The effect of particle size on the determinability of maize composition in reflection mode

Sándor Turza and Mária Váradi
Central Food Research Institute, Herman O. út 15, H-1022 Budapest, Hungary

Introduction

Near infrared (NIR) spectra contains information on the shape, size and probably the surface characteristics of particles.[1] The particles in any kind of food sample have a distribution of sizes. The effect of particle size on NIR spectra has been discussed many times, but its importance was first recognised and documented by Williams.[2,3] For smaller particles, the depth of sample penetration of NIR radiation decreases and the scattering coefficient increases. For a long time it was thought that grinding finely was essential to obtain accurate and precise NIR results but some works have shown that cereal grains are a special case because of the heterogeneous distribution of absorbers through the kernel. Among cereals, the effect of particle size has been investigated mostly for wheat and wheat products. However, this could be equally important in the case of maize. Therefore, our aim was to investigate the effect of different particle sizes on the accuracy of maize composition determination and also that of some pre-treatment methods to eliminate it.

Materials and methods

Samples and chemical analysis

Forty-seven maize samples (300–400 g) were obtained from the Hungrana Starch and Iso-sugar Factory, Szabadegyháza, Hungary. They were all last year's crops. Chemical analyses for moisture, starch, protein and oil were performed at the company's laboratory. The determinations were carried out according to the relevant Hungarian Standards and the results are shown in Table 1.

Sample preparation and NIR spectral recording

The samples were ground to three different particle sizes, 1.3 mm, 1.8 mm and 2.0 mm, respectively, using a Labmill QC-114(Labor-MIM, Hungary) grain grinder and then stored in small glass containers with screw-caps until measurement. The sample cell was a standard powder cuvette and the same one was used for every sample. Spectra were recorded using an NIRSystems 6250 spectrophotometer in the 1100–2498 nm range with 2 nm increments. Every sample spectrum was the average of fifty

Table 1. Basic statistical features of the constituents analysed.

	Starch	Protein	Oil
Sample number	47	47	47
Minimum %	69.05	8.02	3.15
Maximum %	73.95	9.95	5.55
Average %	71.59	8.99	4.27
Variance	1.93	0.27	0.21
Standard deviation %	1.39	0.52	0.46

Near Infrared Spectroscopy: Proceedings of the 10th International Conference

scans and the cuvette was rotated twice at 120 degrees. Three spectra were taken for all samples, which were then averaged to give one log1/*R* spectrum.

Spectral pre-processing

Four different pre-treatments methods were tried and tested (Figure 1). The first was full MSC and second derivative with different gap sizes, the second was full MSC with a 10 nm boxcar smoothing, followed by second derivative, the third was second derivative only and the last one was 10 nm boxcar smoothing with second derivative. The gap size was set at 10, 20, 30, 40, 50 and 60 nm and the segment size was fixed at a value of 2 nm.

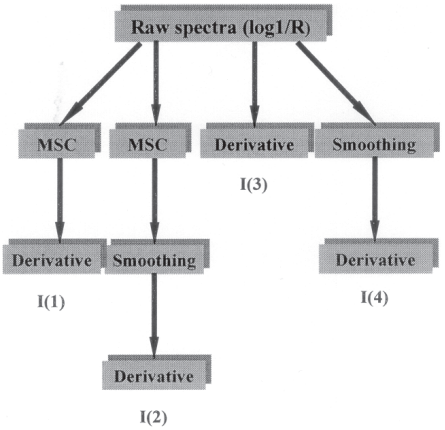

Figure 1. Flowchart of the pre-processing treatments applied.

Qualitative analysis

By taking the average of all 47 spectra for the three mesh sizes after pre-treatments, some wavelengths were selected that are perhaps related to particle size differences. With these selected wavelengths, linear discriminant analysis (LDA) was performed to prove their relevance to particle size.

Calibration and validation procedure

The pre-processed spectra were fed to an MLR algorithm in NSAS. The standard regression option was used with four terms. Linear additive models were calculated, so derivative ratios were not used. As the number of samples was not large enough to have a separate test set, full cross-validation was used. However, there is no possibility in NSAS to carry out cross-validation, so only calibration was done. Those calibrations which had the highest multiple regression coefficient (*R*) and the lowest standard error (*SE*) were selected to be used in Unscrambler. Here, the selected wavelengths were used as inputs for MLR regression, where full cross-validation was applied to test model validity. Regression models were built for starch, protein and oil for the different particle sizes. A so-called "mixture model" was also built, where the 47 samples were made up from the different mesh size groups, i.e. 15 from the 1.3 mm group, 17 from the 1.8 mm group and 15 from the 2.0 mm group, respectively. The difference between the three best "pure" calibrations and the "mixture model" calibrations before outlier removal (same sample number) were tested by a method first described by Pitman.[4]

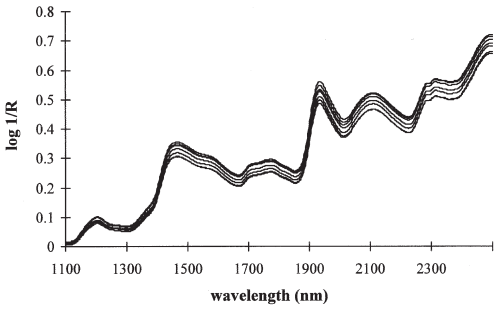

Figure 2. Raw spectra of some maize samples for the 1.8 mm group in the 1100–2498 nm region.

Results

The raw spectra indicated that beside the additive there was a multiplicative type of noise as well. This was manifested through the scissor -like opening of the spectral swarm (Figure 2). When taking the average of all 47 spectra for the three mesh sizes it appeared that this phenomenon was more pronounced in the longer region, in our case from 1932 nm to 2498 nm. When the

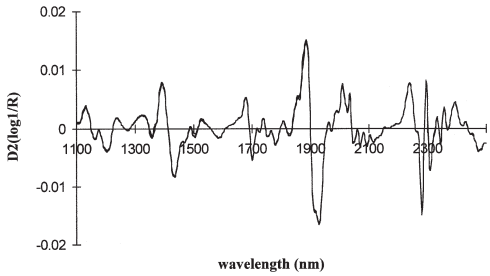

Figure 3. Second derivative spectra of the average of 47 samples for the three mesh size groups.

spectra were preprocessed in some way, the differences disappeared. However, a much closer examination revealed some wavelengths which still exhibit minute differences between different mesh size groups in the appropriate order (Figure 3). With these wavelengths LDA was performed and the results are summarised in Table 2. The wavelength selection was conducted with a 10 nm gap size treated spectra in all four preprocessing ways, for this gap size had the best calibration results in the majority of the cases. The highest classification accuracy was between the 1.3–2.0 mm groups, followed by the 1.3–1.8 mm and the 1.8–2.0 mm groups. In the last case, the proportion of correctly classified samples decreased, especially for the MSC treated spectra. For the first two groups the most dominant wavelength was 1202 nm, which is a relatively weak starch band. For the third group, the 2092 nm proved to be the most important. When the derivative was applied only, a second discriminant variable was necessary to have better classification.

The best regression results are summarised in Table 3. The values are not so good, but it has to be kept in mind that whole maize kernels were used, i.e. the germ was not separated. It is obvious that all results involve MSC treatment, which supports its use under conditions like this. The best figures for starch and protein were attained when the 2.0 mm group was used. For starch, in every pre-processing case, the 2.0 mm group was found to be the best. The oil content was best determined if the 1.3 mm group was selected for regression. The "mixed group" calibrations were not as good as the "pure" ones for starch and oil. The protein determination was not effected by the mixing. The goodness of the models, before outlier removal, was compared and it was found that there were no significant differences between them. After outlier removal the differences changed a little, but not so much that it could effect the significance to a considerable extent. The test method was not applied after outlier removal, as it is only applicable if there are an equal number of samples in each validation set.

Table 2. LDA results of the different mesh size groups with the wavelengths used and reclassification percentages.

	1.3–1.8 mm	1.3-2.0 mm	1.8-2.0 mm
Variables used I(1)	1202 nm 90.4%	1202 nm 93.6%	2092 nm 71.3%
Variables used I(2)	1202 nm 92.6%	1202 nm 93.6%	2092+1202 nm 76.6%
Variables used I(3)	1202 + 1506 nm 80%	1202 + 1506 nm 87.2%	2092 + 1202 nm 75%
Variables used I(4)	1202 nm 86.2%	1202 + 1504 nm 91.5%	2092 nm 79.8%

Table 3. The best calibration and validation results of the "pure" and the "mixed group" samples.

	Rcal	*SEC*	*Rval*	*SEP*	Bias
Starch (2.0 mm), I(2)	0.93	0.51	0.91	0.57	-0.00
Protein (2.0 mm), I(1)	0.93	0.18	0.91	0.20	0.00
Oil (1.3 mm), I(2)	0,88	0.20	0.85	0.23	-0.00
Mixed mesh sizes					
Starch I(1)	0.86	0.66	0.82	0.75	0.00
Protein I(1)	0.93	0.18	0.91	0.21	-0.00
Oil I(1)	0.79	0.24	0.76	0.29	0.00

Conclusions

The combination region is more sensitive to minor differences in particle size. There are some wavelengths that are less affected by the pre-treatments applied. It can also be concluded that the oil content could be best determined using the smallest mesh size group and for starch, the 2.0 mm group seemed to be the best choice. MSC treatment was helpful for every constituent. The small number of samples and the narrow calibration range can be held partially responsible for the relatively poor results. More samples, wider calibration range and more mesh size groups are necessary to prove the validity of results.

References

1. V. Chapelle, J.P. Melcion, P. Robert and D. Bertrand, *Science des Aliments* **9,** 387 (1989).
2. P.C. Williams, *Cereal Science Today* **19,** 280 (1974).
3. P.C. Williams and B.N. Thompson, *Cereal Chemistry* **55,** 1014 (1978).
4. Tom Fearn, *NIR news* **7,** 5 (1996).