A comparison of local and non-linear calibration methods for determination of moisture content in wood samples

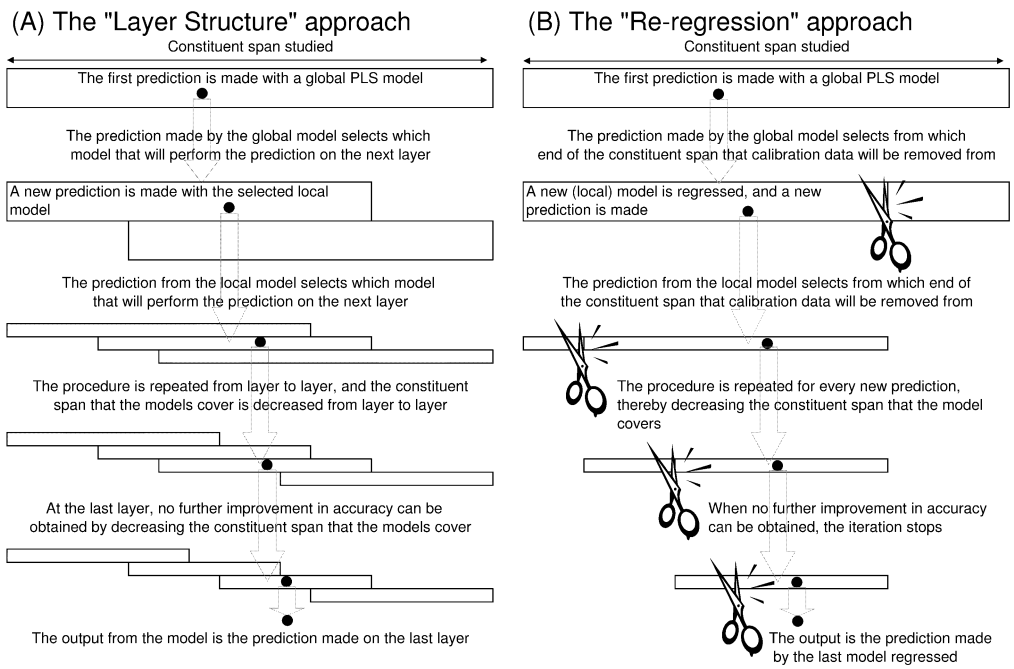
John Dahlbacka[1]* and Tom Lillhonga[2]

[1]Novia University of Applied Sciences, Research and Development, Vaasa, FI-65201, Finland
[2]Novia University of Applied Sciences, Technology and Communications, Vaasa, FI-65201, Finland
*Corresponding author: John.Dahlbacka@novia.fi

Introduction

In this work, a comparison of the measurement accuracy between global, local and non-linear models for the purpose of measurement of moisture content in wood samples was made. PLS[1] (Partial Least Squares) model regression is commonly used in NIR spectroscopy applications but for this particular application it was recently shown that the accuracy could be significantly increased by using a local calibration model variant called ML-PLS (Multi-Layer PLS).[2] It was shown that the measurement error with this method was roughly half of that obtained with a global PLS model. Although this can be seen as a significant improvement of the measurement accuracy, there remained an uncertainty regarding the usefulness of the ML-PLS method in comparison to established local or non-linear methods. Therefore, this comparison between global PLS, LWR[3] (Locally Weighted Regression), ML-PLS, and SVMR[4] (Support Vector Machine Regression) is presented.

The use of local calibration models inevitably requires a method to select a local calibration data set. This can be done in numerous ways. In LWR, this selection is generally based on a distance in principal component score space. However, variations of the method can also, for instance, include the use of the distance in the constituent space as well. The spectral similarity can also be evaluated based on, for instance, spectral correlation,[5] or by the score on the first Orthogonal Signal Correction (OSC) factor.[6] ML-PLS can be described as a stepwise procedure, in which the selection of the local calibration data set is directly based on the PLS prediction obtained from the previous step. Basically the ML-PLS model consist of layers of PLS models, organised so that the models on the current layer cover a narrower constituent span than on the preceding layer. The PLS prediction on the preceding layer selects which model performs the prediction on the current layer. An iterative variation of the method is to regress a model, perform the prediction and, thereafter based on the prediction, remove a portion of the constituent span and regress a new model, etc. These two variants are illustrated in Figure 1.



Reference paper as:
J. Dahlbacka and T. Lillhonga (2012).A comparison of local and non-linear calibration methods for determination of moisture content in wood sampless, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 123-126.

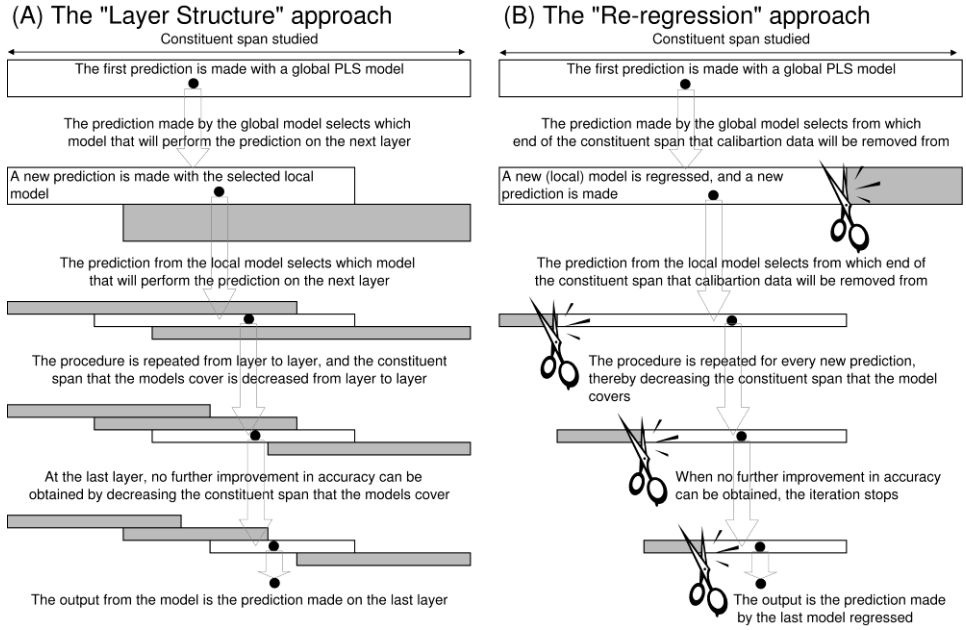


Figure 1. Illustration of ML-PLS.

Materials and Methods

The spectroscopic measurements were carried out with a GetSpec spectrometer, model #: NIR-256L-1.7T1 (Sentronic GmbH, Gostritzer Str. 61-63, D-01217 Dresden, Germany). This diode array instrument has an Indium-Gallium-Arsenide (InGaAs) detector with 256 elements and was equipped with a SentroHead measurement head. The spectral range is 900-1700 nm with 3.125 nm/pixel linear dispersion. Diffuse reflectance absorbance spectra were collected from 905 to 1682 nm (3 nm step size) using the Spec32 v. 1.5.6.8 software as interface; all spectra consisted of 32 co-added scans. Calibration models were calculated using the PLS_Toolbox v. 6.2.1 (Eigenvector Research, Inc. 3905 West Eaglerock Drive Wenatchee, WA 98801, USA) together with Matlab R2008b (The MathWorks AB Box 1023, SE-164 21 Kista, Sweden).

The spectra were collected from seven 0.4 m long samples of Norway spruce. The initially air-dry wood samples were rehydrated and 220 spectra were collected from each sample as the samples were drying, yielding a total of 1540 spectra representing 154 moisture content levels. The spectra from six samples were used for model calibration and those from the remaining wood sample served as an independent validation data set. The standard error of the reference method was determined to 0.1%, based on the weight of the samples directly before and after the collection of the spectra. The moisture content in the calibration data ranged from 22 to 57%, and in the validation data from 24 to 57%.

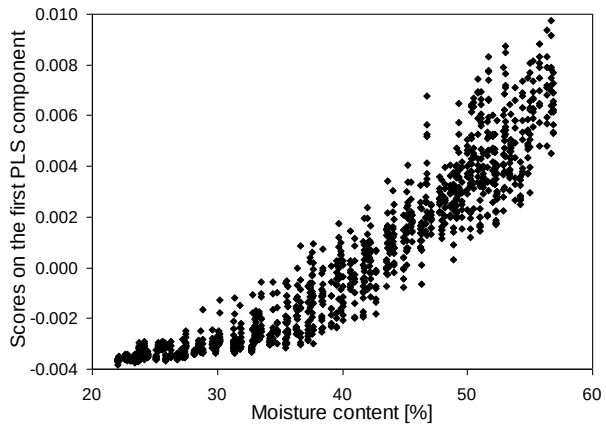


Figure 2. Scores on the first PLS component vs. moisture content.

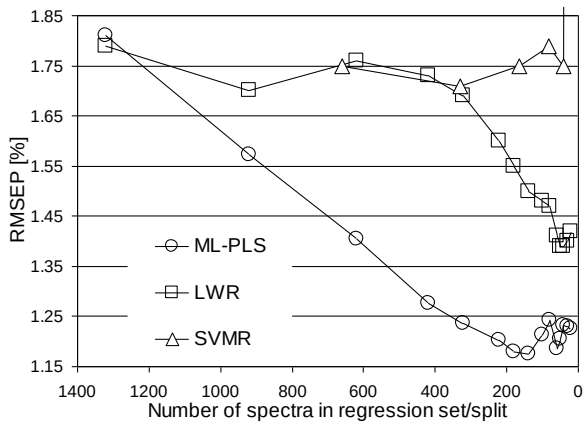


Figure 3. RMSEP plotted against the number of spectra in the regression set/split.

Reference paper as:
J. Dahlbacka and T. Lillhonga (2012).A comparison of local and non-linear calibration methods for determination of moisture content in wood sampless, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 123-126.

Results and Discussion

It is perhaps not fully understood why ML-PLS has been proven to significantly increase the measurement accuracy on the data set studied. However, one answer could be non-linearity. In Figure 2, the score on the first PLS component is plotted against the moisture content. As can be seen, the relationship between the first PLS component (which explains 87% of the moisture content variance) and the moisture content displays a non-linear behaviour. Thus, the non-linear method SVMR could perhaps also have been expected to perform well on this data. Numerous models with different pre-processing alternatives and different model parameters were regressed during the study. However, a second order Savitzky-Golay derivative with a second order 7 point polynomial and mean centering pre-processing generally proved to be a feasible method for all regression methods. The highest accuracy for LWR was obtained with PLS regression and 7 PLS components, whereas the use of constituent data and the iterative selection procedure option in the PLS Toolbox resulted in lower accuracy. Seven PLS components was also close to optimal for the global PLS model and a reasonable alternative for ML-PLS. For SVMR, PLS compression gave slightly better results than PCA compression and 7 PLS components gave almost the same accuracy as with 5 components, which would have been the best alternative found.

Thus, a fairly reasonable comparison (at least in terms of model parameters and pre-processing of the data) between the RMSEP value (root mean square error of prediction) obtained with the different calibration methods could be performed using PLS model regression in LWR, PLS compression in SVMR, 7 PLS components for all methods and the above mentioned pre-processing. This comparison was done with different sizes of the model regression data set or number of data splits; results are shown in Figure 3. With these settings, the (lowest) RMSEP for ML-PLS was 1.17%, for LWR 1.39%, for SVMR 1.71% and for the global PLS 1.81%. According to Figure 3, the optimal number of local calibration points is roughly the same for LWR and ML-PLS. With SVMR, the number of data splits only had a limited impact on the accuracy. Figure 4 shows the scatter plots from the validation data set for the global PLS model, LWR and ML-PLS.

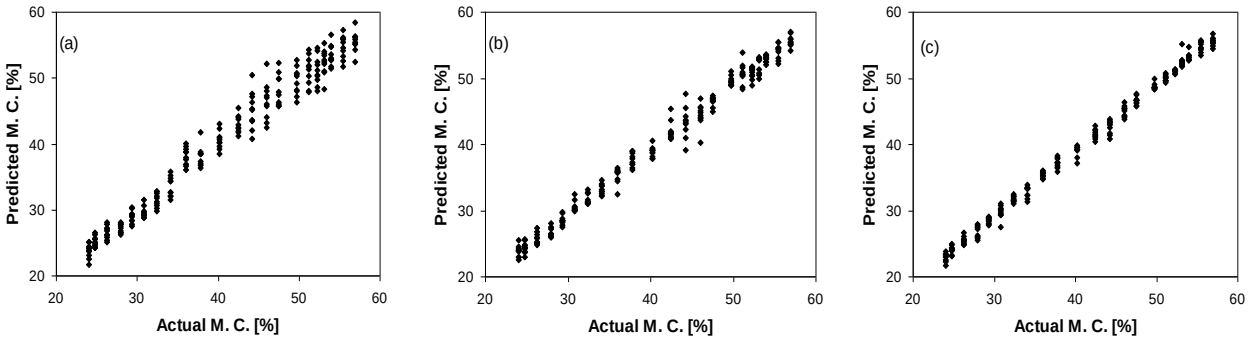


Figure 4. Actual vs. Predicted Moisture Content with (a) a global PLS-model, (b) LWR, and (c) ML-PLS.

Conclusion

The ML-PLS calibration procedure is a powerful method when determining moisture content in wood samples, at least on this particular data set studied. Results from LWR showed that local models can significantly improve prediction accuracy and, by using ML-PLS, the accuracy was further enhanced. In comparison to LWR, the higher accuracy obtained with ML-PLS is believed to be a result of the stepwise approach. Although LWR, as it was implemented in the software used in this study, enables an iterative procedure in selecting the local regression data set, the number of spectra in the local data set remains fixed. In ML-PLS, the number of spectra in the local regression data set is reduced from layer to layer and it is perhaps this approach that gave the advantage over LWR. This result should not be seen as generally applicable but it does suggest that ML-PLS can be an alternative for data where the accuracy is increased by using LWR instead of a global PLS model.

Acknowledgements

This work was made possible through the FIELD-NIRce project, funded by the European territorial cooperation programme Botnia-Atlantica. The authors also want to thank Professor Paul Geladi for scientific advice.

Reference paper as:
J. Dahlbacka and T. Lillhonga (2012).A comparison of local and non-linear calibration methods for determination of moisture content in wood sampless, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 123-126.

References

1. P. Geladi and B. Kowalski, *Anal. Chim. Acta* **185**, 1-17 (1986).
2. J. Dahlbacka and T. Lillhonga, *J. Near Infrared Spectrosc*. **18**, 425-433, (2010).
3. T. Naes, T. Isaksson, T. Fearn and T. Davies, *A user-friendly guide to multivariate calibration and classification*. NIR Publications, Chichester, UK (2002).
4. P.-H. Chen, C.-J. Lin and B. Schölkopf, *Appl. Stochastic Models Bus. Ind*. **21**, 111–136 (2005).
5. J.S. Shenk, P. Berzaghi and M.O. Westerhaus, *J. Near Infrared Spectrosc*. **5**, 223–232 (1997).
6. T. Fearn and A.M.C. Davies, *J. Near Infrared Spectrosc*. **11**, 467–478 (2003).

Reference paper as:
J. Dahlbacka and T. Lillhonga (2012).A comparison of local and non-linear calibration methods for determination of moisture content in wood sampless, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 123-126.