

EVALUATION OF THE HEATING VALUE OF BIOMASS FUEL FROM ELEMENTAL COMPOSITION AND INFRARED DATA

K. Varmuza, B. Liebmann, A. Friedl

*Vienna University of Technology, Institute of Chemical Engineering
Getreidemarkt 9/166, A-1060 Vienna, Austria,
kvarmuza@email.tuwien.ac.at*

ABSTRACT

The heating value of 35 biomass samples (wood, cereals) has been modeled from mass% data for carbon, hydrogen, and nitrogen, as well as from infrared data. Regression models have been obtained by application of a genetic algorithm for variable selection and PLS for creation of models. Standard deviation of prediction errors is typically 0.3 % for models from IR data, and 0.7 % for models from elemental composition data.

Keywords: Fuels, IR, chemometrics, variable selection, PLS regression

INTRODUCTION

Biomass is becoming increasingly important as a renewable source of energy, by incineration as well as production of liquid biofuels, e. g. biodiesel or bioethanol. Compared to the incineration of fossil fuels the overall CO₂ balance is considerably less affected by the usage of biomass. Typical biomass materials used for energy production are wood and so-called energy grass. Special types of cereals have a century old tradition in (bio-)alcohol production, recent developments also consider cereal incineration.

An important property of a fuel is its heating value. The so called "higher heating value" (HHV) is the enthalpy of complete combustion of a fuel with all carbon converted to CO₂, and all hydrogen converted to H₂O. The higher heating value is given for standard conditions (101.3 kPa, 25 °C) of all products and includes the condensation enthalpy of water; it is generally used in the USA. In European countries the "lower heating value" - not used in this study - is more common. It does not include the condensation enthalpy of water [1].

Direct determination of the heating value requires time-consuming calorimetric experiments, which hardly can be automated. Empirical equations have been

published that relate the heating value of a fuel to its elemental composition. Elemental analysis requires tedious laboratory work too, but can be automated. Early mathematical models for coal date back to the late 19th century. Recently, modern chemometric methods have been applied for prediction of heating values of plant biomass from elemental composition [2].

Infrared spectroscopy (IR) and especially near infrared spectroscopy (NIR) became important and routinely used methods for quantitative analyses and for characterization and classification of technological materials and food. IR and NIR can be applied much easier and faster than many other laboratory methods. Chemometric methods allow the generation of multivariate regression models with optimum prediction performance, most used is partial least-squares regression (PLS) [3,4]. Moisture, ash, and heating value have been modeled from NIR data for fuel mixtures of coal, peat, and biofuel [5].

The aim of this study was the development of PLS calibration models for the prediction of higher heating values of wood samples and five different cereals, based on IR reflectance data, and for comparison also on elemental composition data.

EXPERIMENTAL

Samples. A total of $n = 35$ samples was available, 20 from wood, 15 from cereals. The wood samples consist of sawdust (spruce, pine and larch) with varying amounts of additives such as bark, rye and maize flour, and starch. Cereal samples are unmixed wheat, rye, barley, maize and triticale flours.

Calorimetry. The higher heating value of biomass, *HHV*, has been determined by the bomb calorimetric method according to DIN 51900 T3 [2]. About 1 g biomass material was used; range of measured *HHV* was 18,143 to 19,125 kJ/kg; typical analysis errors are ± 60 kJ/kg (ca 0.32 %). The experimental *HHV* values are used as the dependent variable y for regression.

Elemental analysis. The contents of C, H, and N in a sample have been measured by standard methods of elemental analysis as described in a previous work [2]. The concentrations are given in mass% of dry material and have been used as basic x -variables for regression.

Infrared spectroscopy. Spectra were recorded using a Bruker Equinox 55 Fourier Transform-Infrared (FTIR) spectrophotometer equipped with an Attenuated Total Reflection (ATR) accessory and a deuterated triglycine sulfate (DTGS) detector. ATR spectroscopy is a contact sampling method in which a crystal of a high refractive index is used as an internal reflection element. The IR spectra in the range of 4000-600 cm^{-1} were obtained at intervals of approximate 2 cm^{-1} giving 1764 data points per spectrum. Preprocessing of spectra was performed by software Unscrambler [6]: first step was averaging four neighboring absorbance values giving an approximate resolution of 8 cm^{-1} ; second step was calculation of the first derivative by the Savitzky-Golay algorithm (quadratic polynomial, three points). The resulting 439 values are used as x -variables in regression.

CHEMOMETRICS

A linear equation that predicts a dependent variable as a function of several independent variables is of the general form

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_p x_p \quad (1)$$

where \hat{y} is the predicted dependent variable (*HHV*), b_0 the intercept, b_1 to b_p are the regression coefficients, and x_1 to x_p the independent variables; p is the number of independent variables. PLS regression [3,4] - as implemented in software Unscrambler [6] - has been used for the development of such models. The objective is a model with high prediction performance for cases not used in model building and model optimization. In this study a repeated cross validation (CV) has been applied for optimization of models and estimation of prediction errors. Because of the rather small number of samples available, independent test sets have not been used. The optimum number of PLS components, a_{PLS} , has been determined by a leave-a-quarter-out CV (four segments). From the CV-predicted values ${}_{CV}y_i$ ($i = 1 \dots n$) the standard error of prediction for CV, SEP_{CV} , has been calculated [3,4,7].

$$SEP_{CV} = [\Sigma(y_i - {}_{CV}\hat{y}_i - bias)^2 / (n - 1)]^{0.5} \quad i = 1 \dots n \quad (2)$$

The *bias* is the arithmetic mean of the prediction errors ($y_i - {}_{CV}\hat{y}_i$) which in all cases was near zero. *SEP* is identical with the standard deviation of the prediction errors. Four different random splits into four segments have been evaluated and the obtained values for SEP_{CV} averaged; this mean, SEP_{CV4} , is used as a criterion for the prediction performance. In the case of normally distributed prediction errors (which was approximately fulfilled with the used data) $\pm 2.5 SEP_{CV4}$ is a reasonable confidence interval.

For the elemental analysis data, the basic variable set consists of three variables (mass% of C, H, and N). An augmented variable set has been created by adding nonlinear transformations and combinations of the variables, defined by C^2 , H^2 , N^2 , $C \cdot H$, C/H , $\ln(C)$, $\ln(H)$, $\ln(N)$, $\ln(C \cdot H)$, and $\ln(C/H)$, resulting in $p = 13$ variables. Because the added variables are only mathematically defined, a variable selection has been performed using a genetic algorithm (GA, see below) resulting in a variable set with $p = 5$.

For IR data the original variable set consists of 439 variables. Application of a genetic algorithm (GA, see below) resulted in a variable set with $p = 21$.

The principles of genetic algorithms (GA) and successful applications to variable selection have been described by others [8,9]. The software used was MobyDigs [10]. The regression method applied in this software is ordinary least squares regression (OLS) and the performance (fitness) of models has been evaluated by the adjusted correlation coefficient [11] between experimental *HHV* and predicted *HHV* using full cross validation (leave-one-out). The adjusted correlation coefficient considers the number of variables used for the model and penalized models with a large number of variables. Population size used was 50, the maximum number of selected features in a model is limited to 15 by the software. The variables selected in the best 10 best models have been considered. In each trial about 650,000 models were tested. Computation time was 15 minutes for the augmented elemental data set ($p = 13$) and 35 minutes for the IR data set ($p = 439$) on a Pentium 2 GHz.

RESULTS

The heating values, *HHV*, of cereal samples are in the range 18,143 to 18,594 (mean 18,389); all wood samples have higher *HHV*, with a range of 18,668 to 19,125, and a mean of 18,915. The content of carbon shows the same trend with smaller values for cereals (range 40.4 to 42.2%, mean 41.3) than for wood samples (range 46.3 to 47.9%, mean 47.4). Hydrogen and nitrogen contents are higher in cereal samples than in wood samples (H: cereals 6.5 to 6.9%, wood 5.9 to 6.3%; N: cereals 0.9 to 2.2, wood 0.1 to 0.3%).

Figure 1 shows IR spectra from a maize sample and a wood sample. The spectra look very similar and show typical absorptions bands, for instance at about 1000 cm^{-1} from C-OH, at about 1640 cm^{-1} from C=O, and at about 2920 cm^{-1} from C-H; the broad band around 3300 cm^{-1} is from free OH (moisture).

The result of exploratory data analysis by principal component analysis (PCA) is shown in Figure 2. For both data sets a clear separation of the cereals from the wood samples appear. The loading plot for the element data (not shown) reflects the different concentrations of C, H, and N in the two sample groups. The same clustering appears in a dendrogram obtained by hierarchical cluster analysis.

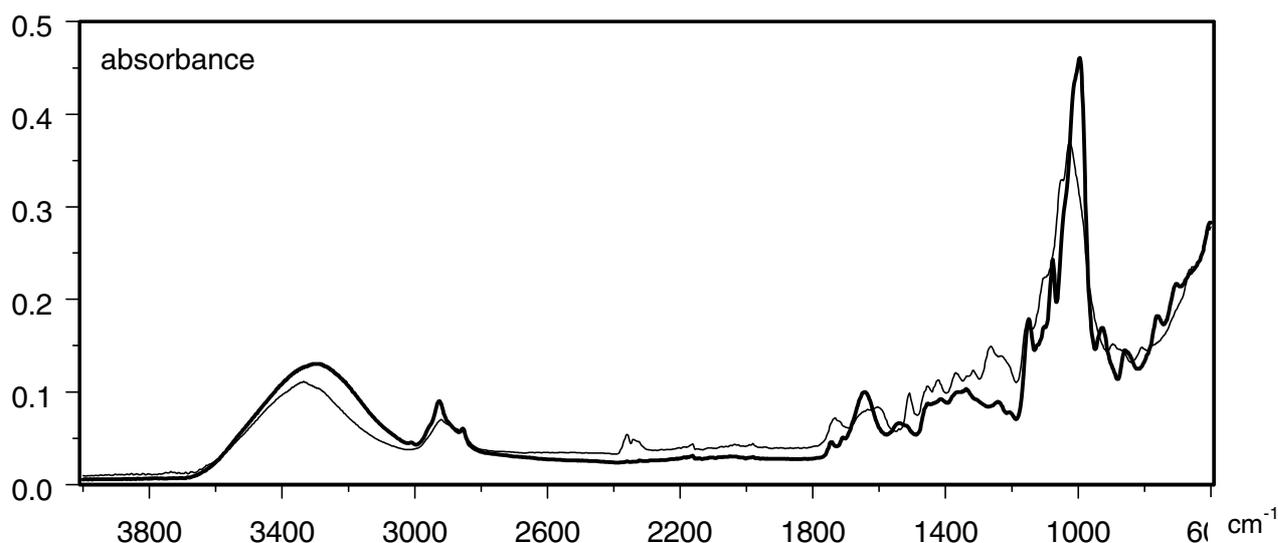


Figure 1. ATR-FTIR spectra of maize (thick line) and sawdust (spruce) with 18 % bark (thin line).

Table 1. Cross-validated PLS models for the higher heating value of biomass derived from elemental composition and infrared data, respectively. *p*, number of features; a_{PLS} , number of PLS components (averaged); SEP_{CV4} , standard error of prediction (average of four runs of 4-fold cross validation).

| <i>p</i> | Features | Methods | a_{PLS} | SEP_{CV4} (kJ/kg) |
|----------|-------------------------------------|----------|-----------|---------------------|
| 3 | C, H, N | PLS | 2 | 130 |
| 13 | C, H, N, and derived features | PLS | 2 | 127 |
| 5 | C, H ² , C·H, C/H, ln(H) | GA + PLS | 2 | 124 |
| 439 | all IR absorbances | PLS | 4 | 125 |
| 21 | selected IR absorbances | GA + PLS | 12 | 62 |

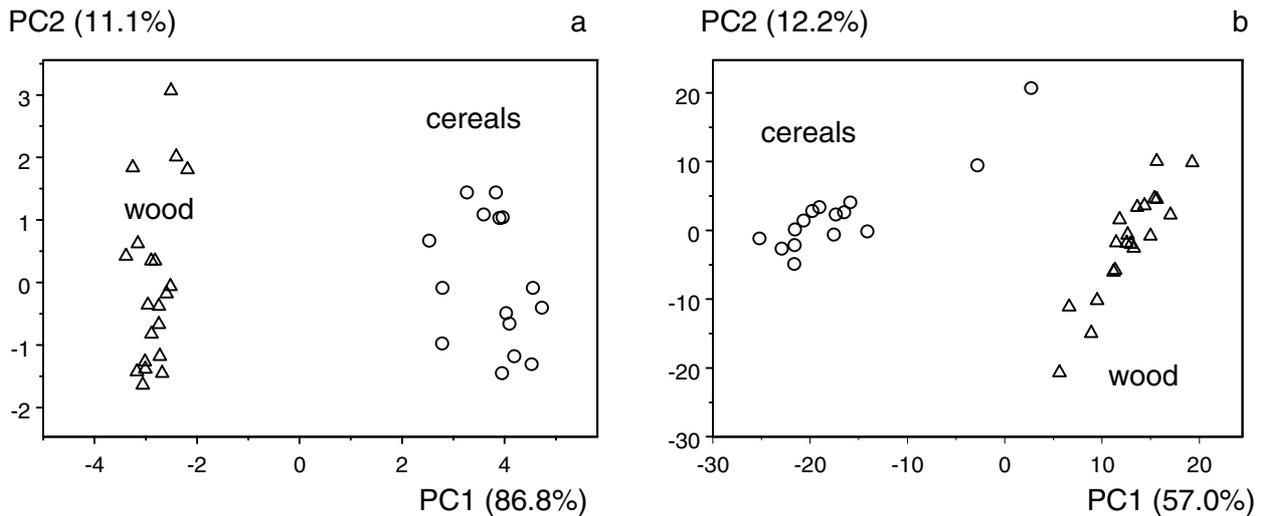


Figure 2. Principal Component Analysis (PCA) of (a) elemental analysis data ($p = 13$), and (b) IR data ($p = 439$) using autoscaled variables, with % preserved variances of first and second principal component scores.

Modeling the *HHV* by only using the carbon content (univariate regression) gives a standard prediction error, SEP_{CV4} , of 132 kJ/kg. Using the three basic variables from the elemental data (%C, %H, %N) gives 130 kJ/kg. Only a non significant improvement is achieved for the augmented variables set ($p = 13$) with a SEP_{CV4} of 127 kJ/kg. The variable set from GA selection with $p = 5$ gives a further small improvement with a SEP_{CV4} of 124 kJ/kg. The best model from elemental data has a standard prediction error of about 0.7 % of the mean of the *HHV* values (Table 1).

A model using all 439 IR variables has a similar prediction performance as models with elemental data with a SEP_{CV4} of 125 kJ/kg. However variable selection by GA resulted in a subset with 21 variables that gives a much better model with a SEP_{CV4} of only 62 kJ/kg. This standard prediction error corresponds to 0.3 % of the mean of the *HHV* values and to a confidence interval of about ± 155 kJ/kg. In Figure 3 experimental *HHV* values are plotted versus predicted values obtained from the best model using 21 IR absorbances; the squared Pearson correlation coefficient between calorimetric determined *HHV* and IR-predicted *HHV* is 0.968.

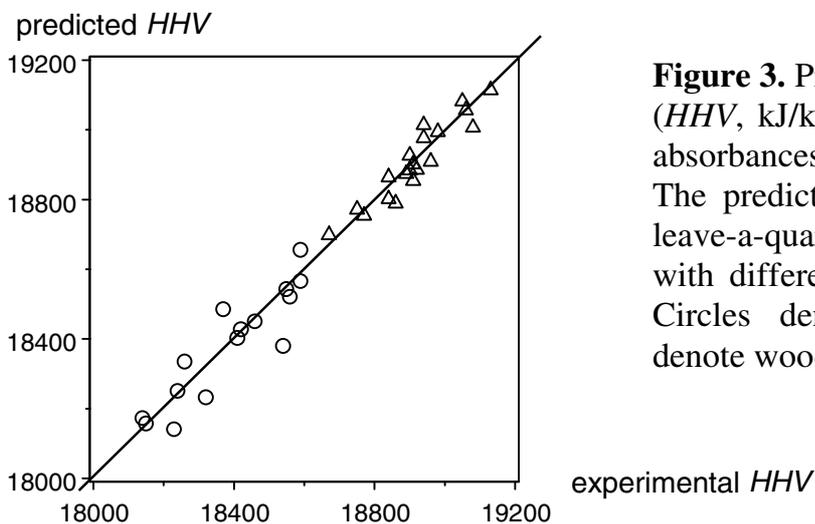


Figure 3. Prediction of higher heating values (*HHV*, kJ/kg) by a PLS model using 21 IR absorbances selected by a genetic algorithm. The predicted values are means from four leave-a-quarter-out cross validation runs with different partitioning of the segments. Circles denote cereal samples, triangles denote wood samples.

CONCLUSIONS

Easily available IR reflection data are promising for the determination of heating values of wood and cereal samples. Feature selection by a genetic algorithm improved the prediction performance in comparison to models with all features. IR data are better suited than elemental composition data. A single model can be used for wood samples and cereal samples.

Acknowledgements. We gratefully acknowledge financial support by Hochschuljubiläumsstiftung der Stadt Wien, Projekt H-7022005. We thank A. Kandelbauer from Competence Center Wood K-Plus (St. Veit an der Glan, Austria) for instrumental support, E. Padouvas for calorimetric measurements, and H. Mikosch for IR support (both Vienna University of Technology), and H. Schausberger (Saatbau Linz, Austria) for providing cereal samples.

REFERENCES

- [1] Corbitt, R. A., Standard handbook of environmental engineering. Mc Graw Hill, New York, 1989.
- [2] Friedl, A., Padouvas, E., Rotter, H., Varmuza, K., *Anal. Chim. Acta* 544 (2005) 191-198.
- [3] Naes, T., Isaksson, T., Fearn, T., Davies, T., A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, 2004.
- [4] Varmuza, K., in: Gasteiger, J. (Eds.), *Handbook of chemoinformatics*, Vol. 3, Wiley-VCH, Weinheim, 2003, p. 1098-1133.
- [5] Geladi, P., Lillhonga, T., Reuter, L., in: Pomerantsev, A. L. (Ed.), *Progress in chemometrics research*, Nova Science Publishers, New York, 2005, p. 193-208.
- [6] Camo Process AS, Software The Unscrambler. <http://www.camo.no>, Oslo, Norway, 2002.
- [7] Kramer, R., *Chemometric techniques for quantitative analysis*. Marcel Dekker, New York, 1998.
- [8] Forina, M., Lanteri, S., Oliveros, M. C. C., Millan, C. P., *Anal. Bioanal. Chem.* 380 (2004) 397-418.
- [9] Leardi, R., Boggia, R., Terrile, M., *J. Chemometrics* 6 (1992) 267-281.
- [10] MobyDigs, 1.0. Talete srl, www.talete.mi.it, Milan, Italy, 2004.
- [11] Frank, I. E., Todeschini, R., *The data analysis handbook*. Elsevier, Amsterdam, 1994.