroscopy. 30 Aug-2 Sept, Kolding , Denmark

Lindgren F, Geladi P, Wold S (1993) The Kernel Algorithm for PLS. J. of Chemometrics. 7, pp 45–59

Martens H, Næs T (1989) Multivariate Calibration., pp 146–147. John Wiley & Sons, Inc., New York

Vogrig R, Karpp B, Portala JF, Adel M, Lefevre R (1993) Scanning for Optimized Lumber Ripping and Crosscutting. Proc. 5th int. Conf. on Scanning Technology & process Control for the Wood Products Industry. 25.27 Oct. Atlanta, USA