Selection and application of chemometric techniques in near infrared evaluation of barley malts

Jenette Y. Wheeler and Peter L. Freeman

Canada Malting Company Ltd, 3316 Bonnybrook Road, SE Calgary, AB, Canada, T2G 4M9.

Debra K. Jones

Great Western Malting Company, PO Box 1529, Vancouver, WA 98668-1529, USA.

Introduction

In todays malting industry there is a need for rapid determination of quality. Traditional methods used to determine malt quality are often time consuming, expensive and may involve the use of hazardous chemicals. These traditional methods include routine measurement of moisture and protein as well as various analysis of malt extracts. The use of near infrared (NIR) analysis provides an alternative to traditional methods which allows rapid non-destructive analysis of whole grain samples.[1,2] Whole grain analysis of malt to predict malt analysis requires the ability to predict the final results of a continuing biological process.

In traditional NIR calibration methods, absorbance/reflectance at specific wavelengths were correlated[3,4] to the laboratory analyses using multiple linear regression.[3,4] The resulting calibrations would involve only a few chosen wavelengths from the electromagnetic spectrum. This worked quite well for constituents such as protein and moisture. However, in the case of barley/malt when trying to predict constituents other than protein and moisture content, it is necessary to use calibration methods which utilize information from the entire electromagnetic spectrum and the associated calibration methods utilizing the principal component procedure.[4]

This paper will discuss the development of NIR calibrations used to measure a range of malt quality constituents using whole grain malt. An acceptable calibration was defined at the outset as one capable of being used in different locations, with many barley varieties and over different crop years.

Experimental

Sample population

A total of 4,386 commercial malt samples were obtained from the six malthouses of Canada Malting Company (Calgary, AB Thunder Bay, ON and Montreal, PQ) and Great Western Malting Company (Vancouver, WA, Pocatello, ID, Los Angles, CA) over a period of three crop years. The samples included 15 different varieties grown in many areas of North America. Micromalt samples were produced either using a Phoenix system (Phoenix Biosystems, Kingswood, S. Australia) or Seeger system (C. Seeger, Maschinenfabrik. Stuttgart-Bad Cannstatt, Germany). The commercial

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

samples used for calibration were obtained from individual malt batches and blended shipment lots. Samples were cleaned and subdivided for wet chemistry[5] and NIR analysis.

Collection of malt NIR spectra

An NIRSystems model 6500 scanning monochromator (Perstorp Analytical, Inc, Silver Spring, MD) was used to generate spectral data. Infrasoft International software (Infrasoft International, Port Matilda, PA) was used to collect and analyse the data, perform the calibration and cross-validation and predict a wide range of malt constituents.

Determination of reference laboratory values

All samples were subdivided and analyzed for moisture, malt protein, fine and coarse grind extract, color, diastatic power, alpha amylase, wort protein, wort viscosity and wort β-glucan. All analysis was carried out using ASBC methodology.[5]

Population structuring

Before calibration methods were applied to the sample population, it was subjected to an analysis of the spectra to determine if all samples were part of the same spectral population. Spectra from each location were grouped using the CENTER algorithm developed by Shenk and Westerhaus,[4,6] which evaluated the global standardized *H* distances (Mahalanobis distances) between each of the spectra and the mean of the population. This method, called CENTERING, was used to eliminate samples that did not fit spectrally within the population.

The SELECT algorithm[4,6] was used to determine the neighbourhood *H* distance of neighbors between spectra within the population. This method, called SELECTING, was used to eliminate one member of a pair of highly correlated spectra from the calibration set, thereby removing redundant spectra from the file.

Selected spectra were combined to generate a calibration set consisting of 2645 malt samples. This combined file was used to generate calibrations for a range of malt constituents. Table 1 shows the calibration ranges for each malt constituent.

Statistical evaluation of calibrations

All calibrations were evaluated using the standard error of cross validation (*SECV*), the coefficient of determination (r^2) and the number of terms. *SECV* is a good indicator of the amount of error within an unknown sample set[4] and cannot be less than the laboratory error associated with the reference measurement of the constituents. The variance that can be explained by the linear relationship between NIR and the wet chemistry is indicated by r^2. The number of terms (n) is the number of principal components needed to describe the spectra for any constituent. A higher number of terms may suggest a possible overfitting of the data.[4,7] The optimal calibration was defined as the calibration with the lowest error (*SECV*), highest fraction of explainable variance (r^2) and the lowest number of terms.[4]

The performance of the optimum calibrations was evaluated using the standard error of prediction (*SEP*) compared to the reproducibility values ($1/2R_{95}$)[5,8] and a comparison of the predicted mean (NIR) to the mean obtained by wet chemistry analysis.

Calibration

A modified partial least square regression (PLS) developed by Shenk and Westerhaus[6] and a PCA method[9,10] were used to develop the prediction equations. All calibrations were performed using the first derivative math treatment.[4,7] A modified version of multiplicative scatter correction (Weighted MSC) was evaluated as a means to reduce variability in the baseline and scale of the

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

Table 1. Constituent ranges for calibration set of 2645 samples.

Constituent	Range of values
Moisture, %	2.9–6.0
Extract, fine grind db, %	75.0–84.1
Extract, coarse grind db, %	73.0–83.4
Color, °ASBC	1.03–3.10
Diastatic power, °DP	69–198
Alpha amylase, DU	15.0–70.7
Malt protein, db, %	8.33–14.87
Wort protein, db, %	3.20–6.90
Wort viscosity. cP	1.29–1.71
wort β-glucan, mg L^{-1}	29–379

spectra.[4,7] The optimal wavelength region was determined by running calibrations on two different wavelength combinations in the vis/NIR region. The robustness of all calibrations was further increased by the use of a repeatability file which accounted for variation between instruments, fluctuations in temperature and differences in sample preparation.

Validation set

A validation set of 479 samples was selected from the overall population in such a manner as to cover all production locations and the entire range of spectral variation. These samples were excluded from all calibrations. This allowed determination of the performance of the optimal calibrations on an unknown sample set. Each sample was scanned once and corresponding wet chemistry analysis was performed in the laboratory.

Results and discussion

Comparison of calibration methods

Table 2 compares the results obtained using PCA and PLS calibration methods. The PLS calibrations showed lower *SECV*s and higher r^2 values for all constituents when compared to the PCA method. This suggests that the impact of the laboratory values must be considered throughout the calibration process. The PLS method was used for all subsequent calibration work.

Spectral region comparisons

Improvement of the PLS calibration required the evaluation of the spectral population. The reflectance of each wavelength was averaged for the calibration set to produce an average spectrum. The standard deviation for each wavelength was determined to produce a graphical representation of the population variation. The region of 400 nm–1100 nm contained a large amount of variation. If the variation in this area was due to random noise, the calibrations would

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

Table 2. Effect of calibration method on the performance of the calibration equation using the first derivative technique.[6,7]

	PCA		PLS	
Constituent	*SECV*	r^2	*SECV*	r^2
Moisture, %	0.28	0.36	0.21	0.66
Extract, fine grind db, %	0.82	0.67	0.54	0.85
Extract, coarse grind db, %	0.92	0.64	0.65	0.81
Color, °ASBC	0.29	0.08	0.23	0.32
Diastatic power, °DP	16.73	0.48	11.56	0.75
Alpha amylase, DU	6.01	0.28	4.33	0.66
Malt protein, db, %	0.68	0.42	0.34	0.86
Wort protein, db, %	0.43	0.19	0.25	0.74
wort viscosity, cP	0.05	0.29	0.03	0.65
wort β-glucan, mg L^{-1}	53.61	0.17	41.88	0.50

be improved by the elimination of this spectral region.[3] Table 3 shows the results of calibrations made using two different spectral segments. The elimination of 400 nm–1100 nm region of the spectrum either had no impact on the performance of the calibrations or resulted in slightly higher *SECV* values (Table 3). The calibration equations using the entire vis/NIR spectrum used a slightly lower number of terms for some constituents which suggests less overfitting of the data. The entire vis/NIR spectrum (400 nm–2500 nm) was therefore used for all subsequent calibration improvements.

Particle size correction

The general shape of the average spectrum and the shape of the population standard deviation was found to be similar. This suggests that the effects of particle size differences may have some effect on the spectra.[4] If particle size effects are a source of significant noise, the calibration performance could potentially be improved by the use of scatter correction before calibration. Table 4 shows the effect of subjecting the population to weighted MSC scatter correction of the spectral data before calibration. The results obtained of calibrations with and without scatter correction were similar with respect to the *SECV* and r^2 values. However, the calibrations using scatter correction utilized slightly fewer terms in the equations, suggesting less noise in the calibrations.

The optimum calibration was obtained using the PLS regression method, using first derivative math treatment, CENTERING, SELECTING, scatter correction and employing wavelengths from 400 to 2500 nm.

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

Table 3. Effect of elimination data from the 400–1100 nm spectral region on the performance of the PLS calibration equation using the first derivative technique.

	400–2500 nm			1100–2500 nm		
Constitutent	*SECV*	r^2	*n*	*SECV*	r^2	*n*
Moisture, %	0.21	0.65	12	0.21	0.69	13
Extract, fine grind db, %	0.54	0.85	12	0.59	0.82	13
Extract, coarse grind db, %	0.65	0.81	11	0.69	0.78	12
Color, °ASBC	0.24	0.34	11	0.23	0.35	12
Diastatic power, °DP	11.65	0.75	12	11.60	0.75	12
Alpha amylase, DU	4.39	0.65	12	4.42	0.63	12
Malt protein, db, %	0.34	0.87	13	0.35	0.86	13
Wort protein, db, %	0.25	0.73	12	0.24	0.74	13
Wort viscosity, cP	0.03	0.64	12	0.03	0.60	11
Wort β-glucan, mg L^{-1}	43.28	0.48	11	43.09	0.47	11

Table 4. Effect of scatter correction on the performance of the PLS calibration equation using the first derivative technique.

	With scatter correction			Without scatter correction		
Constituent	*SECV*	r^2	*n*	*SECV*	r^2	*n*
Moisture, %	0.21	0.65	12	0.21	0.66	13
Extract, fine grind db, %	0.54	0.85	12	0.54	0.85	12
Extract, coarse grind db, %	0.65	0.81	11	0.65	0.81	11
Color, °ASBC	0.24	0.34	11	0.23	0.32	11
Diastatic power, °DP	11.65	0.75	12	11.56	0.75	12
Alpha amylase, DU	4.39	0.65	12	4.33	0.66	13
Malt protein, db, %	0.34	0.87	13	0.34	0.86	13
Wort protein, db, %	0.25	0.73	12	0.25	0.74	13
Wort viscosity, cP	0.03	0.64	12	0.03	0.65	13
Wort β-glucan, mg L^{-1}	43.28	0.48	11	41.88	0.50	11

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

Table 5. Statistical evaluation of optimal calibration on the validation set (479 samples) for a variety of malt constituents.

Constituent	Wet chemistry mean	NIR mean	r^2	*SEP*	$1/2R_{95}$
Moisture, %	4	4	0.57	0.24	0.22
Extract, find grind db, %	80.4	80.4	0.8	0.62	0.56
Extract, coarse grind db, %	79.1	79.2	0.64	0.91	0.76
Colour, °ASBC	1.89	1.87	0.35	0.26	0.25
Diastatic Power, °DP	128	127	0.73	12.03	7.5
Alpha Amylase, DU	47.9	47.9	0.38	5.09	6.5
Malt Protein, db, %	11.42	11.4	0.79	0.39	0.44
Wort Protein, db, %	5.08	5.07	0.63	0.28	0.3
Wort viscosity, cP	1.49	1.49	0.79	0.04	0.07
Wort β-glucan, mg L^{-1}	156	147	0.46	61.92	15

Validation

The optimal calibration was used to predict a range of malt constituents in the validation set. Table 5 shows the means obtained using NIR prediction are similar to the means obtained by wet chemistry for all constituents. The standard error of prediction (*SEP*) is acceptable for most parameters given the standard error of the laboratory methods, which is shown by the reproducibility values ($l/2R_{95}$).

The calibration shows a high error of prediction for wort β-glucan. Part of the difficulty in developing a calibration for wort β-glucan is the lack of precision of the laboratory method.[4,7] Further refinement of the laboratory methods will be required in order to reduce the prediction error seen with NIR.

Wort color has an acceptable standard error of prediction, however, the r^2 value is low. This is an indication that wort color does not have a linear relationship with spectral reflectance measurements and alternative methodology may be required.

Conclusions

The PLS calibration method was superior to the PCA method for all constituents. The PLS calibration method performed best when utilizing all wavelengths from 400 to 2500 nm in the equation and benefitted from the use of scatter correction.

The strength of NIR technology becomes apparent when considering that it is possible to predict malt quality from whole grain malt, rapidly and non-destructively. The analysis is rapid, inexpensive and does not require the use of hazardous chemicals, thereby reducing laboratory and environmental control costs.

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996

Acknowledgements

The authors thank Canada Malting Company Ltd for permission to publish this work.

References

1. P.C. Williams in "Commercial Near Infrared Reflectance Analyzers", *in Near-Infrared Technology in the Agricultural and Food Industries*, Ed by P.C. Williams and K.H. Norris. American Association of Cereal Chemists Inc., St Paul, MN, pp. 35–55 (1987).
2. P.C. Williams and S.G. Stevenson, *Trends in Food Science & Technology* **August,** pp. 44–49 (1990).
3. W.R. Hruschka, "Data Analysis: Wavelength Selection Methods", in *Near Infrared Technology in the Agricultural and Food Industries*, Ed by P.C. Williams and K.H. Norris. American Association of Cereal Chemists, Inc., St Paul, MN, pp. 35–55 (1987).
4. J.S. Shenk and M.O. Westerhaus, *Analysis of Agriculture and Food Product by Near lnfrared Reflectance Spectroscopy*. Infrasoft International, PA (1993).
5. *Methods of Analysis*, 8th Edn. American Society of Brewing Chemists, The Society, St Paul, MN (1992).
6. J.S. Shenk and M.O. Westerhaus, *Crop Science* **31,** 469 (1991).
7. J.S. Shenk, J.J. Workman Jr and M.O. Westerhaus, "Application of NIR Spectroscopy to Agricultural Products", in *Handbook of Near Infrared Analysis*, Ed by D.A. Burns and E.W. Ciurczak. Marcel Dekker Inc., NY, pp. 383–431 (1992).
8. *Analytical-EBC*, 4th Edn. European Brewery Convention, EBC, Zurich, Switzerland (1987).
9. K.H. Norris, R.F. Barnes, J.E. Moore and J.S. Shenk, *J. Animal Sci.* **43,** 889 (1976).
10. J.S. Shenk and M.O. Westerhaus, *Crop Sci.* **31,** 469 (1991).

From Near Infrared Spectroscopy: The Future Waves
© IM Publications Open LLP 1996