Reduction of the number of wavelengths in near infrared spectral images of physically-damaged mushrooms

Carlos Esquerre[1], Aoife A. Gowen[1], Gerard Downey[1,2] and Colm P. O'Donnell[1]*

[1]Biosystems Engineering, University College Dublin, School of Biosystems Engineering, Dublin 4, Ireland
[2]Teagasc Food Research Centre Ashtown, Ashtown, Dublin 15, Ireland
*Corresponding author: colm.odonnell@ucd.ie

Introduction

Near infrared (NIR) spectral imaging produces large amounts of data e.g. a spectral image of 320 × 500 pixels with 121 spectral bands represents 73.8 MB of data to be transferred, processed and stored. In addition, irrelevant and unreliable bands adversely affect the performance of regression/classification models. Therefore, selection of a reduced number of relevant bands is highly desirable.[1] The present study employed the robust ensemble Monte Carlo uninformative variable elimination (REMCUVE) method to reduce the number of wavelength bands in spectral images of mushrooms.

Materials and Methods

Mushrooms

Closed cap *Agaricus bisporus* strain Sylvan A15 (Sylvan Spawn Ltd., Peterborough, United Kingdom) mushrooms grown in the usual commercial way at Teagasc Kinsealy Research Centre (Dublin, Ireland) were used in this study. For each of the 3 repetitions of the experiment, 120 mushrooms were selected at random and transported to the laboratory in special trays which prevented contact between mushroom caps. . Batches of mushrooms were collected on 4th September 2009 (B1), 6th October 2009 (B2) and 23rd March 2010 (B3). The objective of the present study was to evaluate the feasibility of discriminating physically-damaged tissue from sound tissue in mushrooms using a reduced number of NIR wavelengths when slight physical damage was induced at two times points: (a) day of collection and (b) one day after.

Treatment

Experiments commenced within 3 hours of picking the mushrooms. Samples were split into sub-groups of 10 mushrooms, 6 subgroups for the undamaged class and 6 subgroups for the damaged class. Damage was induced in mushrooms by shaking each sub-group separately for 30 s in a plastic box at 400 rpm immediately before scanning - sub-groups of day 0 (collection day), and before packing in punnets - sub-groups of day 1. Three sub-groups of undamaged and 3 of damaged mushrooms were analysed on day 0; each remaining sub-group (3 undamaged and 3 damaged) was packed in a punnet, covered with PVC film and stored for 1 day at 4ºC before analysis. The PVC film was pierced with a pen at each corner of the punnet to reduce the development of condensation.

Near infrared spectral imaging system

Spectral images of samples were acquired with a line scanning system (DV Optics, Italy) over the wavelength range 880 to 1720 nm at 7 nm intervals and stored in ENVI format using Spectral Scanner (DV Optics, Italy) software. The main parts of the system were: (a) illumination source, (b) diffuser, (c) moving base, (d) optics, (e) spectrograph, and (f) camera (Figure 1). The illumination source for the two first repetitions of the experiment consisted of a single tungsten halogen bulb (12 V, 150 W), with the illuminating light being conducted through a fibre optic assembly to the diffuser; this set-up was changed for the third repetition to a system in which five tungsten halogen bulbs were located equally-spaced inside the diffuser in order to increase the dynamic range of the system. The spectrograph was ImSpector N17E with a 50 μm slit width. The camera was an uncooled InGaAs SU320M-1.7RT (Sensors Unlimited, Inc, Princeton, NJ, USA) which operated at 50 MHz with 320 × 240 pixels (spatial dimension × spectral dimension); the moving base speed was set at 10 $mm.s^{-1}$ to obtain pixels of approximately 0.3 × 0.3 mm and the scanning line was approximately 16 cm long.

Reference paper as:
Esquerre, C., Gowen, A.A., Downey, G. and O'Donnell, C.P. (2012). Reduction of the number of wavelengths in near infrared spectral images of physically-damaged mushrooms, 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. 68-71.

The system was turned on and allowed to stabilise for around 30 minutes before calibration at the start of every scanning session. A two-point calibration procedure was followed; first, a black reference (I_b) was obtained by calculating the mean of 50 lines acquired after covering the spectrograph lens with a cap, and then a white tile, provided by the instrument company, was placed on the moving base and the average of 50 lines used as a "white" reference (I_w); the signal from the sample (I_s) was converted and stored as reflectance (R) according to:

$$R = \frac{(I_s - I_b)}{(I_w - I_b)} \quad (1)$$

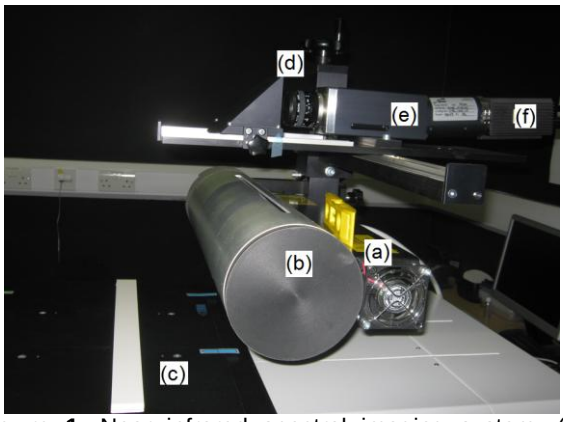

Figure 1. Near infrared spectral imaging system; (a) Illumination source, (b) diffuser, (c) moving base, (d) optics (mirror and lens), (e) spectrograph, and (f) camera.

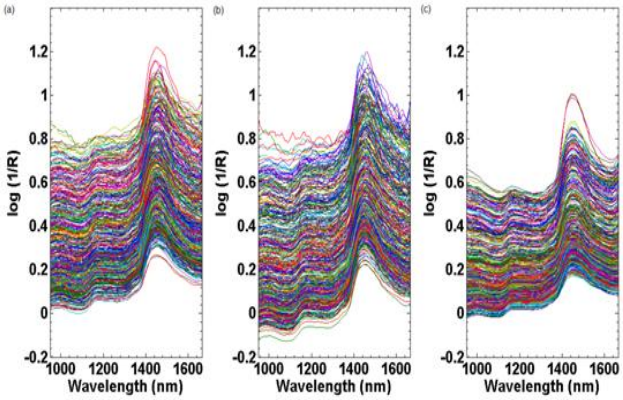

Figure 2. Near infrared spectra of a random sub-sample of 500 spectra from (a) batch one, (b) batch two and (c) batch three mushrooms.

Chemometrics

ENVI files were imported into Matlab version 2008a (The Math Works, Natick, MA, USA); all further data analysis was performed using in-house functions and scripts written in this software, including some functions from the Statistics Toolbox (The Math Works, Natick, MA, USA) and PLS Toolbox (Eigenvector, Wenatchee, WA, USA). Image background was removed from NIR spectral images by creating a mask on the basis of a suitable threshold on the minimum of each reflectance spectrum; selection of the thresholds was made iteratively by analysing the corresponding histogram and drawing a tentative mask image.

One thousand spectra were randomly selected from masked images of each sub-group of mushrooms, generating 12000 spectral samples for each batch (6000 spectra each for undamaged and damaged mushrooms).

Pretreatment

Asymmetric least squares baseline correction was applied to spectra after smoothing with a Savitzky-Golay filter (using 9 points window and second order polynomial). Spectra from each class were scrutinised for outliers; a spectrum was labelled as an outlier and removed from the dataset for any particular pre-treatment if its associated Hotelling statistic (T^2) was greater than the critical value (T^2_{crit}) calculated by Equation 2, where *lv* is the number of components, m is the number of spectra in the dataset and F(0.05,*lv*,300) is the F statistic with α = 0.05, *lv* and 300 degrees of freedom. The number of components (*lv*) was the minimum between 6 and the lowest number of components explaining more than 90 % of the variability in the dataset; from previous experience, score images became unacceptably noisy after the sixth principal component regardless of the pre-treatment or the amount of cumulative explained variance.

$$T^2_{crit} = \left(lv * \frac{m\text{-}1}{m\text{-}lv} \right) * F_{(0.05,lv,300)} \quad (2)$$

Reference paper as:
Esquerre, C., Gowen, A.A., Downey, G. and O'Donnell, C.P. (2012). Reduction of the number of wavelengths in near infrared spectral images of physically-damaged mushrooms, 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. 68-71.

Variable elimination

The robust ensemble Monte Carlo uninformative variable elimination (REMCUVE) method was applied to remove uninformative wavelength bands based on the median of normalised regression coefficients (C_{median}) generated in a Monte Carlo procedure. Mushrooms from the first batch of samples (B1) were used as a training set to build the model; samples from the second batch were used as a tuning set to identify the most suitable number of latent variables while samples from the third batch (B3) were used to validate the final model. The geometric mean (G) of the correctly identified fraction of both classes was used to evaluate the performance of the models.

Results and Discussion

Spectra

Mushroom spectra (Figures 2 and 3) display peaks and other features which are similar to spectra described in previous work.[2,3] A peak with maximum at 1454 nm in log(1/*R*) was the most prominent characteristic and this could be attributed to water, corresponding to the first overtone of O-H stretching ($2\ \nu_1$) plus the combination of the first overtone of the O-H bending and antisymmetric O-H stretching band ($2\ \nu_2 + \nu_3$). In addition, it was possible to observe (a) a small peak at 978 nm which may be ascribed to the second overtone of the O-H stretching band ($3\ \nu_1$), (b) overlapping peaks at 1160 nm and 1230 nm that may be attributed to the combination of the first overtone of the O–H stretching and the OH-bending band ($2\ \nu_{1,3} + \nu_2$) and second overtone of the C-H stretching band of CH (carbohydrate content), (c) a shoulder at 1342 nm that may be related to the first overtone of antisymmetric O-H stretching band ($2\ \nu_3$) plus a combination of C-H stretching with C-H deformation (carbohydrate content) and (d) a shoulder at 1580 nm attributable to strongly-bound water.[4-7]

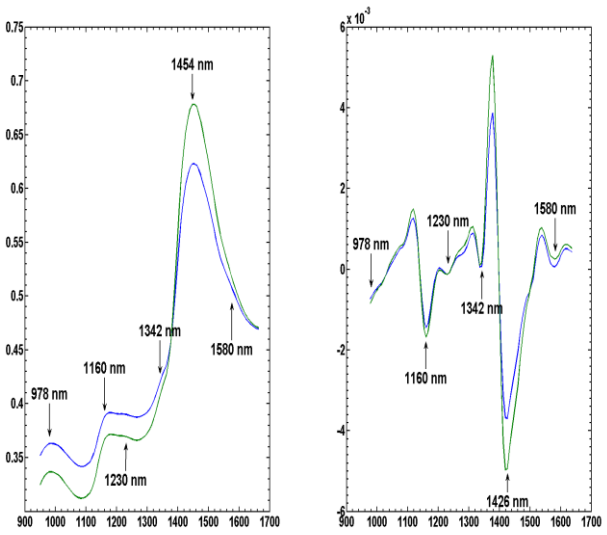

Figure 3. Mean spectra for 6000 randomly-selected pixels of each class of mushrooms from batch 1 (B1); blue and green lines correspond to undamaged and damaged mushrooms respectively. (a) log(1/R) spectra (b) second derivative spectra (window size of 9 data points, second order polynomial smoothing).

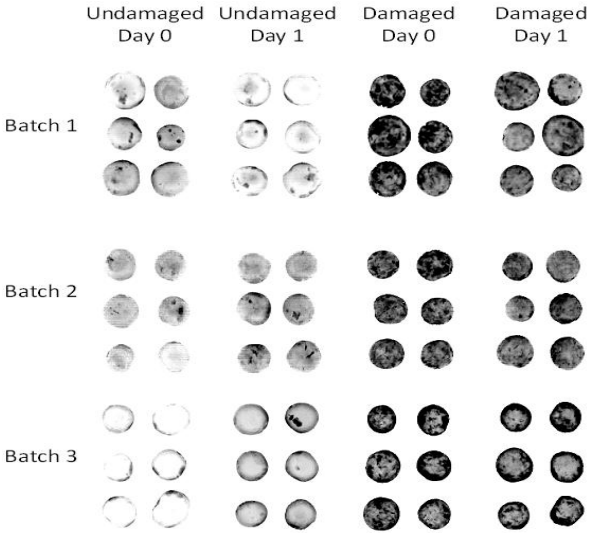

Figure 4. Image of prediction values for some undamaged and damaged mushrooms from batch 1, 2 and 3 (calibration set, tuning set and validation set respectively).

Changing the illumination source improved the quality of B3 spectra, reducing the noise and spread of the data (Figure 2). However, in Figure 2c it is also possible to observe some spectral aberrations in B3 spectra, with different slopes at the extremes related to the offset of each spectrum. These characteristics increase the difficulty for a model built with B1 mushrooms and optimised with B2 samples to predict properly the damage in the B3 dataset.

REMCUVE

Wavelength bands around the 1450 nm band presented the most stable regression coefficients as assessed by the median of normalised absolute values of regression coefficient. A model built with only 2 wavelength bands in this range (1447 and 1454 nm) achieved a good performance (G_2 = 0.813, G_3 = 0.754) in classification of the tuning and validation sets respectively (Table 1). This was in agreement with previous work in which differences in water-related bands were the most prominent features for discriminating

Reference paper as:
Esquerre, C., Gowen, A.A., Downey, G. and O'Donnell, C.P. (2012). Reduction of the number of wavelengths in near infrared spectral images of physically-damaged mushrooms, 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. 68-71.

between undamaged and damaged mushroom tissue in the NIR spectral range.[2,3] Bias in the tuning set may be attributed to inter-batch differences and the large bias in the validation set may be attributed to changes in illumination conditions in addition to possible differences in characteristics of the B1 and B3 samples.

Response variable images calculated with the optimised PLS model are shown in Figure 4. Undamaged and damaged mushrooms from all batches were easy to identify. In the same figure, it is possible to observe some dark patches in the response image of undamaged mushrooms. Those on the borders of each mushroom image may correspond to contact points between mushrooms during growth while the other blotches may originate from handling (during harvest, transport and cleaning), soil residues, condensation or damage caused by pests (flies). More detailed features observed in the B3 prediction images may be explained by the larger dynamic range achieved by the increased illumination consequent on the illumination system upgrade.

Table 1. Partial least squares discriminant analysis model performance for: (a) raw log (1/R) and (b) wavelength bands retained by robust Monte Carlo uninformative variable elimination method (2 retained variables).

Model	Dataset	Correctly classified fraction		G	Bias
		Undamaged	Damaged		
Raw (4 LV)	Training (B1)	0.950	0.825	0.885	0.000
	Tuning (B2)	0.836	0.792	0.814	-0.179
	Validation (B3)	0.817	0.631	0.718	0.288
Selected (1 LV)	Training (B1)	0.955	0.822	0.886	0.000
	Tuning (B2)	0.853	0.796	0.813	-0.217
	Validation (B3)	0.860	0.720	0.754	0.528

Conclusions

The selected (log[1/R]) features of mushrooms were tentatively assigned to water molecule vibration modes due to the high moisture content (ca. 90%) of samples. The first overtone of the O-H stretching bond in water was the main feature in log(1/R) spectra of mushrooms with a maximum at 1454 nm at which the combination of the first overtone of the O-H bending and antisymmetric O-H stretching band ($2\nu_2 + \nu_3$) is present. This band corresponded to the most noticeable difference between undamaged and damaged spectra in both log(1/R) and second derivative spectra. Spectral quality was improved (reduced noise and variability in spectra) when the illumination flux was increased by the use of five bulbs in the light source.

Good discrimination between undamaged mushrooms and physically-damaged mushrooms was possible with a two-wavelength model identified using the robust ensemble Monte Carlo uninformative variable elimination (REMCUVE) method on NIR spectral images. This simple NIR model may form the basis of an automatic grading system for use in the mushroom industry.

Acknowledgements

Financial support from the Irish Department of Agriculture, Fisheries and Food under the FIRM Programme is acknowledged.

References

1. C.H. Spiegelman, M.J. McShane, M. Goetz, M. Motamedi, Q. Yue and G. Cote, *Anal. Chem.* **70,** 35-44 (1998).
2. C. Esquerre, A.A. Gowen, C.P. O'Donnell and G. Downey, *J. Agric. Food. Chem.* **57,** 1903-1907 (2009).
3. C. Esquerre, A.A. Gowen, C.P. O'Donnell and G. Downey, *J. Near Infrared Spectrosc.* **17,** 353-361 (2009).
4. H. Maeda and Y. Ozaki, *J. Near Infrared Spectrosc.* **3,** 191-201 (1995).
5. B. Osborne, T. Fearn and P.H. Hindle, *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman, Singapore (1993).
6. J.S. Shenk, J.J. Workman, 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, CRC Press, Boca Raton FL, p. 347 (2001).
7. H.W. Siesler, Y. Ozaki, S. Kawata and H.M. Heise (Eds), *Near-infrared spectroscopy: Principles, instruments, applications*. Wiley, Weinheim. p. 348 (2002).

Reference paper as:
Esquerre, C., Gowen, A.A., Downey, G. and O'Donnell, C.P. (2012). Reduction of the number of wavelengths in near infrared spectral images of physically-damaged mushrooms, 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. 68-71.