IDENTIFYING TIMBER PERFORMANCE CLASSES USING LATENT CLASS REGRESSION*

RÜDIGER MUTZ†
ETH Zürich, Zähringerstr.24,
CH-8092 Zürich, Switzerland

and FRANKA BRÜCHERT‡
Institute for Forest Utilisation and Work Science,
Albert-Ludwigs-University Freiburg, Germany

(Received for publication 1 November 2005; revision 26 April 2006)

ABSTRACT

Latent class regression is a statistical method which is not well known in wood science for predicting distortion of sawn timber from structural wood characteristics. The method identifies unknown subgroups in a dataset and allows more accurate regression models to be derived. We identified two separate classes to describe the relationship between bow_{dry}, $spring_{dry}$, or $twist_{dry}$ and the predictors: initial distortion bow_{fresh}, $spring_{fresh}$, or $twist_{fresh}$, wood density, ring width, ring orientation, wood type (juvenile or adult), percentage compression wood measured separately on the four faces, and the contagion index which is a measure for the distribution of compression wood. The latent class regression models developed for the separate classes explained the variation in bow, spring, and twist to a higher degree than a single regression model over the entire dataset. For bow, R^2 increased from 0.13 to 0.24 and 0.41 for Class 1 and Class 2, for spring from 0.24 to 0.45 and 0.67, and for twist from 0.15 to 0.38 and 0.33. For individual regression models, the predictors showed a varying effect. In classes with significant compression wood on the faces, the effect of wood type seemed weaker, and vice versa. It was concluded that latent class regression analysis allows a more detailed explanation of the effects of wood structure on sawn timber distortion for heterogeneous datasets.

Keywords: timber performance; wood distortion; latent class regression; mixture distribution models; *Picea abies*.

---

* Based on a paper presented at IUFRO WP S5.01.04 Fifth Workshop on Wood Quality Modelling, 20–27 November 2005, Waiheke Island, New Zealand

† Corresponding author: mutz@gess.ethz.ch

‡ Present address: Forest Research Institute Baden-Württemberg, Wonnhaldestr. 4 D-79100 Freiburg,, Germany

INTRODUCTION

This study applied a statistical method that is not well known in wood science for predicting distortion of sawn timber from structural wood characteristics for *Picea abies* (L.) Karst. (Norway spruce). Conifers such as fir, pine, and spruce exhibit great variability in their wood formation due to genetic differences (Cown *et al*. 2004; Livingston *et al*. 2004), growth conditions, or silvicultural treatment (Mutz 1998; Mutz *et al*. 2004; Seeling 2001; Beaulieu *et al.* 2003) and this affects the mechanical and physical properties of the sawn timber produced. The presence of compression wood is one reason for between-tree as well as within-tree variation in wood properties (Cown *et al.* 2004; Donaldson *et al*. 2004; Wernsdörfer *et al.* 2004; Timell 1986). Timber performance is also strongly affected by reaction wood formation (Beard *et al.* 1993; Seeling 1999; Johansson & Kliger 2002) as compression wood affects wood density and moisture absorption strongly, and subsequently the physical and elasto-mechanical properties of the wood (Burgert *et al.* 2004; Bengtsson 2001a, b; Johansson 2004). Research into the effects of wood structure on technological properties has a long tradition. For Norway spruce, the effect of wood density, and of structural features such as ring width, ring orientation, grain angle, presence of juvenile wood, and compression wood has been subject to extensive research in the last decade, especially the influence of wood structure on the dimensional stability of sawn timber (Bengtsson 2001b; Johansson 2004; Perstorper *et al*. 1995; Kliger *et al*. 1997, 2003; Woxblom 1999; Johansson & Kliger 2002; Öhman & Nyström 2002; Öhman 1999). The hygroscopic and orthotropic nature of wood, the consequences for shrinkage and swelling, and the resulting distortion, i.e., twist, spring, and bow, were investigated by Ormarsson *et al*. (1999), Ormarsson, Dahlblom, & Petersson (2000), Ormarsson, Petersson, & Dahlblom (2000), Bengtsson (2001 a, b), and Johansson *et al*. (2003).

Despite the well-known physiological and technological relationship, the statistical correlations between reaction wood and timber quality criteria often vary, depending on the heterogeneity of the dataset. Apart from the type of measurement used for assessing reaction wood, the statistical modelling method might be a reason for the low correlations.

The formation, distribution, and effect of compression wood are highly variable because of its anatomical structure. Low correlations might derive from the different relationships between compression wood and wood quality in different groups of trees. Such relationships are modelled separately for different known groups — for example, stands — but often information about such groups is missing. However, neither stands nor stem sections nor other natural groups are homogeneous. The variability of wood properties is often higher within groups (e.g., stands) than between groups and it is assumed that the relationships between wood properties are also different within certain groups. Linear regression analyses with variable selection methods (e.g., stepwise) are used in wood science to find an optimal model to predict a wood quality criterion (Miller 1990). However, this statistical method assumes that all data are generated by the same underlying model. We applied a cluster analytical method, latent class regression (LCR) which is new in wood science, to disentangle wood characteristic data in order to identify classes or clusters defined by the relationship between wood quality criteria and any reaction wood present (Wedel & DeSarbo 1995). Unlike common cluster analysis, latent class regression classifies pieces with similar relationships between wood properties instead of similar wood properties. Latent class regression models belong to finite mixture models (FMM) (McLachlan & Peel 2000). The aim of this study was to use latent class regression to identify sawn timber specimens characterised by different regression relationships between wood formation characteristics and timber quality features. To date no data have been published using finite mixture models for modelling the influence of reaction wood on timber quality criteria. However, finite mixture models were used by Liu *et al.* (2002) to model the diameter frequency distributions of mixed-species stands.

MODELLING APPROACH

The statistical idea behind the latent class regression approach is to assume that the distribution of a wood quality criterion is the sum of several unknown distributions, called components, in a mixture distribution framework (Wedel & DeSarbo 1995; Vermunt & Magidson 2003, 2005). Each of these distributions represents one class of different relationships. This allows different classes of regression relationships between timber quality and wood formation to be distinguished. As in common cluster analysis, the meaning of the cluster is unknown and must be inferred from the estimated class-specific relationships between the variables included in the model and further external variables (e.g., stand), which might explain the class membership. If one supposes a mixed distribution consisting of k components, then the total probability density distribution of the outcome variable y (e.g., timber quality criterion) can be expressed as the sum of the specific conditioned probability density distribution $f_i(y|x)$ of the ith individual component (McLachlan & Peel 2000):

$$y \sim f(x) = \sum_{i=1}^{k} \lambda_i f_i(y|x = \lambda_1 f_1(y|x) + \ldots + \lambda_k f_k(y|x) \qquad (1)$$

where $0 \leq \lambda_i \leq 1 \quad (i = 1, 2, \ldots, k)$

$\sum_{i=1}^{k} \lambda_i = 1$

λ_i = the proportions of the *i*th components or classes

x = the indicators of wood formation (e.g., reaction wood).

The proportion of components and the parameters of the density functions $f_i(y|x)$ must be estimated, whereas the number of clusters must be given. The model presented here is univariate. Each wood distortion variable is treated separately. Tree or stand variables can be used as covariates to explain the class proportions l_i. If the relationships are homogeneous within stands and differ between stands, then the latent class proportion will be perfectly predicted by the stand factor. To estimate the parameter a certain density distribution function is assumed for the probability density function of the components. In this study a normal distribution function of f(y|x) was used (DeSarbo & Cron 1988, p. 254f):

$$y \sim f(x) = \sum_{i=1}^{k} \lambda_i f_i(y \mid \mathbf{X}, \sigma_i^2, \mathbf{b}_i) \tag{2}$$

$$= \sum_{i=1}^{k} \lambda_i (2\pi\sigma_i^2)^{-1/2} \exp\left[\frac{-(\mathrm{y}-\mathbf{X}\mathbf{b})^2}{2\sigma_i^2}\right]$$

where σ_i = the variance term of the *i*th cluster
b = vector of regression coefficients for the *i*th cluster
π = 3.14

Given a sample of n (n=1…N) independent observations, the likelihood expression is as follows:

$$L = \prod_{n=1}^{N} \left[\sum_{i=1}^{k} \lambda_i (2\pi\sigma_i^2)^{-1/2} \exp\left[\frac{-(\mathrm{y}-\mathbf{X}\mathbf{b})^2}{2\sigma_i^2}\right]\right] \tag{3}$$

The parameter b, σ_i, λ_i can be estimated simultaneously by using an EM-algorithm (Dempster *et al*. 1977). Additionally, for each piece the probability of joining a particular class can be calculated individually. The maximum of these probabilities indicates the class into which a certain specimen should be classified. If the latent classes are well separated, the probability of joining a certain class approximates to 0 or 1.

To find a model with an adequate number of clusters, information criteria such as Akaike's Information Criterion (AIC) (Bozdogan & Sclove 1984) or Schwarz-Bayes-Criterion (BIC) (Schwartz 1978) are used. The smaller the AIC or BIC, the better the model. Another measure for model comparison is the number of classification errors. If the class-specific distributions of the outcome variable overlap, the individuals cannot be classified unambiguously and this results in classification errors. In general, the latent class model for use for prediction in wood science stands alone, as the model fit cannot be assessed by standard approaches which rely on comparing fitted to observed values. In latent class models, the true class membership of a specimen is not observed and therefore the evaluation of the model fit is elusive (Formann 2003; Garrett *et al*. 2003). Nevertheless, the coefficient of determination as an indicator of the quality of prediction can be calculated separately for each latent class. The coefficient of determination might be higher in some classes than in a simple multiple regression using the whole data set. However, it is also possible to find latent classes with no relationship at all between the variables. Considering the probability of joining a certain class as a measure, conditional cross validation can be used to validate the prediction of the class-specific regression models. In the simplest case of cross validation the sample is split into a test sample and a validation sample. In the test sample, a model is estimated. The validation sample is used to compare fitted values to observed values using a measure of correlation.

The statistical analysis in a finite mixture framework has three objectives. Firstly, latent classes of different relationships between wood quality criteria and other wood-related properties have to be identified, if there are any. If the mixing distributions overlap slightly, individuals can be unambiguously classified into certain classes. Secondly, external variables which can be measured easily (e.g., tree or stand properties) have to be found to explain the class probabilities λ_i. This class probability will make it possible to predict the class membership of each individual. Thirdly, typical validation strategies can be used as cross validation to examine the stability of the new model.

MATERIAL AND METHODS

The data used originate from an EC-funded research project on the characterisation of compression wood and its relevance to sawn timber quality (Gardiner & McDonald 2004). Norway spruce trees from four different sites were sampled, measured, and assessed for log quality and sawn timber quality (67 trees (253 logs) sawn into approx. 800 cants, 391 of them included in the analysis,). The stand characteristics are presented in Table 1.

The four stands represented different degrees of wind exposure, slope, and silvicultural management which is thought to cause the formation of compression wood (Timell 1986). Depending on the site and stand characteristics, the age of the trees, and the social status of the individual trees in the stand structure, a slightly different distribution of the sample trees in three stem-form classes was found in the four stands (Table 2). The stem-form class was defined by the degree of lean or sweep in the 4-m butt log of each tree and was considered an indicator for differing internal wood structure, i.e., variable amount and different distribution of compression wood in the stem, which could affect the sawn wood produced from these trees. The logs were visually assessed, graded, and classified according to EN V DIN 1927-1 (Anon. 1998). The sawn timber (dimension: 50 × 100 mm^2) was visually assessed according to DIN 4074 (Anon. 1989) taking into account knots, ring width, slope of grain, distortion, and boxed pith. In a deviation from DIN 4074, compression wood was quantified separately for each face of the piece using a grid; each grid area was assigned to a class by its compression wood coverage (0%, 1–

25%, 26–50%, 51–75%, 76–100%). Spring, bow, and twist were measured using the FRITS measuring frame (Seeling & Merforth 2000) and the worst 2-m section was identified to calculate the distortion variables. Directly after sawing in fresh

TABLE 1–Stand characteristics

Stand code	IFU-01	IFU-02	IFU-03	IFU-06
Species	*Picea abies*	*P. abies*	*P. abies*	*P. abies*
Number of trees	19	16	16	16
Age in 2001/2002 (years)	69–110	90–120	45–60	117–124
Geographic location	47°54’ 8°4’	48°10’ 8°7’	48°2’ 7°59’	47°46’ 7°45’
Top height (m)	34.1	33.3	37.1	34.9
Dbh (cm)				
mean	46.2	42.3	33.3	25.8
upper quartile	49.7	47.0	40.0	45.0
median	45.8	42.5	35.3	14.0
lower quartile	42.1	37.0	27.0	9.0
Stand density (stems/ha)	766	864	822	850
Canopy closure	Moderately closed	Closed	Moderately closed	Light canopy
Stand structure	Single storey even-aged spruce	Single storey even-aged spruce	Single storey even-aged spruce with beech and fir	Multi-storey uneven-aged spruce with mixture

TABLE 2–Tree characteristics

Stand code	IFU-01	IFU-02	IFU-03	IFU-06
Number of trees in tree class*	d : 19 co : 0 s : 0	d : 2 co : 11 s : 3	d : 4 co : 11 s : 1	d : 2 co : 12 s : 2
Tree height (m) ±std dev.	30.4 ±2.8	31.9 ±2.7	32.6 ±1.8	32.7 ±2.7
Mean dbh (cm) ±std dev.	45.6 ±7.0	47.8 ±9.0	42.0 ±5.2	44.5 ±5.3
Max deviation in first 4-m stem (cm) ±std dev.	5.2 ±2.2	5.1 ±1.4	1.6 ±1.4	2.9 ±1.7
Number of trees in stem form class (1–3)†	1 : 1 2 : 9 3 : 9	1 : 0 2 : 9 3 : 7	1 : 12 2 : 4 3 : 0	1 : 7 2 : 7 3 : 2
Number of logs	62	64	64	63

* Tree class: d = dominant; co = co-dominant; s = suppressed
† Stem form class: 1 = stem dev. <2%; 2 = 2%–5%; 3 = >5%)

condition and after kiln drying to 12–14%, two measurements were taken for each piece. The characteristics of the material are summarised in Table 3.

TABLE 3–Cant characteristics

Stand code	IFU-01	IFU-02	IFU-03	IFU-06
Number of cants (n)	180	94	62	55
Dimension (mm × mm)	50×100	50×100	50×100	50×100
Position in cross section (n)				
Adult	87	39	62	55
Juvenile	93	51	0	0
Boxed pith	0	4	0	0
Spring fresh (mm)				
Mean ±std dev.	1.96 ±1.39	2.45 ±1.32	3.59 ±1.49	2.62 ±1.14
Max / min	9.90 / 0.50	7.80 / 0.60	6.50 / 1.10	7.00 / 1.10
Spring dry (mm)				
Mean ±std dev.	2.06 ±1.18	2.64 ±1.15	3.20 ±1.33	2.25 ±0.83
Max / min	7.70 / 0.00	5.80 / 0.70	8.50 / 1.40	4.70 / 0.80
Bow fresh (mm)				
Mean ±std dev.	2.70 ±1.54	2.59 ±1.39	3.31 ±0.98	3.53 ±1.54
Max / min	9.90 / 0.60	9.60 / 0.60	5.60 / 1.50	9.90 / 1.60
Bow dry (mm)				
Mean ±std dev.	2.19 ±1.35	2.29 ±1.14	2.96 ±1.32	2.79 ±0.99
Max / min	9.50 / 0.00	6.00 / 0.60	8.70 / 1.40	5.00 / 0.80
Twist fresh (mm)				
Mean ±std dev.	2.49 ±3.49	1.63 ±1.37	4.18 ±1.17	3.49 ±1.27
Max / min	28.40 / 0.90	8.3 / 0.50	8.00 / 2.00	7.00 / 2.00
Twist dry (mm)				
Mean ±std dev.	4.28 ±3.46	4.09 ±3.36	4.50 ±2.44	3.38 ±1.30
Max / min	17.70 / 0.00	21.20 / 0.70	14.90 / 2.00	8.90 / 1.00
Wood density RHO (kg/m^3)				
Mean ±std dev.	390.2 ±58.8	452.1 ±44.6	380.2 ±38.3	437.0 ±37.1
Max / min	771.0 / 284.8	615.1 / 380.1	559.9 / 299.1	507.5 / 356.0
Ring width (mm)				
Mean ±std dev.	2.70 ±0.89	2.75 ±0.78	4.14 ±0.89	1.67 ±0.42
Max / min	5.67 / 1.08	5.95 / 1.49	7.08 / 2.75	2.85 / 0.69
Ring orientation (°)				
Mean ±std dev.	44.43 ±24.83	50.52 ±22.31	29.76 ±10.45	31.28 ±10.00
Max / min	90.00 / 2.00	90.00 / 4.00	52.50 / 9.00	50.50 / 10.00
Cw (%)				
Mean ±std dev.	18.85 ±17.49	58.65 ±26.31	41.88 ±21.82	33.20 ±25.14
Max / min	100.00 / 0.00	100.00 / 8.59	87.50 / 5.47	100.00 / 3.13
CON -Index				
Mean ±std dev.	0.84 ±0.58	0.97 ±0.64	1.61 ±0.76	0.90 ±0.57
Max / min	2.75 / 0.00	2.93 / 0.00	3.80 / 0.31	2.58 / 0.00

Principal component analysis, carried out to identify the appropriate variables for latent class regression, included 68 variables which were tested against the six distortion variables spring fresh, spring dry, bow fresh, bow dry, twist fresh, and twist dry. The following parameters were included in the cluster analysis: stand, tree code, position of cant in longitudinal axis of the tree (log position in tree), compass position of cant in tree cross-section, radial position in cross-section (juvenile or adult wood), presence of boxed pith, oven-dry wood density, orientation of annual rings (angle between wide face and tangential line to the ring through the geometrical centre of the cross-section), ring width, slope of grain, compression wood percentage (mean and individual for all four faces), eight different indices to describe the spatial distribution of compression wood in the cant — evenness, dominance, contrast, Shannon index, entropy, Contagion index, relative Contagion index, angular second moment, inverse difference moment (Farina 1998) (mean and individual for all four faces). The following variables were selected as predictors for latent class regression: bow fresh, spring fresh, twist fresh, oven-dry density, ring width, ring orientation, radial position in cross-section (juvenile or adult wood), individual compression wood percentage on all four faces, and the contagion index. The contagion index represents the deviation of the entropy from its possible maximum value and measures the degree of clumping on the surface.

$$C = 2^{*}\ln\cdot(n\text{–}ENT) \quad (4)$$

where *ENT* is the entropy,
n is the number of possible compression wood classes, and
ln is natural logarithm.

The variables "stand" and "log position in tree" were tested as covariates to explain the class probabilities. The significance of parameters is tested by z-tests using the standard errors of parameters. All analysis was carried out using the statistical software package Latent Gold ver. 4.0 (Vermunt & Magidson 2005). Alternatively, an SAS-Macro of a latent class regression model without covariates was applied with similar results. Additionally, cross-validation was performed to validate the stability of the prediction. However, the stands could not be used to separate the sample into test and validation samples. Instead, the sample was randomly split into two subsamples; one subsample (20% of total sample) was used to estimate the regression equation and the other (80%) was a validation sample for predicting values using the estimated regression parameters. The correlation between predicted values and observed values of the validation sample serves as a stability criterion (Efron 1982). This cross-validation procedure was calculated separately for the two latent classes using the individual probability to join the class as weights in the regression (conditional cross-validation) and for the total sample without weighting and variable selection. Due to the fact that the stability might depend on the actual split of the sample, the procedure of drawing a sample and cross-validation was repeated 500 times to estimate a mean value and the confidence interval using Fisher-Z-transformation. This analysis was done using the SAS-MACRO.

RESULTS

Information criteria such as Akaike's Information Criterion (AIC) or Schwarz-Bayes-Criterion (BIC) and the level of classification errors were used to identify the number of latent classes (Table 4). The smallest values on these criteria indicate the optimal number of classes. Whereas the AIC decreased successively for all wood distortion variables, the BIC for two latent classes showed the smallest value for all wood distortion variables. The classification errors for two classes varied between 6% (spring dry) and 15% (twist dry). In contrast to the weak correlations between distortion parameters and compression wood variables over the entire dataset, we could therefore identify two classes with strong but different relationships between the tested variables. The linear regression models described the different effects of compression wood proportion, compression wood distribution index, and other wood features on the tested distortion variables for the two classes identified.

TABLE 4–Information criteria† for determining the optimal number of latent classes

Wood distortion	Number of latent classes	–2LogLik	AIC	BIC	CL-Error
Bow dry	1	–624.38	1274.77	1326.36	0
	2	–561.78	1189.57	1320.53	0.11
	3	–541.66	1189.37	1399.66	0.23
	4	–516.35	1178.70	1468.41	0.24
Spring dry	1	–578.79	1183.70	1235.18	0
	2	–503.72	1073.44	1204.41	0.06
	3	–427.29	1050.58	1260.92	0.18
	4	–432.01	1010.01	1299.73	0.17
Twist dry	1	–831.00	1691.99	1743.15	0.00
	2	–741.84	1549.67	1679.53	0.15
	3	–723.94	1553.89	1762.44	0.20
	4	–700.80	1547.59	1834.84	0.22

† –2LogLik = –2*Loglikelihood
AIC = Akaike's Information Criterion
BIC = Schwarz-Bayes Criterion
CL-Error = classification error.

For bow, the entire dataset split up in 79% (estimated) of the pieces in Class 1 and 21% (estimated) in Class 2 (Table 5). In Class 1, the tested variables accounted for 24% of explained variation, whereas in Class 2 the same predictors explained 41% of the variation in bow of the kiln-dried pieces. In contrast, over the entire dataset only 13% of the variation in bow was explained by single regression (R^2_{tot}). For spring, the model classified 89% of the pieces of the dataset in Class 1 and 11% in

Class 2. The variation by the different regression models explained 45% and 67% respectively of the variation, in contrast to 24% by single regression analysis. For twist, 61% of the pieces were classified in Class 1 (R^2=0.38) and 39% in Class 2 (R^2=0.33) by the latent class regression model (single regression over the whole dataset, R^2=15).

TABLE 5–Parameter† estimates of latent class regression of wood distortion

		Latent class regression for –					
		Bow dry		Spring dry		Twist dry	
R^2tot		0.13		0.24		0.15	
		Class1	Class2	Class1	Class2	Class1	Class2
R^2		0.24	0.41	0.45	0.67	0.38	0.33
Intercept	b_0	0.5916	1.4984	–0.1161	2.2486	1.9664	–0.3595
Predictors							
BOW fresh	b_1	0.1374*	0.4029	0.4205*	–0.1852	0.3381*	0.1789
RHO mean	b_2	0.0007	–0.0005	0.0018*	–0.0025	0.0034	0.0079
RINGWIDTH	b_3	0.1827*	0.2229	0.1317*	0.4351	0.0805	0.6642*
RING_ORIENT	b_4	–0.0035	–0.0156	0.0014	–0.0228	–0.0073	0.0254*
JUVAD							
0	b_5	–0.6604*	–1.3650	–0.0833	–1.8926	3.0178*	1.8158
1	b_6	0.2646	1.4483	0.0493	0.4848	–1.5318*	–0.3568
2	(b_7)	0.3958*	–0.0833	0.0340	1.4078	–1.4860*	–1.4590
CW1	b_8	0.0137*	–0.0261	–0.0064	0.0759*	0.0113	0.1171*
CW2	b_9	0.0303*	–0.0183	0.0009	0.0455	0.0205	0.0864*
CW3	b_{10}	–0.0169*	0.0290	0.0022	–0.0657*	–0.0176	–0.0976*
CW4	b_{11}	–0.0206*	0.0222	0.0036	–0.0284	–0.0141	–0.0967
CON	b_{12}	–0.1357	0.7169	0.1560*	0.7433	0.0427	–0.1735
Error variances							
BOW dry		0.5416	1.7562	0.5214	1.4610	0.8584	5.1712
Class size		0.79	0.21	0.89	0.11	0.61	0.39

* significant at $p < 0.05$

† RHO mean = ovendry density
RING_ORIENT = ring orientation
JUVAD — 0 = boxed pith, 1 = juvenile wood, 2 = adult wood)
CW1= wide cant face orientated towards bark
CW2 = narrow cant face 1
CW3= wide cant face orientated towards pith
CW4 = narrow cant face 2
CON= Contagion index (compression wood distribution index, over all four faces).

Due to effect coding of JUVAD one dummy variable and its coefficient (b_7) is redundant and can be expressed out of the other parameters (b_5, b_6). In each column the estimated raw regression parameters are shown with the variances of residuals (error variance), the class proportion (class size), whereas R^2tot = coefficient of determination of the multiple regression for the whole sample and R^2 the coefficients of determination for the two latent classes.

For all three variables (bow, spring, and twist), we found the tested predictors with opposite effects in the two classes and of varying significance ($p \leq 0.05$). Compression wood was always a statistically significant predictor in one class, but not the other (bow – Class 1; spring, twist – Class 2). For this second class the initial bow_{fresh}, $spring_{fresh}$, or $twist_{fresh}$ before drying was then a statistically significant predictor. The effect of compression wood on the outer face (CW1) and one narrow face (CW2) (mainly positive effect) was qualified by the effect of compression wood on the inner cant face (CW3) and the second narrow face (CW4) (mainly negative effect). However, the spatial distribution of compression wood (clumping of patches) calculated as contagion index contributed significantly to the explained variation of spring only in Class 1. Overall, the normalised effects, bow, spring, and twist increased with increasing compression wood.

According to the different models for bow, spring, and twist in Class 1 and Class 2, wood density, ring width, ring orientation, and the position of the piece in the cross-section (juvenile or adult wood) affected the distortion differently. For twist, boxed pith (JUVAD Class 1) had a very strong positive (i.e., inclining) effect, whereas pieces sawn more peripherally without pith (only juvenile or adult wood) seemed less prone to heavy twisting. For spring and bow, boxed pith seemed less influential. We also tested the contribution of “stand” and “log position in the tree” as covariates for their contribution to the classification. “Stand” contributed only to the classifications in twist, but not in bow and spring. The position of the log in the tree did not have any effect on the distortion variables.

In each class the estimated distribution of bow, spring, and twist showed large deviations from the distribution predicted from the models for some of the stands, especially for stand IFU-06 (Fig.1). For bow, in stands IFU-01, IFU-02, and IFU-03 the estimated frequency was slightly over-estimated by 2% to 8%, whereas for IFU-06 the frequency was strongly under-estimated by 21%. For spring, the frequency in Class 1 was under-estimated ($\Delta = -6\%$) only for IFU-01, whereas for IFU-02, IFU-03, and IFU-06 the frequency in Class 1 was over-estimated between 2% and 11% in comparison to the predicted frequency; the opposite held true for Class 2. For twist, we found the largest deviations in predicted frequency from the regression models and estimated frequency for the stands. For stand IFU-01, the distribution of the pieces in both classes was more even with 48% (Class 1) to 52% (Class 2) greater than predicted ($\Delta = +/-13\%$). For all other stands, frequency in Class 1 was over-estimated by 5% to 23%, and vice versa, in Class 2 frequency was under-estimated to the same extent.

In the last step of the analysis the stability of the given regression solution is examined using cross-validation (Table 6). For all distortion variables the strength of the mean correlation between observed and predicted values decreased somewhat in the validation sample, indicating a loss of predictability of the regression models,

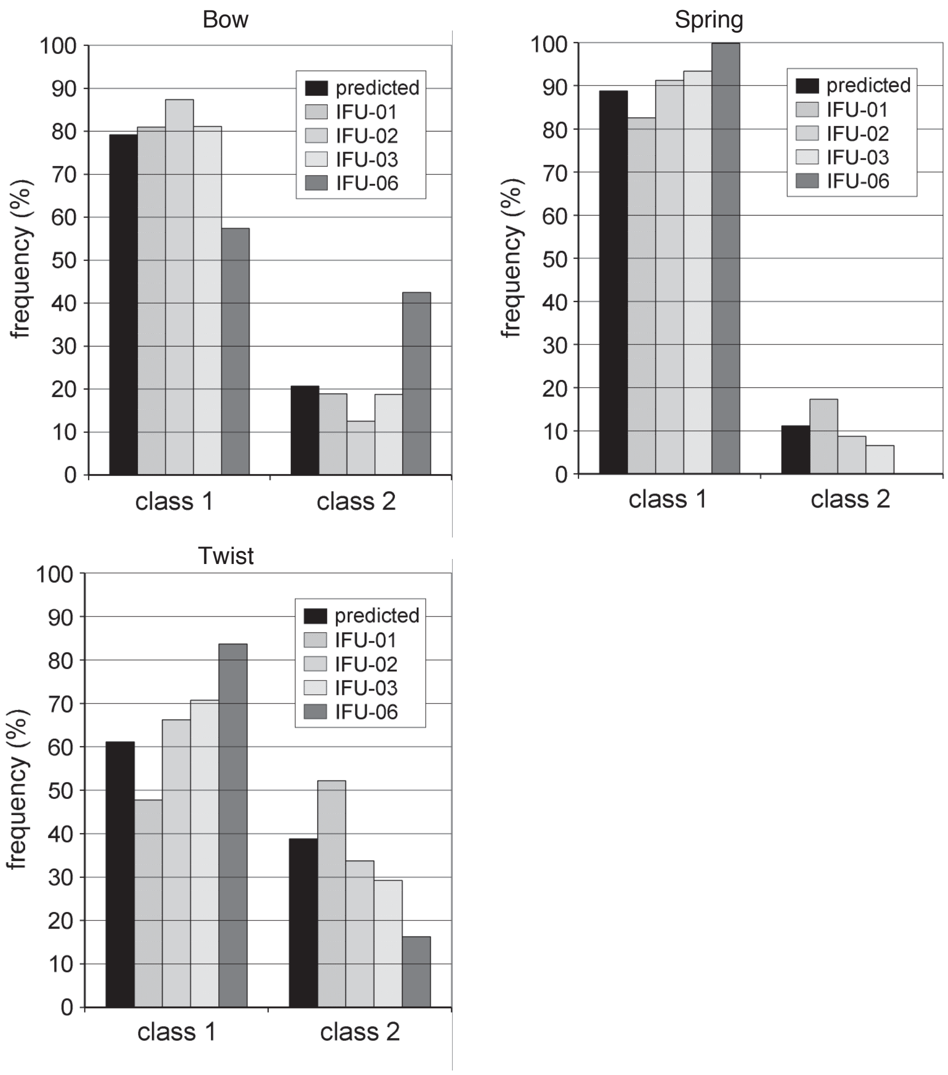

FIG. 1–Class frequency of regression models for bow, spring, and twist; predicted frequency in regression models compared to the estimated distribution of pieces in Classes 1 or 2 separate for each stand

especially for bow in latent Class 2 (from 0.65 to 0.56). The 95% confidence intervals for the validation sample were much higher than for the test sample for all conditions. However, the mean correlations in test and validation sample were always higher for the latent classes than for the total sample. The regression models for the latent classes seem to be as stable as the regression models for the total sample without weighting, under the assumption that the individual class probabilities are known (conditional cross-validation).

TABLE 6–Mean correlations of test and cross-validation samples (n=500 random samples)

Wood distortion	Weighting*	Test sample	cl95T	cu95T	Validation sample	cl95V	cu95V
Bow	No weighting	0.37	0.32	0.41	0.30	0.09	0.48
	Latent Class 1	0.49	0.45	0.53	0.45	0.26	0.60
	Latent Class 2	0.65	0.59	0.70	0.56	0.25	0.76
Spring	No weighting	0.50	0.45	0.54	0.45	0.24	0.61
	Latent Class 1	0.67	0.64	0.70	0.64	0.49	0.75
	Latent Class 2	0.82	0.78	0.86	0.76	0.44	0.91
Twist	No weighting	0.35	0.30	0.40	0.27	0.05	0.47
	Latent Class 1	0.56	0.52	0.60	0.52	0.33	0.66
	Latent Class 2	0.57	0.52	0.62	0.52	0.28	0.69

* Each value is weighted either with 1 (no weighting) or the probability of a specimen belonging to a certain class (Latent Class1 or 2) cl95/cu95 = lower and upper 95% confidence interval of the mean value of test sample (T) or validation sample (V)

DISCUSSION

The results of this study showed that it is possible to improve regression models and increase the predictability of the model (e.g., R^2) in comparison to a single regression analysis by applying latent class regression to highly heterogeneous datasets. Latent class regression improved R^2 for predicting the distortion variables bow, spring, and twist from structural wood characteristics by a factor of 2 to 3 for the individual regressions of identified classes. Cross validation can further support the findings. The same predictors (distortion parameters in fresh condition, wood density, ring width, ring orientation, wood type (juvenile or adult), compression wood percentage, compression wood distribution index) might affect the distortion variables in a different way for different subgroups in the dataset. The common thesis, that wood properties are homogeneous within stands, cannot be held. There are different classes of wood relationships within the same stand.

This method presents the opportunity to analyse heterogeneous datasets in a way that allows the relationships for these subgroups to be characterised in more detail. The objectives of the statistical analysis in a finite mixture framework are as follows. Firstly, it has to be determined whether there are latent classes of different relationships between wood quality criteria and other wood-related properties. If the mixing distributions overlap only slightly, individuals can be unambiguously classified into certain classes. Secondly, external variables that can be measured easily (e.g., tree or stand properties) have to be found to explain the class probabilities $\lambda_{i,}$ which makes it possible to predict the class membership of each individual. Thirdly, certain kinds of validation strategies such as cross validation can be used to examine the stability of the given solution. In our study, the effects of the tested predictors were different in the contrasting classes, i.e., sometimes they were opposite. Thus, it must be concluded that not all the predictors tested affected distortion in the same ways for all sawn pieces of the dataset that were tested. For the dataset and the predictors tested we found that compression wood had a significant effect only on the variation for one class, for bow, spring, and twist. For the other class, either the initial bow (or spring or twist) in fresh condition was highly significant, or the wood type (juvenile or adult) as determined by the radial position of the piece in the cross-section of the log. Initial distortion and wood type on one side, or compression wood on the side, seem to explain a large proportion of the variability of sawn timber with respect to distortion. These results are in accordance with findings reported by Öhman & Nyström (2002), Johansson *et al*. (2001), and Perstorper *et al*. (2001).

REFERENCES

ANON. 1989: DIN 4074 – Sortierung von Nadelholz nach der Tragfähigkeit (Nadelschnittholz). [Strength sorting of coniferous wood; coniferous sawn timber]. *Deutsches Institut für Normung e.V.*

——1998: EN V DIN 1927-1. Qualitative classification of softwood round timber — Part 1: Spruces and firs. *Deutsches Institut für Normung e.V.*

BEARD, J.S.; WAGNER, F.G.; TAYLOR, F.W.; SEALE, R.D. 1993: The influence of growth characteristics on warp in two structural grades of southern pine lumber. *Forest Products Journal 43(6)*: 51–56.

BEAULIEU, J.; GIRARD, B.; FORTIN, Y. 2003: Study of geographical variation in kiln-drying behaviour of plantation-grown white spruce. *Wood Fibre Science 25*: 45–67.

BENGTSSON, C. 2001a: Mechano-sorptive bending creep of timber — influence of material parameters. *Holz als Roh- und Werkstoff 59(4)*: 229–236.

——2001b: Variation of moisture induced movements in Norway spruce (*Picea abies*). *Annals of Forest Science 58(5)*: 569–581.

BOZDOGAN, H.; SCLOVE, D.L. 1984: Multi-sample cluster analysis using Akaike's information criteria. *Annals of the Institute of Statistics and Mathematics 36*: 163–180.

BURGERT, I.; FRÜHMANN, K.; KECKES, J.; FRATZL, P.; STANZL-TSCHEGG, S. 2004: Structure-function relationships of four compression wood types: micromechanical properties at the tissue and fibre level. *Trees: Structure and Function 18(4)*: 480–485.

COWN, D.J.; BALL, R.D.; RIDDELL, M.J.C. 2004: Wood density and microfibril angle in 10 *Pinus radiata* clones: Distribution and influence on product performance. *New Zealand Journal of Forestry Science 34(3)*: 293–315.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, R.B. 1977: Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B39*: 1–38.

DeSARBO, W.S.; CRON, W.L. 1988: A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification 5*: 249–282.

DONALDSON, L.A.; GRACE, J.; DOWNES, G.M. 2004: Within-tree variation in anatomical properties of compression wood in radiata pine. *IAWA Journal 25(3)*: 253–271.

EFRON, B. 1982 : "The Jacknkife, the Bootstrap and Other Resampling Plans". SIAM, Philadelphia.

FARINA, A. 1998: "Principles and Methods in Landscape Ecology". First ed. Chapham & Hall.

FORMANN, A.K. 2003: Latent class model diagnosis — A review and some proposals. *Computational Statistics & Data Analysis 41*: 549–559.

GARDINER, B.; McDONALD, E. (Ed.) 2004: Compression wood in conifers – Characterisation of its formation and its relevance to timber quality. Final report EC project, Contract: QLRT-2000-00177, Quality of Life & Management of Living Resources, FP5.

GARRETT, E.; MIECH, R.; OWENS, P.; EATON, W.W.; ZEGER, S.L. 2003: Checking assumptions in latent class regression models via a Markov chain estimation approach: An application to depression and socio-economic status. *John Hopkins University, Department of Biostatistic, Working Paper 17*.

JOHANSSON, M. 2004: Prediction of bow and crook in timber studs based on variation in longitudinal shrinkage. *Wood and Fiber Science 35(3)*: 445–455.

JOHANSSON, M.; KLIGER, R. 2002: Influence of material characteristics on warp in Norway spruce studs. *Wood and Fiber Science 34(2)*: 325–336.

JOHANSSON, M.; NYSTROM, J.; ÖHMAN, M. 2003: Prediction of longitudinal shrinkage and bow in Norway spruce studs using scanning techniques. *Journal of Wood Science 49(4)*: 291–297.

JOHANSSON, M.; PERSTORPER, M.; KLIGER, R.; JOHANSSON, G. 2001: Distortion of Norway spruce timber. Part 2. Modelling twist. *Holz als Roh- und Werkstoff 59(3)*: 155–162.

KLIGER, R.; JOHANSSON, M.; PERSTORPER, M.; JOHANSSON, G. 2003: Distortion of Norway spruce timber. Part 3: Modelling bow and spring. *Holz als Roh- und Werkstoff 61(4)*: 241–250.

KLIGER, R.; PERSTORPER, M.; JOHANSSON, G.; BENGTSSON, C. 1997: Mechanical properties of sawn timber: influence of stand type and raw material characteristics. Pp. 132–141 *in* Nepveu, G. (Ed.) IUFRO WP S501 04 Proceedings of second workshop: "Connection between Silviculture and Wood Quality through Modelling Approaches and Simulation Software". Berg en Dal Kruger National Park, South Africa, 26–31 August 1996.

LIU, C.; ZHANG, L.; DAVIS, C.J.; SOLOMON, D.S.; GOVE, J.H. 2002: A finite mixture model for characterizing the diameter distributions of mixed species forest stands. *Forest Science 48(4)*: 653–661.

LIVINGSTON, A.K.; CAMERON, A.D.; PETTY, J.A.; LEE, S.L. 2004: Effect of growth rate on wood properties of genetically improved Sitka spruce. *Forestry 77(4)*: 325–334.

McLACHLAN, G.; PEEL, D. 2000: "Finite Mixture Models". John Wiley, New York. 419 p.

MILLER, A.J. 1990: "Subset Selection in Regression". Chapman & Hall, New York.

MUTZ, R. 1998: "Inhomogenität des Roh- und Werkstoffs Holz. Konzeptuelle, methodische und empirische Implikationen für holzkundlichen Untersuchungen". Dr. Kovac, Hamburg. 352 p.

MUTZ, R.; GUILLEY, E.; SAUTER, U.H.; NEPVEU, G. 2004: Modelling juvenile-mature wood transition in Scots pine (*Pinus sylvestris* L.) using nonlinear mixed effects-models. *Annals of Forest Sciences* 61: 831–841.

ÖHMAN, M. 1999: Growth-stress related warp of green sawn timber as an indicator of the amount of compression wood. Pp. 61–68 *in* Berti, S; Macchioni, N.; Negri, M.; Rachello, E. "Industrial End Uses of Fast Grown Species". Eurowood Technical Workshop Proceedings, Florence, Italy, 31 May–1 June.

ÖHMAN, M.; NYSTRÖM, J. 2002: Measurement of green plank shape for prediction and elimination of compression wood. *Scandinavian Journal of Forest Research 17(4)*: 377–384.

ORMARSSON, S.; DAHLBLOM, O.; PETERSSON, H. 1999: Influence of spiral grain on stiffness grading of structural timber. Pp. 63–69 *in* Walford, G.B.; Gaunt, D.J. (Ed.) Proceedings of the Pacific Timber Engineering Conference Volume 1, 14–18 March, Rotorua, New Zealand. *New Zealand Forest Research Institute, Forest Research Bulletin No. 212*.

——2000: A numerical study of the shape stability of sawn timber subjected to moisture variation. Part 3: Influence of annual ring orientation. *Wood Science and Technology 34(3)*: 207–219.

ORMARSSON, S.; PETERSSON, H.; DAHLBLOM, O. 2000: Numerical and experimental studies on influence of compression wood timber distortion. *Drying Technology 18(8)*: 1897–1919.

PERSTORPER, M.; JOHANSSON, M.; KLIGER, I.R.; JOHANSSON, G. 2001: Distortion of Norway spruce. *Holz als Roh- und Werkstoff 59(1/2)*: 94–103.

PERSTORPER, M.; PELLICANE, P.J.; KLIGER, I.R.; JOHANSSON, G. 1995: Quality of timber products from Norway spruce. Part 2. Influence of spatial position and growth characteristics on warp. *Wood Science and Technology 29(5)*: 339–352.

SCHWARTZ, G. 1978: Estimating the dimensions of a model. *Annals of Statistics 6*: 461–464.

SEELING, U. 1999: Einfluss von Richtgewebe ("Druckholz") auf Festigkeit und Elastizitat des Fichtenholzes [Influence of compression wood for strength and elasticity of timber of spruce]. *Holz als Roh- und Werkstoff 57(2)*: 81–91.

——2001: Transformation of plantation forests — expected wood properties of Norway spruce (*Picea abies* (L.) Karst.) within the period of stand stabilisation. *In* Cameron, A.D.; Mason, W.L.; Malcolm, D.C. (Ed.) *Forest Ecology and Management 151(1/3)*: 195–210.

SEELING, U; MERFORTH, C. 2000: FRITS — a new equipment to measure distortion. *Holz als Roh- und Werkstoff 58(5)*: 338–339.

TIMELL, T.E. 1986: "Compression Wood in Gymnosperms". Vol. I, II, III. Springer Verlag.

VERMUNT, J.K.; MAGIDSON, J. 2003: Latent class models for classification. *Computational Statistics & Data Analysis 41(3-4)*: 531–537.

VERMUNT, J.K.; MAGIDSON, J. 2005: "Technical Guide for Latent GOLD 4.0: Basic and Advanced". Statistical Innovations Inc., Belmont, Massachusetts.

WEDEL, M.; DeSARBO, W. 1995: A mixture likelihood approach for generalized linear models. *Journal of Classification 12*: 21–55.

WERNSDÖRFER, H.; RECK, P.; SEELING, U.; BECKER, G.; SEIFERT, T. 2004: Erkennung und Messung des Reaktionsholzes bei Fichte (*Picea abies* (L.) Karst.) mittels Verfahren der digitalen Bildanalyse. *Holz als Roh- und Werkstoff 62(4)*: 243–252.

WOXBLOM, L. 1999: Warp of sawn timber of Norway spruce in relation to end-user requirements: quality, sawing pattern and economic aspects. *Acta Universitatis Agriculturae Sueciae Silvestria 126*: 138.