Application of Near-Infrared Spectroscopy and Chemometrics in Bioethanol Production

Bettina Liebmann, Anton Friedl, Kurt Varmuza*

Vienna University of Technology, Institute of Chemical Engineering
Getreidemarkt 9/166-2, A-1060 Vienna, Austria
www.lcm.tuwien.ac.at, www.vt.tuwien.ac.at
liebmann@mail.zserv.tuwien.ac.at, afriedl@mail.zserv.tuwien.ac.at,
kvarmuza@email.tuwien.ac.at

1. Objective

Application of easily available near-infrared (NIR) spectroscopy for the

- monitoring of alcoholic fermentation processes (also with infections) and
- quantification of ethanol, glucose and glycerol - from single component solutions to multiconstituent fermentation substrates.

The PLS models' prediction performance was tested with a repeated double cross validation before and after variable selection by a Genetic Algorithm (GA).

2. Experimental

Samples

Clear Standards: Dilution series with each 10 aqueous solutions of glucose and ethanol, resp. (0-100 g/l)

Complex Standards: Addition of typical fermentation by-products to clear standards described above

→ Each 10 g/l fructose, maltose and xylose to glucose samples

→ 8 g/l glycerol, 2 g/l isobutanol and 2 g/l isoamylalcohol to ethanol samples

Mash: Samples from alcoholic fermentation by yeast

→ Options: glucose mash, wheat mash or rye mash

→ Glucose mashes sampled during the course of fermentation (every 30 min.)

→ Wheat and rye mashes withdrawn after completion of fermentation; sample centrifugation and stepwise addition of glucose

Stillage: Samples of distillation residue after separation of ethanol by means of a rotary vacuum evaporator

→ Samples available from experiments with feedstock wheat and rye

→ Stepwise addition of ethanol to the centrifuged samples

Data

NIR Absorbance Data: AOTF-NIR spectrometer for transflexion measurements (*Brimrose Luminar 5030*)

- 1100-2300 nm, $\Delta\lambda$ = 5 nm, 1st derivative Savitzky-Golay results in 235 variables
- Variable reduction by GA [1,2] to each 15 relevant *x*-variables for glucose, ethanol, and glycerol quantification, resp.

Reference Values: Concentrations (g/l) determined by weight, HPLC analysis or a combination of HPLC value with weight added by the stepwise addition method

- Ethanol: 0-99.8 g/l
- Glucose: 0-180.2 g/l
- Glycerol: 0.1-14.4 g/l

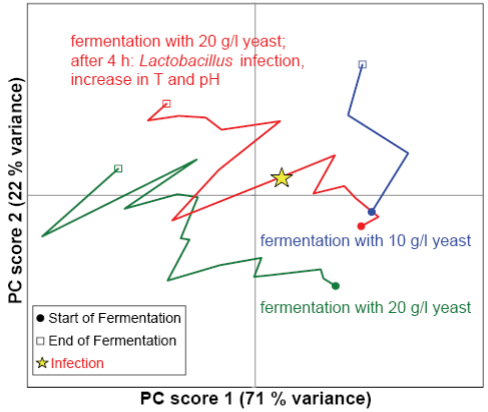
3. Chemometrics

- Principal Component Analysis (PCA) as available in software The Unscrambler [3] for fermentation monitoring.
- Development and validation of PLS models by repeated double cross validation (rdCV) - package *chemometrics* in R [4-6].
- Prediction performance of models with / without variable selection are assessed by criteria derived from rdCV (100 repetitions, up to 12,000 test set predicted values):

SEP_{test}	standard deviation of prediction errors y-$\hat{y}$
TI_{90}	90 % tolerance interval of prediction errors
a_{opt}	optimum number of PLS components [7]
R^2	squared Pearson correlation coefficient

4. Fermentation Monitoring

Principal Component Analysis (PCA) score plot for three different alcoholic fermentations of glucose with yeast (duration 8 hours).



The small data set (36 samples from three fermentations, 235 NIR variables) allows only preliminary conclusions: Differences in yeast concentration are visible in the PCA plot. However, an infection by lactic acid bacteria is not clearly identifiable. Both PCA scores together cover 93 % of total variance in the original data set.

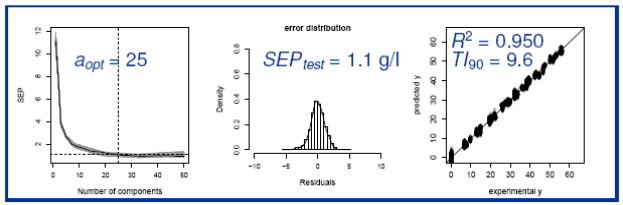
5. PLS Model Evaluation

PLS model comparison - without/with variable selection

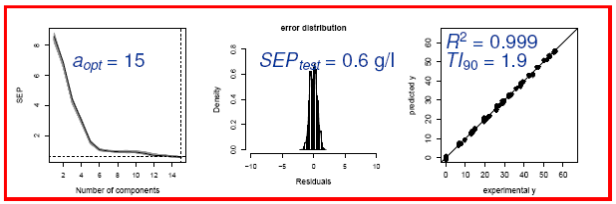
Ethanol Concentration in Stillage

86 samples from wheat and rye stillages
ethanol concentration range: 0.0-56.0 g/l ethanol

Ethanol model using all available 235 NIR variables



Ethanol model using 15 GA selected NIR variables

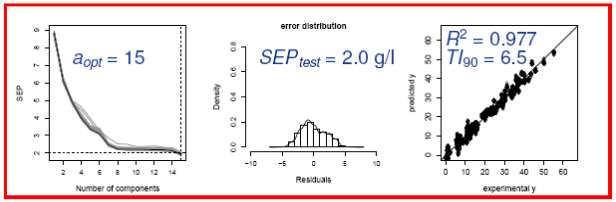


PLS models after variable selection

Glucose Concentration in Mash

120 samples of wheat and rye mashes
glucose concentration range: 0.1-55.3 g/l

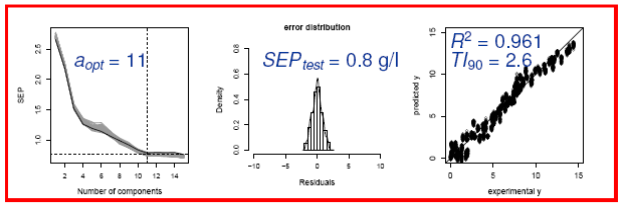
Glucose model using 15 GA selected *x*-variables



Glycerol Concentration in Mash

85 samples of glucose mashes
glycerol concentration range: 0.1-14.4 g/l

Glycerol model using 15 GA selected *x*-variables



6. PLS Models' Prediction Performance

Glucose

Sample Set	*n*	SEP_{test} NIR all	SEP_{test} NIR GA	Range Glucose	
clear standards	10	5.5	1.1	0-100.1	a
complex standards	10	5.2	0.7	0-100.1	a
glucose mashes	85	5.6	3.5	0-180.2	b
wheat mashes	64	2.9	1.0	0.9-43.7	c
rye mashes	56	4.4	1.2	0.1-55.3	c
wheat+rye mashes	120	2.9	2.0	0.1-55.3	c

Ethanol

Sample Set	*n*	SEP_{test} NIR all	SEP_{test} NIR GA	Range Ethanol	
clear standards	10	4.2	0.2	0-99.8	a
complex standards	10	4.8	0.2	0-99.7	a
glucose mashes	85	2.6	1.9	1.8-72.3	b
wheat stillages	30	2.1	0.4	0-44.6	c
rye stillages	56	1.5	0.4	0-56.0	c
wheat+rye stillages	86	1.1	0.6	0-56.0	c

n	number of samples
SEP_{test}	standard deviation of prediction errors (g/l)
NIR all	all 235 NIR absorbance values available
NIR GA	GA selected NIR absorbance values (15 variables different for each sample set given above)
a	reference concentration by weight
b	reference concentration by HPLC
c	concentration by HPLC and weight added

7. Summary

- Rapid, non-destructive NIR analysis allows **quantification of glucose, ethanol and glycerol** in multiconstituent substrates of the bioethanol production process. Samples included different feedstock options, variations in enzymatic pretreatment as well as infections in the fermentation.
- A **good analytical reference method** is mandatory for PLS model creation (see glucose determination in glucose mash models).
- **Variable selection** by Genetic Algorithm **improves prediction performance** for all investigated PLS models.
- Repeated double cross validation offers a **sophisticated optimization strategy** for model complexity (number of PLS components). Furthermore, prediction performance can be reasonably estimated.
- In comparison, **full cross validation** (not shown here) yields **higher prediction errors**, as the optimum number of PLS components is chosen more conservatively.

8. Acknowledgments

This work was funded by the *Austrian Research Promotion Agency (FFG), BRIDGE program*, project no. 812097/11126, and supported by W. Krenn, Vogelbusch GmbH, Vienna. We thank S. Bozic and C. Sanchis Grob for fermentation experiments and P. Filzmoser for collaboration and development of a repeated double cross validation software in R.

9. References

1. Software MobyDigs, v 1.0. Talete srl, www.talete.mi.it, Milan, Italy, 2004.
2. Leardi, R.: J. Chromatogr. A 1158 (2007) 226-233.
3. Software The Unscrambler v 9.0. Camo Process AS, www.camo.no, Oslo, Norway, 2004.
4. Varmuza, K., Filzmoser, P.: Introduction to Multivariate Statistical Analysis in Chemometrics. CRC-Press, Boca Raton, USA. (In print 2009)
5. Software R, v 2.7.0. R Development Core Team, www.r-project.org, 2008.
6. Mevik, B. H., Wehrens, R.: J. Statistical Software 18 (2007), issue 2.
7. Hastie, T., Tibshirani, R. J., Friedman, J.: The elements of statistical learning. Springer, New York, USA, 2001.