

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Application of Principal Components Regression for Analysis of X-Ray Diffraction Images of Wood

Joshua C. Bowden¹ and Robert Evans²

¹CSIRO Information Management and Technology,

²CSIRO Materials Science and Engineering,
Australia

1. Introduction

We report on the use of Principal Component Regression (PCR), a least squares calibration method employing Principal Component Analysis (PCA) for the analysis of large sets of images obtained by wide angle X-ray diffractometry (WAXD) of wood. PCR multivariate models were created for prediction of microfibril angle (MFA), an important wood structural parameter. Multivariate regression techniques have previously been applied to X-ray diffraction data for a variety of analytical problems. These include analysis of complex mineral mixtures (Karstang and Eastgate 1987) and for automated high-throughput analysis of powder diffraction profiles obtained by azimuthal averaging of 2D diffraction patterns (Gilmore, Barr et al. 2004). Multivariate classification from X-ray diffraction data has also been used for discrimination of insulin microcrystals (Norrman, Stahl et al. 2006) and brucite particle morphology (Matos, Xavier et al. 2007). The models created in this study used pixels from across the full area of the diffraction image as predictor variables, a technique which has not been found applied to diffraction data in the literature. A related technique, multivariate analysis of congruent images (MACI), was introduced by Eriksson, Wold et al. (2005) for discrimination and quantification of events within video image sequences. MACI uses wavelet analysis to decompose images and obtain predictive information which is correlated with frame-by-frame differences and that can describe properties such as the movement of an object or change in composition of a feature.

This chapter describes four predictive PCR models, two that are species-specific, and two that incorporate multiple species. Differences between these models, and their relative accuracy, will be discussed. SilviScan will be described and the issues relating to management of the large amounts of data collected by the system will be examined.

2. Multivariate calibration

Principal component regression is a widely used multivariate calibration technique for creating predictive models by relating multivariate sets of measurements (X) to sets of dependent variables (y) (Hoskuldsson 1995). PCR uses the properties of principal

component analysis to reduce the dimensionality of the original data to form linearly independent factors (\mathbf{T}) required for least squares regression analysis. In the case of X-ray diffraction analysis, modelling of a dependent variable using PCR relies on the proportionality that exists between diffracted X-ray energy and the quantity of a compound or structural feature present within a sample. If there is a linear relationship with the predictand, then PCR will be a good choice for modelling the variable using X-ray diffraction images. A linear relationship between a dependent variable and diffraction data is valid providing the diffraction experimental parameters are within prescribed limits of a materials density (these limits are dependent on X-ray wavelength) and providing all other experimental conditions are kept constant (Tripp and Conrad 1972; Cave and Robinson 1998).

PCR involves three main steps. Step 1 is data reduction using PCA to obtain scores (\mathbf{T}) and normalised eigenvectors (\mathbf{P}) (Equation 1 and 2). Step 2 uses \mathbf{T} as the independent variables matrix in the inverse least squares (ILS) calibration (Kramer 1998) (Equation 3 and 4), and step 3 produces the regression coefficients (in terms of the original \mathbf{X} variables) required for predicting the properties ($\hat{\mathbf{y}}$) of unknown samples (Equation 5). PCA is represented by Equation 1 (Brereton 2003). The eigenvector and score pairs are termed principal components (PCs) and are numbered from PC1 to PCn, in decreasing order of the variance associated with the vectors in the original data matrix. The matrix \mathbf{R} is the residual data matrix which is the variance not associated with the principal components chosen, and which would generally describe noise. Equation 2 represents the projection of the eigenvectors onto the original data to obtain the reduced data (scores) \mathbf{T} matrix:

$$\mathbf{X}_{m,n}^{cal} = \mathbf{T}_{m,c} \mathbf{P}_{c,n} + \mathbf{R}_{m,n} \quad (1)$$

$$\mathbf{T}_{m,c} = \mathbf{X}_{m,n}^{cal} \mathbf{P}_{c,n}' \quad (2)$$

where m is the number of samples, n is the number of variables - in our case the total number of pixels in the image, c is the number of components and $'$ indicates the transpose. For analysis involving images, the term 'eigenimage' is used to denote the 2D version of the eigenvector that is obtained from PCA.

The equations required for the inverse least squares regression calibration step are:

$$\mathbf{y}'_{m,1} = \mathbf{b}_{1,c} \mathbf{T}'_{m,c} \quad (3)$$

$$\mathbf{y}'_{m,1} \cdot \mathbf{T} [\mathbf{T}' \mathbf{T}]^{-1} = \hat{\mathbf{b}}_{1,c} \quad (4)$$

$$\boldsymbol{\beta}_{1,n} = \hat{\mathbf{b}}_{1,c} \mathbf{P}_{c,n} \quad (5)$$

$$\hat{\mathbf{y}}_{x,1} = \mathbf{X}_{x,n} \boldsymbol{\beta}'_{1,n} + \bar{y}_{est} \quad (6)$$

where $\mathbf{y}_{m,1}$ is the vector of dependent variables, \mathbf{b} and $\hat{\mathbf{b}}_{1,c}$ are the true and best estimates for the least squares coefficients, $\boldsymbol{\beta}_{1,n}$ are the regression coefficients in terms of the original \mathbf{X}

variables, \hat{y} are the predicted (best estimates of) y values from any other set of data and \bar{y}_{est} is the best estimation of the average y data.

The ILS calibration requires that the number of factors used (c) is less than or equal to the number of samples (m). Once the calibration is performed, y data may be predicted from unknown samples (\mathbf{X}) using the regression coefficients β of the original \mathbf{X} variables and the best estimation (\bar{y}_{est}) of the average y data. The \bar{y}_{est} value can be obtained using an added column vector of 1's in the \mathbf{T} matrix and will equal the average of the y data if the \mathbf{X} matrix was mean centred prior to the PCA step.

The scores obtained from PCA are orthogonal and rule out modelling techniques that use cross-terms ($\mathbf{T}_{m,c1} \times \mathbf{T}_{m,c2}$). However, both cross and quadratic ($\mathbf{X}_{m,n} \times \mathbf{X}_{m,n}$) terms may be introduced into the initial \mathbf{X} data, which in the case of large data sets comes at the risk of excessive memory use. Variables can be scaled prior to the PCA step by one of a number of techniques such as mean centering (subtracting the mean of a variable), variance scaling (equalising variance) and range scaling (normalising data to lie within a nominated range). Scaling techniques are used to improve the predictive capability of the data (Berg, 2006) and can highlight low variance variables that may otherwise be left in the \mathbf{R} matrix during the PCA step. Vector normalisation (VN) is another scaling technique that normalises between samples (rather than variables) and involves the following transformation:

$$\mathbf{x}_{vn} = \frac{\mathbf{x}}{\sqrt{\sum_{i=1}^n x_i^2}} \quad (7)$$

where \mathbf{x}_n is the normalised vector, \mathbf{x} the original vector and x_i the i th variable of the sample.

$$\mathbf{H} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}' \quad (8)$$

Removal of high leverage samples from the model data set before inclusion in the model training data is good practice as they influence the model in a disproportionate fashion. The leverages for each sample are the values of the diagonal of the 'hat' matrix (\mathbf{H}) as seen in equation 8 (Dobson, 2002).

A useful measure of the accuracy of a multivariate regression model is the root mean square standard error of prediction (RMSEP) (Faber, Song et al. 2003):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \quad (9)$$

and is a measure of the average error between a test set of reference samples and a models predicted values (\hat{y}). This measure is biased high when using imprecise reference values (y) and can be corrected if the standard deviation of the reference data is known (Bro, Rinnan and Faber, 2005). Tests are available for determination of the per-sample prediction error, based on the leverage of the sample (see Bro, Rinnan and Faber, 2005).

The optimum number of components to keep in a model can be determined from the minimum in RMSEP versus the number of components (Martens and Naes 1989). Another common measure of prediction error is RMSEP from cross calibration. Cross calibration requires generation of multiple models with exclusion of a small number of samples (usually one) from each model and calculating the error as the average of the RMSEP of the excluded samples for all models. This technique is not practical or necessary when data sets are large and in our case where there would be a large number of models to be made (each potentially requiring PCA on the subset of data), the computational requirements would become excessive.

3. X-ray diffraction of wood and the SilviScan system

Important factors affecting the diffraction patterns of wood samples include cell shape and microfibril angle (MFA) (Evans 1999; Lichtenegger, Reiterer et al. 2001), cellulose crystallite dimensions and cellulose crystallinity (Wada, Okano et al. 1997; Washusen and Evans 2001; Andersson, Serimaa et al. 2003; Garvey, Parker et al. 2005). These properties have been shown to influence many of the bulk properties of wood including modulus of elasticity (MoE) for structural timber (Evans 2006) and important properties such as pulp yield, a useful metric for production capacity in the pulping industry (Downes, Evans et al. 2003). The fast and accurate estimation of these attributes is useful in large scale surveys of wood properties for genetic selection trials and forestry management (Lindstrom, Harris et al. 2004; Thumma, Nolan et al. 2005).

Wide angle x-ray diffractometry has been developed as an integral part of the high-throughput instrumentation known as SilviScan, now in its 3rd generation. SilviScan consists of three complimentary techniques, transverse (radial-tangential) surface microscopy and image analysis, X-ray densitometry and WAXD. These three techniques are used to analyse 2mm by 7mm radial sections, typically cut from 12mm diameter increment cores taken from standing trees (Fig. 1).

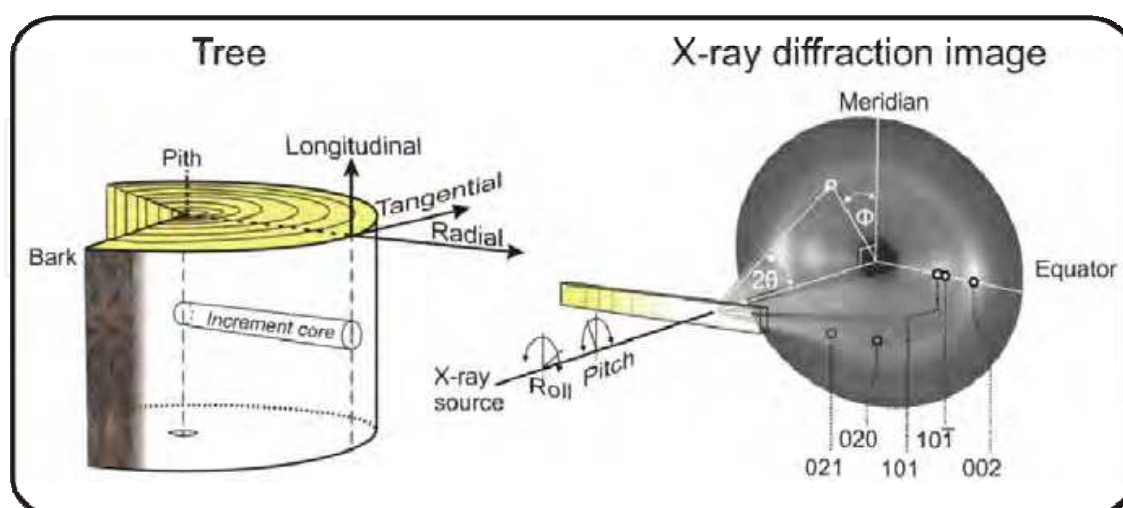


Fig. 1. Typical geometry of wood samples used in relation to the tree and the diffraction pattern produced. Miller indices associated with the observed reflections are indicated. The commonly used 040 meridional reflection is outside the range of the camera and is unreliable when fibre orientation varies as it does in core samples.

The interaction of an X-ray beam with the wood fibre structure gives diffraction images that contain information about the helical orientation of the fibres (MFA). A geometrical relationship between MFA and the variance of the observed 002 azimuthal peaks (Φ direction, Fig.1) is currently used by SilviScan software to determine MFA (Evans 1999). In addition, 3D cellular orientation is estimated (roll and pitch relative to the x-ray beam). Roll is calculated from the rotation of the diffraction pattern around the beam direction and pitch is calculated from the distortion of equatorial reflections towards the meridional axis. Pitch (fibre tilt in the direction of the X-ray beam) is important as it causes asymmetric peak broadening and an overestimation of MFA (Stuart and Evans 1995; Evans, Stringer et al. 2000), which has to be corrected for during SilviScan analysis.

This work demonstrates the applicability of multivariate statistical calibration methods for determination of MFA from WAXD images obtained on SilviScan. The same method can be used to extract information on fibre orientation, crystallite size and any other property that affects the images. Occasionally, when MFA is very high, the diffraction patterns have almost no azimuthal intensity variation, making them very difficult to analyse by any method that relies on the determination of azimuthal peak widths. The consequences of this problem are data 'dropouts' and an example of them are seen later in the results section. Thus overcoming this problem through calibrated multivariate techniques for estimation of MFA is of interest.

4. Material and methods

4.1 SilviScan

The SilviScan 3 instrument scans samples in three stages. First, automated microscopy and image analysis of the polished top transverse surface (radial-tangential plane) gives growth-ring boundary orientation, ray cell orientation, and average radial and tangential fibre diameters binned in 25 micron intervals. Second, automated X-ray scanning densitometry (customised 1392×1040 pixel camera, Photonic Science, UK) gives density variation binned in 25 micron intervals and normalised to the actual average sample density obtained by independent gravimetric analysis. Third, automated scanning X-ray diffractometry (customised 1392×1040 pixel intensified camera, Photonic Science, UK) using a nickel-filtered Cu source and 200 micron diameter capillary-focussed beam, gives microfibril angle (MFA) and fibre axis orientation in 100 micron steps. MFA was corrected for fibre orientation in the beam direction.

4.2 Wood samples

All wood samples (two softwoods and one hardwood) were taken from previous projects for which the SilviScan data was available. The samples were in the form of 2mm wide (tangential dimension) and 7mm high (longitudinal dimension) pith-to-bark (radial) strips. Samples varied in length from 20 mm – 200 mm.

30-year-old *Pinus caribaea* var. *hondurensis* (*P. caribaea*) from Beerburrum, in south-east Queensland, Australia. Twelve radial samples from an original 520 (65 trees, 8 heights per tree) were selected for high resolution analysis on SilviScan-3 at the 4m sampling height. 18486 diffraction images were taken of these samples, 10000 used for creating the models and 6925 images were used for model verification.

23-year-old *Pinus radiata* (*P. radiata*) from Tallaganda, Canberra, Australia were taken from a set of previously well characterised samples as 2mm tangential, 7 mm longitudinal and 20 mm radial strips (Schimleck 2002). The resulting diffraction dataset consisted of 11822 images from 116 paired ends of the same sample.

Eucalyptus delegatensis (*E. delegatensis*) from East Gippsland, Victoria, Australia were taken as an example of a hardwood (Evans and Ilic 2001). The resulting diffraction image dataset consisted of 10638 images from 102 paired ends of the same sample.

4.3 Data analysis

Principal component analysis was undertaken using in-house software developed using C++ and OpenCL and making use of high end graphics processing units (Bowden 2010). Images were reduced to 128 by 128 pixels by selecting 1 out of 8 pixels in each dimension. This resulted in a 16384 element vector for each image in the data sets. Inverse least squares software was written in-house and the combined system used hierarchical data format (HDF) files for data storage and management. HDF Viewer version 2.7 and Microsoft Excel 2003 was used for visualisation and data filtering (HDF 2000).

The modelling procedures involved the following: Two datasets were used to create the models, one from the *P. caribaea* data set (10000 images, ~7 wood samples) and another from a combination of *P. radiata* (5000 images), *E. delegatensis* (5000 images) and *P. caribaea* (10000 images), making a 20000 image data set. These are referred to as the 'single-species' and 'multi-species' data sets. From each data set, models were made using mean centred image data of both vector normalised (referred to as 'VN' models) and unmodified data. Models were then tested with 6925 *P. caribaea* and 5000 samples of either *P. radiata* or *E. delegatensis*. RMSEP model verification was carried out on a per species basis with models from both data sets.

5. Results

Fig. 2 shows two diffraction images at opposite ends of the range of MFA values seen in *P. caribaea* samples. Fig. 2a is a low MFA sample (9°), with very narrow azimuthal peaks (101, $10\bar{1}$ and 002 in order from the centre) close to the equator. Fig. 2b is an example of a high MFA sample (30°) and the azimuthal peaks are much broader. The arrows (Fig. 2a) on the 002 reflections indicate the direction of azimuthal broadening in high MFA samples.

Fig. 3 shows the first 18 eigenvectors/eigenimages from PCA of the single-species dataset. The eigenimages show various symmetry features involving rotation, reflection and inversion. These symmetry features are characteristics of the diffraction pattern derived from wood samples and result from the major sources of variance in the diffraction images. For example, the first eigenimage can be interpreted as the primary contribution to variation in equatorial intensity with complementary variation in MFA (azimuthal peak width).

The eigenimage PC2 arises from differential absorption of equatorially scattered X-rays. This differential is caused mainly by rotation of the samples in the horizontal plane and strong horizontal variations in wood density (annual rings) in the radial direction. PC3 is the result of changes due to tilting (roll) of the fibre around the X-ray beam, causing the main

diffraction peaks to be rotated away from the equator and described by rotational symmetry combined with a lack of reflection symmetry. Some higher order PCs such as PC7 do not show rotational symmetry and probably arise from fibre tilt in the direction of the X-ray beam (pitch) and rotation around the fibre axis.

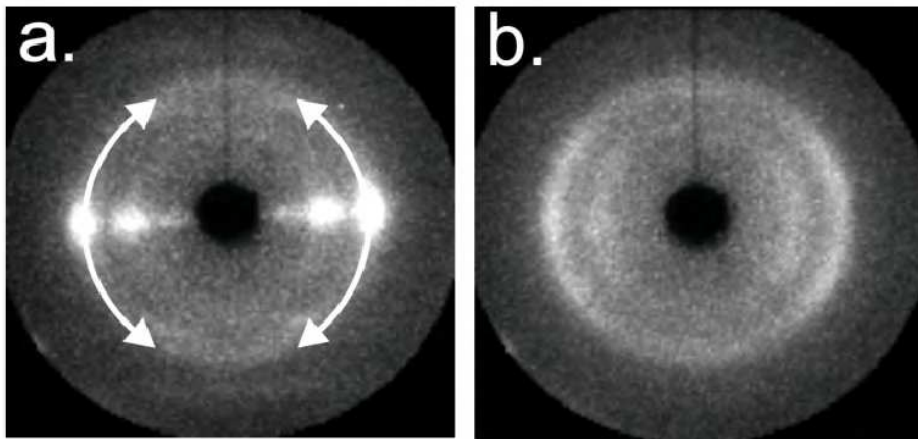


Fig. 2. Example of a. low (9°) and b. high (30°) MFA diffraction images from the *P. caribaea* samples.

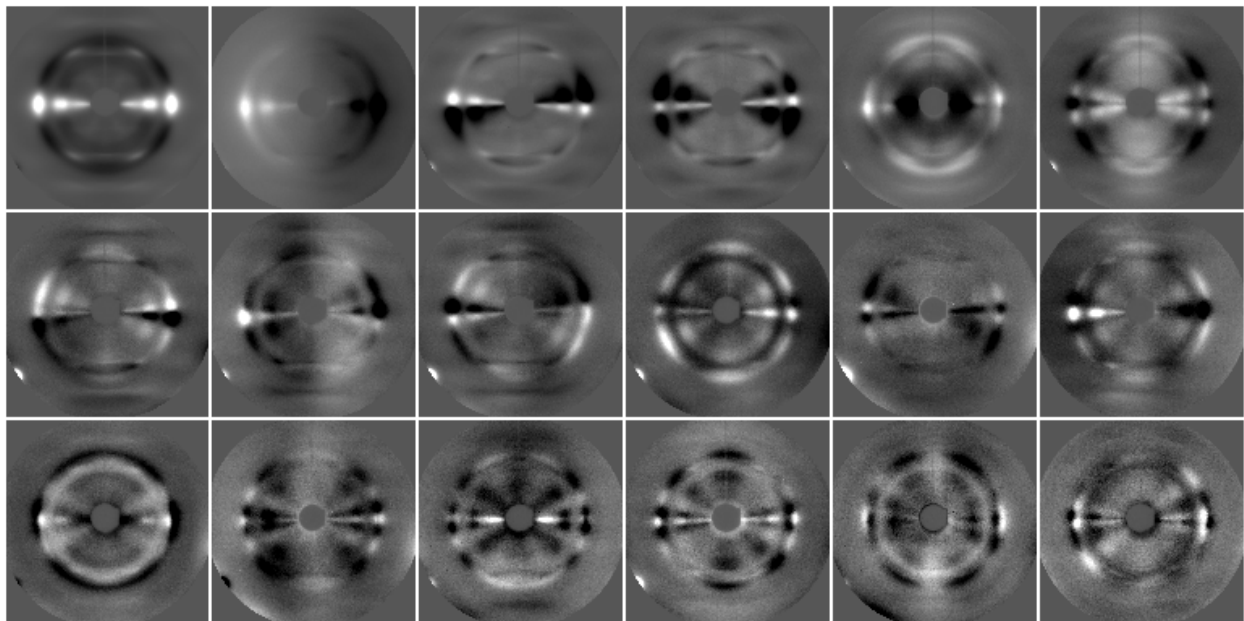


Fig. 3. The first 18 eigenimages of 10000 vector normalised diffraction images obtained from the single-species dataset. PCs are in numeric order going left to right, top to bottom. Bright areas indicate positive values, dark areas negative and mid-grey zero (The absolute sign of the pixels is irrelevant). These images are the 'unit', unweighted components of the modelled image.

Fig. 4 shows a logarithmic plot of eigenvalue (the dot product of each scores vector with itself) for successive PCs from the mean-centred, vector-normalised single-species data. The eigenvalues rapidly decrease by over 3 orders of magnitude.

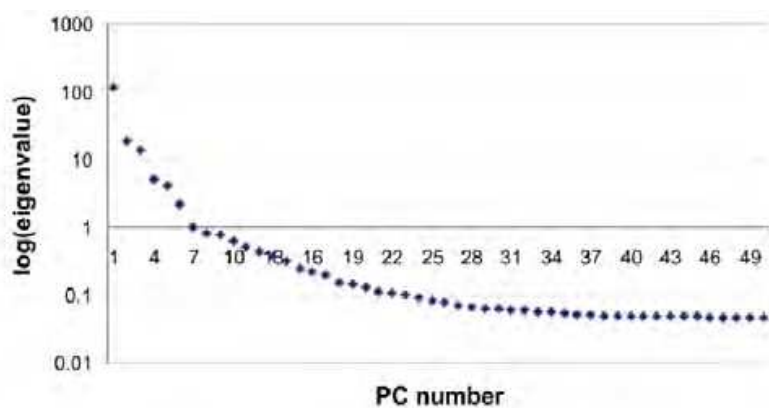


Fig. 4. Eigenvalue for each PC for the mean-centred, vector-normalised single-species PCA model.

As discussed in section 2 above, the ILS modelling procedure incorporates the eigenimage vectors into a predictive model of the dependent variable using the proportionality found within the T matrix (Equation 4). The regression coefficients (β) such as those seen in Fig. 5 are also generated. These coefficients show a distinct pattern of strongly negative values at the equatorial positions and strongly positive values moving into the equatorial-meridian arc.

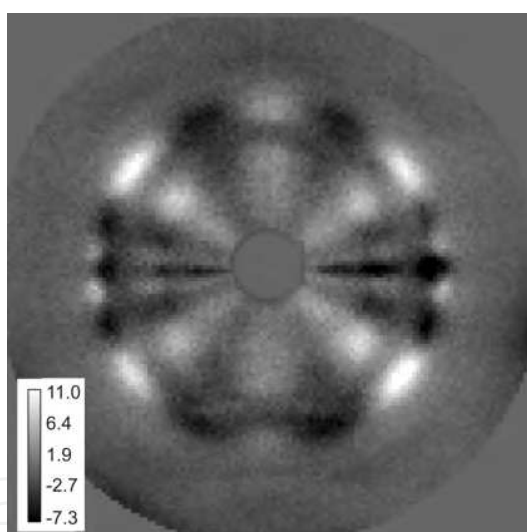


Fig. 5. Regression coefficients obtained from the single-species model using 17 PCs. The image shows distinct positive bands placed diagonally across the image and extending to the limit of the 002 peak 2θ value. For these images both the intensity and the sign of the values is important.

This pattern follows the behaviour of the original data (Fig. 2), where high MFA samples exhibit broad, low intensity equatorial peaks and low MFA samples show more intense, narrow peaks. Thus the model can predict the MFA of unknown samples by summing the product of each pixel value in the unknown image and the regression coefficient at that pixel position, and adding the best estimate for the mean MFA of the calibration set. It is understood that the unknown image is collected on a similar X-ray camera, with matching effective experimental geometry and pre-processed in the same way as the calibration images.

Fig. 6 shows that the RMSEP of the *P. caribaea* test data is reduced only when specific eigenimages are added to the model. The contributing PCs are seen to be PC1, 4, 5, 6, 10, 14, 16, and 17. When referenced against the eigenimages in Fig. 3, the influential PCs are seen to have a dominant characteristic of two orthogonal mirror planes, one through the meridian and another at the equator. This symmetry arises from the vertically aligned wood fibres and the horizontal scattering of x-rays from the crystalline cellulose c-axis planes which, on average are aligned with the fibre axis.

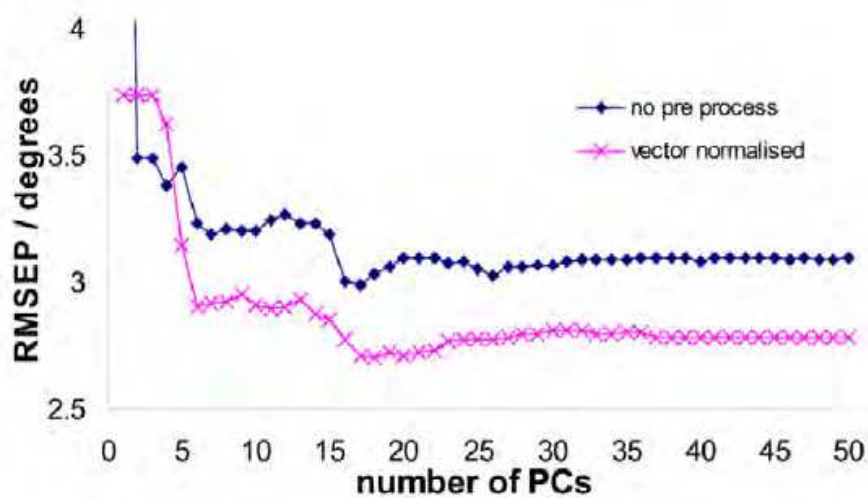


Fig. 6. RMSEP derived from the 10000 image single-species model using 6925 *P. caribaea* verification samples shows that vector normalisation gives consistently better prediction than non-vector normalised (mean centred only) data. The vector normalised samples reach a minimum on the addition of PC17 to the model.

Fig. 7 shows the scatter plots of predicted MFA versus the reference data MFA for the three species as obtained using the single-species data set. It is seen that the single species model predicts *P. caribaea* reference data well, with the predicted data following the line of perfect correlation adequately (RMSEP = 2.7° at 17 PCs for VN samples). The prediction of *P. radiata* samples is also found to be accurate (RMSEP = 1.7° using 19 PCs for the VN samples), with a deviation in the slope producing an overestimation of high MFA samples.

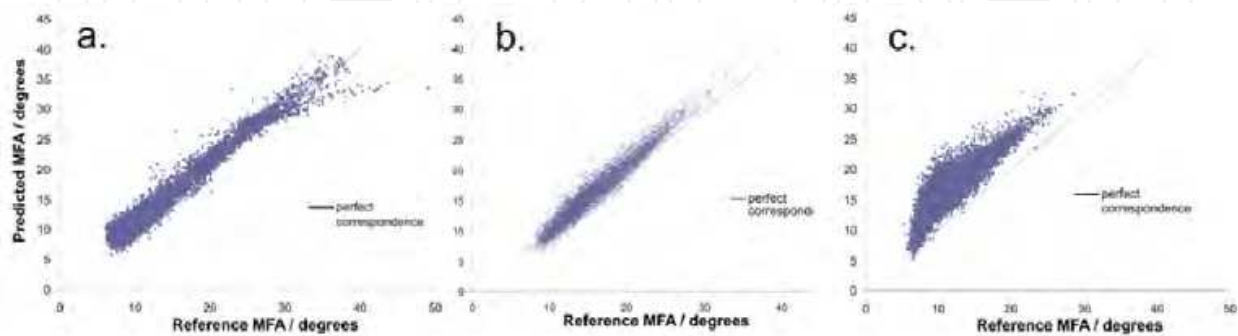


Fig. 7. Reference versus predicted scatter plots of MFA from the single-species model using 17 PCs and VN samples for the predictive model. a. *P. caribaea*, b. *P. radiata* and c. *E. delegatensis* reference samples.

The *E. delegatensis* samples do not fair so well, with a consistent shift in the predicted MFA values over the reference values (RMSEP = 5.4° using 37 PCs in the VN samples). Analysis shows that much of the error in the *E. delegatensis* samples is due to the \bar{y} value obtained from the *P. caribaea* (17.1°) not being a good estimate for the *E. delegatensis* \bar{y} value (11.9°). Correcting for this difference and using it in the model (i.e. offsetting the *E. delegatensis* predicted data by 17.1-11.9) results in a reasonable RMSEP of 2.1° using 18 PCs.

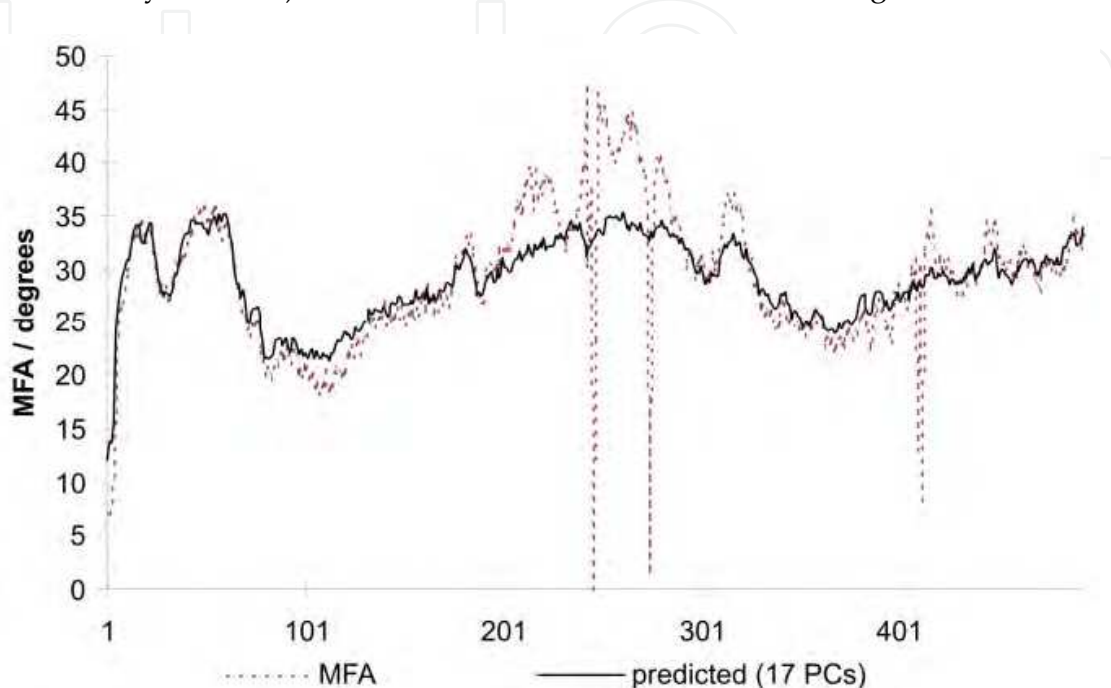


Fig. 8. Plot of predicted MFA versus the reference MFA data for a selected region of high MFA within the early growth wood of a *P. caribaea* sample. Model is from 17 PCs using the VN single-species data only. X axis represents sample number moving along a wood sample.

The data in Fig. 8 shows the advantage of using models such as these as a verification tool for data measured by other procedures. Three data 'dropouts' can be observed in a region of high MFA in the reference MFA data in Fig. 8. These problems are possibly due to high fibre pitch, which distorts the diffraction patterns and thus cause the direct determination of MFA to fail occasionally. It is seen however, that the multivariate model has predicted MFA to follow a profile under the reference data and does not have the problem of dropouts.

Fig. 9 shows that improvements can be made to the predictive potential of models by using multiple species in the calibration data. Fig. 9a shows that the multi-species model marginally increases the accuracy of prediction of *P. radiata* samples, notably drawing the high MFA samples onto the line of perfect correspondence (RMSEP = 1.2° using 37 PCs, VN samples). Fig. 9b shows that the *E. delegatensis* samples are still not well modelled (RMSEP = 4.2° using 18 PCs). The offset is predominately due to the difference in the average MFA between the multi-species dataset (15.2°) and the *E. delegatensis* reference samples (12.1°). This lack of fit indicates that the average values used in offsets of the models must be carefully monitored and that it may be necessary to generate species specific models.

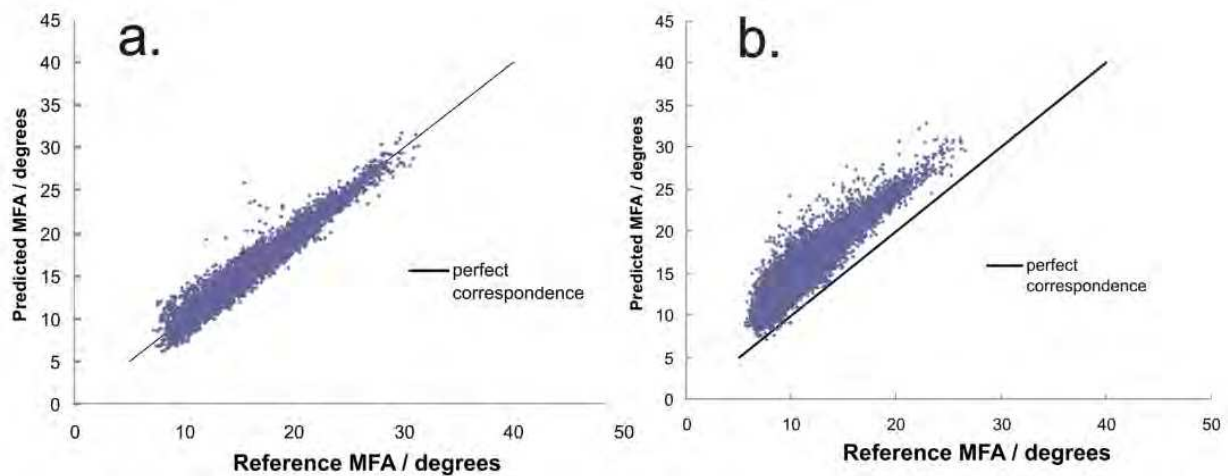


Fig. 9. Reference versus predicted scatter plots of MFA being predicted from the 200000 image multi-species model a. Prediction of *P. radiata* samples MFA from model using 37 PCs and b. Prediction of *E. delegatensis* samples using 17 PCs in the model.

6. Discussion

The PCA modelling procedure reduced the dimensionality of the image dataset from the 16384 pixels present in the original images to a sum of contributions of a smaller number of principal components. Each component found had a characteristic (and mutually orthogonal) distribution of intensities over the 2D image plane. The PCA technique selects the features of these components in order of contributing variance to the (pre-processed) data set. Due to the broad nature of the diffraction features in wood samples, PCA was able to obtain PCs that describe most of the variance in the images. The inverse least squares procedure then used the newly formed variables as predictors to produce a least squares model which can accurately predict MFA. Vector normalised image data produced more accurate models in all experiments than unprocessed, mean centred only data. An R^2 of 0.96 was obtained for *P. caribaea* samples in both single-species and mixed-species models. *P. radiata* R^2 reached a maximum around 0.93 for both models and *E. delegatensis* R^2 increased from 0.84 to 0.91 when the model was produced from the multi-species data set.

Although normalisation procedures are effective, use of regression coefficients with data from other X-ray systems will suffer from errors due to between system differences. These problems are typical of PCR modelling techniques that are sensitive to the precise instrumental system being calibrated.

A disadvantage of this analysis method for large image datasets is the computer memory requirements needed in creation of the models. A single wood sample could generate up to 3000 1.4 Mpixel images, and there may be hundreds of such samples in a project. These difficulties can be overcome through use of high performance cluster-based computing systems and the use of graphics processing units (GPUs) to speed up the PCA data reduction procedure. In the analysis above, images were down-sized to 128 by 128 pixels by discarding pixels. Although not shown, the use of more pixels in the analysis or use of averaging of the pixels should improve the resulting accuracy of models produced. The data analysis system found benefit from use of the HDF data format as it allowed both data and

results to be grouped, with each experiment recorded and similar experiments able to be filed together. The HDF format allows data importation into multiple other popular data analysis programs and thus aides data analysis by researchers with differing software access.

The computational burden of the final prediction part of the procedure is relatively low compared with mapping Φ - θ images into orthogonal axes and segmentation of the image as is presently used by the geometric method of MFA prediction within the SilviScan system. Thus as a technique to increase the robustness of a high throughput system, whole image multivariate property prediction has potential.

A further advantage of multivariate techniques is that standard methods for outlier detection can be used and thus samples that lie outside the calibration can be highlighted and scrutinized by an operator. The ability to emphasize unusual samples is of interest in a high throughput system, where the amount of data collected often precludes a thorough assessment of all results obtained. The technique should tolerate noisy images and allow faster sample throughput.

7. Conclusions

The calibration techniques used here open the way for rapid determination of difficult to obtain wood properties. Measurements from instruments that can accurately measure properties of interest, albeit more slowly can be used for calibration and subsequently used in the high throughput instrument. Models will have to be carefully designed and tested such that they will be valid for a species of interest. These models can be incorporated into the instrument software for checking data integrity. In addition, the signal-to-noise advantage inherent in employing the whole image as a predictor could be used to obtain accurate values at a faster rate.

8. Acknowledgment

Forestry Plantations Queensland are gratefully acknowledged for use of *P. caribaea* samples. Dr F. Chen and Dr M. Hilder for reading of manuscript in early stages, Ms S. Harper for practical help on the SilviScan system and Mr N. Ebdon for running samples.

9. References

- Andersson, S., Serimaa R., Paakkari, T., Saranpaa, P., Pesonen, E. (2003). Crystallinity of wood and the size of cellulose crystallites in Norway spruce (*Picea abies*). *Journal of Wood Science*, Vol. 49, No. 6, pp.531-537.
- Berg, R. A. V. D., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. and Werf, M. J. V. D. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, Vol. 7, No. 142.
- Bowden, J. C. (2010). Application of the OpenCL API for Implementation of the NIPALS Algorithm for Principal Component Analysis of Large Data Sets. e-Science Workshops, 2010 Sixth IEEE International Conference on e-Science.
- Brereton, R. G. (2003). *Chemometrics: data analysis for the laboratory and chemical plant*. Wiley, 9780471489788, Chichester, West Sussex, England; Hoboken, NJ.

- Bro, R., Rinnan, A., Faber, N.M. (2005). Standard error of prediction for multilinear PLS2. Practical implementation in fluorescence spectroscopy. *Chemometrics and Intelligent Laboratory Systems* Vol. 75, pp. 69 - 76
- Cave, I. D. and Robinson, W. (1998). Measuring microfibril angle distribution in the cell wall by means of x-ray diffraction. *Microfibril angle in wood*. B. G. Butterfield. pp. 94-107, Christchurch, University of Canterbury.
- Downes, G., Evans, R., Wimmer, R., French, J., Farrington, A. and Lock, P. (2003). Wood, pulp and handsheet relationships in plantation grown *Eucalyptus globulus*. *Appita Journal*, Vol. 56, No. 3, pp. 221-228.
- Dobson, A.J. (2002) An introduction to generalized linear models 2nd ed. Chapman and Hall/CRC, 1584889500 Boca Raton, Florida.
- Eriksson, L., Wold, S. and Trygg, J. (2005). Multivariate analysis of congruent images (MACI). *Journal of Chemometrics*, Vol. 19, No. 5-7, pp. 393-403.
- Evans, R. (1999). A variance approach to the X-ray diffractometric estimation of microfibril angle in wood. *Appita Journal*, Vol. 52, No. 4, pp. 283.
- Evans, R., Stringer, S. and Kibblewhite, R. P. (2000). Variation of microfibril angle, density and fibre orientation in twenty-nine *Eucalyptus nitens* trees. *Appita Journal*, Vol. 53, No. 6, pp. 450-457.
- Evans, R. and Ilic, J. (2001). Rapid prediction of wood stiffness from microfibril, angle and density. *Forest Products Journal*, Vol. 51, No. 3, pp. 53-57.
- Evans, R. (2006). Wood stiffness by X-ray diffractometry. *Characterization of the cellulosic cell wall*. D. D. Stokke and L. H. Groom. Iowa, Blackwell Publishing.
- Faber, N. M., Song, X. H. and Hopke, P. K. (2003). Sample-specific standard error of prediction for partial least squares regression. *TrAC Trends in Analytical Chemistry*, Vol. 22, No. 5, pp. 330-334.
- Garvey, C. J., Parker, I. H. and Simon, G. P. (2005). On the interpretation of X-ray diffraction powder patterns in terms of the nanostructure of cellulose I fibres. *Macromolecular Chemistry and Physics*, Vol. 206, No. 15, pp. 1568-1575.
- Geladi, P. and Kowalski, B. R. (1986). Partial least squares regression - a tutorial. *Analytica Chimica Acta*, Vol. 185, pp. 1-17.
- Gilmore, C. J., Barr, G. and Paisley, J. (2004). High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles. *Journal of Applied Crystallography*, Vol. 37, No. 2, pp. 231-242.
- The HDF Group. Hierarchical data format version 5, 2000-2010.
<http://www.hdfgroup.org/HDF5>
- Hoskuldsson, A. (1995). A combined theory for PCA and PLS. *Journal of Chemometrics*, Vol. 9, No. 2, pp.91-123.
- Karstang, T. V. and Eastgate, R. J. (1987). Multivariate calibration of an x-ray diffractometer by partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, Vol. 2, No. 1-3, pp.209-219.
- Kramer, R. (1998). *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker, Inc., 9780824701987, New York.
- Lichtenegger, H. C., Reiterer, A., Stanzl-Tschegg, S. E. and Fratzl, P. (2001). Comment about The measurement of the micro-fibril angle in soft-wood by K. M. Entwistle and N. J. Terrill. *Journal of Materials Science Letters*, Vol. 20, No. 24, pp.2245-2247.

- Lindstrom, H., Harris, P., Sorensson, C. T. and Evans, R. (2004). Stiffness and wood variation of 3-year old *Pinus radiata* clones. *Wood Science and Technology*, Vol. 38, No. 8, pp.579-597.
- Matos, C. R. S., Xavier, M. J., Barreto, L. S., Costa, N. B. and Gimenez, I. F. (2007). Principal component analysis of X-ray diffraction patterns to yield morphological classification of brucite particles. *Analytical Chemistry*, Vol. 79, No. 5, pp.2091-2095.
- Martens, H. and Naes, T. (1989) *Multivariate calibration*. Wiley, 0471909793 Chichester [England]; New York.
- Norrman, M., Stahl, K., Schluckebier, G. and Al-Karadaghi, S. (2006). Characterization of insulin microcrystals using powder diffraction and multivariate data analysis. *Journal of Applied Crystallography*, Vol. 39, No. 3, pp.391-400.
- Schimleck, L. R., Evans, R. and Matheson, A. C. (2002). Estimation of *Pinus radiata* D. Don clear wood properties by near-infrared spectroscopy. *Journal of Wood Science* Vol. 48, No. 2, pp.132-137.
- Stuart, S-A and Evans, R. (1995) Effect of fibre tilt on the estimation of the microfibril angle by x-ray diffraction. 49th Appita Annual General Conference, April 3-7, Hobart, Tasmania, Australia.
- Thumma, B. R., Nolan, M. R., Evans, R. and Moran, G. F. (2005). Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*, Vol. 171, No. 3, pp.1257-1265.
- Tripp, V. W. and Conrad, C. M. (1972). X-Ray Diffraction. *Instrumental analysis of cotton cellulose and modified cotton cellulose*. R. T. O'Connor. Marcel Dekker, 9780824715007, New York.
- Wada, M., Okano, T. and Sugiyama, J. (1997). Synchrotron-radiated X-ray and neutron diffraction study of native cellulose. *Cellulose*, Vol. 4, No. 3, pp.221-232.
- Washusen, R. and Evans, R. (2001). Prediction of wood tangential shrinkage from cellulose crystallite width and density in one 11-year-old tree of *Eucalyptus globulus* Labill. *Australian Forestry*, Vol. 64, No. 2, pp.123-126.

IntechOpen



Principal Component Analysis - Engineering Applications

Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0182-6

Hard cover, 230 pages

Publisher InTech

Published online 07, March, 2012

Published in print edition March, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as energy, multi-sensor data fusion, materials science, gas chromatographic analysis, ecology, video and image processing, agriculture, color coating, climate and automatic target recognition.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Joshua C. Bowden and Robert Evans (2012). Application of Principal Components Regression for Analysis of X-Ray Diffraction Images of Wood, Principal Component Analysis - Engineering Applications, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0182-6, InTech, Available from:
<http://www.intechopen.com/books/principal-component-analysis-engineering-applications/application-of-principal-components-regression-for-analysis-of-x-ray-diffraction-images-of-wood>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen