Estimation of Proximate Composition of Quinoa (*Chenopodium quinoa*, Willd.) Flour by Near-Infrared Transmission Spectroscopy

Christian Encina-Zelada[1,2,3]([✉]) [iD], Vasco Cadavez[1] [iD], Jorge Pereda[2],
Luz Gómez-Pando[4], Bettit Salvá-Ruíz[2] [iD], Martha Ibañez[4],
José A. Teixeira[3] [iD], and Ursula Gonzales-Barron[1] [iD]

[1] CIMO Mountain Research Centre, School of Agriculture, Polytechnic Institute
of Braganza, Bragança, Portugal
cencina@lamolina.edu.pe
[2] Department of Food Technology, Faculty of Food Industries,
National Agricultural University La Molina, Lima, Peru
[3] Centre of Biological Engineering, School of Engineering,
University of Minho, Braga, Portugal
[4] Cereals and Andean Crops Programme, Faculty of Agronomy,
National Agricultural University La Molina, Lima, Peru

Abstract. The aim of this study was to develop chemometric models for protein, fat, ashes and carbohydrates contents of quinoa flour using Near-Infrared Transmission (NIT) spectroscopy. Spectra of quinoa flour obtained from grains of 70 different cultivars were scanned while dietary constituents were determined by reference AOAC methods. As a pre-treatment, spectra were subjected to extended multiplicative signal correction (EMSC) with polynomial degree 0, 1 or 2. Next, the Canonical Powered Partial Least Squares (CPPLS) algorithm was applied, and models were compared in terms of accuracy and predictability. For all models, root mean square errors of cross-validation (RMSECV), root meat square errors of prediction (RMSEP) and coefficient of correlation of cross-validation (R_{CV}) were computed. Robust models were obtained when quinoa spectra were pre-processed using EMSC of polynomial degree 2 for both fat (RMSECV: 0.268% and RMSEP: 0.256%) and carbohydrates (RMSECV: 0.641% and RMSEP: 0.643%) following extraction of five CPPLS latent variables. Good coefficients of correlation of prediction (R_P: 0.690–0.821) were found for all constituents when models were validated on a test data set consisting of 13 quinoa flour spectra. Thus, good predictions of the dietary constituents of quinoa flour could be achieved by using NIT technology, as implied by the low coefficient of variation of prediction (CV_P): 5.64% for protein, 3.88% for fat 7.32% for ashes and 0.80% for carbohydrates contents.

Keywords: Quinoa Flour · Calibration · Chemometrics · Bootstrap

1   Introduction

Quinoa (*Chenopodium quinoa*, Willd.) has been considered by FAO as one of the crops destined to offer food security in the 21st century, because the plants show tolerance to frost, salinity and drought, and have the ability to grow on marginal soils [1].

The application of chemometric techniques to the global signature spectra of any food can be used to evaluate compositional characteristics that may be not easily detected by chemical analysis [2]. Considering their advantages over standard methods, near-infrared transmission (NIT) spectroscopy has been shown to be a good rapid method for evaluation of grains and flours [3]. The objective of this study was to develop NIT models to estimate proximate composition of quinoa flour using the novel Canonical Powered Partial Least Squares (CPPLS) multivariate algorithm [4].

2   Methodology

2.1   Samples and Proximate Composition Analysis

The samples utilised in this study were quinoa (*Chenopodium quinoa*, Willd.) flour milled from orange, beige, black and yellow quinoa grains, corresponding to 70 different cultivars. They were all harvested in Peru at the National Agricultural University La Molina (Cereals and Andean Crops Programme) and the Regional Development Centre—Highland (Junin, Peru), in different seasons between 2010 and 2012. Samples of quinoa seeds were first grounded in a Cyclotec 1093 mill, and sieved through a 0.5 mm mesh screen. Moisture, protein, fat and ashes contents were determined using the reference methods 925.10, 920.87 (conversion factor of 6.25), 923.05 and 923.03, respectively [5]. The determinations were done in triplicate and averaged. Total carbohydrates content was calculated by difference. Proteins, fat, ashes and carbohydrate contents were expressed in dry basis (db).

2.2   Near-Infrared Transmission (NIT) Spectra Acquisition

NIT spectra were acquired by placing the whole flour directly in an Infratec 1241 analyser (Module Foss Tecator, Denmark), using 60-mm quartz cuvettes, and scanning the region 850–1050 nm (wavenumber range of 11765–9524 cm^{-1} approximately). The spectra were recorded at scanning step intervals of 2 nm to give 100 data points per sample. A total of 10 frequency scans were performed per sample, and carefully assessed for consistency. To correct for the non-linearity in the measure of transmittance (T), T was transformed into absorbance (A) by taking the base 10 logarithm of the reciprocal of the transmittance values.

2.3   NIT Spectral Pre-processing

To minimise the multiplicative effects of light scattering, quinoa flours' spectra were subjected to extended multiplicative signal correction (EMSC) with polynomial degrees of 0, 1 or 2. EMSC attempts to separate physical light scattering effects from

chemical (vibrational) light absorbance, which helps in correcting for wavelength dependent baseline effects in the spectra [6]. For each of the constituents analysed, the CPPLS algorithm was fitted to these EMSC pre-processed spectral data with polynomial degree 0 (EMSC–0), 1 (EMSC–1) or 2 (EMSC–2), thereby producing three combinations or models (i.e., EMSC–0/CPPLS, EMSC–1/CPPLS, EMSC–2/CPPLS) which were compared in terms of accuracy and predictability.

2.4 Chemometric Multivariate Data Analysis

The extraction of information from quinoa flour's pre-processed spectra to estimate protein, fat, ashes and carbohydrates contents was performed by the CPPLS chemometric algorithm. As Mevik [4] explained, the CPPLS algorithm can employ additional variables (AV) to further improve the predictive performance of the model. Thus, the AVs determined by trial-and-error were: carbohydrate ("cho") for the moisture model; moisture ("mois") and variety ("var") for the fat model; "var" for the ashes model; and "var" and ashes ("ash") for the carbohydrate model (Table 2).

As a first step, the full data set was divided into a subset for calibration ($\sim$81% data, 57 samples) and the remaining $\sim$19% (13 samples) for prediction or validation, by means of random split stratified by variety. After EMSC scatter-corrected spectra (polynomial degree 0, 1 or 2), the CPPLS algorithm was fitted to each of them. The performance of the different models was determined by cross-validation as an internal calibration method using the calibration data set. In our case, the leave-one-out (LOO) method was used. The performance of the model was assessed by the root mean square error of cross-validation (RMSECV), which is deemed as the best single estimate of the prediction capability of the model [7]. Then, the optimal number of components of a model was selected at the first RMSECV local minimum, rather than the absolute minimum (to avoid overfitting). For such a number of components, the root mean square error of calibration (RMSEC) was computed. In addition, the coefficients of correlation between reference values and values fitted by cross-validation (R_{CV}) and the calibration model (R_C) were computed.

Following completion of the calibration, models were validated using the prediction data set to determine the accuracy of prediction. Model performance was evaluated by obtaining the root mean square error of prediction (RMSEP) and the coefficient of correlation (R_P) between reference values and those predicted by the model. To assess the best model (i.e., EMSC-0/CPPLS, EMSC-1/CPPLS or EMSC-2/CPPLS) for each dietary constituent, the model had to present not only a low RMSEP but also a high R_P. The spectra analysis was conducted using the "mass" [8], "caret" [9], "pls" [4], "emsc" [10], and the "prospectr" [11], packages implemented in the R software version 3.3.1 [12].

3 Results and Discussion

3.1 Proximate Composition Analysis of Quinoa Flour

The means, ranges and standard deviations of the major compounds of the quinoa flours (g/100g db) are summarised in Table 1. Quinoa samples from Peru, Bolivia and

Brazil, evaluated by Ferreira [13], presented higher protein content (11.40–36.10% db) and lower carbohydrates content (43.64–76.37% db). They also encountered substantially higher fat (6.19–15.52% db) and ashes (3.07–9.15% db) contents than those of our study. The variation in ashes are influenced by the dependence of the mineral content on type of soil and fertiliser application. Samples of quinoa cultivated in Chile [14] also presented higher protein content (16.0–20.2% db) but a fat content (4.4–7.5% db) comparable to that of our samples. In general, our levels are in agreement with those found by Vega-Gálvez [15], who determined fat content averages between 5.0 to 7.2%. Moisture is the compound having the highest variability among published studies (from 8.26–11.51% in Repo-Carrasco-Valencia [16] up to 25.66–33.16% in Ferreira [2]) because it depends upon drying and storage of seeds.

Table 1. Summary statistics of the major dietary compounds of quinoa flour samples in % dry basis, except for moisture (% wet basis).

Compound	Minimum	Maximum	Mean	St. deviation
Moisture	9.17	11.91	10.58	0.71
Protein	8.51	11.16	9.89	0.70
Fat	5.67	7.78	6.61	0.45
Ashes	2.51	4.11	3.17	0.37
Carbohydrates	78.48	82.12	80.33	0.88

3.2   Assessment of Chemometric Models

It was observed that, as a whole, a higher order polynomial of EMSC led to fewer optimal latent variables (Table 2). Except for the protein's chemometric models, the polynomial degree 1 or 2 of the EMSC signal correction of quinoa's flour spectra generally improved the accuracy of the models (when compared to polynomial degree 0), yet to different extent: the reduction in RMSECV (5.1–7.2%) in the model for carbohydrates was comparable to that of fat (5.9–7.9%), while it was lower for ashes (4.7–5.9%). In the case of RMSEC, the greatest reductions were attained for the fat model (5.5–24.4%) and ashes model (5.1–12.0%). Likewise, the polynomial degree 1 or 2 of the EMSC filter improved the correlation statistics of calibration: the increase in R_{CV} and R_C values was notorious in the models for fat (increase in 3.8–5.0% and 1.6–6.2%, respectively) and ashes (5.0–6.3% and 2.6–5.9%), whereas the improvement for the carbohydrates model only occurred in the R_{CV} values (13.5–18.7%) (percentage differences not shown but calculated from Table 2).

The predictability statistics of the models were also improved by using the polynomial degrees 1 or 2. Comparing RMSEP values, the greatest reductions were achieved for the carbohydrates model (19.2–21.3%) and ashes (10.6–17.0%), while the improvement for the fat model was lower (4.3–6.2%). The correlation statistics of prediction R_P improved using polynomial degrees 1 or 2 in the models for estimating ashes (22.4–33.1%), carbohydrates (17.7–19.3%) and fat (5.2–7.4%) content when compared to the "0" base.

Table 2. Accuracy of prediction of NIT chemometric models for quinoa flour constituents defined by signal correction (EMSC: extended multiplicative signal correction with polynomial degree 0, 1 or 2) and CPPLS (canonical powered partial least squares) algorithm, as measured by the root mean square errors of cross-validation (RMSECV), calibration (RMSEC) and prediction (RMSEP), and the coefficients of correlation between reference values and those estimated by cross-validation (R_{CV}), calibration (R_C) and prediction (R_P), all of them computed at the minimum number of components. The most robust model for protein, fat, ashes and carbohydrates are shown in bold.

Compound	Algotithm	Signal Correction	Degree	Variable	Number components	RMSECV (%)	RMSEC (%)	RMSEP (%)	R_{CV}	R_C	R_P
Protein	CPPLS	EMSC	**0**	cho	**10**	**0.5839**	**0.4347**	**0.5575**	**0.5053**	**0.7664**	**0.6899**
			1		7	0.6205	0.4882	0.5023	0.3987	0.6924	0.7581
			2		8	0.6250	0.4706	0.4954	0.3833	0.7185	0.7657
Fat	CPPLS	EMSC	0	mois + var	7	0.2915	0.2202	0.2732	0.7722	0.8773	0.6640
			1		11	0.2744	0.1664	0.2614	0.8014	0.9319	0.6987
			2		**5**	**0.2683**	**0.2082**	**0.2562**	**0.8111**	**0.8911**	**0.7128**
Ashes	CPPLS	EMSC	0	var	7	0.2739	0.2234	0.2794	0.6867	0.8053	0.5364
			1		**8**	**0.2577**	**0.1967**	**0.2319**	**0.7297**	**0.8530**	**0.7139**
			2		5	0.2610	0.2121	0.2497	0.7213	0.8265	0.6567
Carbohydrates	CPPLS	EMSC	0	ash + var	9	0.6910	0.4939	0.8169	0.5047	0.7869	0.6883
			1		7	0.6559	0.5049	0.6599	0.5730	0.7759	0.8103
			2		**5**	**0.6409**	**0.5131**	**0.6429**	**0.5990**	**0.7675**	**0.8210**

Variable: mois = moisture, ash = ashes, cho = carbohydrates, var = variety

Taking the three models together (Table 2), the models estimating fat and ashes presented generally the highest predictive capacity, as deduced from the ranges of R_{CV} (0.772–0.811 and 0.687–0.730, respectively) and R_P (0.664–0.713 and 0.536–0.714, respectively). The models for carbohydrates (R_{CV}: 0.505–0.599; R_P: 0.688–0.821) and protein (R_{CV}: 0.383–0.505; R_P: 0.690–0.766) were still of good predictive performance.

3.3 Validated Chemometric Models for Quinoa's Flour Dietary Constituents

Considering that a good model should bear low values of RMSECV and RMSEP, and high values of R_{CV} and R_P, the final model for each quinoa's constituent was selected among those presented in Table 2. Robust models were obtained when quinoa flour spectra were pre-processed using EMSC of polynomial degree 2 for fat (RMSECV: 0.268% and RMSEP: 0.256%) and carbohydrates (RMSECV: 0.641% and RMSEP: 0.643%) following extraction of five CPPLS latent variables each. For protein (RMSECV: 0.584% and RMSEP: 0.558%), the best model was that of polynomial degree 0 with ten CPPLS latent variables, and for ashes (RMSECV: 0.258% and RMSEP: 0.232%) it was reached with a polynomial degree 1 with eight CPPLS latent variables. High coefficients of correlation of prediction (R_P: 0.690–0.821) were found for protein, fat, ashes and carbohydrates when models were validated on a test data set consisting of 13 quinoa flour spectra.

Coefficients of variation of prediction (CV_P) of the final models were also estimated by dividing RMSEP by the observed mean value for every constituent (values calculated from Tables 1 and 2). This resulted in 5.64% for protein; 3.88% for fat; 7.32% for ashes; and 0.80% for carbohydrates contents.

Finally, in order to further characterise the prediction performance of each of the final models, uncertainty about the correlation coefficient of prediction (R_P) was built by bootstrapping. At each of the 1000 iterations for bootstrapping, a new 80% calibration/20% validation data partition was randomly obtained, the chosen model was fitted to the calibration data with the pre-determined number of components, and R_P was extracted from the test data. The histograms of R_P built for each of the final models (Fig. 1, left) show that the NIT models for estimating ashes and fat had the lowest uncertainty (i.e., narrower spread) about R_P, and therefore were the most robust chemometric models. Thus, the bootstrap results confirmed what was initially suggested by the R_{CV} and R_C values, which were the highest for ashes and fat (Table 2).

The wider spread of the R_P histogram for protein corroborated that, among the four dietary constituents studied, the model for protein presented the lowest precision.

The degree of fitting and predictability of the final models can be appreciated from the scatter plots between the reference values and those fitted (Fig. 1, middle) and predicted (Fig. 1, right) from the NIT calibration models. Judging from the validation and bootstrapping results, it can be said that the predictability of the NIT models is acceptable, bearing in mind that chemical analyses also have associated errors. For proximate data estimation, it is inevitable that NIT errors are higher than those of the

reference techniques, as the error of NIT technique contains both laboratory and NIT errors, although the main advantage of the NIT procedure is its higher repeatability, speed and non-destructive nature.
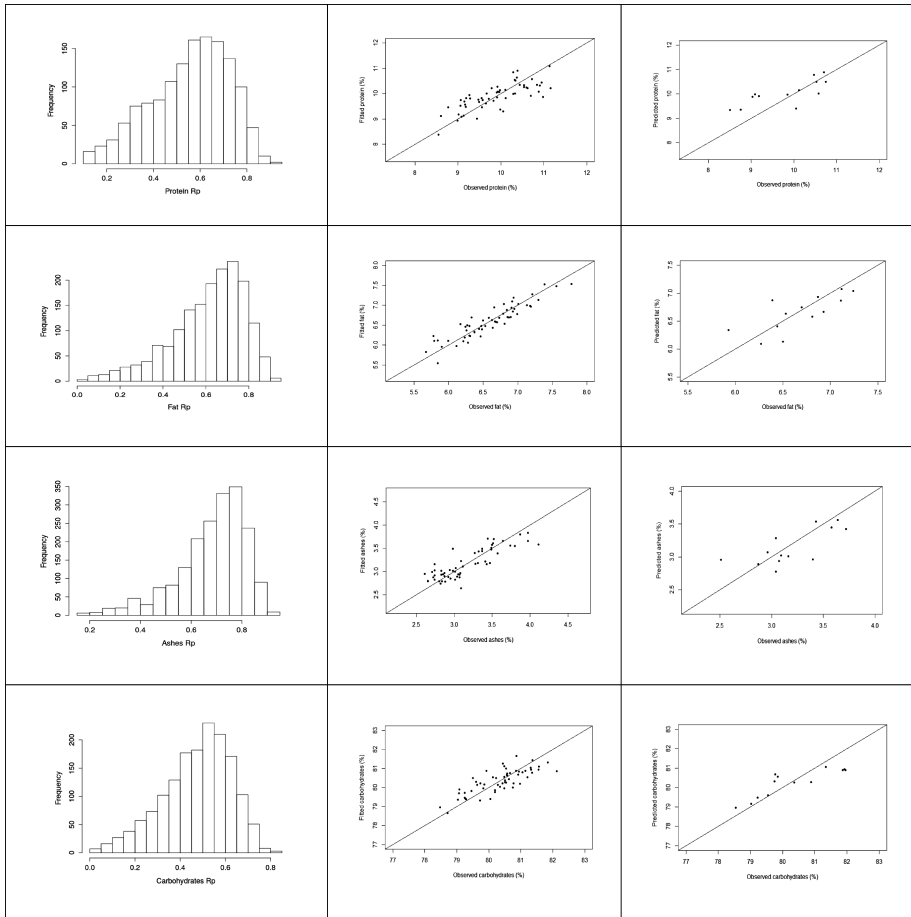

Fig. 1. Prediction performance of NIT chemometric models for protein, fat, ashes and carbohydrates contents in quinoa flour, as evaluated by the uncertainty about the correlation coefficient of prediction (R_P) built by bootstrapping (left), and the scatter plots between chemical reference values and those fitted to the calibration data set (middle) and predicted using the validation data set (right).

4 Conclusions

As a whole, a higher order polynomial of the EMSC signal correction led to fewer optimal latent variables. Except for the protein contents' model, the polynomial degree 1 or 2 improved the accuracy statistics of the models (when compared to polynomial

degree 0). The best combinations in terms of accuracy and predictability for estimating fat and carbohydrates contents in quinoa flour were attained by correcting light scattering with polynomial degree 2 following extraction of five CPPLS latent variables. For ashes, best results were achieved with a polynomial degree 1 with eight CPPLS latent variables, while for protein, a polynomial degree 0 with ten CPPLS latent variables. Although the most robust models could be developed for fat (90% CI RMSEP: [0.479–0.654] as determined by the bootstrap method) and ashes (90% CI RMSEP [0.420–0.575]), the predictability was still good for carbohydrates (90% CI RMSEP: [0.741–0.973]) and protein (90% CI RMSEP: [0.636–0.845]). Thus, in this study, good predictions of the dietary constituents of quinoa flours may be achieved by using NIT technology. The main advantages of this technique are the rapid determination for routine analysis, reduced costs and absence of sample preparation and waste generation.

References

1. Jacobsen, S.E., Mujica, A., Jensen, C.R.: The resistance of quinoa (Chenopodium quinoa, Willd.) to adverse abiotic factors. Food Rev. Int. **19**(1–2), 99–109 (2003)
2. Oliveri, P., Di Egidio, V., Woodcock, T., Downey, G.: Application of class-modelling techniques to near infrared data for food authentication purposes. Food Chem. **125**, 1450–1456 (2011)
3. Sudar, R., Jurković, Z., Galonja, M., Turk, I., Arambašić, M.: Application of Near Infrared Transmission for the determination of ash in wheat flour. Agriculturae Conspectus Scientificus **72**(3), 233–238 (2007)
4. Mevik, B.H., Wehrens, R., Liland, K.H.: Pls: partial least squares and principal component regression. R package version 2.5-0. Available online at https://cran.r-project.org/web/packages/pls/ (2015). Accessed 04 Feb 2017
5. AOAC: Official methods of analysis of the association of analytical chemists international. In Horwitz, W. (ed.), 17th ed., AOAC International, Gaithersburg, MD, USA (2000)
6. Panero, P.S., Panero, F.S., Panero, J.S., Silva, H.E.B.: Application of extended multiplicative signal correction to short-wavelength near infrared spectra of moisture in marzipan. J. Data Anal. Inf. Process. **1**(3), 30–34 (2013)
7. Mevik, B.H., Wehrens, R.: The pls package: principal component and partial least squares regression in R. J. Stat. Softw. **18**(2), 1–24 (2007)
8. Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T.: Classification and regression training. Package "caret". Date 2016-11-10. Repository CRAN. Available online at https://cran.r-project.org/web/packages/caret/index.html (2016). Accessed 04 Feb 2017
9. Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D.: Modern applied statistics with s. package "caret". Date 2016-04-21. Repository CRAN. Available online at https://cran.r-project.org/web/packages/MASS/index.html (2016). Accessed 04 Feb 2017
10. Liland, K.H.: Extended multiplicative signal correction. package "EMSC". Date 2016-04-24. Repository CRAN. Available online https://cran.r-project.org/web/packages/EMSC/index.html (2016). Accessed 04 Feb 2017

11. Stevens, A., Ramirez-Lopez, L.: An introduction to the prospectr package. Vignette R package version 0.1.3. Available online at https://github.com/antoinestevens/prospectr (2013). Accessed 04 Feb 2017
12. R Core Team: R, a language and environment for statistical computing. R Foundation for Stastistical Computing, Vienna, Austria. Available online at http://www.R-project.org/ (2016). Accessed 04 Feb 2017
13. Ferreira, D.S., Pallone, J.A.L., Poppi, R.J.: Direct analysis of the main chemical constituents in Chenopodium quinoa grain using Fourier transform near-infrared spectroscopy. Food Control **48**, 91–95 (2015)
14. González-Martín, M.I., Moncada, G.W., Fischer, S., Escuredo, O.: Chemical characteristics and mineral composition of quinoa by near-infrared spectroscopy. J. Sci. Food Agric. **94**(5), 876–881 (2014)
15. Vega-Gálvez, A., Miranda, M., Vergara, J., Uribe, E., Puente, L., Martínez, E.: Nutrition facts and functional potential of quinoa (Chenopodium quinoa willd.), an ancient Andean grain: a review. J. Sci. Food Agric., **90**(15), 2541–2547 (2010)
16. Repo-Carrasco-Valencia, R., Hellström, J.K., Pihlava, J.M., Mattila, P.H.: Flavonoids and other phenolic compounds in Andean indigenous grains: Quinoa (Chenopodium quinoa), kañiwa (Chenopodium pallidicaule) and kiwicha (Amaranthus caudatus). Food Chem. **120** (1), 128–133 (2010)