Application of Partial Least Squares regression to relate tastiness of boiled potatoes to chemical and physical tests

Marie F. Smith[1*], Christine S. Leighton[2], Liesl Morey[1] and Hettie C. Schönfeldt[3]

[1]ARC-Biometry Unit, P O Box 8783, Hatfield, Pretoria, 0001 South Africa
[2]ARC-Sensory Analysis Unit, Animal Production Institute, Private Bag X2, Irene, 0062 South Africa
[3]School of Agricultural and Food Sciences, University of Pretoria, Pretoria, 0001 South Africa

Accepted 14 January 2009

Partial Least Squares regression (PLS) was used to understand the relationship between 15 sensory attributes of potatoes boiled in their skins as dependent variables (Y), and six objective (chemical and physical) measurements as the independent variables (X). The ARC-Sensory Analysis Unit, at Irene, South Africa, conducted the study. A trained sensory panel (n=10) was used to determine the texture, aroma and flavour attributes of five potato cultivars; Mondial, BP1, Up-to-Date, Van der Plank and Caren. Four repetitions of each sample were used. The PLS regression formed three distinct groups of the cultivars: (1) Mondial, (2) BP1 and Van der Plank, which contrasted with (3) Caren and Up-to-Date. These groupings were confirmed by the culinary uses of the cultivars and their ability to retain shape after boiling. Mondial is favoured for making potato salad or boiling as it retains its shape during cooking. BP1 and Van der Plank are suitable for most uses of boiled potatoes as they retain their shape, whereas Caren and Up-to-Date do not retain their shape during cooking and are suitable for baking and frying.

Keywords: Multivariate regression, partial least squares regression, principal component analysis, sensory analysis

* To whom correspondence should be addressed (E-mail: smithmf@arc.agric.za)

Introduction

Multivariate regression is applied to understand the effect of many measured independent or uncorrelated variables on a response variable of interest, whereas Partial Least Squares (PLS) is a multivariate regression method that can be used to compare blocks of variables, or relate one or several response variables (Y) to several correlated explanatory variables (X). This method aims to identify the underlying factors, or linear combinations of the explanatory variables, which best model the dependent variables. PLS regression can deal efficiently with data sets where there are large numbers of variables that are highly correlated, involving substantial random noise, and where the number of rows are less than the number of columns. MacFie and Hedderley (1993) state that PLS regression can be seen as a hybrid of multiple regression and principal component analysis (PCA). Relating sensory and instrumental variables to determine the effects of cultivar differences on hop flavour in beer (Peppard *et al.* 1989) and on meat quality (Toscas *et al*. 1999), are early references to PLS regression use with regard to sensory evaluation of foods.

PLS regression consists of relating X and Y matrices in one single estimation procedure. In contrast to the PCA approach where each component is a linear combination of the X variables only, PLS regression is based on the component scores using both Y and X matrices. PLS1 is used to predict one dependent variable, and PLS2 to predict several dependent variables, from a set of independent variables. In PLS1 uncorrelated components of independent variables are computed (as in PCA), such that they maximally predict the dependent variable. In the case of PLS2 uncorrelated components, which are latent to both data sets, are computed such that they maximally predict the set of dependent variables. Furthermore, a cross-validation procedure assists in the determination of the correct number of dimensions (principal components as in PCA) to include in the model (Helland, 1988). Details of the PLS regression theory and its similarities with PCA and multiple linear regression analysis are described in Aastveit and Martens (1986).

PLS regression was used to understand the relationship between sensory attributes of potatoes boiled in their skins to six objective test measurements, with the objective of determining the most important chemical and physical measurements that influence the sensory profiles of the potatoes.

Material and methods

Experimental data

The Sensory Analysis Unit of the Agricultural Research Council (ARC) Animal Production Institute, South Africa, conducted the study. A trained sensory panel (n=10) was used to determine the texture, aroma and flavour attributes of five different potato cultivars (Mondial, BP1, Up-to-Date, Van der Plank and Caren) from the Limpopo region in South Africa. Four repetitions of each sample were measured. The samples were evaluated on an eight-point intensity scale (Meilgaard *et al.* 1991), with one denoting the least intense (e.g. no earthy potato aroma) and eight denoting the most intense condition (e.g. extremely intense earthy potato aroma). For each sample 15 sensory attributes were taken: three aroma attributes (buttery, earthy and vegetable water), three flavour attributes (buttery, earthy and vegetable water), one texture attribute (hardness, the force required to slice through the potato with a knife), one oral first bite attribute (fracturability in mouth), and seven chewing attributes (mealiness in mouth after chewing, waxy or sticky feeling on pallet after pressing on tongue, moistness of potato in mouth, coarse texture, content of grainy particles in mouth after chewing, and compactness in mouth before swallowing).

In addition, six objective tests were applied in order to

assist in the characterisation of different potato cultivars. These comprised four physical tests (shear force resistance (kg) of each sample for raw and cooked potatoes using an Instron Universal Testing Machine, specific gravity (SG) that is a measurement of density and determines the starch weight in potatoes, and % softening of a sample from raw to cooked), and two chemical analyses (% dry matter content and % starch content).

Statistical analyses

Analysis of variance (ANOVA) was performed to determine whether significant differences existed among the means of the five cultivars on all sensory attributes and objective tests (Table 1). Fisher's protected t-test least significant difference was applied to separate cultivar means at the 5% level of significance. All attributes and objective tests were significantly different ($P \leq 0.05$), except for the three aroma, three flavour and grainy texture attributes. The statistical program GenStat® (Payne *et al*., 2007) was used for all statistical analyses mentioned in this paper.

Results and discussion

PLS regression was applied to relate the sensory attributes to the objective tests and to identify the most important attributes differentiating between the potato cultivars after boiling. PLS regression is a technique whereby components, similar to those in PCA, are derived by taking into account the variation in the sensory attributes (Y matrix) that is relevant for explaining variation in the objective measurements of interest (X matrix) in the original samples. These components are constructed as linear combinations of the original variables. Usually the first few components capture or explain most of the variation in the entire original data set.

The PLS regression generates output similar to PCA, which provides the percentage variance that each component explains within each data set for each dimension included in the model, as well as the scores and loadings of the samples on the latent variables. A biplot of the loadings of both the X and Y variables for the first two PLS components (Figure 1), allows a researcher to conclude: (1) on the relationship among variables of each data set, and (2) on the relationship among X and Y variables (Payne *et al*., 2007). In such plots points closest together are most similar, and those far apart dissimilar. Variable points in opposite directions from each other are negatively correlated and those close together are positively correlated to each other.

A PLS regression was performed for the case where the Y matrix comprised 15 sensory attributes. The explanatory vari-

Table 1 Means of sensory attributes and objective tests (physical and chemical measurements) for five potato

	Cultivars				
Sensory attributes	Mondial	BP1	Van der Plank	Up-to-Date	Caren
Hardness to cut	3.1 ab	3.4 a	3.5 a	2.4 b	2.7 ab
Fracturability	5.9 bc	5.5 cd	5.3 d	6.4 a	6.0 ab
Mealiness	3.0 b	3.0 b	3.7 a	4.3 a	4.1 a
Waxiness	1.9 a	1.9 a	1.5 b	1.4 b	1.5 b
Stickiness	2.0 ab	2.3 a	2.3 a	1.5 c	1.9 bc
Moistness	3.6 a	3.5 ab	2.8 d	3.2 bc	3.0 cd
Coarse texture	2.6 b	2.5 b	2.6 b	3.4 a	3.4 a
Compactness	2.8 ab	3.0 ab	3.1 a	2.3 c	2.7 bc
Grainy content	2.3 a	2.3 a	2.4 a	2.4 a	2.9 a
Aroma, buttery	1.4 a	1.2 a	1.3 a	1.2 a	1.4 a
Aroma, earthy	2.3 a	2.7 a	2.5 a	2.6 a	2.2 a
Aroma, vegetable water	2.2 a	2.2 a	2.1 a	2.6 a	2.4 a
Flavour, buttery	1.4 a	1.3 a	1.3 a	1.2 a	1.3 a
Flavour, earthy	2.2 a	2.3 a	2.4 a	2.4 a	2.3 a
Flavour, vegetable water	2.2 a	2.1 a	2.3 a	2.5 a	2.3 a
Physical measurements					
Specific gravity	1.05 c	1.06 bc	1.08 a	1.06 b	1.06 b
Raw shear force	2.97 c	3.2 c	4.0 a	4.4 a	3.6 b
Cooked shear force	0.42 ab	0.36 b	0.44 ab	0.34 b	0.46 a
% Softening	86.0 b	88.6 b	89.0 b	92.3 a	87.0 b
Chemical measurements					
% Dry matter content	14.65 e	18.31 d	18.67 c	18.98 b	22.04 a
% Starch content	10.34 c	13.30 b	13.88 b	14.36 ab	16.14 a

Means per row followed by the same letter (a,b,c,d) did not differ significantly at the 5% level, applying Fisher's protected t-test least significant difference

ables matrix X comprised six objective tests and indicator variables for cultivars. The indicator variables for cultivars are five variables consisting of one for those samples that came from a particular cultivar and zero for those that did not (Toscas *et al*., 1999). All sensory attributes and objective test values used in the PLS regression application were taken as the means for each cultivar. Not all objective tests were measured on the same scale, thus these were standardised to have the same variance. This was not done for the sensory attributes as these were all assessed using the same evaluating scheme, but the values were mean centred. Cross-validation was used to determine the number of components to include in the analysis; one dimension was indicated, but two were included so that a biplot could be constructed.

Table 2 indicates the percentage of the variance explained by the dependent variables (sensory attributes) and independent variables (objective tests), ranked according to the significant first dimension. In this dimension (cf. horizontal axis of Figure 1) of the sensory attributes, mealiness (Mc) explained most of the variance (86.6%) and contrasted with waxiness (Wm), which explained second most of the variance (85.6%). In the second dimension (cf. vertical axis of Figure 1) fracturability (Fm) explained most of the variance (85.2%) and contrasted with hardness to cut (Hk, 71.8%). In terms of the objective tests, % softening (%S) explained most of the variance (82.6%) and contrasted most with cooked shear force (Sc, 73.0%).

It can be seen in Figure 1 that cultivars Mondial, BP1 and Van der Plank contrasted with Caren and Up-to-Date on the first PLS component. Furthermore, it clearly contrasted cooked shear force (Sc) with all other objective tests (Sr, St, DM, %S and SG), thus they are negatively correlated with Sc. Most of the aroma (Ab, Ae & Av), flavour (Fe, Fb & Fv) attributes and grainy content (Gm) cluster around the midpoint of the biplot, which is an indication of their small contribution to the model.

Cultivars BP1 and Van der Plank were mostly associated with hardness to cut (Hk), stickiness (Sm) and compactness (Cm), and contrasted mainly with fracturability (Fm), coarse texture (Ct), and mealiness (Mc). Mondial was mostly associated with moistness (Mt), cooked shear force (Sc) and waxiness (Wm), and contrasted mostly with mealiness (Mc) and the other objective tests SG, % softening (%S), % dry matter content (DM), % starch content (St) and raw shear force (Sr). On the other hand, Up-to-Date was mostly associated with fracturability (Fm), coarse texture (Ct) and mealiness (Mc), and contrasted mainly with hardness to cut (Hk), stickiness (Sm) and compactness (Cm). Cultivar Caren was mostly associated with the aroma and flavour of cooked vegetable water (Av & Fv), grainy texture (Gm) and contrasted most with hardness to cut (Hk).

The PLS regression contrasted cultivars Mondial, BP1 and Van der Plank with Caren and Up-to-Date on the first component (Figure 1). This explains the cultivar characteristic whereby the first group retains its shape when boiled and the second group does not. Furthermore, the PLS regression grouped Van der Plank and BP1 together (bottom left quadrant) and contrasted them with Caren and Up-to-Date (top right quadrant). Mondial was dissimilar from both these groups (top left quadrant). Apart from the PLS loadings for each data set, PLS regression calculates scores for each set. The generated X component scores of the objective tests provide a ranking of the cultivars in the order: Mondial (score = -3.3), BP1 (score = -1.1), Van der Plank (score = 0.7), Caren (score = 1.5), and lastly Up-to-Date (score = 2.2). Table 3 is a summary of the five cultivars ranked according to their PLS scores, subjective cooking and tasting observations made by the laboratory staff during preparation of the samples, culinary uses, as well as the PLS interpretation.

Table 2 Percentage variance explained, ranked according to the first dimension, after PLS application to sensory attributes and objective tests (chemical and physical measurements)

Y matrix (Dependent variables)	% of the Y variances explained		X matrix (Independent variables)	% of the X variances explained	
Sensory attributes	1st dim	2nd dim	Chemical & physical variables	1st dim	2nd dim
Mealiness	86.6	6.5	% Softening	82.6	4.4
Waxiness	85.6	0.7	% Starch content	79.3	1.4
Flavour, earthy	65.6	15.2	Raw shear force	77.8	0.0
Coarse texture	58.8	35.8	Cooked shear force	73.0	2.6
Moistness	52.9	20.6	% Dry matter content	68.1	1.5
Flavour, buttery	50.7	13.1	Specific gravity	45.2	37.1
Aroma, veg. water	43.0	56.6			
Flavour, veg. water	41.5	9.0			
Grainy content	28.3	0.9			
Hardness to cut	28.1	71.8			
Stickiness	27.7	68.2			
Aroma, buttery	23.7	1.3			
Compactness	20.7	75.3			
Fracturability	13.5	85.2			
Aroma, earthy	1.8	19.2			

Table 3 PLS ranking of potato cultivars by independent variable component scores, subjective cooking and tasting observations, main culinary uses and PLS loading biplot interpretation

Cultivar (PLS group)	PLS score	Subjective cooking and tasting observations & culinary uses	PLS biplot results (Figure 1)
Mondial (1)	-3.3	Tasty and buttery. Favoured for making potato salad or boiling as it retains its shape, and makes moist mash, as it is very moist and waxy, least mealy, grainy and low starch content.	Mostly associated with moistness, cooked shear force, waxiness, and contrasted with all other objective tests, mealiness and has low starch content.
BP1 (2)	-1.1	Earthy and pleasant tasting. Most cooking uses. Retains its shape when cooked. Makes good, slightly sticky mash, is hard to cut, sticky, moist, waxy and compact.	Mostly associated with stickiness, compactness and hardness to cut. It contrasted most with fracturability, coarse texture and mealiness.
Van der Plank (2)	0.7	Earthy and pleasant tasting. Most cooking uses. Retains its shape when cooked. Makes drier, slightly grainy mash, is hard to cut, slightly sticky, compact, with high SG, thus high starch content.	Mostly associated with hardness to cut, compactness and stickiness. It contrasted most with fracturability, coarse texture and mealiness.
Caren (3)	1.5	Tasty. Best fried or baked. Breaks up most when cooked. Makes dry, grainy mash, is coarse and has very high dry matter and starch content.	Mostly associated with cooked vegetable water aroma and flavour, and graininess and contrasted with hardness to cut, stickiness and compactness.
Up-to-Date (3)	2.2	Not very tasty. Best fried or baked. Does not retain its shape when cooked. Makes good, slightly dry and smooth but slightly grainy mash, is most fracturable, mealy, coarse, has high softening, dry matter and starch content.	Mostly associated with coarse texture, fracturability, mealiness and contrasted mostly with stickiness, compactness and hardness to cut.

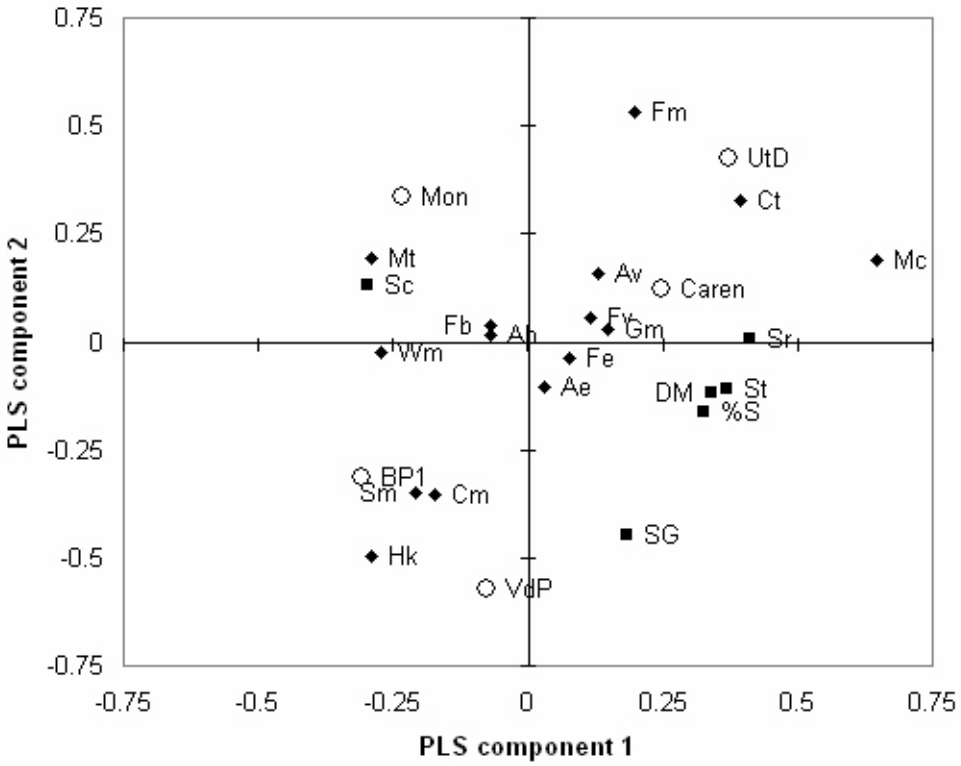

Figure 1 Loadings for the first two partial least squares regression components of five boiled potato cultivars on 15 sensory attributes and six objective tests.

- Sensory attributes are: Mealiness (Mc), Waxiness (Wm), Hardness (Hk), Fracturability (Fm), Stickiness (Sm), Moistness (Mt), Coarse texture (Ct), Compactness (Cm), Grainy content (Gm), Aroma buttery (Ab), Aroma earthy (Ae), Aroma vegetable water (Av), Flavour buttery (Fb), Flavour earthy (Fe), Flavour vegetable water (Fv).
- Physical measurements: Specific gravity (SG), Raw shear force (Sr), Cooked shear force (Sc), % Softening (%S).
- Chemical measurements: % Dry matter content (DM) and % Starch content (St).

○ Cultivars: Mondial (Mon), BP1, Van der Plank (VdP), Caren, Up-to-Date (UtD).

Conclusions

The PLS regression contrasted Mondial with the other four cultivars. Mondial retains its shape, is tasty and buttery, makes very moist mash and is favoured for boiling and potato salad. BP1 and Van der Plank retain their shape when boiled, were found to be earthy and pleasant tasting, make good mash which is slightly sticky, and can be used in any cooking method. The culinary uses of Caren and Up-to-Date are similar as they do not retain their shape when boiled, make a dryer, grainy mash and are favoured for frying and oven baking; Caren breaks up easily and was found to be tasty, but Up-to-Date was not very tasty.

Acknowledgements

The authors wish to thank the Sensory Analysis Unit, ARC-Animal Production Institute, for the use of this data.

References

AASTVEIT, H. & MARTENS, H., 1986. ANOVA interactions interpreted by partial least squares regression. *Biometrics* 42, 829-844.

HELLAND, I.S., 1988. On the structure of partial least squares regression. *Commun., Statist.-.Simula.Comput.*, 17, 581-607.

MACFIE, H.J.H. & HEDDERLEY, D., 1993. Current practice in relating sensory perception to instrument measurements. *Food Quality & Preference* 4, 41-49.

MEILGAARD, M., CIVILLE, G.V. & CARR, B.T., 1991. Sensory evaluation techniques, 2nd edn, CRC Press. New York.

PAYNE, R.W., MURRAY, D.A., HARDING, S.A., BAIRD, D.B. & SOUTAR, D.M., 2007. GenStat for Windows® (10th Edition) Introduction. VSN International, Hemel Hempstead, UK.

PEPPARD, T.L., RAMUS, S.A., WITT, C.A. & SIEBERT, K.J., 1989. Correlation of sensory and instrumental data in elucidating the effect of varietal differences on hop flavour in beer. *Journal of the ASBC* 47, 18-26.

TOSCAS, P.J., SHAW, F.D. & BEILKEN, S.L., 1999. Partial least squares (PLS) regression for the analysis of instrument measurements and sensory meat quality data. *Meat Science* 52, 173-178.