Prediction of Main Potato Compounds by NIRS

Ainara López-Maestresalas*, Claudia Pérez, Roberto Tierno, Silvia Arazuri, Jose Ignacio Ruiz de Galarreta, Carmen Jarén

Department of Projects and Rural Engineering, Universidad Pública de Navarra, Campus de Arrosadia 31006, Navarra, Spain
ainara.lopez@unavarra.es

Potato (*Solanum tuberosum*, L) compounds are generally determined by analytical methods including gas-liquid chromatography (GLC), HPLC and UV-VIS spectrophotometry. These methods require a lot of time and are destructive. Therefore, they seem to be not suitable for in-line applications in the food industry. Near-infrared spectroscopy (NIRS) is a technique that presents some advantages over reference methods for quantitative analysis of agricultural and food products since it is fast, reliable and non-destructive.

For this reason, in this study, quantitative analyses were carried out to determine main compounds in potatoes using NIRS.

Potato tubers grown in two consecutive years were used for the analyses. NIR spectral acquisition was acquired on lyophilized samples. In year 1, a total of 135 samples were used while 228 samples were used in year 2. Lyophilized samples were also scanned by NIRS, two replicates per samples were acquired and the mean spectrum of each sample was used for the analysis.

Different chemical analyses were carried out each year. Thus, in year 1 the following parameters were quantified: reducing sugars (RS) and nitrogen (N), whereas in year 2, total soluble phenolics (TSP) and hydrophilic antioxidant capacity (HAC) were extracted and quantified. Then, chemometric analyses were performed using Unscrambler X (version 10.3, CAMO software AS, Oslo, Norway) to correlate wet chemical analysis with spectral data. Quantitative analyses based on PLS regression models were developed in order to predict the above chemical compounds of tubers in a non-destructive manner.

Good PLS regression models were obtained for the prediction of nitrogen and TSP with coefficients of determination (R^2) above 0.83. Moreover, PLS models obtained for the estimation of HAC could be used for screening and approximate calibrations.

1. Introduction

Potato (*Solanum tuberosum*, L.) is one of the most important crops in the world, occupying the fifth position in terms of production. Despite of being a highly appreciated product, potato industry faces the continuously growing demand of quality products from consumers and regulatory bodies. The acceptance of these products in the market depends on several factors as the general aspect, texture and internal quality. The analytical methods commonly employed to determine the main compounds of potatoes in order to control their quality require a lot of time and are destructive. These include gas-liquid chromatography (GLC), HPLC and UV-VIS spectrophotometry. Therefore, they seem not to be suitable for in-line applications in the food industry (Chen et al., 2010). In this respect, near-infrared spectroscopy (NIR) presents some advantages over reference methods for quantitative analysis of agricultural and food products. For these reasons, the aim of this study is to estimate potato compounds by NIR spectroscopy in a wide range of samples.

2. Materials and methods

2.1 Vegetal material

A total of 363 tubers were used in this study. Tubers were selected from potato accessions (Potato Germplasm Collection, NEIKER) grown during the years 2012 and 2013 in a precise field trial in Arkaute (Álava) in the north-east of Spain. Some of the varieties included in this work are currently undergoing

breeding programs. In year 1 (2012), 135 tubers were scanned by NIRS (n_{year1}=135) while 228 tubers were scanned in year 2 (2013) (n_{year2}=228). The range of samples studied covers white-, red-, yellow- and purple-fleshed varieties.

Tubers were lyophilized prior to spectral acquisition. For this, potatoes were cut lengthwise in order to obtain representative samples of the different tissues. Pieces from 5 to 8 different tubers were then lyophilized in a freeze-dryer Alpha d1-4 (CHRIST, Germany) until they reached 250 g of fresh weight. This process was carried out at -50ºC and 0 atmospheres until the samples lost their whole water content. After that, they were ground with liquid nitrogen up to fine dust and stored at -20ºC until their use.

2.2 Spectral acquisition

After lyophilisation, spectral data were adquired with a Luminar 5030 Miniature 'Hand held' AOTF-NIR (Acousto-Optic Tunable Filter-Near Infrared) Analyzer (Brimrose, Baltimore, MD, USA). This is a rugged portable spectrometer capable of conducting non-destructive and contact/non-contact tests. This system comprises an InGaAs detector and pre-aligned lamp assembly. This spectrometer allows diffuse reflectance measurements and liquid transmission with probe attachment in the wavelength range from 1100 to 2300 nm. Moreover, it has a scanning speed of 16,000 wavelength/s and a wavelength increment adjustable from 1 to 10 nm. This system can acquire 26 spectra per second. In this study, the signals were acquired with Brimrose Analytical Software SNAP32! with Brimrose MACRO language and transferred later on Unscrambler X (version 10.3, CAMO software AS, Oslo, Norway).

Two NIR measurements were acquired per lyophilized sample and the mean spectrum was used for further studies. Data from every cultivar was used to build a representative calibration set containing all the variability of samples under study with the objective to improve the prediction capabilities.

The 1100-2300 nm spectral range with 601 points (2 nm steps) was used to obtain the spectra at room temperature. Each spectrum measured was the average of 50 spectra.

2.3 Chemical analysis

The extraction and quantification of the chemical compounds were carried out in NEIKER-Tecnalia.

Estimation of reducing sugars (RS) was performed following the dinitrosalicylic acid method. For reagent preparation, first 4.8 g of NaOH were diluted in 60 ml of distilled water. Then, 3 g of dinitrosalicylic acid and 150 ml of distilled water were added to the solution. After complete dilution, 90 g of Rochelle salt were added to a total volume of 330 ml. After that, 0.3 g of ground potato were diluted in 1 ml of distilled water in a test tube. Next, 2 ml of the dinitrosalicylic acid solution were added to the test tube. Samples were placed in a bain-marie for 5 min. Then, they were introduced in a cold water bath, shaken for 2 min and left to rest for another 10 min. Later on, 1 ml of sample solution was diluted in 5 ml of distilled water. The absorbance of samples at 546 nm was calculated with a spectrophotometer. Finally, content of reducing sugars was calculated by the relationship between absorbance and percentage in sugars, and reported as percentage of DW.

The estimation of total nitrogen content of the samples was carried out by Kjeldahl method. It consists of three basic steps: 1) digestion of the sample in sulphuric acid with a catalyst, which results in conversion of nitrogen to ammonia; 2) distillation of the ammonia into a trapping solution; and 3) quantification of the ammonia by titration with a standard solution. It is reported as percentage of DW.

The extraction of TSP and its quantification was made from lyophilized samples (0.5 g) with 10 ml methanol: H_2O (70:30). The solid was re-suspended by shaking in a vortex for 5 min. The mixture was centrifuged at 8000 rpm for 10 min at 4ºC, and the supernatant containing extracted phenolics was collected. The extraction operation was repeated once again with the pellet and the mixture was centrifuged at 8000 rpm for 10 min at 4ºC once again. The methanolic extracts were pooled, transferred to 20 ml volumetric flasks, and made up to the 20 ml mark. Total phenolics were determined by an adapted microscale protocol for the Fast Blue BB spectrophotometric method using gallic acid as standard. The extract solution (1 ml) was taken in a cuvette, then 0.1 ml of 0.1% FBBB [4-benzoylamino-2,5-dimethoxybenzenediazonium chloride hemi(zinc chloride) salt] dissolved in methanol was added, mixed for 30 s, followed by 0.1 ml 5% NaOH, mixed, and the resulting mixture allowed to incubate for 90 min at room temperature. Absorbance was measured at 420 nm using methanol 70% as blank. Total phenolic were calculated by interpolating the absorbance data in a calibration curve prepared with gallic acid (concentration range: 0.0-500.0 µg/ml) and reported as mg GAE g^{-1} DW.

Hydrophilic antioxidant capacity (HAC) was analysed in year 2 (2013) following the DPPH method. For this, 30 µl of sample was pipetted into 1.17 ml of DPPH• solution to initiate the reaction. The absorbance was read every minute at 515 nm for 180 min using a spectrophotometer. Under these conditions, the decrease in absorbance reached a plateau within the 2 h sampling period. Therefore, a reaction time of 2 h was used for all the DPPH• assays. Methanol (70%) was used as a blank. Trolox with concentrations from 0 to 1000 µM was used as a standard. The antioxidant activity was reported in µmoles of Trolox equivalents per g sample (µmol TE/g DW).

2.4 Chemometric analysis

In this study, spectral and chemical data of potatoes were correlated by a Partial Least Squares (PLS) regression method. Samples were divided into calibration and prediction datasets corresponding to 70% and 30%, respectively. The calibration dataset was used to build the PLS model and then, it was externally evaluated using the prediction dataset. Different pre-treatments of the spectral data were carried out including techniques for the scatter correction and derivatives, such as Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC) and First Derivative (1st der). SNV is a method of spectral normalization, which establishes a common scale for all spectra by centering each spectrum on its mean value and scaling it by its standard deviation and MSC reduces scatter effects in the data (Rinnan et al., 2009). Both first and second derivatives remove baseline offsets in the data, while the latter is also useful for separating overlapping peaks. Both first and second derivatives remove baseline offsets in the data (Burger & Geladi 2007). Venetian Blinds cross-validation method was applied to define subsets of the calibration set of samples. Both pre-processing of data and PLS regression models were carried out in Unscrambler X (version 10.3, CAMO software AS, Oslo, Norway). The accuracy of the models for the estimation of chemical compounds, was evaluated on the basis of the following parameters: the coefficients of determination (R^2) obtained in cross-validation and prediction; the prediction error of the model defined as the root mean square error for cross-validation (RMSECV); and, the root mean square error for prediction (RMSEP) (Naes et al., 2002).

The coefficient of determination indicates the percentage of the variance in the Y variable that is accounted for by the X variable. An R^2 value between 0.50 and 0.65 indicates that discrimination between high and low concentrations of the parameter measured can be made. A value for R^2 between 0.66 and 0.81 indicates that model can be used for screening and "approximate" calibration, while, an R^2 between 0.82 and 0.90 can be used for most applications. Calibration models with an R^2 value higher than 0.91 are considered to be excellent (Williams, 2003); and a value for R^2 above 0.98 can be used in any application (Williams, 2001).

The standard error of cross-validation (SECV) and the standard error of prediction (SEP) were also reported.

3. Results and discussion

An overview of the concentration of the different parameters analyzed in calibration and prediction datasets is given in table 1.

Table 1: Overview of concentration of the different parameters and number of samples (n) in the calibration and prediction datasets

Year	Parameter	Calibration dataset				Prediction dataset			
		n	Range	Mean	SD	*n*	Range	Mean	SD
2012	RS	90	0.22-1.83	0.76	0.35	45	0.20-1.34	0.73	0.31
	N	90	1.05-2.26	1.62	0.27	45	1.12-2.33	1.60	0.26
2013	TSP	152	0.75-14.28	3.95	3.12	76	0.91-14.44	4.20	3.38
	HAC	152	0.90-40.37	9.55	8.53	76	1.40-39.82	9.78	8.31

Table 1 shows that the range of TSP content was wide as it was the range of potatoes used in this study. As stated before, purple- and red-fleshed potatoes were used and it is known that those have much more range of TSP than yellow-fleshed ones. Same behaviour was observed for HAC values.

In table 2, the results obtained in cross-validation and prediction datasets from the two years analysed are shown. The latent variables (LV) used was different for each PLS model carried out. Thus, 6 LV were needed for HAC prediction, while 7 LV were used for RS estimation, 8 LV for the prediction of TSP and finally 11 LV were necessary for the estimation of N. The number of samples (*n*) shown on the table was computed after identification and elimination of a few outliers.

Table 2: Cross-validation and prediction statistics for PLS models developed to predict different compounds in potatoes

Year	Parameter	Pre-process	*n*	LV	R^2CV	RMSECV	R^2P	RMSEP	SEP
2012	RS	1st der+SNV	113	7	0.49	0.25	0.42	0.24	0.25
	N	MSC+1st der	128	11	0.90	0.08	0.86	0.10	0.11
2013	TSP	1st der (9p)	219	8	0.84	1.20	0.83	1.41	1.33
	HAC	1st der (9p)	228	6	0.67	4.84	0.70	4.52	4.51

It can be seen that in year 1 (2012) a good PLS model was obtained for nitrogen with an R^2 value of 0.86 for the prediction dataset and a low SEP value of 0.11. According to the R^2 value, this model could be used for most applications. Figure 1 shows the PLS model develop for the prediction of Nitrogen content of samples.
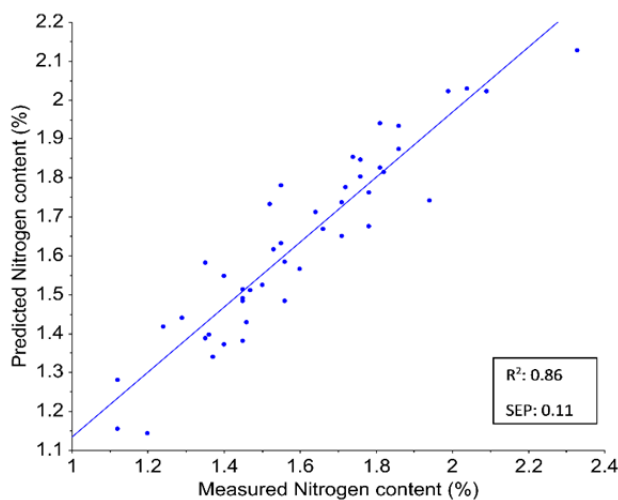

Figure 1: PLS regression plot of predicted versus measured Nitrogen content (%) of lyophilized samples in the 1100-2300 nm spectral range pre-processed with MSC + 1st derivative.

Figure 2 shows the plot of weighted regression coefficients for the PLS regression model developed to estimate Nitrogen content of lyophilized potatoes. This plot indicates that multiple spectral wavelength regions along the NIR spectral range played an important role in the PLS model for prediction of Nitrogen.
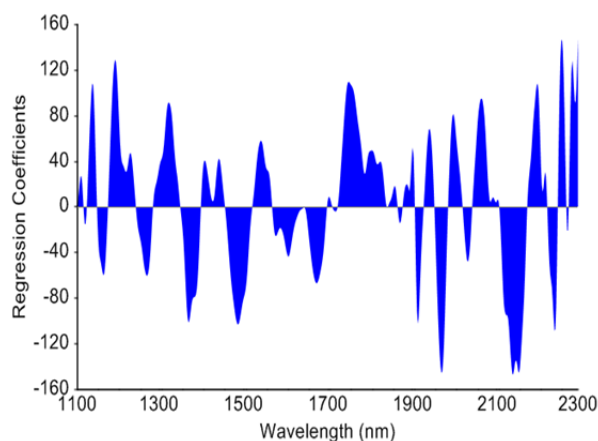

Figure 2: Weighted regressions plot showing important variables in modelling Nitrogen content in lyophilized potato samples from year 1 (2012).

Other authors have investigated the correlation between NIRS and nitrogen absorption of potato plants to calculate proportions of supplemental nitrogen fertilizer in potato crops obtaining low SEP values (0.09%) (Young et al., 1997).

On the other hand, a good model for RS was not obtained, probably due to the small presence of those compounds in potato tubers. The results obtained in this study are similar to those obtained by Mehrubeoglu and Cote (1997) and Haase (2011).

It can also be seen in table 2 that PLS models developed with samples belonging to year 2 (2013) reported good results for the prediction of TSP with an R^2 value of 0.83 for the validation set and a SEP value of 1.33 (figure 3). Therefore, according to these two statistics, this is considered a good model. On the other hand, the estimation of HAC reported an R^2 value of 0.70 and a SEP value of 4.51. These results suggest that this model could be used for screening and approximate calibrations.

In this study, high correlations for cross-validation and prediction were obtained with small RMSECV and RMSEP values regarding phenolic content. Similarly, in a research developed by Shiroma-Kian et al. (2008),

estimation of the polyphenol compounds in lyophilized potatoes using Fourier transform infrared spectroscopy (FTIR) was performed. Authors achieved excellent performance statistics with correlations >0.99 for the cross-validated PLS models with a SECV value of 4.17.

In a related work, Bonierbale et al. (2009) estimated the total and individual carotenoid concentration in *Solanum phureja* potatoes by NIRS obtaining R^2 values between 0.63 and 0.92 with SEP values ranging from 20 to 610 µg/100g DW.

Figure 3 shows that there are two groups of samples according to their TSP content. The largest one with TSP values up to 6 mg GAE/g DW that includes yellow- and red-fleshed cultivars, and a second group of samples with greater TSP values between 6-15 mg GAE/g DW corresponding to some red- and purple-fleshed tubers.
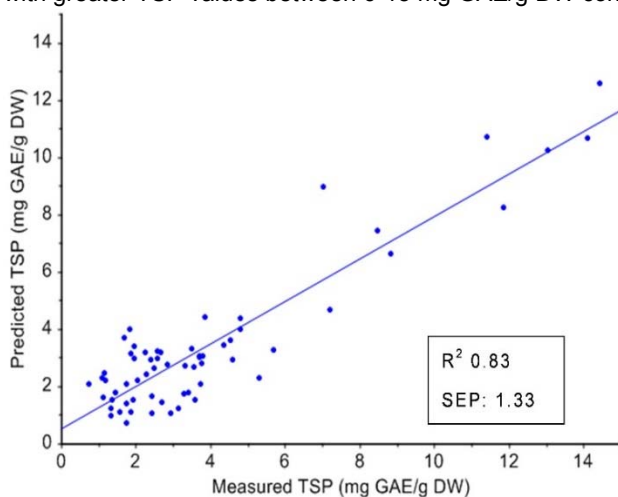

Figure 3: PLS regression plot of predicted versus measured TSP content (mg GAE/g DW) of lyophilized samples in the 1100-2300 nm spectral range pre-processed with 1st derivative.
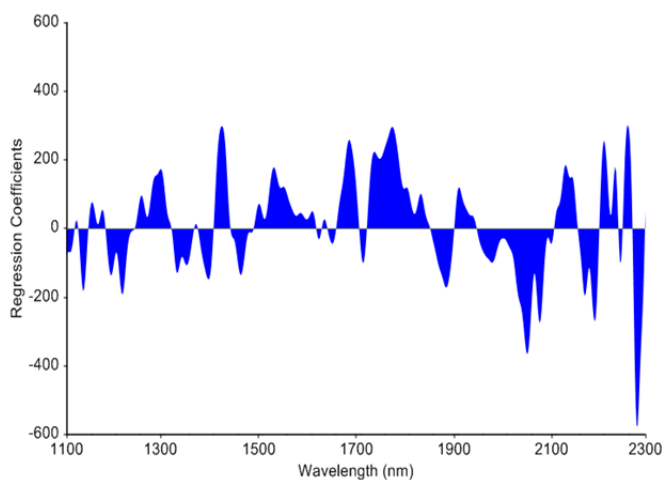

Figure 4: Weighted regressions plot showing important variables in modelling TSP content in lyophilized potato samples from year 2 (2013).

In figure 4 the plot of weighted regression coefficients for the PLS regression model developed to estimate TSP content of lyophilized potatoes is shown. Similar to figure 2, in this plot it can be seen that multiple spectral wavelength regions were important in the PLS model for prediction of TSP, especially the spectral region between 1700 and 1850 nm associated with 1st stretching overtone of C=H and the 2000 to 2100 nm spectral region related to starch absorption bands (Osborne et al., 1993).

4. Conclusions

Quantitative analyses of different potato compounds have shown that good PLS regression models could be obtained for the prediction of nitrogen and TSP in a large collection of potato varieties. Moreover, PLS model

obtained for the estimation of HAC could be used for screening and approximate calibrations. This is a very important point, since most of the varieties used in this study are included in breeding programs where it becomes essential to identify samples with high and low hydrophilic antioxidant capacity.

Acknowledgments

References

Bonierbale M., Grüneberg W., Amoros W., Burgos G., Salas E., Porras E., Zum Felde T., 2009, Total and individual carotenoid profiles in Solanum phureja cultivated potatoes: II. Development and application of near-infrared reflectance spectroscopy (NIRS) calibrations for germplasm characterization, Journal of Food Composition and Analysis 22(6), 509-516.

Burger, J. & Geladi, P., 2007, Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering, Journal of Near Infrared Spectroscopy 15(1), 29-38.

Chen J.Y., Zhang H., Miao Y., Asakura M., 2010, Nondestructive Determination of Sugar Content in Potato Tubers Using Visible and Near Infrared Spectroscopy, Japan Journal of Food Engineering 11(1), 59-64.

Haase N.U., 2011, Prediction of potato processing quality by near infrared reflectance spectroscopy of ground raw tubers, Journal of Near Infrared Spectroscopy 19(1), 37-45.

Mehrubeoglu M., Cote G.L., 1997, Determination of total reducing sugars in potato samples using near-infrared spectroscopy, Cereal Foods World 42(5), 409-413.

Naes T., Isaksson T., Fearn, T., Davies T., Eds., 2002, A user friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.

Osborne B., Fearn T., Hindle P., Eds., 1993, Practical NIR spectroscopy with applications in food and beverage analysis. Longman Scientific and Technical, Harlow, UK.

Rinnan, Å., Van den Berg, F. & Engelsen, S.B., 2009, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends in Analytical Chemistry, 28(10), 1201–1222.

Shiroma-Kian C., Tay D., Manrique I., Giusti M.M., Rodriguez-Saona L.E., 2008, Improving the screening process for the selection of potato breeding lines with enhanced polyphenolics content, J. Agric. Food Chem. 56(21), 9835-9842.

Williams P., Eds., 2003, Near-infrared technology—Getting the best out of light. PDK Grain, Nanaimo, Canada.

Williams P.C., Eds, 2001, Implementation of Near-Infrared Technology. Near-Infrared technology in the agricultural and food industries. 2nd ed. American Association of Cereal Chemists, Inc., Minesota, USA.

Young M.W., MacKerron D.K.L., Davies H.V., 1997, Calibration of near infrared reflectance spectroscopy to estimate nitrogen concentration in potato tissues, Potato Research 40(2), 215-220.